STUDIES IN *FUZZINESS* AND *SOFT COMPUTING*

Jorge Casillas
Francisco J. Martínez-López (Eds.)

Marketing Intelligent Systems Using Soft Computing

Managerial and Research Applications

Springer

Jorge Casillas and Francisco J. Martínez-López (Eds.)

Marketing Intelligent Systems Using Soft Computing: Managerial and Research Applications

Studies in Fuzziness and Soft Computing, Volume 258

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 243. Rudolf Seising (Ed.)
Views on Fuzzy Sets and Systems from Different Perspectives, 2009
ISBN 978-3-540-93801-9

Vol. 244. Xiaodong Liu and Witold Pedrycz
Axiomatic Fuzzy Set Theory and Its Applications, 2009
ISBN 978-3-642-00401-8

Vol. 245. Xuzhu Wang, Da Ruan, Etienne E. Kerre
Mathematics of Fuzziness – Basic Issues, 2009
ISBN 978-3-540-78310-7

Vol. 246. Piedad Brox, Iluminada Castillo, Santiago Sánchez Solano
Fuzzy Logic-Based Algorithms for Video De-Interlacing, 2010
ISBN 978-3-642-10694-1

Vol. 247. Michael Glykas
Fuzzy Cognitive Maps, 2010
ISBN 978-3-642-03219-6

Vol. 248. Bing-Yuan Cao
Optimal Models and Methods with Fuzzy Quantities, 2010
ISBN 978-3-642-10710-8

Vol. 249. Bernadette Bouchon-Meunier, Luis Magdalena, Manuel Ojeda-Aciego, José-Luis Verdegay, Ronald R. Yager (Eds.)
Foundations of Reasoning under Uncertainty, 2010
ISBN 978-3-642-10726-9

Vol. 250. Xiaoxia Huang
Portfolio Analysis, 2010
ISBN 978-3-642-11213-3

Vol. 251. George A. Anastassiou
Fuzzy Mathematics: Approximation Theory, 2010
ISBN 978-3-642-11219-5

Vol. 252. Cengiz Kahraman, Mesut Yavuz (Eds.)
Production Engineering and Management under Fuzziness, 2010
ISBN 978-3-642-12051-0

Vol. 253. Badredine Arfi
Linguistic Fuzzy Logic Methods in Social Sciences, 2010
ISBN 978-3-642-13342-8

Vol. 254. Weldon A. Lodwick, Janusz Kacprzyk (Eds.)
Fuzzy Optimization, 2010
ISBN 978-3-642-13934-5

Vol. 255. Zongmin Ma, Li Yan (Eds.)
Soft Computing in XML Data Management, 2010
ISBN 978-3-642-14009-9

Vol. 256. Robert Jeansoulin, Odile Papini, Henri Prade, and Steven Schockaert (Eds.)
Methods for Handling Imperfect Spatial Information, 2010
ISBN 978-3-642-14754-8

Vol. 257. Salvatore Greco, Ricardo Alberto Marques Pereira, Massimo Squillante, Ronald R. Yager, and Janusz Kacprzyk (Eds.)
*Preferences and Decisions,*2010
ISBN 978-3-642-15975-6

Vol. 258. Jorge Casillas and Francisco J. Martínez-López (Eds.)
Marketing Intelligent Systems Using Soft Computing: Managerial and Research Applications, 2010
ISBN 978-3-642-15605-2

Jorge Casillas and
Francisco J. Martínez-López (Eds.)

Marketing Intelligent Systems Using Soft Computing: Managerial and Research Applications

Springer

Editors

Dr. Jorge Casillas
Department of Computer Science
and Artificial Intelligence
Computer and Telecommunication
Engineering School
University of Granada
Granada E-18071
Spain
E-mail: casillas@decsai.ugr.es

Dr. Francisco J. Martínez-López
Department of Marketing
Business Faculty
University of Granada
Granada, Spain E-18.071
E-mail: fjmlopez@ugr.es

and

Department of Economics and
Business – Marketing Group
Open University of Catalonia
Barcelona, Spain E-08.035
E-mail: fmartinezl@uoc.edu

ISBN 978-3-642-15605-2 e-ISBN 978-3-642-15606-9

DOI 10.1007/978-3-642-15606-9

Studies in Fuzziness and Soft Computing ISSN 1434-9922

Library of Congress Control Number: 2010934965

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

Dr. Jay Liebowitz

Orkand Endowed Chair in Management and Technology
University of Maryland University College
Graduate School of Management & Technology
3501 University Boulevard East
Adelphi, Maryland 20783-8030 USA
jliebowitz@umuc.edu

When I first heard the general topic of this book, Marketing Intelligent Systems or what I'll refer to as Marketing Intelligence, it sounded quite intriguing. Certainly, the marketing field is laden with numeric and symbolic data, ripe for various types of mining—data, text, multimedia, and web mining. It's an open laboratory for applying numerous forms of intelligentsia—neural networks, data mining, expert systems, intelligent agents, genetic algorithms, support vector machines, hidden Markov models, fuzzy logic, hybrid intelligent systems, and other techniques. I always felt that the marketing and finance domains are wonderful application areas for intelligent systems, and this book demonstrates the synergy between marketing and intelligent systems, especially soft computing.

Interactive advertising is a complementary field to marketing where intelligent systems can play a role. I had the pleasure of working on a summer faculty fellowship with R/GA in New York City—they have been ranked as the top interactive advertising agency worldwide. I quickly learned that interactive advertising also takes advantage of data visualization and intelligent systems technologies to help inform the Chief Marketing Officer of various companies. Having improved ways to present information for strategic decision making through use of these technologies is a great benefit. A number of interactive advertising agencies have groups working on "data intelligence" in order to present different views of sales and other data in order to help their clients make better marketing decisions.

Let's explore the term "marketing intelligence". The *Marketing Intelligence & Planning* journal, published by Emerald Publishers, "aims to provide a vehicle that will help marketing managers put research and plans into action." In its aims and scope, the editors further explain, "Within that broad description lies a wealth of skills encompassing information-gathering, data interpretation, consumer psychology, technological resource knowledge, demographics and the marshalling of human and technical resources to create a powerful strategy." Data interpretation seems to be at the intersection of "marketing" and "intelligence". By applying advanced technologies, data can be interpreted and visualized in order to enhance the decision making ability of the marketing executives. Certainly, blogs and social networking sites are rich forums for applying mining techniques to look for hidden patterns and relationships. These patterns may enrich the discovery process and allow different views, perhaps those unexpected, from those initially conceived.

In Inderscience's *International Journal of Business Forecasting and Marketing Intelligence*, the focus is on applying innovative intelligence methodologies, such as rule-based forecasting, fuzzy logic forecasting, and other intelligent system techniques, to improve forecasting and marketing decisions. In looking at the Winter 2010 Marketing Educator's American Marketing Association Conference, there are a number of tracks presented where the use of intelligent systems could be helpful: Consumer behavior, global marketing, brand marketing, business-to-business marketing, research methods, marketing strategy, sales and customer relationship management, service science, retailing, and marketing & technology. Digital-centered marketing where one takes advantage of such digital marketing elements as mobile, viral, and social marketing channels is a growing field that can apply the synergies of marketing and intelligent systems. Positions for Directors of Marketing Intelligence are also appearing to be the champions of new marketing methods. Gartner Group reports, such as the August 2008 report on "Social Media Delivers Marketing Intelligence", are further evidence of this evolving field.

In a recent report of "hot topics" for college undergraduates to select as majors in the coming years, the fields of service science, sustainability, health informatics, and computational sciences were cited as the key emerging fields. Certainly, marketing intelligence can play a key role in the service science field, as well as perhaps some of the other fields noted. In May 2008, there was even a special issue on "Service Intelligence and Service Science" published in the Springer *Service-Oriented Computing and Applications* Journal. In July 2009, there was the 3rd International Workshop on Service Intelligence and Computing to look at the synergies between the service intelligence and service sciences fields. In the years ahead, advanced computational technologies will be applied to the service science domain to enhance marketing types of decisions.

In 2006, I edited a book titled Strategic Intelligence: Business Intelligence, Competitive Intelligence, and Knowledge Management (Taylor & Francis). I defined strategic intelligence as the aggregation of the other types of intelligentsia to provide value-added information and knowledge toward making organizational strategic decisions. I see strategic intelligence as the intersection of business intelligence, competitive intelligence, and knowledge management, whereby business intelligence and knowledge management have a more internal focus and competitive intelligence has a greater external view. Marketing intelligence seems to contribute to both business and competitive intelligence—helping to identify hidden patterns and relationships of large masses of data and text and also assisting in developing a systematic program for collecting, analyzing, and managing external information relating to an organization's decision making process.

I believe that this book sheds important light on how marketing intelligence, through the use of complementary marketing and intelligent systems techniques, can add to the strategic intelligence of an organization. The chapters present both a marketing and soft computing/intelligent systems perspective, respectively. I commend the editors and authors towards filling the vacuum in providing a key reference text in the marketing intelligence field. Enjoy!

Preface

The development of *ad hoc* Knowledge Discovery in Databases (KDD) applications for the resolution of information and decision-taking problems in marketing is more necessary than ever. If we observe the evolution of so-called Marketing Management Support Systems (MkMSS) through time, it is easy to see how the new categories of systems which have appeared over the last two decades have led in that direction. In fact, during the eighties, the inflection point was set that marked a transition stage from what are known as Data-driven Systems to Knowledge-based Systems, i.e. MkMSS based on Artificial Intelligent (AI) methods. The popular Marketing Expert Systems were the first type in this MkMSS category. Then, other new types within this category appeared, such as Case-based Reasoning Marketing Systems, Systems for the Improvement of Creativity in Marketing, Marketing Systems based on Artificial Neural Networks, Fuzzy Rules, etc.

Most of these systems have been recent proposals and, in any case, their application is still scarce in marketing practical and, specially, academic domains. Anyhow, we have noticed a clear greater interest and use of these Knowledge-based Systems among marketing professionals than among marketing academics. Indeed, we perceive a notable disconnection of the latter from these systems, who still base most of their analytical methods on techniques belonging to statistics. Doubtless, this fact contributes to these two dimensions of marketing—i.e. the professional and the academic—grow apart.

During the years that we have been working on this research stream, we have realized the significant lack of papers, especially in marketing journals, which focus on developing *ad hoc* AI-based methods and tools to solve marketing problems. Obviously, this also implies a lack of involvement by marketing academics in this promising research stream in marketing. Among the reasons that can be argued to justify the residual use that marketing academics make of AI, we highlight a generalized ignorance of what some branches of the AI discipline (such as knowledge-based systems, machine learning, soft computing, search and optimization algorithms, etc.) can offer. Of course, we encourage marketing academics to show a strong determination to approximate AI to the marketing discipline. When we talk about approximation, we refer to going far beyond a superficial knowledge of what these AI concepts are. On the contrary, we believe that multidisciplinary research projects, formed by hybrid teams of marketing and artificial intelligence people, are more than necessary.

In essence, the AI discipline has a notable number of good researchers who are interested in applying their proposals, where business in general, management and, in particular, marketing are target areas for application. However, the quality of such applications necessarily depends on how well described the marketing problem to be solved is, as well as how well developed and applied the AI-based methods are. This means having the support and involvement of people belonging to marketing, the users of such applications.

Considering the above, this editorial project has two strategic aims:

1. Contribute and encourage the worldwide take-off of what we have called Marketing Intelligent Systems. These are, in general, AI-based systems applied to aid decision-taking in marketing. Moreover, when we recently proposed this term of Marketing Intelligent Systems, we specifically related it to the development and application of intelligent systems based on Soft Computing and other machine-learning methods for marketing. This is the main scope of interest.
2. Promote the idea of interdisciplinary research projects, with members belonging to AI and marketing, in order to develop better applications thanks to the collaboration of both disciplines.

This book volume presented here is a worthy start for these purposes. Next, we briefly comment on its structural parts.

Prior to the presentation of the manuscripts selected after a competitive call for chapters, the first block of this book is dedicated to introducing diverse leading marketing academics' reflections on the potential of Soft Computing and other AI-based methods for the marketing domain.

Following these essays, the book is structured in five main parts, in order to articulate in a more congruent manner the rest of the chapters. In this regard, the reader should be aware of the fact that some of the chapters could be reasonably assigned to more than one part, though they have been finally grouped as follows.

The first part deals with segmentation and targeting. Duran *et al.* analyze the use of different clustering techniques such as k-means, fuzzy c-means, genetic k-means and neural-gas algorithms to identify common characteristics and segment customers. Next, Markic and Tomic investigate the integration of crisp and fuzzy clustering techniques with knowledge-based expert systems for customer segmentation. Thirdly, Van der Putten and Kok develop predictive data mining for behavioral targeting by data fusion and analyze different techniques such as neural networks, linear regression, k-nearest neighbor and naive Bayes to deal with targeting. Finally, Bruckhaus reviews collective intelligent techniques which allow marketing managers to discover and approach behaviors, preferences and ideas of groups of people. These techniques are useful for new insights into firms' customer portfolios so they can be better identified and targeted.

The second part contains several contributions grouped around marketing modeling. Bhattacharyya explores the use of multi-objective genetic programming to derive predictive models from a marketing-related dataset. Orriols-Puig *et al.* propose an unsupervised genetic learning approach based on fuzzy association rules to extract causal patterns from consumer behavior databases. Finally, Pereira and Tettamanzi introduce a distributed evolutionary algorithm to optimize fuzzy rule-based predictive models of various types of customer behavior.

Next, there are two parts devoted to elements of the marketing-mix, specifically applications and solutions for Communication and Product policies.

In the third part, Hsu *et al.* show how a fuzzy analytic hierarchy process helps to reduce imprecision and improve judgment when evaluating the preference of customer opinions about customer relationship management. López and López propose a distributed intelligent system based on multi-agent systems, an analytic hierarchy process and fuzzy c-means to analyze customers' preferences for direct marketing. Wong also addresses direct marketing but using evolutionary algorithms that describe Bayesian networks from incomplete databases.

The fourth part consists of two chapters directly related to Product policy, plus a third dealing with a problem of consumer's choice based on diverse criteria, mainly functional characteristics of products, though this contribution also has implications for strategic and other marketing-mix areas. Genetic algorithms have proved to be effective in optimizing product line design, according to both Tsafarakis-Matsatsinis and Balakrishnan *et al.* in their chapters. A dynamic programming algorithm is also used in the second case to seed the genetic algorithm with promising initial solutions. In Beynon *et al.*'s chapter, probabilistic reasoning is hybridized with analytic hierarchy processes to approach the problem of consumer judgment and the grouping of the preference criteria that drive their product/brand choices.

The final part is a set of contributions grouped under e-commerce applications. Sun *et al.* propose a multiagent system based on case-based reasoning and fuzzy logic for web service composition and recommendation. Dass *et al.* investigate the use of functional data analysis for the dynamic forecasting of price prediction in simultaneous online auctions. Finally, Beynon and Page deploy probabilistic reasoning and differential evolution to deal with incomplete data for measuring consumer web purchasing attitudes.

This book is useful for technicians who apply intelligent systems to marketing, as well as for those marketing academics and professionals interested in the application of advanced intelligent systems. Synthetically, it is especially recommended for the following groups:

- Computer Science engineers working on intelligent systems applications, especially Soft-Computing-based Intelligent Systems.
- Marketers and business managers of firms working with complex information systems.
- Computer Science and Marketing academics, in particular those investigating synergies between the AI and Marketing.
- PhD students studying intelligent systems applications and advanced analytical methods for marketing.

Finally, we wish to thank Springer and in particular Prof. J. Kacprzyk, for having given us the opportunity to make real this fascinating and challenging dream. We are also honored and privileged to have received help and encouragement from several notable world marketing academics; we thank you for your support, smart ideas and thoughts. Likewise, we offer our most sincere acknowledgment and gratitude to all the contributors for their rigor and generosity in producing such high quality papers. Last but not least, we especially thank the team of reviewers for their great work.

March 2010

Granada (Spain)
Jorge Casillas and Francisco J. Martínez-López
University of Granada, Spain

Contents

Essays

Segmentation and Targeting

Marketing Modelling

Communication/Direct Marketing

Product

E-Commerce

Marketing and Artificial Intelligence: Great Opportunities, Reluctant Partners

Berend Wierenga

Professor of Marketing
Rotterdam School of Management, Erasmus University
e-mail: bwierenga@rsm.nl

1 Introduction

Marketing managers make decisions about products, brands, advertising, promotions, price, and distribution channels, based on deep knowledge about customers. The outcomes of marketing decisions are dependent on the behavior of other actors such as competitors, suppliers and resellers. Furthermore, uncertain factors such as the overall economy, the state of the financial sector and (international) political developments play an important role. Marketing decision making not only refers to tactical marketing mix instruments (the well-known 4Ps), but also to strategic issues, such as product development and innovation and long term decisions with respect to positioning, segmentation, expansion, and growth.

This short description illustrates that marketing is a complex field of decision making. Some marketing problems are relatively well-structured (especially the more tactical marketing mix problems), but there are also many weakly-structured or even ill-structured problems. Many marketing phenomena can be expressed in numbers, for example sales (in units or dollars), market share, price, advertising expenditures, number of resellers, retention/churn, customer value, etc. Such variables can be computed and their mutual relationships can be quantified. However, there are also many qualitative problems in marketing, especially the more strategic ones. Therefore, besides computation, marketing decision making also involves a large degree of judgment and intuition in which the knowledge, expertise, and experience of professionals play an important role. It is clear that marketing decision making is a combination of analysis and judgment.

As we will see below, the analytical part of marketing decision making is well served with a rich collection of sophisticated mathematical models and procedures for estimation and optimization that support marketing decision making. However, this is much less the case for the judgmental part where knowledge and expertise play an important role. The question is whether the acquisition and use of knowledge and expertise by marketing decision makers and their application to actual marketing problems can also benefit from appropriate decision support technologies. In this book on marketing intelligent systems, it is logical to ask what the field of Artificial Intelligence can contribute here. Artificial Intelligence (AI) deals

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 1–8.
springerlink.com

with human intelligence and how this can be represented in computers. Important topics in AI are knowledge, knowledge representation, reasoning, learning, expertise, heuristic search, and pattern recognition. All these elements are relevant in the daily life of marketing decision makers who constantly use their knowledge, expertise and intuition to solve marketing problems. Therefore, potentially AI can make an important contribution to marketing decision making. However, so far this potential has only been realized to a very limited extent. This contribution takes a closer look at the opportunities for AI in marketing, takes stock of what has been achieved so far, and discusses perspectives for the future.

2 Marketing Problem-Solving Modes

We start with a discussion about marketing problem-solving modes. These are specific ways of making marketing decisions. Basically, decision making is dependent on three factors: the marketing problem, the decision maker, and the decision environment. This results in four different marketing problem-solving modes: Optimizing, Reasoning, Analogizing, and Creating (ORAC) (Wierenga and Van Bruggen 1997; 2000).

The ORAC model is depicted in Figure 1 and shows the full continuum of how marketing decision makers deal with problems. At the one extreme we have hard calculation (“clocks of mind”), and at the other we have free flows of thought, mental processes without a clear goal (“clouds of mind’). We briefly discuss the four marketing problem-solving modes.

O	=	Optimizing	Clocks of Mind
			↑
R	=	Reasoning	.
			.
A	=	Analogizing	.
			↓
C	=	Creating	Clouds of Mind

Fig. 1 The ORAC model of marketing problem-solving modes (Wierenga and van Bruggen 1997; 2000)

Optimizing implies that there is an objectively best solution that can be reached by proper use of the marketing instruments. This is only possible if we have precise insight in the mechanism behind the variable that we want to optimize (e.g. sales, market share or profit). Once this mechanism is captured in a mathematical model, the best values for the marketing instruments (dependent on the objective function) can be found by applying optimization or simulation. An example of optimizing is deciding on the best media plan (i.e. the allocation over media such as TV, press, internet) for an advertising campaign, once the advertising budget and the relevant reach and costs data of the media are known.

Reasoning means that a marketer has a representation (mental model) of certain marketing phenomena in mind, and uses this as a basis for making inferences and drawing conclusions. For example, the decision maker may have a mental model of the factors that determine the market share of his brand. Suppose that this is relatively low in one particular geographical area. The manager might then reason (if ….then…) that this can be due to several possible causes, (i) deviant preferences of consumers; (ii) low efforts of salespeople; or (iii) relatively strong competition (Goldstein 2001). Market research can help to verify each of these possible causes and will result in decisions about possible actions (e.g. change the taste of the product; increase the salesforce). The outcomes of market research may also lead to the updating of managers' mental models.

Analogizing takes place when marketing decision makers, confronted with a problem, recall a similar problem that previously occurred and was solved in a satisfactory way. Decision makers often organize their experiences in the form of "stories". New cases are easily interpreted using existing stories, and solutions are found quickly, often automatically. This type of analogical reasoning occurs very often in marketing. For example, when a new product is introduced, experiences with earlier product introductions act as points of reference.

Creating occurs when the decision maker searches for novel and effective ideas and solutions. This means mapping and exploring the problem's conceptual space and involves divergent thinking. In marketing, creating is a very important marketing problem-solving mode. Marketers are always looking for innovative product ideas, catchy advertising themes and imaginative sales promotion campaigns.

3 Marketing Problem-Solving Modes and Decision Support Technologies

Over time a rich collection of decision aids have become available that can support marketing managers to improve the effectiveness and efficiency of their decisions. The complete set is referred to as *marketing management support systems (MMSS)*. (Wierenga and van Bruggen 2000). Figure 2 shows how the decision support technologies used in these MMSS are related to the marketing problem-solving modes. The mapping to marketing problem-solving modes is not exactly one-to-one, but Figure 2 shows the overall tendencies.

Marketing management support systems can be divided in two categories, data-driven and knowledge-driven. Marketing data have become available abundantly over the last few decades (e.g. scanner data; internet data) and data-driven decision support technologies are very prominent in marketing. They are particularly important for optimizing and reasoning. Methods from operations research (OR) and econometrics play an important role here. For example, OR methods can be used to optimally allocate the advertising budget over advertising media and econometric analysis can help to statistically determine the factors that affect market share. As we have just seen, the latter information is useful for reasoning about possible marketing actions, and for the updating of managers' mental models.

<table>
<tr><th>Marketing Problem-Solving Modes</th><th>Decision Support Technologies</th></tr>
<tr><td>Optimizing</td><td rowspan="4"><u>Data-driven</u>
• Operations Research (OR)
• Econometric Modeling
• Predictive Modeling/NN

<u>Knowledge-driven</u>
• Knowledge-Based Systems/ Expert Systems
• Analogical Reasoning/ Case-Based Reasoning
• Creativity Support Systems</td></tr>
<tr><td>Reasoning</td></tr>
<tr><td>Analogizing</td></tr>
<tr><td>Creating</td></tr>
</table>

Fig. 2 Marketing problem-solving modes and decision support technologies

Predictive modeling techniques used in Customer Relationship Management (CRM) and direct marketing are also data-driven. (Neural nets-NN is a predictive modeling technique that has its roots in AI).

Knowledge-driven decision support technologies are particularly useful for marketing problem-solving modes that deal with weakly structured problems, parts (i.e. the qualitative element) of reasoning, analogizing, and creating. Knowledge-based systems (KBS) and expert systems (ES) are important examples. The latter, in particular, can also be used for reasoning about the factors behind particular marketing phenomena, for example the success of new products, or the effect of an advertising campaign. Decision support technologies based on analogical reasoning, such as case-based reasoning (CBR) have great potential for the analogizing and creating modes. This is also a potential application area for creativity support systems (Garfield 2008).

4 The State of Artificial Intelligence (AI) in Marketing

Figure 2 shows that the potential for knowledge-driven decision support technologies in marketing is high. Contributions from AI are possible for three of the four marketing problem-solving modes. However, reality does not reflect this. To date, data-driven approaches, mostly a combination of operations research and econometric methods are dominant in marketing management support systems. It is safe to say that data-driven, quantitative models (i.e. the upper-right corner of Figure 2) make up over 80% of all the work in decision support systems for marketing at this moment. Compared to this, the role of artificial intelligence in marketing is minor[1]. The number of publications about AI approaches in marketing literature is limited and the same holds true for the presence of marketing in AI literature.

In 1958 Simon and Newell wrote that "the very core of managerial activity is the exercise of judgment and intuition" and that "large areas of managerial activity have hardly been touched by operations and management science". In the same paper (in Operations Research) they foresaw the day that it would be possible "to handle with appropriate analytical tools the problems that we now tackle with judgment and guess". Strangely enough, it does not seem that judgment and intuition in marketing have benefitted a lot from the progress in AI since the late fifties. It is true that AI techniques are used in marketing (as we will see below), but only to a limited degree.

There are several (possible) reasons for the limited use of AI in marketing.

- Modern marketing management as a field emerged in the late 1950s. At that time, operations research and econometrics were already established fields. In fact, they played a key role in the development of the area of marketing models (Wierenga 2008), which is one of the three academic pillars of contemporary marketing (the other pillars are consumer behavior and managerial marketing). Artificial intelligence as a field was only just emerging at that time.
- OR and econometrics are fields with well-defined sets of techniques and algorithms, with clear purposes and application goals. They mostly come with user-friendly computer programs that marketers can directly implement for their problems. However, AI is a heterogeneous, maybe even eclectic, set of approaches, which often takes considerable effort to implement. Moreover, most marketing academics are not trained in the concepts and theories of AI.
- The results of applications of OR and econometrics can usually be quantified, for example as the increase in number of sales or in dollars of profit. AI techniques, however, are mostly applied to weakly-structured problems and it is often difficult to measure how much better a solution is due to the use of AI, for example a new product design or a new advertising campaign. Marketers seem to be better at ease with rigorous results than with soft computing.

There may also be reasons on the side of AI.

- There seems to be little attention for marketing problems in AI. A recent poster of the "The AI Landscape" (Leake 2008) shows many (potential) applications of AI, ranging from education, logistics, surgery, security, to art, music, and entertainment, but fails to mention marketing, advertising, selling, promotions or other marketing-related fields.

[1] Here we refer to the explicit use of AI in marketing. Of course, AI principles may be imbedded in marketing-related procedures such as search algorithms for the Internet).

- Perhaps the progress in AI has been less than was foreseen in 1958. In general, there has been a tendency of over-optimism in AI (a point in case is prediction about when a computer would be the world's chess champion). The promised analytical tools to tackle judgmental marketing problems may come later than expected.

4.1 Applications of AI in Marketing

The main applications of AI in marketing so far are expert systems, neural nets, and case-based reasoning. We discuss them briefly.

4.1.1 Expert Systems

In the late eighties, marketing knowledge emerged as a major topic, together with the notion that it can be captured and subsequently applied by using knowledge-based systems. In marketing, this created a wave of interest in expert systems. They were developed for several domains of marketing (McCann and Gallagher 1990). For example: systems (i) to find the most suitable type of sales promotion; (ii) to recommend the execution of advertisements (positioning, message, presenter); (iii) to screen new product ideas, and (iv) to automate the interpretation of scanner data, including writing reports. Around that time, over twenty expert systems were published in marketing literature (Wierenga & van Bruggen 2000 Chapter 5).An example of a system specially developed for a particular marketing function is BRANDFRAME (developed by Wierenga, Dalebout, and Dutta 2000; see also Wierenga and van Bruggen 2001). This system supports a brand manager, which is a typical marketing job. In BRANDFRAME the situation of the (focal) brand is specified in terms of its attributes, competing brands, retail channels, targets and budgets. When new marketing information comes in, for example from panel data companies such as Nielsen and IRI, BRANDFRAME analyzes this data and recommends the marketing mix instruments (for example: lower the price; start a sales promotion campaign). It is also possible to design marketing programs in BRANDFRAME, for example for advertising or sales promotion campaigns. The system uses frame-based knowledge representation, combined with a rule-based reasoning system. In recent years, marketing literature has reported few further developments in marketing expert systems.

4.1.2 Neural Networks and Predictive Modeling

Around 2000, customer relationship management (CRM) became an important topic in marketing. An essential element of CRM (which is closely related to direct marketing) is the customer database which contains information about each individual customer. This information may refer to socio-economic characteristics (age, gender, education, income), earlier interactions with the customer (e.g. offers made and responses to these offers, complaints, service), and information about the purchase history of the customer (i.e. how much purchased and when). This data can be used to predict the response of customers to a new offer or to predict customer retention/churn. Such predictions are very useful, for example for selecting the most promising prospects for a mailing or for selecting customers in need of special attention because they have a high likelihood of leaving the company. A large set of techniques is available for predictive modeling. Prominent techniques are neural networks (NN) and classification and regression trees (CART), both with their roots in artificial intelligence. However, also more classical statistical techniques are used such as discriminant analysis and (logit) regression (Malthouse and Blattberg 2005; Neslin et al 2006). CRM is a quickly growing area of marketing. Companies want to achieve maximum return on the often huge investments in customer databases. Therefore, further sophistication of predictive modeling techniques for future customer behavior is very important. Fortunately, this volume contains several contributions on this topic.

4.1.3 Analogical Reasoning and Case-Based Reasoning (CBR)

Analogical reasoning plays an important role in human perception and decision making. When confronted with a new problem, people seek similarities with earlier situations and use previous solutions as the starting point for dealing with the problem at hand. This is especially the case in weakly structured areas, where there is no clear set of variables that explain the relevant phenomena or define a precise objective. In marketing we have many such problems, for example in product development, sales promotions, and advertising. Goldstein (2001) found that product managers organize what they learn from analyzing scanner data into a set of stories about brands and their environments. Analogical reasoning is also the principle behind the field of case-based reasoning (CBR) in Artificial Intelligence. A CBR system comprises a set of previous cases from the domain under study and a set of search criteria for retrieving cases for situations that are similar (or analogous) to the target problem. Applications of CBR can be found in domains such as architecture, engineering, law, and medicine. By their nature, many marketing problems have a perfect fit with CBR. Several applications have already emerged, for example CBR systems for promotion planning and for forecasting retail sales (see Wierenga & van Bruggen 2000, Chapter 6). A recent application uses CBR as a decision support technology for designing creative sales promotion campaigns (Althuizen and Wierenga 2009). We believe that analogical reasoning is a fruitful area for synergy between marketing and AI.

4.2 Perspective

Although there is some adoption of AI approaches in marketing, the two areas are almost completely disjoint. This is surprising and also a shame, because the nature of many marketing problems makes them very suitable for AI techniques. There is a real need for decision technologies that support the solution of weakly-structured marketing problems. Van Bruggen and Wierenga (2001) found that most of the existing MMSS support the marketing problem-solving mode of optimizing, but that they are often applied in decision situations for which they are less suitable (i.e. where the marketing problem-solving modes of reasoning, analogizing or creating are applicable). Their study also showed that a bad fit between the marketing-problem-solving mode and the applied decision support technology results in significantly less impact of the support system.

It would be fortunate if further progress in AI can help marketing to deal with the more judgmental problems of its field. Reducing the distance between marketing and AI also has an important pay-off for AI. Marketing is a unique combination of quantitative and qualitative problems, which gives AI the opportunity to demonstrate its power in areas where operations research and econometrics cannot reach. Marketing is also a field where innovation and creativity play an important role. This should appeal to the imaginative AI people.

Hopefully the current volume will be instrumental in bringing marketing and AI closer together.

References

Althuizen, N.A.P., Wierenga, B.: Deploying Analogical Reasoning as a Decision Support Technology for Creatively Solving Managerial Design Problems. Working paper. Rotterdam School of Management, Erasmus University (2009)

Garfield, M.J.: Creativity Support Systems. In: Burnstein, F., Holsapple, C.W. (eds.) Handbook on Decision Support Systems. Variations, vol. 2, pp. 745–758. Springer, New York (2008)

Goldstein, D.K.: Product Manager's Use of Scanner Data: a Story of Organizational Learning. In: Desphandé, R. (ed.) Using Market Knowledge, pp. 191–216. Sage, Thousand Oaks (2001)

Leake, D.B.: AI Magazine Poster: The AI Landscape. AI Magazine 29(2), 3

Malthouse, E.C., Blattberg, R.C.: Can we predict customer lifetime value? Journal of Interactive Marketing 19(1), 2–16 (2005)

Mc Cann, J.M., Gallagher, J.P.: Expert Systems for Scanner Data Environments. Kluwer Academic Publishers, Boston (1990)

Neslin, S.A., Gupta, S., Kamakura, W., Lu, J., Mason, C.H.: Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. Journal of Marketing Research 43, 204–211 (2006)

Simon, H.A., Newell, A.: Heuristic Problem Solving: the Next Advance in Operations Research. Operations Research 6, 1–10 (1958)

Van Bruggen, G.H., Wierenga, B.: Matching Management Support Systems and Managerial Problem-Solving Modes: The Key to Effective Decision Support. European Management Journal 19(3), 228–238 (2001)

Wierenga, B. (ed.): Handbook of Marketing Decision Models, p. 630. Springer Science + Business Media, New York (2008)

Wierenga, B., van Bruggen, G.H.: The Integration of Marketing Problem-Solving Modes and Marketing Management Support Systems. Journal of Marketing 61(3), 21–37 (1997)

Wierenga, B., van Bruggen, G.H.: Marketing Management Support Systems: Principles, Tools, and Implementaiton, p. 341. Kluwer Academic Publishers, Boston (2000)

Wierenga, B., Van Bruggen, G.H.: Developing a Customized Decision Support System for Brand Managers. Interfaces 31(3) Part 2(2), 128–145 (2001)

Wierenga, B., Dalebout, A., Dutta, S.: BRANDFRAME: A Marketing Management Support System for the Brand Manager. In: Wierenga, B., van Bruggen, G. (eds.) Marketing Management Support Systems: Principles, Tools, and Implementation, pp. 231–262. Kluwer Academic Publishers, Boston (2000)

Data Mining and Scientific Knowledge: Some Cautions for Scholarly Researchers

Nick Lee[1] and Gordon Greenley[2]

1 Professor of Marketing and Organizational Research and
Marketing Research Group Convenor
Aston Business School, Birmingham, UK
Co-Editor: European Journal of Marketing
2 Professor of Marketing and Marketing Subject Group Convenor
Aston Business School, Birmingham, UK
Co-Editor: European Journal of Marketing

1 Introduction

Recent years have seen the emergence of data analytic techniques requiring for their practical use previously unimaginable raw computational power. Such techniques include neural network analysis, genetic algorithms, classification and regression trees, v-fold cross-validation clustering and suchlike. Many of these methods are what could be called 'learning' algorithms, which can be used for prediction, classification, association, and clustering of data based on previously-estimated features of a data set. In other words, they are 'trained' on a data set with both predictors and target variables, and the model estimated is then used on future data which does not contain measured values of the target variable. Or in clustering methods, an iterative algorithm looks to generate clusters which are as homogenous within and as heterogeneous between as possible.

Such analytic methods can be used on data collected with the express purposes of testing hypotheses. However, it is when such methods are employed on large sets of data, without *a priori* theoretical hypotheses or expectations, that they are known as data mining. In fact, it appears that such is the explosion in use of such methods, and in particular their use in commercial contexts such as customer relationship management or consumer profiling, that it is the methods themselves which are considered to be 'data mining' methods. However, it should be made clear at the outset of this essay that it is the use that they are put to which should be termed 'data mining', not the tools themselves (Larose 2005). This is despite the naming of software packages like 'Statistica Data Miner', which sell for sums at the higher end of 6-figures to commercial operations. In fact, a technique as ubiquitous as multiple regression can be used as a data mining tool if one wishes.

It is the aim of this essay to place the recent exponential growth of the use of data mining methods into the context of scientific marketing and business research, and in particular to sound a note of caution for social scientific researchers about the over-use of a data-mining approach. In doing so, the fundamental nature

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 9–15.
springerlink.com

of data mining is briefly outlined. Following this, data mining is discussed within a framework of scientific knowledge development and epistemology. Finally, the potential use of data mining in certain contexts is noted. We will conclude with some important points that business and marketing scholars should consider when considering the use of data mining approaches.

2 The Data Mining Method

Data mining is one part of a wider methodology termed Knowledge Discovery in Databases (KDD). Within this process, the term *data mining* refers to the uncovering of new and unsuspected relationships in, and the discovery of new and useful knowledge from, databases (e.g. Adriaans and Zantinge, 1996; Hand et al, 2001). While it should be natural for the scholar to immediately consider the question of what exactly is knowledge, this will be dealt with in the next section. In a more practical context, scientists and businesspeople deal with large databases on a day-to-day basis. In many cases, they use the data to answer questions that they pose in a very structured way – such as 'what is the difference between the mean level of job satisfaction across high and low-stress salespeople' or 'which customers bought brand *x* during August'. Such structured queries have been common practice for many years. The difference between the data mining approach and a normal structured interrogation of a data set is that, when data mining, one starts without such a structured question, but instead is interested in *exploring* the database for any potential 'nuggets' of interest.

Another key point of interest is that – while this is not an essential component of a data-mining approach – most methods of data mining involve *learning* algorithms. Unlike traditional analysis of data, learning algorithms (such as genetic algorithms or neural networks) are able to be 'trained' to create rules which are able to describe a data set, that are then able to work on new data. While humans could of course train themselves to do this, the advantage of the learning algorithm is that it can work with far larger data sets, in far less time, than humans – as long as the data set contains at least some structure.

3 Data Mining and Scientific Knowledge

The characteristics of the data mining approach introduced above have significant relevance to its use to generate scientific knowledge. Of course, as shall be seen subsequently, data mining has use in many contexts outside of science as well. However, as Editors of the *European Journal of Marketing*, a major journal dedicated to advances in marketing theory, it is their use in scientific knowledge development in marketing (and by extension more general business or social research) which is our primary concern in this essay[1]. Marketing has long debated

[1] It is important to note that – while we were invited to write this essay as Editors of EJM – the opinions expressed here should not be taken to represent a formal editorial policy or direction for EJM in any manner.

its status as a science (e.g. Buzzell, 1963, Hunt, 1983; 1990; 2003, Brown, 1995), with various scholars taking different viewpoints on both the nature of science itself, and whether marketing can be considered to be a science. Hunt's (e.g. 1983) work is arguably the most articulate and significant corpus regarding this issue, and it is defensible to say that – working with the premise that one wants to class marketing as a science – Hunt's delineation of the nature of science (e.g. 1983) can be taken as broadly coherent in the context of marketing research.

Hunt defines three essential characteristics of a science (1983: pp. 18); "(1) a distinct subject matter, (2) the description and classification of the subject matter, and (3) the presumption that underlying the subject matter are uniformities and regularities which science seeks to discover". Hunt also adds (pp. 18-19) that to be termed a science, a discipline must employ the *scientific method*; which he defines as a "set of procedures". Like the nature of science itself, the scientific method has been subject to considerable debate and controversy over the last century (e.g. Feyerabend, 1993). One of the key areas of misunderstanding is whether the method refers to the practical techniques used for discovery, or the conceptual/theoretical method used to justify a discovery as knowledge (Lee and Lings 2008). Hunt (1983) points out that the scientific method is not dependent on the use of particular data collection methods, tools, or analysis techniques, since it is of course the case that different sciences use different tools as appropriate. Instead, the scientific method should more accurately be perceived as a method for *justifying* the knowledge claims uncovered by investigation (Lee and Lings 2008). In this sense, there are multiple (perhaps infinite) ways of discovery, and of making knowledge claims about the world, but at present only one scientific method of justifying those claims as actual *knowledge*.

This method – termed more formally the hypothetico-deductive method – has proven to be the foundation of scientific research since its formal articulation by Karl Popper. Thus, in exploring the usefulness of data mining for scientific research, it is naturally necessary to do so in relation to where it may sit within a hypothetico-deductive approach to research. While a full explication of the hypothetico-deductive method is outside the scope of this short essay, it is the term *deductive* which is of most relevance to the present discussion. Of course, deduction refers to the creation of theory from logical argument, which may then be tested through empirical observation. While it is often characterized as a cycle of induction and deduction, the essence of the hypothetico-deductive method is the idea that one should create theoretical hypotheses through deductive reasoning, and then collect empirical data in an attempt to *falsify* those hypotheses. Certainly, one may begin the cycle by using inductive reasoning from some empirical observation or discovery, but until formal hypotheses are generated and subsequently tested, one should not claim to have created any scientific knowledge.

The idea of falsification is of critical importance in this context. Current definitions of the nature of scientific research depend on the assumption that empirical data alone can never prove a hypothesis, but only disprove it. Thus, the hypothetico-deductive method can be seen as a way of systemizing the search for falsifying evidence about our hypotheses. This is in direct opposition to the pure empiricist or logical positivist position which was heretofore dominant in scientific research. Such approaches instead considered empirical observations not just to be sufficient proof alone, but in fact that all other types of evidence (e.g. rational thought and logical deduction) were of no use in knowledge generation.

If one considers the hypothetico-deductive method to be the foundation of scientific research within marketing (cf. Hunt 1983), then the place of data mining is worthy of some discussion. Drawing from the nature of data mining as defined above, it would seem that data mining may have some use in a scientific knowledge creation process, but that this use would be limited. More specifically, the data mining approach is fundamentally an *inductive* one, in which a data set is interrogated in an exploratory fashion, in the hope of turning up something of interest. Surely, if one is working within a hypothetico-deductive paradigm of knowledge generation, any findings from a purely data mining study could never be considered as actual knowledge. Instead, such findings should be treated as *knowledge claims*, and used to help generate explicit theoretical hypotheses which can then be tested in newly-designed empirical studies, which collect new data. Only when these theoretical hypotheses fail to be falsified with additional empirical work can the knowledge claim then be considered as knowledge.

It is certainly the case that the use of theory to help explain these empirical discoveries is also worthy of significant discussion. Or in other words whether purely empirical results from a data mining study *only* are enough to justify hypothesis generation and further testing. However, a full discussion of this is outside the present scope, given the short space available. Even so, our short answer to this question would be that the emergent inductive empirical relations would need theoretical deductive explanation as well, in order to justify them as testable hypotheses in a scientific context. In this sense, empirical data mining results are but a single strand of evidence or justification for a hypothesis, rather than sufficient alone.

It is important to make clear however that this position refers to the data mining *method*, not to any particular technique or algorithm. Certainly, many algorithms commonly of use in data mining applications can and have been usefully employed in a deductive manner in scientific studies – such as multiple regression, principle components analysis, clustering, classification trees, and the like. However, the critical issue is not one of technique, but of the underlying epistemological position of the task employing the technique.

4 Data Mining in a Practical Context

Notwithstanding the above, it is not the intention of this essay to decry the use of data mining approaches in general, since they are clearly of major potential use in both commercial and some scientific applications. Beginning with commercial applications, it is clear that marketing-focused firms can employ data mining methods to interrogate the huge customer databases which they routinely generate. Such work is common, and can include such tasks as market segmentation, customer profiling, and auditing. For example, it is well-known that Google utilizes data mining methods to predict which advertisements are best matched to which websites. Thus, without any actual knowledge (as we would term it) of *why*, Google can predict an advertisement's likely success depending on how it is matched (Anderson 2008). Considering the terabytes of data Google collects constantly, such methods are likely to be the most effective way to predict success. Yet the question of whether raw prediction is actually scientific knowledge is moot in this and most other practical situations. As most business researchers know, few business organizations are particularly interested in *explaining* the theory of why things are related, but only in *predicting* what will happen if variables are changed. In other words, what 'levers' can be manipulated to improve performance? Data mining is an ideal tool for this task.

However, raw data mining is also of significant use in many scientific fields outside of the business or social sciences. For example, sciences such as biochemistry work with massive data sets in many cases. In these situations, data mining can be usefully employed in uncovering relationships between for example genetic polymorphisms and the prevalence of disease. There is a significant difference between such scientific work and a typical business research study. In such biosciences, researchers often work within a very exploratory, or descriptive, context, and they also often work within contexts without large amounts of competing or unmeasured variables. For example, if one is working within a database of the human genome, then this is *all* the data. Conversely, if one is working within a database of customer characteristics, there may be many hundreds of unmeasured variables of interest, and any description of that database will be able to incorporate but a tiny subset of the possible explanatory variables. Even so – as is the case within neuroscientific research at present – purely exploratory or descriptive approaches (which data mining is useful for) must eventually be superseded by theory-driven hypothetico-deductive approaches (e.g. Senior and Russell, 2000).

5 Discussion and Conclusions

The aim of this invited essay was to explore the implications and uses of data mining in the context of scientific knowledge generation for marketing and business research. In doing so, we defined both data mining and scientific knowledge. Importantly, data mining was defined as a method of exploration, not as a set of particular tools or algorithms. Knowledge was defined as distinct from a knowledge claim, in that a knowledge claim had not been subject to a hypothetico-deductive attempt at falsification. The importance of this distinction is that in most cases one cannot claim data mining approaches as tools of knowledge generation in a scientific context. At best, they are highly useful for the generation of hypotheses from data sets, which may previously have been unexplored. In this way, it is interesting to draw parallels with qualitative research approaches such as grounded theory (e.g. Glaser and Strauss, 1967). Glaser's approach to grounded theory instructs that no appreciation of prior theories should be made before either collecting or analyzing data, in order to 'let the data speak' (Glaser 1992). Any argument in favor of data mining as a knowledge generation tool must therefore look to such approaches as justification. However, it is our view that – while such methods can result in truly original findings which would be unlikely to emerge from any other method – those findings should always be considered preliminary knowledge claims until further confirmatory testing.

This is because, without *a priori* theoretical expectations (i.e. hypotheses), one is always at risk of over-interpreting the data. In fact, many data mining techniques use the term 'overfitting' to refer to this situation (Larose, 2005). In such an instance, one's findings are essentially an artifact of the data set, and may not bear relation to the rest of the world. In other words, the training set is explained increasingly exactly, but the results are increasingly less generalizable to other data. Of course, if your data set is all of the relevant data in the world (as is the case in some scientific contexts), this is not a problem. However in most situations, and particularly within the business and social research contexts, our data contains only a subset of the available data, in terms of both subjects and possible variables. Overfitting in this case results in findings which are likely to have low external validity.

Thus, we urge business and social researchers to exercise caution in the application of data mining in scientific knowledge generation. Even so, this is not to say that we consider it to be of no use at all. Just as many other exploratory techniques are of use in the hypothetico-deductive cycle, data mining may provide extremely valuable results in the context of the knowledge generation process as a whole. However, researchers would be well advised to avoid presenting the findings of pure data mining as anything other than preliminary or exploratory research (although of course this may be of significant use in many cases). Although we did not specifically discuss it here, we would also urge researchers to make sure they are knowledgeable in the appropriate use of various data mining algorithms, rather than using them as a 'black box' between their data and results. Such an approach runs the risk of being characterized as 'data-driven' and therefore should be given little time at top-level journals. In this way, we also urge editors and reviewers at journals to think carefully about the actual contribution of such studies, despite their often complex and impressive technical content.

In conclusion, it is our view that explanatory theory is the key contribution of scientific research, and this should not be forgotten. Theory allows us to explain situations and contexts beyond our data, in contrast to pure prediction, which may have no real explanatory value whatsoever. While it may be of interest in many situations, it should not be characterized as scientific knowledge.

References

Adriaans, P., Zantinge, D.: Data Mining. Addison-Wesley, Harlow, England (1996)

Anderson, C.: The End of Science: The Data Deluge Makes the Scientific Method Obsolete. In: WIRED, vol. 16 (7), pp. 108–109 (2008)

Buzzell, R.D.: Is Marketing a Science? Harvard Business Review 41(1), 32–40 (1963)

Brown, S.: Postmodern Marketing, Thompson, London (1995)

Feyerabend, P.K.: Against Method, 3rd edn. Verso, London (1993)

Glaser, B.G.: Basics of Grounded Theory Analysis. Sociology Press, Mill Valley (1992)

Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine, Chicago (1967)

Hand, D., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press, Cambridge (2001)

Hunt, S.D.: Marketing Theory: The Philosophy of Marketing Science. Irwin, Homewood, IL (1983)

Hunt, S.D.: Controversy in Marketing Theory: For Reason, Realism, Truth and Objectivity. M.E. Sharpe, Armonk (2003)

Hunt, S.D.: Truth in Marketing Theory and Research. Journal of Marketing 54, 1–15 (1990)

Lee, N., Lings, I.: Doing Business Research. Sage, London (2008)

Senior, C., Russell, T.: Cognitive neuroscience for the 21st century. Trends in Cognitive Science 4, 444–445 (2000)

Observations on Soft Computing in Marketing

David W. Stewart

Dean of and Professor of Management and Marketing,
A. Gary Anderson Graduate School of Management,
University of California, Riverside, California, USA

Marketing managers make use of a variety of computer-based systems to aid decision- making. Some of these models would be considered "hard" models in the sense that they are based on quantitative data, usually historical data related to some type of market response and some empirically derived functional form of the relationships among actions in the market and market response (Hanssens, Parsons and Schultz 2008). Such models have been widely employed to decisions involving pricing and promotion, advertising scheduling and response, product design, and sales call scheduling, among others (Lilien and Rangaswamy 2006). These models, while very useful, require very rich data, as well strong assumptions about the generalizability of historical data to future events. These are assumptions likely to be less and less valid in an increasingly volatile world that includes regular introduction of new means of communications and product/service distribution, as well as new product and service innovations.

Quantitative models in marketing are also limited by two other factors. First, it is often not possible to obtain certain types of data that would be desirable for making marketing decision - for example, experimental data on customer response to different product characteristics, advertising levels, product prices, etc. Although some data may be available from interviews, test market experiments, and the like, it is often necessary to supplement them with the judgment of experienced marketing managers. The second limitation is related to the complexity of many marketing factors, many of which are unquantifiable. The decision environment may simply to be too complex to develop a quantitative model that captures all of the relevant parameters.

As a result of these limitations marketers have sought to build models that not include hard quantitative components, but also soft models that incorporate managerial judgment. These models are not "expert systems" in the classic sense of the term because they do not capture a set of replicable rules that would be characteristic of the use of artificial intelligence (Giarratano and Riley 2004, Little and Lodish 1981). Rather, decision calculus attempts to capture the subjective judgments, and hence, the experience of a decision maker within the context of a predictive model.

For at least four decades the marketing literature has documented the development and commercial use of models that incorporate the judgments of experienced managers. Models have been published which assist managers in making decisions about a wide range of marketing variables, including prices, couponing

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 17–19.
springerlink.com

efforts, advertising spending, media selection, and sales force scheduling. Many of these systems require that managerial expertise be used to set parameters and even to specify model forms. The way in which these models use judgmental data, the nature of their construction, and the purposes to which they are put differ in important ways from those of the typical expert system.

Montgomery and Weinberg (1973) describe the typical decision calculus modeling exercise as:

- Managers first verbalize their implicit model of the situation or issue of interest, specifying factors that influence a criterion variable of interest and the relationships of factors to one another and to the criterion variable;
- This verbal model is translated to a formal mathematical model. In most applications the response function has two components, a current component and a delayed (or lagged) component. Lilien and Kotler (1983) provide a useful overview of typical forms these models take in marketing applications;
- Parameters associated with the mathematical model are estimated; and
- An interactive procedure is implemented that allows managers to examine the influence of variations of particular factors on the criterion. By examining the model outputs obtained by changing model inputs, managers can examine alternative decisions and determine the sensitivity of the criterion to changes in particular input factors;

Obviously, the development of a useful decision support tool is a primary benefit of model building involving decision calculus. Little and Lodish (1981) argue that numerous additional benefits also accrue from the model building exercise. Among these additional benefits are:

- Model building facilitates the collection of data;
- It makes assumptions explicit;
- It identifies areas of disagreement and the nature of that disagreement; and
- It helps identify needs for information that have a high payoff potential.

Decision calculus models have a great deal in common with soft computer, though soft computing clearly brings a broader array of tools and methods to the task of informing decision-making. Soft computing also takes advantage of the enormous increase in computational power and the new techniques in biological computation that have emerged since the development of decision calculus models (Abraham, Das, and Roy 2007). Nevertheless, there is a common philosophical and methodological history that unites these different types of models. The underlying notion is that complex problems can be solved at a molar level as an alternative computational models that seek to fit quantitative models at a more micro level.

Although it has demonstrate its utility in a host of venues, soft computing has yet to demonstrate its utility in solving practical marketing problems. It seems only a matter of time before it does so given the complex data sets now available to marketing organizations. It is also likely that these tools will carry benefits similar to those already demonstrated for decision calculus models in marketing.

References

Abraham, A., Das, S., Roy, S.: Swarm Intelligence Algorithms for Data Clustering. In: Maimon, O., Rokach, L. (eds.) Soft Computing for Knowledge Discovery and Data Mining, pp. 279–313. Springer, New York (2007)

Giarratano, J.C., Riley, G.: Expert Systems, Principles and Programming, 4th edn. PWS Publishing, New York (2004)

Hanssens, D.M., Parsons, L.J., Schultz, R.L.: Market Response Models: Econometric and Time Series Analysis, 2nd edn. Springer, New York (2008)

Lilien, G., Kotler, P.: Marketing Decision Making: A Model-building Approach. Harper and Row, New York (1983)

Lilien, G.L., Rangaswamy, A.: Marketing Decision Support Models. In: Grover, R., Vriens, M. (eds.) The Handbook of Marketing Research, pp. 230–254. Sage, Thousand Oaks (2006)

Little, J.D.C., Lodish, L.M.: Judgment-based marketing decision models: Problems and possible solutions/commentary on Judgment-Based Marketing Decision Models. Journal of Marketing 45(4), 13–40 (1981)

Montgomery, D., Weinberg, C.: Modeling Marketing Phenomena: A Managerial Perspective. Journal of Contemporary Business (Autumn), 17–43 (1973)

Soft Computing Methods in Marketing: Phenomena and Management Problems

John Roberts

Professor of Marketing
College of Business and Economics,
Australian National University (Australia)
and
London Business School (UK)
e-mail: john.roberts@anu.edu.au

1 Introduction

Soft computing techniques gained popularity in the 1990s for highly complex problems in science and engineering (e.g., Jang et al. 1997). Since then, they have slowly been making their way into management disciplines (Mitra et al. 2002). In order to understand the potential of these methods in marketing, it is useful to have a framework with which to analyze how analytical methods can provide insight to marketing problems.

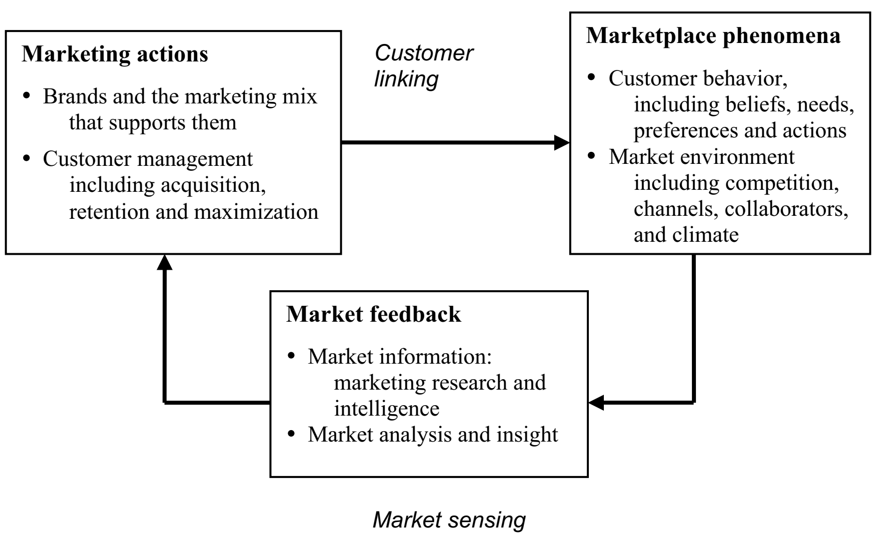


Fig. 1 A Model of the Market Decision Making Process

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 21–26.
springerlink.com

Marketing may be regarded as harnessing the resources of the organization to address the needs of its target customers, given the marketplace environment in which it competes (the top arrow in Figure 1). George Day calls this process "customer linking" (Day 1994). Actions in the top right box can be analyzed either from an internal perspective in terms of the products and services of the organization and the marketing mix that supports them, or externally in terms of its customers: how it attracts, retains and maximizes the value it provides to and captures from them. However, in order to focus the organization's actions, an understanding of the environment is necessary, and feedback from the marketplace helps the manager better target her actions to where they will be most effective (the bottom arrows in Figure 1). Day calls this function "market sensing." Market sensing has the dual elements of gathering data from the market and transforming those data into insights for action, by using suitable analytical tools.

Soft computing tools form one weapon in the marketing analyst's toolkit to provide that insight. In understanding the potential (and limitations) of soft computing tools, it is useful to analyze this environment. This chapter specifically examines the management actions for which the suite of techniques is well-suited, and the phenomena on which it can throw insight (the two top boxes in Figure 1). Details of the techniques of soft computing that belong to the bottom box are covered elsewhere in this volume.

2 Marketplace Phenomena

Soft computing has particular strengths in the case of large databases and complex phenomena. To understand where these are most likely to occur it is useful to decompose the consumer decision. One traditional method of analyzing consumer decisions is by use of Lavidge and Steiner (1961)'s Hierarchy of Effects model (also known as the demand funnel and a variety of other names). This model is illustrated in Figure 2:

One major driver of complexity of models (in terms of number of observations, parameters, and interactions between variables) is that of heterogeneity. When we have to allow for differences between individual consumers (or a large number of groups of consumers), the tractability of traditional models is likely to come under threat. In marketing, in reference to Figure 2, we do see situations where consumers vary in their proclivity to enter the category (need arousal). Both the diffusion of innovation and hazard rate literatures address this problem (for example, see Roberts and Lattin 2000). Similarly, Fotheringham (1988) has used a fuzzy set approach to modeling consideration set membership in the information search stage to probabilistically describe whether a brand will be evoked. Next, it is in the modeling of beliefs (perceptions), preferences, and choice that probabilistic representations of consumer decision processes have really come into their own, with Hierarchical Bayes now used as a standard approach to modeling consumer differences (see, Rossi and Allenby 2003 for a review). Finally, as we move from the acquisition stages to the retention and value maximization ones, customer satisfaction models have used a variety of soft computing techniques to identify individual or segment-level threats and opportunities.

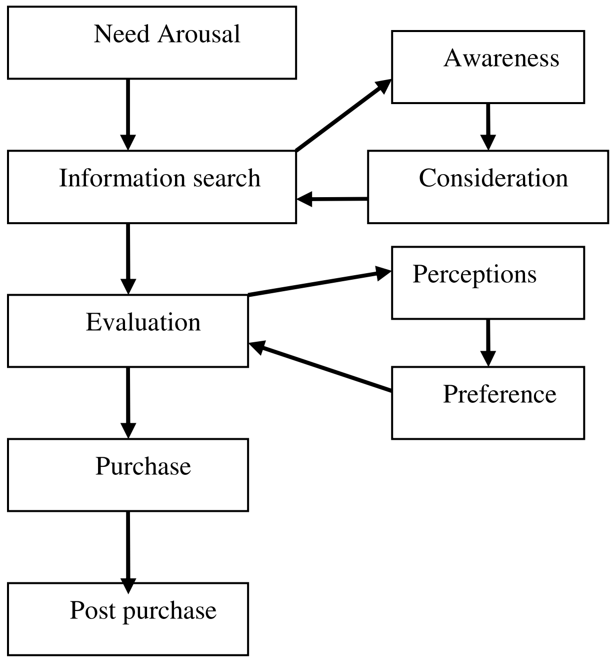


Fig. 2 Lavidge and Steiner (1961)'s Hierarchy of Effects Model

While soft computing has much to recommend it in each stage of the hierarchy of effects, where it has the most to offer is when these complexities are compounded. That is, while we can encounter large scale problems in each of these areas, it is the convolution of these large scale problems that particularly lends itself to the approach. Typical examples of such multi-level problems include the following:

- *Multidimensional consumer differences*. We may have to segment on more than one basis (either within or between the levels of Figure 2). For example, within levels we may need to segment on the application to which the consumer is putting a service and her socio-economic profile. Between levels we may have to segment on the susceptibility of a consumer to an innovation at the need arousal level and the firm's competitive capability at the purchase level.
- *Multiple consumer purchases*. The consumer may make multiple purchases within the category (suggesting a need to study share of wallet) or across categories (requiring estimation of cross-selling potential across multiple products).
- *Interactions between consumers*. Consumer networks may be critical, necessitating a study of order of magnitude of n^2 with respect to customers, rather than just n (where n is the number of customers).

- *Interactions between members of the market environment.* Interactions between members of the channel, collaborators, competitors, and other groups (such as government regulators) may further compound the complexity of the problem.

3 Management Problems

While multidimensional differences may exist at the level of the consumer or in the climate, they may not require complex models on the part of the manager to understand that variance. Before advocating a move to complex computing and modeling approaches, we must understand where just looking at the mean of distributions of heterogeneity is not going to lead to appropriate decisions, relative to a study of the entire distribution (or some middle approach such as looking at variances).

Sometimes demand side factors alone may lead to a requirement to study the distribution of consumer tastes. The fallacy of averages in marketing is often illustrated by the fact that some people like iced tea, while others like hot tea. The fallacy of averages would suggest (incorrectly) that generally people like their tea lukewarm.

In other situations, it is the context of managerial decision making in Figure 1 that makes complexity in the marketplace phenomena intractable to simplification and the use of means. The most obvious example of when modeling averages is problematic is when asymmetric loss functions exist: the benefit to the manager of upside error is not equal to the loss of downside. This will occur in a variety of situations. With lumpy investment decisions based on forecasts of consumer demand, over- and under-forecasts are likely to lead to very different consequences. Over-estimating demand is likely to lead to idle equipment and capital, while under-estimation will cause foregone contribution and possible customer dissatisfaction. Risk aversion on the part of the manager is another factor that will lead to asymmetric loss functions (Wehrung and Maccrimmon 1986). Finally, the presence of multiple decisions will lead to a requirement to study the whole distribution of customer outcomes, not just the mean. For example, in the ready to eat cereal and snacks market, Kellogg's website lists 29 sub-brands[1]. Obviously, there are major interactions between these various sub-brands, and category optimization across them is an extremely complex problem. It is impossible to address without reference to the total distribution of beliefs, preferences and behaviors. Averages will not enable to answers to such portfolio management problems.

4 Summary

Soft computing techniques have a number of advantages. Primarily, their ability to handle complex phenomena means that restrictive and potentially unrealistic assumptions do not need to be imposed on marketing problems. Balanced against

---

[1] `http://www2.kelloggs.com/brand/brand.aspx?brand=2`

this advantage is the loss of simplicity and parsimony, and this may incur associated costs of a loss of transparency and robustness. The mix of situations that favor soft computing techniques is increasing for a variety of reasons which may be understood by reference to Figure 1. Perhaps the primary drivers are trends in the market feedback tools available. Digital data capture means that large data sets are becoming available, enabling the modeling (and estimation) of consumer behavior in considerably finer detail than was previously possible. Computing power has obviously increased, and Moore's law now enables calculations that would have been impossible, excessively onerous, or time intractable to be readily available. However, developments in both marketplace phenomena and managerial actions have also increased the potential application of soft computing approaches. Markets have become increasingly fragmented with the advent of highly targeted media and mass customization of products. For example, the U.K.'s largest retailer, Tesco, addresses over four million segments (Humby et al. 2008). In the top left box of Figure 1, managers have become increasingly sophisticated, with many firms employing sophisticated data mining techniques to address their customers. The emergence of specialist consulting and software firms (such as salesforce.com, dunhumby, and SAP) to support them has accelerated adoption in this area. Digitization has also increased the ability of the manager to experiment with a variety of strategies, leading to much richer mental models of the market, favoring the use of soft computing methods.

Soft computing has the ability to lead us along paths that, as Keynes said, are more likely to be "vaguely right" rather than "precisely wrong" (e.g., Chick 1998). It is important that the migration path towards its use does not come at the cost of transparency or credibility. One way to ensure that this does not occur is to apply the techniques to environments that need the explanatory power they afford, and which influence management decisions for which the distribution of outcomes is critical, as well as the mean.

References

Chick, V.: On Knowing One's Place: The Role of Formalism in Economics. The Economic Journal 108(451), 1859–1869 (1998)

Day, G.S.: The Capabilities of Market-Driven Organizations. Journal of Marketing 58(4), 37–52 (1994)

Stewart, F.A.: Consumer Store Choice and Choice Set Definition. Marketing Science 7(3) (Summer), 299–310 (1988)

Humby, C., Hunt, T., Phillips, T.: Scoring Points: How Tesco Continues to Win Customer Loyalty. Kogan Page, London (2008)

Jang, J.-S.R., Sun, C.-T., Mizutani, E.: Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence. Matlab Curriculum Series, Boston (1997)

Lavidge, R.J., Steiner, G.A.: A Model for Predictive Measurements of Advertising Effectiveness. Journal of Marketing 25(6), 59–62 (1961)

Mitra, S., Pal, S.K., Mitra, P.: Data Mining in Soft Computing Framework: A Survey. IEEE Transactions On Neural Networks 13(1), 3–14 (2002)

Roberts, J.H., Lattin, J.M.: Disaggregate Level Diffusion Models. In: Mahajan, V., Muller, E., Wind, Y. (eds.) New Product Diffusion Models, pp. 207–236. Kluwer Academic Publishers, Norwell (2000)

Rossi, P.E., Greg, M.: Bayesian Statistics and Marketing. Marketing Science 22(3), 304–328 (Summer, 2003)

Wehrung, D., Maccrimmon, K.R.: Taking Risks. The Free Press, New York (1986)

User-Generated Content: The "Voice of the Customer" in the 21st Century*

Eric T. Bradlow

K.P. Chao Professor, Professor of Marketing, Statistics, and Education,
Editor-in-Chief of Marketing Science, and Co-Director of the Wharton
Interactive Media Initiative,
University of Pennsylvania, Pennsylvania, USA

1 Introduction

It doesn't take an academic paper to point out the prominence that companies like Facebook, MySpace, YouTube, etc. have had on our popular culture today. Many see it as an efficient communication mechanism (Web 2.0 if you will) in comparison to email and static content postings which now, remarkably only 15 years later after the internet 'launch', some people see as "old school". In some sense, Andy Warhol's prediction of "15 minutes fame" for each and any one of us can now be "self-generated" through our own hard work and user-generated content. Thus, with all of this impact (societally) as backdrop, where does it leave many of us, as academics? The answer, and I hope this essay provides some impetus for that, is not on the sidelines.

As a good sign, recently a joint call for funded research proposals between the Wharton Interactive Media Initiative (WIMI, www.whartoninteractive.com) and the Marketing Science Institute (MSI, www.msi.org) on the impact and modeling of user-generated content (UGC) generated an overwhelming response with over 50 submissions. Even better news was that these submissions were broad in their scope. As a non-random sampling of ideas generated, consider the following.

- What is the impact of UGC on customer satisfaction and stock prices? Is there information contained in UGC that can help predict supra-normal returns? This can be considered, if you will, an update to the work of Fornell and colleagues (e.g. Anderson et. al, 1994), but now one based on UGC.
- How does the quantity and valence of UGC impact the diffusion (Bass, 1969) of new products? Note, that while the "scraping" of quantity information for the 'amount' of UGC may be somewhat simple, the valence of that information ('quality') is less so. While this may make the

* Financial support for this work was provided by the Wharton Interactive Media Initiative (www.whartoninteractive.com).

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 27–29.
springerlink.com

timid shy away, this is one example of an opportunity where data mining and marketing scientists can partner together.

- Conjoint Analysis (Green and Rao, 1971) has long been a mainstay of marketing researchers as a method to understand consumer preferences for product features. But, how does one know that one has the right attributes in the first place – i.e. the classic "garbage in garbage out"? In a recent paper, Lee and Bradlow (2009) utilize feature extraction and clustering techniques to discover attributes that may be left off via standard focus group or managerial judgment methods.

While these three examples are different in spirit, they all share a common theme, what can really be extracted from UGC that would aid decision-makers? In the next section, I discuss some thoughts, supportively encouraging and otherwise.

2 Marketing Scientists Should Care about UGC or Should They?

Forecasting is big business. The ability to predict consumer's actions in the future allows marketing interventions such as targeted pricing (Rossi et al, 1996), target promotions (Shaffer and Zhang, 2002), and the like. The promise that UGC can improve these marketing levers is certainly one reason that firms are investing heavily in data warehouses that can store this information and without this does UGC really have a "business future"?

While it might seem tautological that UGC can help predict "who wants what" and "when", it becomes less obvious when one conditions on past behavioral measures such as purchasing, visitation, etc… (Moe and Fader, 2004). In addition, what is the cost of keeping UGC at the individual-level? Thus, a new stream of research on data minimization methods, i.e. What is the least amount of information that needs to be kept for accurate forecasting? (Musalem et. al, 2006 and Fader et al, 2009) will soon, I believe, be at the forefront of managerial importance. Fear, created by the loss of not keeping something that may somehow, someday be useful, will be replaced by the guiding principles of parsimony and sufficiency (in the statistical sense and otherwise).

Or, let us consider another example of UGC, viral spreading through social networks (Stephen and Lehmann, 2009). Does having content that users provide, knowing who their friends are, and how and to what extent they are sharing that information provide increased ability for targeted advertising? Does it provide the ability to predict "customer engagement" which can include pageviews, number of visits (Sismeiro and Bucklin, 2004), use of applications (now very popular on websites) and a firm's ability to monetize it? These are open empirical questions which marketing scientists likely can not answer alone because of the widespread data collection that is necessary. We conclude next with a call for Data Mining and Marketing Science to converge.

3 Marketing Scientists and Data Mining Experts Need Each Other Now More Than Ever

With all of the data that is abundant today, theory is now needed more than ever. Yes, I will say it again, theory is need now more than ever despite the belief of some that the massive amounts of data available today might make "brute empiricism" a solution to many problems. Without theory, all we are left with is exploration and sometimes massively unguided at that. Through the partnering of data mining/KDD experts in data collection and theory, and marketing scientists who can help link that data and theory to practice, UGC presents the next great horizon for "practical empiricism". While the lowest hanging fruit might be including UGC covariates as predictors in models of behavior, hopefully our scientific efforts will move beyond that towards an understanding of its endogenous formation (the whys of people's creation of it) and also an understanding of when it is truly insightful.

References

Anderson, E.W., Fornell, C., Lehmann, D.R.: Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. Journal of Marketing 58(3), 53–66 (1994)

Bass, F.M.: A New Product Growth Model for Consumer Durables. Management Science 15, 215–227 (1969)

Fader, P.S., Zheng, Z., Padmanabhan, B.: Inferring Competitive Measures from Aggregate Data: Information Sharing Using Stochastic Models, Wharton School Working Paper (2009)

Green, P.E., Rao, V.R.: Conjoint measurement for quantifying judgmental data. Journal of Marketing Research 8, 355–363 (1971)

Griffin, A., Hauser, J.R.: The Voice of the Customer. Marketing Science 12(1), 1–27 (Winter 1993)

Lee, T.Y., Bradlow, E.T.: Automatic Construction of Conjoint Attributes and Levels From Online Customer Reviews, Wharton School Working Paper (2009)

Moe, W.W., Fader, P.S.: Dynamic Conversion Behavior at E-Commerce Sites. Management Science 50(3), 326–335 (2004)

Musalem, A., Bradlow, E.T., Raju, J.: Bayesian Estimation of Random-Coefficients Choice Models using Aggregate Data. Journal of Applied Econometrics (2006) (to appear)

Rossi, P.E., McCulloch, R.E., Allenby, G.M.: The Value of Purchase History Data in Target Marketing. Marketing Science 15, 321–340 (1996)

Shaffer, G., Zhang, Z.J.: Competitive One-to-One Promotions. Management Science 48(9), 1143–1160 (2002)

Sismeiro, C., Bucklin, R.E.: Modeling Purchase Behavior at an E-Commerce Web Site: A Task Completion Approach. Journal of Marketing Research, 306–323 (August 2004)

Stephen, A.T., Lehmann, D.R.: Is Anyone Listening? Modeling the Impact of Word-of-Mouth at the Individual Level, Columbia University Working Paper (2009)

Fuzzy Networks

Dawn Iacobucci

E. Bronson Ingram Professor in Marketing,
Owen Graduate School of Management, Vanderbilt University,
Nashville, TN, USA

Knowledge discovery and fuzzy logic have great potential for social network models. Networks are currently extraordinarily popular, and while it's fun to work in an area that people find interesting, there is such a thing as a topic being too popular. Networks are too popular in the sense that they are not widely understood by users, hence they are thought to be the new, new thing, capable of answering all questions, from "Will my brand's presence on Facebook help its equity?" to " Will a network bring peace to the Middle East?" Fuzzy logic should help new users proceed from naïve enthusiasm to thoughtful application, because fuzzification embraces approximation; huge questions cannot be answered with simple, precise estimates, but a fuzzy approach can put the inquirer in the rough vicinity of an answer (Martínez-López and Casillas, 2008).

This essay considers three popular uses of networks: word-of-mouth, brand communities, and recommendation agents, and the application of fuzziness in each realm. The first of these, **word-of-mouth**, has long been recognized as a powerful marketing force. Marketers routinely consider the diffusion of a new product or idea into and throughout the marketplace using models that posit the mechanism of customers informing each other. Those who adopt early are thought to influence the choices of those who adopt later. Hence, currently, the marketing question that seems to be the "holy grail" takes this form, "How can networks help me identify my influential customers?"

This question is remarkably easy to answer via social network techniques. Actors in the network are assessed for their volume and strength of interconnections. Actors that are more interconnected with others are said to be "central" compared to more "peripheral" in the network. Depending on what the network ties reflect, centrality may manifest an actor's importance, power, communication access, and the like. In a word-of-mouth network such as those sought in diffusion studies, these central players are the very essence of an influential opinion leader.

There are several criteria to assess centrality, and as a result, indices abound (Knoke and Yang, 2007). For example, some measures reflect the sheer number of connections, or their weighted strengths or frequencies of connections. Other indices capture the extent to which actors are key in bridging multiple parts of the network map. Still other centrality measures reflect a sense of closeness among the network players, as in the number of steps between pairs of actors, or their "degrees of separation." Nevertheless, the centrality indices share the property

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 31–34.
springerlink.com

that each captures the extent to which an actor has more connections to others, or stronger, or more frequently activated ties to others. These ties may be primarily inbound, and then the actor is said to be popular. The ties may be predominately outward bound, and then the actor is said to be expansive (e.g., extroverted).

So where does fuzziness come in? Marketers understand that just because a customer engages in high activity, whether they claim many friends on a mobile phone plan, or are a frequent blogger, or actively recruit many friends on their Facebook page, it does not necessarily translate into their being an influential. But for all practical purposes, isn't this status "close"? If someone posts to a blog, and some readers dismiss the posting as being uninformed, the marketer may be disappointed that this blogger isn't as influential as first thought. Yet given their blogging volume and sheer statistical incidence, would it not likely be the case that their postings would impact some readers? Their blogging activity is presumably motivated by high customer involvement, thus may convey credibility, or at least passion. Thus, managing brand perceptions in the eyes of these frequent posters, frequent frienders, or frequent callers would be a social marketing activity whose result would be sufficiently close to the strategic aims of identifying and leveraging the influential customer. It is close enough.

Brand communities are a contemporary marketing and social network phenomenon. Brand communities exist in real life and frequently online. People gather to share and learn and simply enjoy like-minded others.

Some scholars claim they comprise a marketing or business strategy. I disagree. Marketing managers can try to launch such a community, and they can certainly insert marketing materials (brands, services, information) into the community in the hopes of effective persuasion. However, most authentic brand communities are grass-roots efforts, created by the love of a common element. For example, Harley riders got together long before some marketer coined the term, "brand community."

Marketing managers of brands that create such buzz and fondness can only hope to leverage the resulting community. Marketing managers of brands that create a collective yawn could persevere to eternity and not be successful in creating a community.

When brand communities do exist, marketing phenomena such as diffusion can occur relatively quickly for two reasons. First, while the brand community can be rather large and its membership absolutely informal (e.g., no list of such actors exists), the community is still better defined and smaller to manage than the amorphous set of customers sought in the first application (of finding influentials among all customers). The marketer needs simply to be present at the auto / bike / beer / quilting event or website, and the community itself will take care of the information management, if it perceives value in the market offering.

In addition, brand communities are largely democratic. In social network parlance, this egalitarian status shows itself distinctively in highly reciprocal or mutual ties. The ties create a clique of relatively highly interconnected actors comprising a subgroup within the network. Unlike the hierarchical relations between an early adopter exerting influence over a later adopter, customer elements in brand communities share mutual respect and communication. In such structures, those actors who extend ties in great volume tend to also receive them proportionally frequently.

There is a lot to be learned from the patterns of social networks of brand communities—how do the Saturn and Harley communities compare? How do communities of brands whose customers are predominately women compare with those for men's brands? How do Latin American constituted communities compare with networks for British brands and customers? And of course, is there a structural network distinction between communities of highly profitable brands and those that are less so?The very egalitarian nature of the brand community is related to the fuzziness principle for this social network phenomenon. Specifically, while it is true that members of a brand community are not created equal in terms of facility and likelihood of becoming a brand champion, it is not important. The marketing manager's actions can be somewhat imprecise. If the marketer gets the brands and communications into the hands of the brand champion, diffusion will be rapid within the community. But even if the marketer misses, and the materials reach a proxy actor, doing so will eventually affect the same result, with the simple delay of the proxy communicating to the real community leaders. It is close enough.

Finally, the third marketing phenomenon that can benefit from a fuzzy application of social networks is that of **recommendation agents** (Iacobucci, Arabie and Bodapati, 2000). Current data-based algorithms for suggesting new products to purchase or new hyperlinks to follow for related articles to read are based on clustering techniques. Social networks models can contribute to this pursuit in lending the concept and techniques of structural equivalence. Two actors are said to be structurally equivalent if they share the same pattern of ties to others. If we are willing to fuzz up this criterion, then two customers would be said to be stochastically equivalent if they share similar searches, purchases, or preference ratings.

This third application is different from the first two in that they had been true social networks—the entities in a word-of-mouth network or in a brand community are predominately human, and the ties between these actors, social, be they communication links or ties of liking, respect, sharing, etc. The recommendation agency problem is contextualized in a mixed network of ties among entities that are human, electronic, tangible goods and brands and intangible services. The nonhuman actors may be said to be connected to the extent they are similar, bundled, complementary, etc. The human actors may be interconnected via the usual social ties, but may not be; the recommendation system in Amazon uses no friending patterns, but that in Netflix allows for others to make direct suggestions to people they know.

For this phenomenon, like seeking influentials and seeding brand communities, fuzzy networks should suffice to yield good marketing results. Browsing book titles, music CDs, or movie DVDs in stores is easier than doing so online, yet a model-derived suggestion can put the customer in the ball-park of a new set of titles that may be of interest. Amazon's statistics indicate that recommendations are not mindlessly embraced; e.g., the website offers indices such as, "After viewing this item, 45% of customers purchased it, whereas 23% of customers purchased this next item." When one item is viewed, and another is suggested, the suggested item need not be embraced for the tool to be approximately useful. The suggested item puts the user down a new search path, restarting a nonrandom walk. The user begins with a goal, which may be achieved immediately upon initial search, or it may be more optimally achieved upon corrected iteration based on inputs of recommendations resulting in successive approximations.

Thus, we see that the system's recommendation need not be "spot on." Rather, the system only needs to be close enough.

The study of network structures is a huge enterprise, and the application of networks to marketing and business phenomena is only in its infancy. These three examples were meant to be illustrative, drawing on popular and contemporary uses with which most readers will be familiar. Other examples at the nexus of fuzzy and networks will also benefit from the advantages of both. What was highlighted with the three exemplar fuzzy networks was the good news—that the application of networks does not need to be super precise for there to be great benefits realized.

References

Iacobucci, D., Arabie, P., Bodapati, A.: Recommendation Agents on the Internet. Journal of Interactive Marketing 14(3), 2–11 (2000)

Knoke, D., Yang, S.: Social Network Analysis, 2nd edn. Sage, Thousand Oaks (2007)

Martinez-Lopez, F.J., Casillas, J.: Marketing Intelligent Systems for Consumer Behavior Modeling by a Descriptive Induction Approach Based on Genetic Fuzzy Systems. Industrial Marketing Management Press (2008), doi:10.1016/j.indmarman.2008.02.003

KDD: Applying in Marketing Practice Using Point of Sale Information

Adilson Borges[1] and Barry J. Babin[2]

[1] Reims Management School
IRC Professor of Marketing
Reims Cedex, France
[2] Louisiana Tech University
Reims Management School
Max P. Watson, Jr. Professor of Business
Chair, Department of Marketing and Analysis
Louisiana Tech University, Ruston, LA, USA

1 Introduction

The dramatic increase in computing power that has emerged over the past two to three decades has revolutionized decision making in most business domains. In particular, point of sale data has been recorded by retailers now since the time of scanner technology. However, the great volumes of data overwhelmed conventional computation routines until more recently. Although the basic principles of data mining can be found in automatic interaction detection routines dating back to the 1960s, the computational limitations of those days prevented a thorough analysis of all the possible combinations of variables.

Today, KDD procedures are commonplace as data mining hardware and software provides power to search for patterns among practically any imaginable number of combinations. No longer do we talk about computer capacity in terms of megabytes, but more commonly, data storage is discussed in terms of terabytes (1000 gigabytes) or petabytes (1000 terabytes). Thus, although this may seem like an overwhelming amount of data and to be less theory driven than is appropriate for conventional multivariate data analysis procedures, it is clear that multivariate data analysis is applicable within soft computing and other data mining procedures (see Hair, Black, Babin and Anderson 2010). In particular, routines such as cluster analysis, multidimensional scaling, and factor analysis can be integrated into these routines to help establish patterns that can be validated and reduce the risk of identifying patterns based on randomly occurring generalizations.

Retail management, like all marketing efforts, deals with decision making under conditions of uncertainty. This paper describes a KDD application from a retail setting. Managers constantly seek the best arrangement of products to maximize the value experience for consumers and maximize sales revenues for the retailer. Can KDD procedures assist in the store layout question? Here is a description of one attempt to do so.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 35 – 41.
springerlink.com

This paper proposes a new grocery store layout based on the association among categories. We use the buying association measure to create a category correlation matrix and we apply the multidimensional scale technique to display the set of products in the store space. We will imply that the buying association, measured through the market basket analysis, is the best way to find product organization that are best suited to one stop shopping.

2 The Store Layout

Increasing space productivity represents a powerful truism in retailing: customers buy more when products are merchandised better. By careful planning of the store layout, retailers can encourage customers to flow through more shopping areas, and see a wider variety of merchandise (Levy and Weitz, 1998).

There are at least two layout approaches: the traditional grid layout and the consumption universe layout. The traditional approach consists in repeating the industrial logic implementation, which means putting products that share some functional characteristics or origins in the same area. So we will find the bakery area (with bread, cakes, biscuits, etc), the vegetable area (with carrots, beans, etc), and so on.

This traditional approach has been improved by the use of cross-elasticities, which should measure use association. Retailers have changed some categories and put more complementary in use items together. If a consumer wants take photos at a family party, s/he needs at least the camera and the film. In these cases, both products are complementary, because consumers need both at same time to achieve a specific goal (Walters, 1991).

The nature of the relationship among products could be twofold: the use association (UA) or the buying association (BA). UA is the relationship among two or more products that meet specific consumer need by their functional characteristics. We can classify the relationship among different categories by their uses: the products can be substitutes, independent and complementary (Henderson and Quandt, 1958 ; Walter, 1991). The BA is the relationship established by consumers through their transaction acts and it will be verified in the market basket. While UA is not a necessary condition for BA, because UA depends much more on the products functional characteristics, BA depends on buying and re-buying cycles as well as on store marketing efforts.

Despite improvements, the store remains organized in “product categories” as defined by the manufacturers or category buyers. This approach is company oriented and it fails to respond to the needs of the time pressured consumer. Some retailers are trying to move from this organization to something new, and are trying to become ¨consumer oriented¨ in their layout approach. Tesco has rethought their store layout with ¨plan-o-grams¨ to try to reflect local consumers needs (Shahidi, 2002). Other French retailers have used consumption universe layouts to make it easier for consumers to find their product in a more hedonic environment.

This approach allows supermarkets to cluster products around meaningful purchase opportunities related to use association. Instead of finding coffee in the beverage section, cheese in fresh cheese, ham in the meat section, and cornflakes in the cereal section, we could find all those products in the breakfast consumption universe. Other universes, such as the baby universe or tableware universe, propose the same scheme to cluster different product categories. It is too soon to foresee the financial results of such applications, but it shows, however, the retailer's desire to improve in store product display.

These new layout applications do not take the one stop shop phenomenon into account. In fact, this approach is based on the principle that conjoint use of products will unconditionally produce conjoint buying. The main problem with this rationale is that use association alone cannot be used to explain the associations carried out in the buying process (the market basket), because it fails to take buying time cycles into account. For example, bread and butter should be classified as occasional complements, and then they should be found in the same market basket (Walters, 1991). However, this could be not true, since the products have different buying and re-buying cycles. In that case, buying association may be weak, because bread is usually bought on a daily basis, and butter once every week or two.

On the other hand, 'independent products' don't have any use relationship, so they should not have any stable buying association. Meanwhile, Betancourt and Gautschi (1990) show that some products could be bought at the same time as a result of the store merchandising structure, store assortment, the marketing efforts and consumption cycles. So, the fact that two products are complementary is not a guarantee that those products will be present in the same market basket. In addition, some researchers have found that independent products have the same correlation intensity as complementary ones in the market baskets (Borges *et alli*, 2001). So, the store layout construction has to incorporate the market basket analysis to improve the one stop shopping experience. This allows retailers to cluster products around the consumer buying habits, and then to create a very strong appeal for today's busy consumers.

3 The Buying Association: A Way to Measure the Relationship among Products

The relationship between categories has always been articulated through their use, but this is not enough to explain conjoint presence in the market basket. These two kinds of relationships were clear for Balderston (1956), who presented it as (1) use complementary, if products are used together, and (2) buying complementary, if products are bought together.

BA can be computed from supermarket tickets, and indicates real consumer behavior (it is not based on consumers' declaration or intention). Loyalty cards and store scanners have produced a huge amount of data that is stored in data warehouses and analyzed by data mining techniques. Data Mining is regarded as the analysis step in the Knowledge Discovery in Databases (KDD) process, which is a "non-trivial process of extracting patterns from data that are useful, novel and comprehensive". In data mining, BA is considered as an association rule. This association rule is composed of an antecedent and consequence set : A $\Rightarrow$ B, where A is an antecedent and B a consequent; or A,B $\Rightarrow$ C, where there are two antecedents and one consequence (Fayyad *et alli*, 1996). The BA is calculated by the following formula:

$$\delta_{AB} = \frac{f(AB)}{f(A)}, \tag{1}$$

where f(AB) represents the conjoint frequency of both products A and B and f(A) represents the product A frequency in the database. This equation is similar to the conditional probability that could be written as $(A \cap B)/A$, given that A intersection B represents the market baskets where both products, A and B, are present at same time.

The buying association represents the percentages of consumers that buy product A and who also buy product B. It shows the relationship strength between products, considering *only the relationships carried out on buying behavior*. This can be represented as a percentage: a BA of 35% between coffee and laundry is interpreted as 35% of consumers have bought coffee also bought laundry in the same shopping trip.

In the same way that cross-elasticity is not symmetric, BA is also not symmetric. The BA_{FC} can be different from BA_{CF} (this relationship depends mainly on the category penetration rates over the total sales). Mathematically:

$$\forall\ F>C, \text{ so } (F \cap C)/F < (F \cap C)/C.$$

So, if A frequency is different from B frequency, then the relationship among those products will always be asymmetric. For example, "F" represents the film and "C" the camera. Suppose the condition F>C is confirmed, then the film has a larger penetration in the market baskets than camera. If this condition is satisfied, then $BA_{FC}<BA_{CF}$.

BA does not measure casual relationships, but only a correlated presence. It gives us a probability P (AB) to find a product B since we have found the product A in a market basket. Hence the sample in the data mining applications is usually large ($N \rightarrow \infty$), we can consider this measure as a conditional probability by Bernoulli's theorem. Therefore, we can state that $BA_{AB} = P_{B|A}$, which allows us to use the entire mathematical arsenal from conditional probability on BA (Hays, 1977).

4 Method and Results

The first step toward a store layout map is to measure the relationship among products. To do so, we use a two year database from three French supermarkets. The database has 1,700,000,000 transactions during the period. Each transaction has the consumer identification, the data, the EAN codes, the quantities of each product, and the total value. We have chosen 20 different categories to construct a correlation matrix, which are: water, bread, cornflakes, ham, detergent, cheese, pasta, butter, wine, sauce, mayonnaise, coffee, beverage, milk, yogurt, toothpaste, deodorant, shampoo, chips, beer. For space reasons, we will not show the correlation matrix, because this is not the main point of the article.

Once we have established the correlation matrix, we are able to calculate the spatial representation of these relationships through the multidimensional scaling technique (MDS). We have used data as distance and an asymmetric matrix to produce the results.

In order to use the correlation matrix as distances among categories, we have inversed the values by subtracting 1 from all values. So, if two products have a strong correlation (say 0,95) the proximities will be small (1-0,95 = 0,05), which means that those categories are similar and should be represented in a nearby space on the map.

To assess validity we have make individual MDS analyses for each store, and we found the same structural results for the map representation. We will not show each map here for reasons of space. However, the model stress is 0,41110 and the square correlation is 0,20971, which means that we have to be cautious about accepting this model. We represent all categories in the multidimensional space as showed in the Figure 1.

The first cluster in the Figure 1 is comprises cornflakes, ham, butter and cheese. These products are usually bought together and buyers who require one stop shopping might get a better experience from this categories layout. One can see in the cluster 1 a breakfast consumption universe. This can be true, even if we have other breakfast products, as coffee or milk (cluster 5) and bread (cluster 3) placed in other areas of the map. We have to stress that the buying association and the consumption universes are not incompatible. Common use products could be present in the same basket, since they have the same time cycle purchase.

Clusters 2 and 3 are of great interest in identifying the limits in the consumption universe approach. Cluster 2 shows beer, water and pasta being bought together in many shopping trips, as well as wine, beverage, bread and detergent (cluster 3). The use relationships among those products are less obvious than for cluster 1, even if they are bought together frequently.

Cluster 4 comprises chips, mayonnaise and sauce. These products could be displayed in the aperitif area, even if we would have expected to see sauce with pasta (cluster 2). It is important to say that sauce category is composed of pasta and tomato sauces, which have been showed as being complementary with pasta in the use association approach. Cluster 5 presents milk, coffee and yogurt, with can be considered a coffee break time consumption section.

Cluster 6 represents the personnel care products, with toothpaste, deodorant and shampoo. That is probably the better-fitted cluster in terms of consumption universes approach. They have strong correlations and the shopping occasions for those categories are frequently the same. At same time, they share cognitive meaning with personnel care family.

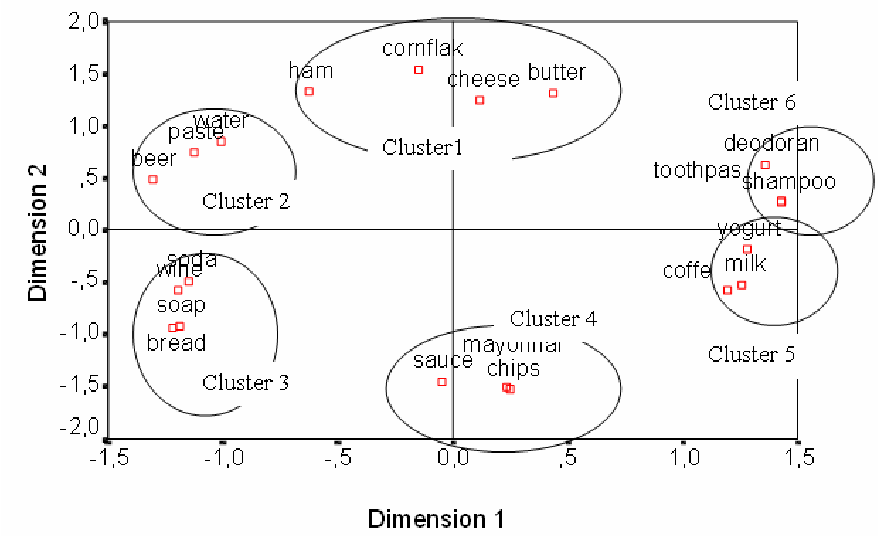


Fig. 1 MDS on Buying Association Matrix – General Results

5 Discussion

By introducing the buying association as a market basket measurement, we would incorporate both use association and one stop shop principle into the merchandise organization. By assembling categories with strong buying associations, we have tried to propose a new store layout, where consumers find everything they want in the same store area, maximizing the consumer's use of time spent in the store.

This is descriptive research, and we have not tested the impact of possible layouts on consumer behavior or store sales. New research should measure the layout impact on shopping satisfaction and impulse buying, which can be done through in-store or laboratory experiments. But, new KDD processes, such as some described in this volume, can be automated and include routines using scaling techniques such as these to help produce optimal merchandising arrangements. These routines may also help retailers determine not only how to merchandise goods physically in the store, but assist with timing so that they know how often to scramble merchandise. Thus, this paper illustrates a mechanism through which retailers can better take advantage of KDD processes.

References

Agrawal, R., Imieliski, T., Swami, A.: Database Mining: A performance perspective. IEEE transactions on Knowledge and Data Engineering 5(6) (1993)

Balderston, F.E.: Assortment choice in Wholesale and Retail Marketing. Journal of Marketing 21, 175–183 (1956)

Betancourt, R., Gautschi, D.: Demand complementarities, household production, and retail assortments. Marketing Science 9(2), 146–161 (1990)

Borges, A., Cliquet, G., Fady, A.: L'association des produits dans les assortiments de supermarchés : critiques conceptuels et nouvelle approche. In: 17*ème* Congrès de l'Association Française du Marketing, Deauville (May 2001)

FMI, Convenience is Key for Consumers. Supermarket Research, 2(8), November/December, 1-2 (2000)

Hays, W.L.: Statistics for the social sciences, 2nd edn., Holt International Edition (1977)

Henderson, J.B., Quandt, R.: Micro Economics Theory : a mathematical approach. McGraw-Hill, New York (1958)

Levy, M., Weitz, B.: Retailing Management, 3rd edn. Irwin/McGraw-Hill (1998)

Shahidi, A.: The End of Supermarket Lethargy: Awakened Consumers and Select Innovators to Spur Change. Supermarket Industry Perspective (2002) `http://www.bearingpoint.com/industries/consumer_and_industrial_markets/pdfs/Supermkt_Industry_POV_Final.pdf`

Walters, R.G.: Assessing the impact of retail price promotions on product substitution, complementary purchase, and interstore sales displacement. Journal of Marketing 55(1), 17–28 (1991)

Marketing – Sales Interface and the Role of KDD

Greg W. Marshall

Charles Harwood Professor of Marketing and Strategy
Crummer Graduate School of Business, Rollins College,
Winter Park, Florida, USA
Editor: Journal of Marketing Theory and Practice
President of the Academy of Marketing Science

1 Marketing Versus Sales

The phenomenon of less-than-harmonious organizational interface between marketing and sales is not news. Cespedes (1995) seminal treatment of "concurrent marketing" provided strong evidence of customer suboptimization due to the inability of marketing, sales, and (Cespedes argued) also customer service to properly integrate people, processes, systems, and strategies such that both the customer experience and return-on-customer-investment (ROCI) are maximized. The topic reached the "boardroom level" with the publication in 2006 of a special double issue of the *Harvard Business Review* on sales. The majority of the articles therein were focused wholly or in part on the marketing/sales problem, providing insights for executives on how to break down the barriers to create a more customer-centric enterprise model.

In the *HBR* special issue, Kotler, Rackham, and Krishnaswamy (2006) suggest that the two most important reasons for the friction between marketing and sales are economic and cultural. Economic in the sense that sales tends to view marketing as an expense of doing business – in part, a "marketing supports sales" mindset. Marketing, on the other hand, may view the sales force as a budget black hole for the organization – particularly when CEOs need a quick extra infusion of revenue to bring home the quarter's goal. Typically, in such cases sales is incented to go out and generate the business and is compensated purely on the revenue.

Inherent cultural conflict between the two functions is profound and deep-seated. Marketers are stereotyped as the thinkers, the planners, and the big picture/long term folks who are out of touch with what the customer on the street needs and wants *today*. Salespeople on the other hand are classically cast as immediate- action oriented, possibly projecting the needs and wants of one or a few customers into their definition of strategic action. This approach is driven largely by the direct link between behavior and compensation.

Paradigmatic differences such as these cannot be expected to bode well for the customer, which is why Bosworth and Holland (2004) laid out a vision and

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 43–48.
springerlink.com

systematic step-wise approach for changing the sales role through an integrated organizationwide performance management system centered on shared metrics among marketing, sales, and customer service. Riesterer and Emo (2006) provide ideas to further enhance the organizational capability of marketing to impact sales force (and thus customer) success via enabling technologies that allow broader branding strategies to be translated for salesperson use at the level of call planning and customer contact. (The reader is encouraged to explore Riesterer and Emo's conceptualization of "customer experience management.")

Ultimately, though, providing a roadmap and tools for sales to enhance performance against organizational customer goals, while quite useful, is not in-and-of-itself paradigm changing. Ten years from now hopefully the paradigm will have changed and much of this marketing – sales debate will be moot. The centrality of KDD – knowledge discovery in databases – can be a key driver in affecting this outcome.

2 A Call for Change

Fundamentally, for the first time in modern business history, the meaning of "selling" in the context of modern business-to-business (B2B) relationships is being questioned in the executive suite. Although much of the prior writing in this domain has laid the onus for change agency squarely on the shoulders of marketers (i.e., marketers enable salespeople for success), that approach is neither a sustainable nor truly paradigmatic solution. Cespedes' vision of concurrency of direction among marketing, sales, and service can only be realized by completely rethinking the entire business model, recasting old roles, and integrating internal firm components with a total customer focus. Colletti and Fiss (2006) argue that five fundamental macro-level changes are affecting customer-related activities in most major companies:

- **Customers have gained power.** Supply now generally outstrips demand. Customers have more knowledge, more choices, and more capability to dictate offerings, channel paths, and usage recourse.
- **Customers have gone global.** This fact more than any other has contributed to the "corporatization" of selling – that is, the trend for organizations to approach each other as business and channel partners at an *organizational level* (rather than a traditional salesperson/purchasing agent level). In the global marketplace, the destructive narrow view of what comprises the selling function and the supposed dichotomy between marketing and sales roles is exacerbated.
- **Channels have proliferated.** Channels are now networks, strategic partnerships and alliances, and integrated systems. Knowledge and information – adding value to both parties as an integral part of the offering – are at the core of what makes a B2B relationship "sticky." This concept is central to the "service-dominant logic in marketing" – ultimately physical products become more and more commoditized, and the value is in the results of KDD.

- **More product companies sell services.** Following up on the channels points above, today's B2B customers essentially are buying your strategies and vision for mutual performance enhancement delivered through the relationship. This viewpoint provides a strong challenge to proliferating traditional marketing and sales roles.
- **Suppliers have adopted a "one company" organizational structure.** The selling firm presents with a single corporate face. A.G. Laffley, former chairman of Procter & Gamble, understood the power and impact of this approach when he took over leadership the firm, which at the time was mired in rigid, old-style marketing and selling processes (Laffley and Charan 2008). Now, P&G's customer enterprise is entirely focused on a multi-functional, KDD-driven approach to the market that involves the customer's voice not just in a reactionary mode after product prototype development but also formally integrates customers as partners in the full cycle of the development and marketing of an offering ("customer" is used here to denote both in-channel and end-user).

The historical challenges and resulting suboptimization brought on by marketing and sales silos, coupled with the profound customer-driven changes outlined above, leave CEOs with a decision on how to reinvent organizational models to become customer-centric. As mentioned at the outset of this essay, by definition this means properly integrating people, processes, systems, and strategies such that both the customer experience and return-on-customer-investment (ROCI) are maximized. Trailer and Dickie (2006) propose that while customers' buying processes have already evolved vis-à-vis the new world of ubiquitous, instant, global communication (think consumer research on Twitter), companies' marketing and selling processes are too-often frozen in time in the 1990s. Technology and knowledge are major aspects of what Trailer and Dickie's research found among the most fruitful investment opportunities for enhancement of the customer experience and ROCI – in short, a Customer Relationship Management (CRM)-enabled KDD.

3 KDD as a Catalyst for Paradigmatic Change

The idea that effectively structured and executed CRM enables knowledge sharing, which allows the customer to maximize benefits from an offering and relationship, is at the core of how KDD can and should be a key intra-organizational facilitator of marketing-sales-customer service integration. The premise is that in today's marketplace knowledge sharing is the crux of the value of engaging in a B2B relationship. KDD, manifest in the marketing environment through integrated CRM, has three flagship principals that – in order to be accomplished – require permeation of functional organizational barriers: (1) customer value creation through knowledge enhancement; (2) reconceptualization of the "product" as an integrative, shared, two-directional "process" between provider and customer; and (3) a leap forward to a more permeable membrane between provider and purchaser that is driven by mutually beneficial knowledge sharing via investment in information processes (Storbacka and Lehtinen 2001).

Importantly, as a result KDD (also called data mining) holds promise to also provide a critical internal common ground of mutual interest among marketing, sales, and customer service functions. Swift (2001) coined the following definition of data mining that is appropriate for out purposes: a process of analyzing detailed data and extracting and presenting actionable, implicit, and novel information to solve a business problem. That is, data mining discovers knowledge. And knowledge is the shared capital that creates mutual value in a B2B relationship. The longstanding call by Cespedes (1994) and others to use intra-organizational integration of performance management systems and shared metrics to fuel needed changes in the customer enterprise business model -- post-marketing versus sales feud -- is actionable if the centrality of the marketing and sales (and customer service) roles becomes *generating, analyzing, and developing and implementing information-based knowledge discovery*.

If there is uncertainty as to whether an organizationwide KDD focus has promise to rally common ground among disparate silos, consider the following applications proposed by Swift (2001) related to the technology of KDD to solve business problems:

- Customer profitability
- Customer retention
- Customer segmentation
- Customer propensity
- Channel optimization
- Targeted marketing
- Risk management
- Fraud prevention
- Market-basket analysis
- Demand forecasting
- Price optimization

These uses and more are pervasive in strategic importance to a firm, and importantly the metrics associated with their application impact (and are impacted by) all three areas: marketing, sales, and customer service. As Bosworth and Holland (2004) have suggested, finding legitimate common ground in metrics among internal functional groups and codifying those in a shared culture and performance management system is a critical element to developing a customer centric enterprise. A cultural shift to an organizationwide strategic focus on KDD also has the external benefit of directly addressing each of Colletti and Fiss' (2006) five fundamental changes mentioned earlier that are affecting customer-related activities (Colletti and Fiss 2006):

- **Customers have gained power.** Knowledge sharing capabilities, resulting from KDD, as an overt element in the value proposition alleviates traditional angst between buyers and sellers, including distrust and the need for buyers to generate their own data in order to check or refute seller-generated claims ("my data is better than your data").

- **Customers have gone global.** A KDD-based metaphor for marketing and sales roles is inherently strategic. Common goals, metrics, and performance management systems can be developed that reduce suboptimization of the customer enterprise.
- **Channels have proliferated.** Pressure is relieved in fighting commodity (aka "price") based competition inherent to many physical products, as the larger portion of the value-added for clients becomes the deep knowledge shared. Disintermediation concerns are mitigated by this added value.
- **More product companies sell services.** Sensitivity to the service aspects of the relationship are enhanced through the partnership mind-set between client and supplier. KDD must address the service portion of the relationship with equal vigor to other informational areas and needs within the firm.
- **Suppliers have adopted a "one company" organizational structure.** Because KDD is organizationwide, the supplier firm can now effectively approach the buying firm at the *enterprise level*. The traditional sales role is enhanced and upgraded to a role focused on integrative account management, likely with a cross-functional support team in place (physically and virtually). Because the supplier's product line is no longer sold in disparate chunks, opportunities for cross-selling and up-selling across lines is greatly enhanced. An overall profit maximization model can be implemented that involves the firm's broad array of offerings over a longer time frame.

In sum, the challenge to reduce customer suboptimization due to the inability of marketing and sales (and also customer service) to effectively integrate people, processes, systems, and strategies toward maximizing the customer experience and ROCI can be answered through cultural change toward KDD. KDD provides a common ground for goals, metrics, and performance management among these organizational functions. The approach is particularly satisfying because it has the potential both for intra-organizational benefit as well as benefit in the marketplace by addressing critical changes affecting customer-related activities.

References

Bosworth, M.T., Holland, J.R.: Customer Centric Selling. McGraw-Hill, Boston (2004)

Colletti, J.A., Fiss, M.S.: The Ultimately Accountable Job: Leading Today's Sales Organization. Harvard Business Review, 124–131 (July-August 2006)

Cespedes, F.V.: Concurrent Marketing: Integrating, Product, Sales, and Service. Harvard Business School Press, Boston (1995)

Kotler, P., Rackham, N., Krishnaswamy, S.: Ending the War between Sales and Marketing. Harvard Business Review, 68–78 (July-August 2006)

Laffley, A.G., Charan, R.: The Game Changer: How You Can Drive Revenue and Profit Growth with Innovation. Crown Business, New York (2008)

Riesterer, T., Emo, D.: Increasing Marketing's Impact on Selling. South-western Educational Publishing, Cincinnati (2006)

Storbacka, R., Lehtinen, J.R.: Customer Relationship Management: Creating Competitive Advantage through Win-Win Relationship Strategies. McGraw-Hill, Singapore (2001)

Swift, R.S.: Accelerating Customer Relationships: Using CRM and Relationship Technologies. Prentice Hall PTR, Upper Saddle River (2001)

Trailer, B., Dickie, J.: Understanding What Your Sales Manager is Up Against. Harvard Business Review, 48–55 (July-August 2006)

Applying Soft Cluster Analysis Techniques to Customer Interaction Information

Randall E. Duran, Li Zhang, and Tom Hayhurst

Singapore Management University and Catena Technologies Pte Ltd
Catena Technologies Pte Ltd, 30 Robinson Road, Robinson Towers #11-04, Singapore 048546
e-mail: randallduran@smu.edu.sg

Abstract. The number of channels available for companies and customers to communicate with one another has increased dramatically over the past several decades. Although some market segmentation efforts utilize high-level customer interaction statistics, in-depth information regarding customers' use of different communication channels is often ignored. Detailed customer interaction information can help companies improve the way that they market to customers by taking into consideration customers' behaviour patterns and preferences. However, a key challenge of interpreting customer contact information is that many channels have only been in existence for a relatively short period of time, and thus, there is limited understanding and historical data to support analysis and classification. Cluster analysis techniques are well suited to this problem because they group data objects without requiring advance knowledge of the data's structure. This chapter explores the use of various cluster analysis techniques to identify common characteristics and segment customers based on interaction information obtained from multiple channels. A complex synthetic data set is used to assess the effectiveness of k-means, fuzzy c-means, genetic k-means, and neural gas algorithms, and identify practical concerns with their application.

1 Introduction

The number of ways that companies and customers communicate has increased dramatically over the past few decades. For example, retail banking customer interactions have gone beyond branch, mail, and person-to-person phone communications to include interactions through ATMs, bank web sites, email, mobile messaging, internet chat, social networking, and virtual reality environments. Although market segmentation efforts have utilized high-level customer interaction statistics – such as the frequency of interactions with a customer – in-depth information available regarding customers' use of different communication channels is often ignored. Making use of detailed customer interaction information can improve the way that organizations characterize customers' behaviour and preferences. Consequently, this

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 49–78.
springerlink.com

knowledge, either alone or combined with other demographic information, can provide marketing efforts with a competitive advantage.

Unlike most traditional sources of data used for customer segmentation, there is limited historical context for interpreting customer interaction information; many of the channels have only been in existence for a relatively short period of time, and new ones are continuing to evolve. Unsupervised classification techniques, such as cluster analysis, are well suited to help address this challenge because they group data based only on descriptions of the data and their relationships, which are extracted directly from the raw information without requiring advance knowledge of its structure. Furthermore, within the domain of cluster analysis methods, techniques that make use of fuzzy logic and artificial intelligence – such as genetic and neural algorithms – have the potential to provide unique insights into customers' behaviour patterns and achieve superior computational efficiency.

This chapter explores the use of various cluster analysis techniques to identify common characteristics and segment customers based on interaction information, such as the frequency, time, duration, and purpose of each interaction across multiple channels. The effectiveness of k-means, fuzzy c-means, genetic k-means, and neural gas algorithms is assessed to provide an understanding of the techniques' effectiveness and identify practical concerns with their application. Specifically, the goal of this research is to answer four questions: Can customer segments be identified only using customer interaction data? How accurately are the segments drawn? How well do clusters match known customer profiles? How well do soft computing approaches to cluster analysis perform, as compared with traditional methods?

In order to illustrate its relevance, the analysis is presented in the context of supporting the marketing activities of a retail bank. The effectiveness of the clustering is assessed using synthesized data sets that include interaction patterns that represent different retail banking customer groups. Starting with a synthetic data set that has a known composition enables the effectiveness of the cluster analysis to be evaluated independently from variations and uncertainties in the real data to which it is applied. Trying to validate these techniques using data derived from real-world customer interactions would be very difficult. In this case, there might be multiple meaningful customer groupings and the cluster analysis could identify ones that do not correspond to groupings derived using other approaches, making the comparison and validation of different approaches problematic. Furthermore, lack of underlying information could make it more difficult to correlate and verify the groupings, thus raising doubts regarding the validity of the clustering results. For example, distinct clusters might be identified for part-time workers who are also students and part-time workers who are not, but it would be difficult to confirm this distinction if the bank's customer records did not have recent information about customers' school enrolment status. Using synthetic data to support the evaluation of the clustering methods avoids this concern.

The structure of this chapter is as follows. The first section provides a literature review of cluster analysis, discussing its use within the financial services industry and for customer relationship management (CRM). The second section outlines a business context and discusses how the synthetic interaction data were constructed. The third section describes the research approach. The fourth section presents the results. The conclusion identifies practical applications of this research and identifies further areas of investigation.

2 Literature Review

Cluster analysis as a statistical tool has been actively studied in several fields such as statistics, numerical analysis, and machine learning. From a practical perspective, it has played an important role in various data mining applications in the domain of marketing, CRM, and computational biology. The following section provides a brief introduction to basic cluster analysis concepts and lists a few of its applications in the financial services industry.

2.1 Background of Cluster Analysis

Cluster analysis is a collection of techniques for dividing a set of objects into meaningful groups based on features that describe the objects and their relationships (Tan 2005). The desired result is that objects within a group should tend to be similar to one another, and objects in different groups tend to be less similar. This similarity is typically measured as the "distance" between each pair of objects, according to a metric appropriate to the type of data being measured. To help illustrate this concept, Fig. 1 shows a simple example of how a bank could use cluster analysis to segment its customers. Each customer is described by the number of credit cards they possess and how often they have made online bill payments over the past two years. Three clusters can be obtained, as shown in Fig. 1. Cluster 1 is comprised of customers that have a large number of credit cards and make many online payments; cluster 2 contains customers that have an intermediate number of credit cards and make relatively few online payments, and cluster 3 contains all customers with few credit cards who rarely make online payments. This simple example only uses two variables, whereas a cluster analysis will typically involve many more.

Cluster analysis can be regarded as an unsupervised classification method, because it classifies data based on their underlying structure or characteristics. In contrast, supervised classification techniques assign a class label to new objects according to an existing model that is based on objects with known labels. For example, supervised classification methods could be used to label credit card applications as either 'approved' or 'rejected' according to a model derived from a pre-existing set of applications with well-understood characteristics. Conversely, cluster analysis methods would divide a set of credit card applications into multiple groups, whose underlying characteristics could then be used as the basis for deciding whether to approve the applications, group by group. In market research studies, both supervised classification methods and cluster analysis methods are used to divide data into different segments. Supervised classification methods find segments characterized by predicted customer behaviours and are mostly used for targeted labelling, whereas cluster analysis methods are more explorative and are often used to discover unknown groups, without any *a priori* information that is used as a training set.

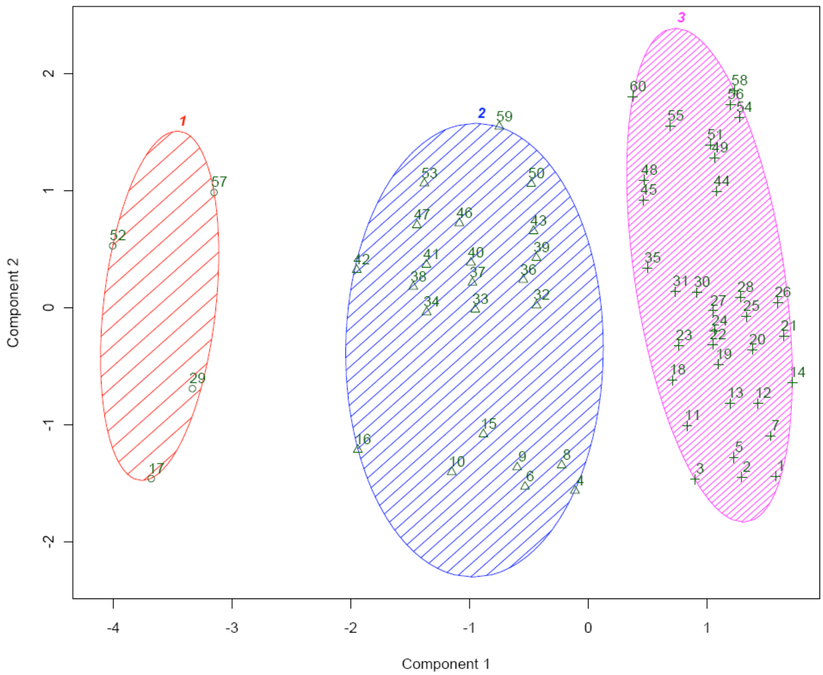


Fig. 1 A sample clustering of bank customers

Clustering methods can be broadly categorized as partitional, hierarchical, or overlapping (Hruschka 2009). Partitional clustering methods divide a set of objects into a number of non-overlapping clusters, the number of which is usually predefined, such that each object is in exactly one cluster. K-means clustering (MacQueen 1967) is a widely used partitional clustering algorithm. The k-means algorithm first allocates a number of randomly selected points to be the initial centre of each cluster, and then assigns each object in the data set to the nearest centre to form clusters. The cluster centres for the next iteration are then assigned to be the centres of the clusters from the previous iteration, and the objects are reassigned to the new centres. This process is repeated until the centres are stable and do not change with subsequent iterations.

Hierarchical clustering is an alternative to partitional methods. These methods distribute the data into a set of nested clusters organized as a tree structure. Agglomerative methods start with as many clusters as objects in the dataset and repeatedly merge the two closest clusters until a single cluster remains. Divisive methods start with a single cluster containing all data and then repeatedly split clusters until a stopping criterion is met (Han 2001). Both partitional and hierarchical methods can be considered to be exclusive, or crisp, clustering methods because each object is placed in exactly one cluster.

Alternatively, overlapping clustering methods can assign objects to more than one cluster. Fuzzy clustering, for example, allows each object to belong to multiple clusters; each object will be associated with a cluster based on a weighting between zero and one. Each object's total weight across all associated clusters is equal to one, or, in other words, each cluster shares a portion of the object. The fuzzy c-means (FCM) algorithm is a popular fuzzy clustering algorithm. It is broadly similar to k-means, but instead of assigning points to their closest cluster centre, a membership degree is defined to describe the proximity of the object to each cluster centre (Nikhil 1996).

Many of these clustering methods depend on randomly generated initial clusters. It is possible that, depending on the starting configuration, some clusters never have points assigned to them and, as a result, the clustering process can terminate with a sub-optimal solution. One way to address this problem is to run the clustering algorithm repeatedly with different, randomly generated, initial values, and then select the solution that minimizes the objective function, which defines the evaluation criterion of the solution. However, this approach can be very time consuming, and there is no guarantee that an optimal solution will be achieved within a given number of iterations.

Another approach to finding an optimal solution is to use evolutionary algorithms. These produce clusters by iteratively sampling clustering solutions from the search space, evaluating them against the objective function, and applying a mutation, crossover, or selection operator to generate new solutions. While evolutionary algorithms do not guarantee that an optimal solution will be found, they tend to generate more promising solutions during the exploration of the search space. Evolutionary algorithms, therefore, have a higher probability of reaching an optimal solution with fewer random initializations than repeatedly running k-means, although each run might take a longer time (Hruschka 2009). Genetic algorithms are a common type of evolutionary algorithm.

Like evolutionary algorithms, competitive learning algorithms can also be used to determine optimal clustering solutions. Some competitive learning methods can also be used to automatically find the optimal number of clusters (Fritzke 1997). Competitive learning algorithms iteratively adapt the locations of cluster centres based on the input data and gradually move towards the optimal solution using neural network methods. There are two categories of competitive learning algorithms: hard competitive learning and soft competitive learning. Hard competitive learning methods, such as k-means, use a "winner-takes-all" approach during the adaptation of the winning cluster centre for each input data point. Soft competitive learning methods address k-means' sensitivity to the initial values' positions by using a "winner-takes-most" approach during the adaptation of cluster centres, so not only the winning cluster centre is adapted, but also some or all the other centres. For example, the neural gas algorithm (Martinetz 1993), is a competitive learning algorithm that ranks the cluster centres according to their distance to each given data point and then adapts them in the ranked order to move towards the optimal solution. Another, similar, technique is the Self-Organizing Map (SOM), originated by Kohonen (2001). The difference between neural gas and SOM is that neural gas does not have a topology imposed on the network, while SOM has a fixed network dimensionality which makes it possible to map the usually large n-dimensional input space to a reduced k-dimensional structure for easy data visualization (Fritzke 1997). Nevertheless, studies have shown that the traditional k-means clustering method has produced higher classification accuracy than neural networks using Kohonen learning (Balakrishnan 1994).

2.2 Applications in the Financial Services Industry and Customer Relationship Management

Cluster analysis has been used as a multivariate statistical modelling technique in the financial services industry for a variety of purposes. Credit risk managers have combined supervised and unsupervised classification techniques to evaluate credit risks. Zakrzewska has investigated the combination of cluster analysis and decision tree models by first segmenting customers into different clusters characterized by similar features and then building decision trees that define classification rules for each group separately (Zakrzewska 2007). Each credit applicant was assigned to the most similar group from the training dataset, and their credit risk was evaluated based on rules defined for the group. Results of the cluster analysis on credit risk datasets demonstrated greater precision than decision tree models.

Cluster analysis has also been used in credit card portfolio management to identify potentially bankrupt accounts, fraudulent transactions, and distressed credit card debt (Allred 2002; Peng 2005). Clusters of accounts can be used to predict credit card holders' behaviours, allowing appropriate policies to be developed for each individual cluster. Likewise, Edelman has applied an agglomerative hierarchical clustering method to group monthly credit card payment transactions so that the groupings can be used to assist in the scheduling of resources allocated to address delinquent accounts (Edelman 1992).

Another growing application area for cluster analysis is customer relationship management (CRM), which utilizes data from various sources, including demographic information, transaction history, and call centre activities. CRM evaluates customer behaviour, such as spending habits, to help optimize and fine-tune marketing and pricing strategies. For example, as part of a CRM survey, a large sample of respondents could be divided into different market segments according to a number of variables related to consumer behaviour. Appropriate services and products can then be tailored to suit each particular market segment and therefore achieve the highest efficiency and profitability. Balakrishnan et al. (1996) have applied both competitive learning and k-means algorithms to generate clusters of coffee brand choice data that supported strategic marketing decisions. A combination of both methods was found to provide useful segmentation schemes.

Most of these applications have made use of traditional clustering methods, and have relied primarily on demographic and transactional data. However there is doubt as to how useful this information is for practical business purposes such as predicting customer profitability (Campbell and Frei 2004). In contrast to previous efforts, the remaining sections of this chapter will examine how fuzzy and artificial intelligence-based clustering methods can be applied to customer interaction information.

3 Business Context and Data Used for Analysis

This section introduces the business context of this study emphasizing on the pervasive multi-channel customer interaction data available and the benefits it can bring to marketing practices with the help of cluster analysis. The synthetic data used for analysis is also described including its structure and characteristics, which are carefully designed to simulate the real customer interaction pattern.

3.1 Business Context

The business context for this analysis is that of a bank attempting to obtain meaningful marketing insights from its interactions with retail customers. Retail customers typically use many different channels, including the call centre, voice recognition units (VRU), text messaging, Internet web sites, dedicated mobile web sites, physical branches, and automated teller machines (ATMs). Use of a combination of these channels by banking customers is common in developed nations. Because of banks' rapid adoption of new channels, this business domain provides good potential for utilizing cluster analysis techniques to analyse and segment customer information. Multi-channel customer interaction data analysis could be performed independently, or in support of existing data mining and segmentation practices.

Electronic channels, particularly text messaging and the Internet, have the potential to supply rich information about customers' behaviour; however, because they are new, their usage is not well understood. Accordingly, Sinisalo et al. recommend that when supporting "next generation" channels, firms should go beyond demographic and psychographic data, and use behavioural data to profile and categorize customers (Sinisalo et al. 2007). Customers that have similar behaviour patterns can then be grouped together for analysis and servicing. Multi-channel interaction information provides a fertile source of data for achieving this objective. Furthermore, while the demographic data that is commonly used for customer marketing analysis is widely available, customer interaction data is a new and relatively untapped resource.

One application of segmentation using interaction information would be to correlate identified customer groups with marketing considerations such as sensitivity to fees, product type preferences, loyalty, and default risk. Such information could be used to help determine the best products and services to offer to those customer groups. Customers who interact primarily via branches and call centres and consistently have lengthy interactions might be classified as "chatters" who highly value human interaction as part of the banking experience. Based on this interaction-based insight, a dedicated relationship manager could be included in a bundled service package that the bank could offer specifically to chatters. Beyond revenue generation purposes, effective classification of customer behaviour patterns can also be useful for risk control purposes. For example, certain customer segments may not be offered credit products if, according to their interaction characteristics, as a group they have a higher propensity to default. Moreover, previous research has shown that banking customers' use of a specific channel can be correlated to their economic value to the enterprise, even when controlling for demographic differences (Hitt and Frei 2002).

3.2 Synthetic Data Structure and Design

A key objective of this study is to evaluate the effectiveness of cluster analysis by running it on a data set where the factors driving interaction behaviour are completely understood. Using real-world customer data for this purpose would be impractical, since it would be extremely difficult, if not impossible, to obtain information about all the pertinent factors that influence each customer's behaviour. Understanding these underlying factors is necessary to effectively evaluate the results of the cluster analysis. Synthetic data sets were, therefore, designed independently from the cluster analysis implementation. To avoid biasing the research approach based on knowledge of the input data, the data set design parameters were only shared with the analysis team towards the end of the analysis.

The data sets were generated algorithmically and represented different types of retail banking customers. The goal was to produce realistic, complex sets of data that characterized different user groups and subgroups, which can then be used to determine how accurately the cluster analysis could identify the underlying customer groups as segments based on their interactions. Specifically, the data generation was driven by the following factors:

- Who – their age range and lifecycle stage
- Why – the purpose of their interaction
- When – time of day and day of week of the interaction
- Where and how – which channel was used for the interaction

Who was the primary driver for determining the interaction pattern. Customers were broken up into three primary groups and eight subgroups, which produced eleven subgroup-category combinations. The timing of customer interactions was generated by sub-group specific functions that took into consideration biases of that group towards times of the day and days of the week when they would contact the bank. Channel access rules were also taken into account, whereby branch access was limited to weekday business hours and from 9am to 1pm on Saturday. The interaction frequency – defined as the average number of interactions per month – was also varied by subgroup.

Detailed interaction profiles were defined for each of the customer subgroups. These profiles describe the purpose of the interaction, in what proportion different channels are used for each interaction type, and the duration of each interaction, according to channel. Table 1 shows the profile summary for one subgroup, Working High School Students.

While the synthetic data were designed to be realistic, some relevant factors were knowingly omitted due to overall project scope. Specifically, the synthetic data had the following limitations: 1) the data only included customer-initiated interactions; 2) only a subset of the available channels and transaction purposes were represented; and 3) interactions were distributed evenly throughout the days of the month. However, it is not expected that expanding the scope and complexity of the data set to address these concerns would significantly affect the results of the cluster analysis.

Table 1 Sub-group data construction parameters for Working High School Students

Interaction Purpose	*Proportion of all sub-group interactions*	*Interaction Channel*	*Proportion of purpose*	*Duration (norm. dist.)*	
				mean	*stdev*
Deposit	20%	ATM	70%	60	30
		Branch	30%	180	60
Withdrawal	40%	ATM	90%	80	30
		Branch	10%	200	60
Acct Application	5%	Internet	50%	800	300
		Branch	50%	300	100
Account inquiry	20%	Internet	40%	200	80
		Branch	10%	300	100
		Call Centre	20%	300	100
		Mobile-SMS	30%	30	2
Marketing inquiry	10%	Internet	30%	200	80
		Branch	30%	300	100
		Call Centre	40%	300	100
Funds transfer	5%	Internet	40%	150	80
		Branch	5%	300	100
		Call Centre	25%	300	100
		Mobile-SMS	30%	30	2

3.3 Synthetic Data Group Characteristics

The synthetic data design was documented in a tabular form that spanned eight pages. A simplified, more qualitative presentation of the customer subgroup characteristics is provided as follows. Abbreviations for each of the subgroups, or customer types, are provided in parentheses for later reference.

- *High school students* – have relatively few interaction purposes and transact exclusively after school and on weekends. As a group, they interact relatively infrequently and favour automated channels. Working high school students (SHW) have proportionately more withdrawals than non-working students (SHN).

- *University students* – have a wider range of interaction purposes, including interactions related to credit cards. They also perform more electronic fund transfers, i.e. bill payments, than high school students and favour automated channels. Working university students (SUW) interact evenly across the 8am to

12pm time period, at a medium frequency. Non-working students (SUN) favour the evening and weekends and interact at a low frequency.

- *Workers* – have the widest range of interaction purposes, including interactions related to loans, such as mortgages. They have a balanced use of different channels, not favouring automated channels over any other, and interact at a medium frequency. Full time workers (WAF) interact mostly before work, during lunch breaks, after work, and on weekends. Part time workers (WAP) interact evenly between 6am and 10pm.
- *Unemployed* – are similar to Workers but have fewer fund transfers and more account inquiries. Unemployed customers (WAU) tend to favour the branch and call centre channels over the Internet. They interact relatively infrequently and do so evenly between 8am and 12pm.
- *Domestic* – are similar to Workers but do relatively more funds transfers and transact evenly between 6am and 12pm. Domestic customers (WAD) interact at a high frequency.
- *Retired-age workers* – favour the branch and phone channels over automated channels, interact at a medium frequency, and have longer interactions than other groups. Otherwise, retired age customers who work full time (RAF) have similar interaction characteristics to those of Workers, except that they have fewer application-related interactions and more withdrawal-related interactions. Retired-age customers who work part time (RAP) are similar except that they do proportionately more deposit interactions and favour the daytime during weekdays to interact.
- *Retirees* – have interaction behaviour that is very similar to retired-age workers, but interact at a low frequency. Like retired-age part-time workers, retirees (RAN) prefer to interact during the daytime on weekdays.

The structure of the synthetic data was designed to simulate interaction patterns of actual subgroups of the general population. While each of the groups had its own unique interaction characteristics, there was also significant overlap between their behaviour patterns. Customer age, the high level partition between the groups, was not included the data sets provided for analysis, since a main objective was to determine whether the clustering techniques could identify meaningful segments without the support of demographic information.

A clean data set, where all the customers consistently followed their prescribed behaviour patterns, was produced to serve as a baseline. However, it is unlikely that in a real-world environment that such consistency would be found. Therefore, data sets with different levels of random noise were also produced. Noise was quantified as the percentage of customers' interactions that would follow a random pattern rather than the prescribed behaviour patterns. Additionally, a data set was generated that included a group of hybrid, or "transitional", customers, who exhibited one group behaviour during the first half of the time period and another group of behaviour during the second half. Specifically, the transitional group's interactions over the time period alternated between unemployed and full time employed behaviour patterns.

4 Research Approach

The research approach applied multiple cluster analysis techniques to simulated multi-channel customer interaction data, as discussed in Section 3. The primary objective was to assess the effectiveness of different algorithms and determine their usefulness in different situations. Whereas cluster analysis can involve a number of different steps (Nargundkar 2000) – such as variable selection, data validation, data standardization, addressing of outliers, algorithm selection, determining the number of clusters, and validation of results – they may not all be relevant depending on the context of the analysis. The general process and how it was applied to the interaction-based customer data are discussed in the following subsections. Additional details are presented in the experimental results section.

4.1 Variable Selection

Variable selection is the first step of the cluster analysis process. It determines the dimensions used in the cluster analysis. The number of suitable variables often depends on the data being analyzed and the granularity of the clusters desired. The selection process can be done either through judgmental selection, which is to choose the variables manually, or by factor analysis, which is to define the selected variables as a set of factors, usually extracted as a linear combination of an initial set of variables (Goldberg 1997). For the purposes of this study, judgmental selection was chosen over factor analysis because the features of the clusters could be easily derived and analyzed from the variables directly, rather than requiring to be extracted through factor analysis.

Judgmental selection of variables requires a good understanding of the data being analyzed and how well the variables reflect the characteristics of the data. When judgmental selection is used, it is beneficial to select more variables than necessary at the beginning and then eliminate redundant ones after performing several iterations of cluster analysis. Assessing the spread of cluster means across all dimensions can be used to determine which of the variables are useful and which ones should be dropped (Nargundkar 2000).

Interpreting customer interaction data derived from multiple channels is a challenging task given the large number of user behaviour variables associated with each channel. For example, when a customer communicates with the bank via a call centre, the bank can record when the communication starts and ends, who initiated it, and its purpose. Customer communications initiated through the bank's web site will produce similar information, which can be obtained from web server logs and user session data (Rho 2004). The customer interaction record was initially defined as the set of characteristics common to all of the communication channels.

Table 2 shows the set of variables selected for the customer interaction data and their defined values. The raw data were then transformed into customer description data, where data points describe a customer's interactions with the bank over a period of time. Details of this transformation are presented in the experimental results section.

Table 2 Selected variables for customer interaction data

Variable Name	*Defined Values*	*Value Type*
Customer ID	e.g. 998831	Nominal
Channel	Branch / ATM / Call-Centre / SMS / Web	Nominal
Purpose	Account Inquiry / CCA Inquiry / Marketing Inquiry / CCA Application / Account Application / Loan Application / Withdrawal / Deposit / Funds Transfer	Nominal
Initiator	Customer / Bank	Nominal
Date and time	e.g. 2009-03-03 12:41:37 PM	Interval
Duration	0 ~ 3600 sec	Ratio

4.2 Data Validation

When preparing data for cluster analysis, it is generally necessary to validate the data. Invalid values should be removed if they cannot be fixed or replaced. However, because the data set analysed was synthesized and flaws were not included by design in the data set, this step was not relevant to the cluster analysis in this particular case. While it could have been possible to include flaws in the synthesized data, doing so would not have yielded any significant benefit, since this effort was mainly focused on assessing the effectiveness of clustering algorithms on the data.

4.3 Data Standardization

It is necessary to map the variables being analyzed to an equivalent scale so that the clustering algorithms can effectively compare different variables, regardless of how they were originally measured. How variables are standardized will depend on their value type. For example, nominal variables may be standardized by creating multiple binary variables for each of the nominal states and grouping them in order to avoid the influence of increased number of predictors. Interval and ratio variables can be standardized by normalizing the values to have a mean of 0 and standard deviation of 1.

When analysing the customer interaction information, standardization was only performed on the aggregated customer description data since these are the data used for cluster analysis instead of the customer interaction data. To illustrate how the standardization was put into practice, consider the following case. The total number of interactions per customer was measured over a given period of time. Sample values ranged from 9 to 116, with a mean of 52, and a standard deviation of 19. Likewise, the proportion of interactions via branch was measured as a ratio ranging from 0 to 1, with a mean of 0.25, and a standard deviation of 0.15. By normalizing both variables to have a mean of 0 and standard deviation of 1, both variables will make an equal contribution to the similarity measurement in cluster analysis.

4.4 Addressing Outliers

Observations that deviate significantly from the rest of the data, referred to as outliers, are common in data sets. It is often the case that outliers represent unusual behaviours or erroneous data. Hence, including outliers can bias cluster analysis results. Once the data have been standardized, outliers can be identified, based on how many standard deviations the points are away from the mean in each dimension. If a data point is too far from the mean, it often indicates an outlier. As was the case with data validation, because the data set used for analysis was synthesized, it was assumed that no erroneous data were present. Furthermore, it was of interest to see how the cluster analysis algorithm would organize the entire data set. Based on this rationale, no outliers were removed from the data set.

4.5 Algorithm Selection

As discussed in Section 2.1, the most appropriate clustering method to use will normally depend on the characteristics of the data being analysed. However, because a key objective of this research was to compare the effectiveness of different types of algorithms, multiple techniques were applied. In order to ensure that the same customer segments can be consistently identified from the same set of data, the chosen clustering algorithm should be as stable as possible. This makes evolutionary algorithms and soft competitive learning algorithms potentially good choices for evaluation, for the reasons discussed in Section 2.1.

Four clustering techniques were applied to customer interaction data, to compare their effectiveness. The first technique was the traditional k-means algorithm. The second technique was the genetic k-means (GKM) algorithm, which uses the same objective function as k-means but with an evolutionary approach to searching the solution space. The third technique was the neural gas algorithm (NG), which is based on soft competitive learning. Subsequently, the most efficient and effective algorithm of these three was then compared with a fuzzy clustering algorithm. In summary, K-means, genetic k-means (GKM) and neural gas algorithm (NG) were selected as crisp techniques; the fuzzy c-means algorithm was selected as fuzzy algorithm.

4.6 Decide Number of Clusters

When applying the above-mentioned clustering algorithms, the number of clusters must be chosen in advance. Determining a suitable number of clusters is important, since using too few clusters is likely to result in very broadly characterized clusters that do not show the complete structure of the data set, while using too many clusters may mistake random noise in the data for actual information. The idea is, therefore, to pick a number that produces a clustering solution that is both statistically good and contains meaningful clusters with respect to the data being analysed.

Compactness and separation are two commonly used criteria to evaluate clustering results. High compactness means that the data points within each cluster are close to each other. High separation means that the clusters themselves are widely spaced. Ideally, a good clustering should have both high compactness and separation. While there is no perfect way to determine the optimal number of clusters, they are commonly chosen by visual inspection or computation of statistical measures.

To visually determine a good cluster number, the selected clustering algorithm was repeatedly run using different numbers of clusters. For each clustering, the clusters were plotted in the dimensions of the two principal components that explain the largest variances of the data. These cluster data plots were then used to assess the compactness and separation of the clustering results. In addition, the normalized cluster means for each variable dimension were computed and each cluster's normalized cluster means sorted according to their absolute values. The highest-ranked means for each cluster showed the dominant features for each cluster. The feature representation of each cluster helped to interpret the "meaning" of each clustering solution. The two methods can be combined to choose a number of clusters that gives a cluster plot with compact and well-separated clusters, and where each cluster has meaningful characteristics.

Statistical methods can also be used to estimate the optimal number of clusters. For crisp methods, the simplest measure of compactness is the within-cluster sum of squares (WSS) metric and the simplest measure of separation is the between-cluster sum of squares (BSS) metric (Tan 2005). The Calinski and Harabasz index (CHI) (Calinski 1974) and the Hartigan index (HI) (Hartign 1975) are both based on WSS and BSS measures. These techniques can be viewed as line charts that compare the number of clusters on the x-axis to the index values on the y-axis as well as the successive differences of the index values. Where the chart of the successive differences is convex, the knee point in the curve is the place where the transition occurs from substantive clusters to erroneous clusters. This provides a good indication of the optimal number of clusters. The Dunn index (DI) can also be used to measure both compactness and separation in terms of intra-cluster and inter-cluster distances (Dunn 1974). The maximum Dunn index value defines the optimal number of clusters. The Silhouette Coefficient (SC) also provides a measure of compactness and separation (Kaufman 1990). The maximum of the average silhouette coefficient of all points determines the optimal number of clusters. In addition, the Hubert gamma statistic evaluates the separation of the clusters, which is maximized at the optimal number of clusters (Halkidi 2001).

In the field of fuzzy clustering analysis, two frequently used cluster validity indexes are partition coefficient (PC) and partition entropy (PE) (Bezdek 1974). Both indexes measure the fuzziness of a partition based on the membership values. A higher partition coefficient value and a lower partition entropy value signify a less fuzzy partition, and, hence, denser clustering. Xie and Beni introduced an XB index that measures the ratio of total variation of the data points with respect to the cluster centres to the minimum total separation between the cluster centres. The smaller the XB index, the better the clustering solution (Xie and Beni 1991).

Both visual inspection and statistical measures were used to analyze the results of using the different clustering algorithms, for ranges of three to sixteen clusters. For visual inspection, cluster plots and cluster feature representations were produced. For statistical analysis, five index values – DI, SC, Hubert gamma, CHI and HI – were computed for the crisp clustering algorithms. Three index values – PC, PE and XB – were computed for the fuzzy c-means algorithm. A subjective combination of the visual and statistical assessment was then used to determine the optimal clustering.

4.7 Validate Clustering Results

Once an optimal set of clusters has been generated, it is important to evaluate how well the clustering algorithm has partitioned the data set. One simple way is to verify visually whether the clusters are well separated. However, this can be rather difficult, especially for high dimensional data sets. Therefore, procedures have been proposed to evaluate the results of a clustering algorithm. There are three ways of evaluating cluster validity (Halkidi 2002). The first approach is based on external criteria, by comparing the result with a pre-defined clustering structure that reflects the a priori characteristics of the data. The second type is based on internal criteria, by evaluating the result against some statistics derived from the data itself such as a proximity matrix. The statistics discussed in Section 4.6 can be used as internal evaluation criteria to determine how good a clustering solution is without comparing with other clustering solutions. The third type is based on relative criteria, which are mainly used to compare clustering solutions resulting from the same algorithm but with different parameters.

To compare different crisp clustering techniques applied to the same data set, the corrected Rand index (Gordon 1999) can be computed, to measure the level of agreement of the class labels. A high value for this measure indicates a high level of agreement. A similar measurement is the contingency table, which computes the number of data points that fall into the same clusters between two clustering solutions. In this study, both measurements were used as relative criteria to evaluate different solutions derived from each crisp algorithm – traditional k-means, genetic k-means, and neural gas – comparing different runs of the same algorithm, and also to compare solutions produced by different algorithms from the same data.

Since the structure of test data set was known, the clustering results could be compared with a known baseline. In the study, comparison against the baseline was used to assess the effectiveness of both crisp and fuzzy clustering algorithms on clean and noisy data. To evaluate how closely the fuzzy clustering results matched the known class labels, the Fuzzy Rand index was calculated which is based on the membership correlation between the data points (Campello 2007). This is a common validation method based on external criteria.

In order to evaluate whether a clustering algorithm was consistently producing the same clustering, it was necessary to compare two clustering solutions derived from different data sets with the same underlying customer interaction behaviour characteristics. This comparison was difficult and there have not been any active studies in this area. A feature matching approach that computes the percentage of matching features in each cluster for the two clustering solutions was used for this purpose. A high matching score indicates a high similarity between the two clustering solutions. Detailed validation procedures are illustrated in the experimental results section.

5 Results

Once the synthetic customer interaction data had been generated, the first step was to generate the customer description data from the interaction events. The customer description data was then input into the clustering algorithms to partition the set of customers into different clusters with unique characteristics. All the data processing and cluster analysis procedures were implemented in the R language and computation environment. R was chosen because it provided off-the-shelf clustering algorithms and was well suited for data manipulation and graphing. All the experiments were run on a Core 2 Duo 2.4 GHz machine with 2GB RAM.

The following subsections discuss the results produced at each stage of the cluster analysis when applied to two similar but distinct synthetic data sets. The final subsection provides a qualitative discussion of the results and their implications.

5.1 Variable Selection

The primary synthetic data set that was analyzed contained 100,442 interactions performed by 1,933 unique customers. Initially 24 variables that describe customers based on their interaction history were identified. Of the 24 variables, five describe the percentage of interactions a customer makes via each of the five different channels; nine describe the percentage of interactions a customer makes for each of the nine different purposes; another eight variables describe the percentage of interactions a customer makes during different time frames of the day, different days of the week and different periods of the month; and the last two variables capture the number of interactions a customer makes and the average interaction duration across all channels. The average interaction duration was defined as the normalized average duration across all channels. For each customer, the mean duration of use of each channel was first normalized to the interval of 0 to 1 with respect to the minimum and maximum duration of all interactions using that channel. The duration variable was then computed as the mean of all the channel mean durations. To help represent usage patterns more accurately, variables were defined as the percentage of total interactions rather than as the absolute number of interactions.

Based on this initial set of variables, clustering results were generated for different numbers of clusters. Cluster means were then calculated across all dimensions for each cluster to help determine which variables were superfluous and should be eliminated. In particular, variables with very small min-max cluster mean spread and small cluster means were considered as non-discriminative, which were eliminated from the cluster analysis.

5.2 Standardize Data

Since all the values were obtained from the synthetic data set, it was assumed that there were no invalid or missing values. Next, to put all variables on an equivalent scale, the values were standardized to have a mean of 0 and a standard deviation of 1 across all customers. Table 3 shows the original and standardized values of

Table 3 Original and standardized values for a subset of variables for one customer

Variable Name	*Original Value*	*Mean*	*Standard Deviation*	*Standardized Value*
Interactions	30	51.96	18.81	-1.167
Branch	0.100	0.2501	0.1509	-0.9973
AcctInquiry	0.1667	0.1329	0.0702	0.4812
Duration	0.1470	0.1940	0.0395	-1.1899

several variables. The original values were specific to a particular customer, whereas the mean and standard deviation were calculated from the number interactions for all the customers.

5.3 Decide Number of Clusters

As discussed in Section 4, because the synthetic data set did not contain any outliers the candidate clustering algorithms were applied direct to the derived customer description data, and the optimal number of clusters was then assessed using both visual inspection and computed statistical measures. The following paragraphs illustrate how different metrics can be applied to the clustering results generated by k-means and the fuzzy c-means algorithms to determine the optimal number of clusters. These two algorithms serve as examples of crisp and fuzzy clustering techniques, respectively.

Since, due to the random initialization problem, both k-means and fuzzy c-means algorithms do not always produce a stable and optimal solution, repeating the clustering process many times helps to increase the likelihood of obtaining a stable solution. Hence, they were both run 3,000 times, and the clustering solution with the minimum value of the objective function was chosen. It was determined that 3,000 repetitions was a sufficiently large number to guarantee a relatively stable, close-to-optimal solution for the data sets examined.

In order to visualise the clustering results, they were displayed using the CLUSPLOT algorithm (Pison 1999), which shows how the points are distributed according to the two principal components, and represents clusters as ellipses of various sizes and shapes. These plots were helpful for seeing where single clusters appeared to be composed of multiple, distinct sub-clusters. When sub-clusters are visually identified in this way, it can be potentially beneficial to increase the number of clusters. Fig. 2 shows an example of the diagram generated for a 6-cluster solution on the clean data set using the k-means algorithm. Note that the appearance of overlapping clusters in the figure is due to the projection of the multidimensional data set onto a two dimensional view.

To gain further insights into the features that each cluster represented, a "feature plot" was generated for each cluster, in the form of a bar chart that shows the cluster's primary characteristics. These characteristics are determined by the highest-ranked cluster means as described in section 4.6. Fig. 3 gives an example of

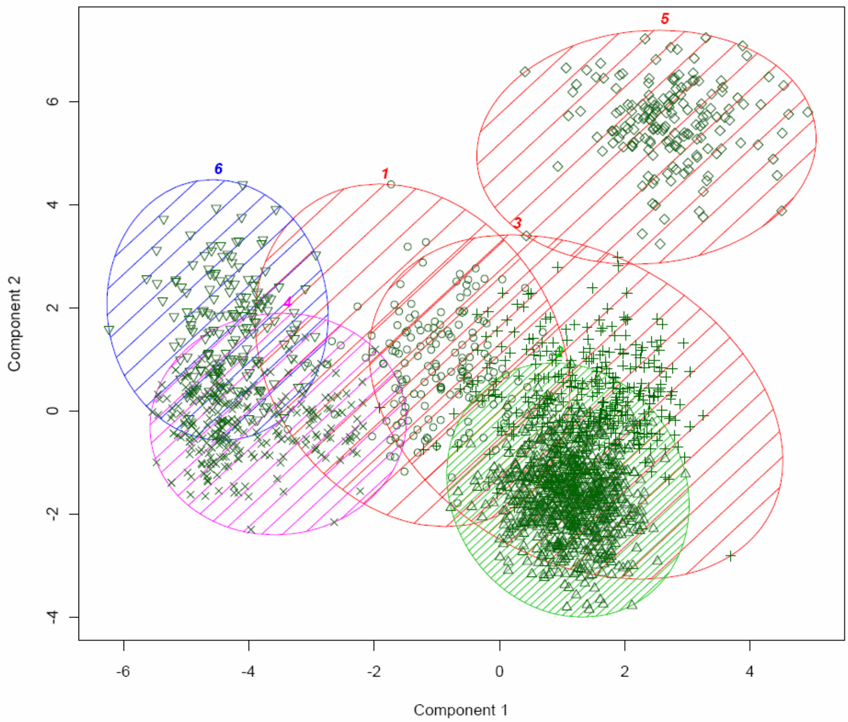


Fig. 2 Diagram of a clustering solution with 6 clusters generated using k-means

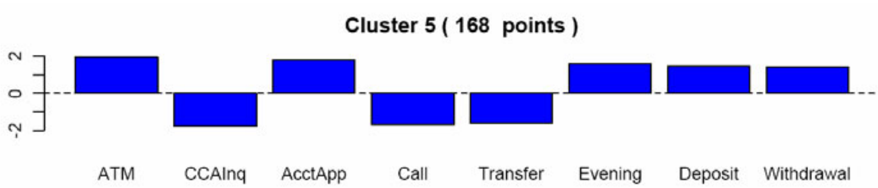


Fig. 3 Feature plot of a sample cluster selected from a 6-cluster solution generated using k-means

the feature plot for one cluster in a six-cluster solution generated using k-means on the clean data set. It shows the top eight features for this cluster are: heavy ATM usage, few CCA inquiries, many account applications, few call interactions, few fund transfers, heavy evening usage, and many deposit and withdrawal transactions. Similarly, Table 4 shows the top six features sorted in decreasing order of their absolute significance for all six clusters. The clusters' nicknames summarize their main characteristics.

Table 4 Top six features in decreasing order of absolute significance for a 6-cluster solution generated using k-means

ID	*Cluster Nickname*	*# of Cust.*	*Top six features in decreasing order of absolute significance*					
			1	*2*	*3*	*4*	*5*	*6*
1	Heavy SMS and Web users for various purposes	854	+LoanApp	+Transfer	+Interaction	+MarketInq	+SMS	+Web
2	Web and SMS users for CCA application	322	+CCAApp	+Web	+SMS	-AcctInq	-LoanApp	-Branch
3	Heavy branch users	252	+Duration	+Branch	+Interaction	-Deposit	-Web	-SMS
4	Infrequent marketing inquiry users via call-centre	174	+MarketInq	-Night	+Call	-Withdrawal	-Interaction	-AcctInq
5	Evening ATM users for account application	168	+ATM	-CCAInq	+AcctApp	-Call	-Transfer	+Evening
6	Weekday night users that prefer branch	163	-Weekends	+Night	-Evening	+Branch	-Web	+Duration

Table 4 shows that, as measured by the number of customers, the first cluster is extremely large compared to the other clusters. Also, interactions for all the purposes of loan applications, funds transfers, and marketing inquiries are the dominant features of this cluster. Based on these two observations, the number of clusters was increased to eight in an effort to partition the first cluster. However, the 8-cluster solution did not subdivide this cluster by interaction purpose, as expected. In fact, it divided the large cluster into two smaller clusters, one representing users who make many loan applications and the other representing users who make many funds transfers. In addition, it also extracted a group of users from the second largest cluster that represents heavy web and SMS users applying for credit card accounts, only at night. This implies that simply analyzing the cluster features is not a satisfactory method for determining the optimal number of clusters. However, the feature plots were useful for interpreting the meanings of the clusters and providing qualitative insights that supported the quantitative assessment methods.

The Hubert gamma, Dunn Index, Silhouette Coefficient (SC), Calinski-Harabasz Index (CHI), and Hartigan Index (HI) statistical measures were also computed for different numbers of clusters to help determine the optimal number of clusters for the crisp clustering algorithms. Fig. 4(a) shows the values for three of these indexes across different numbers of clusters. The plot shows that the Hubert gamma and SC index reach their maximum values with three clusters. However, the indexes also have large values with 4, 5, or 6 clusters. Inspecting the within-cluster sum of squares plot, the optimal number of clusters should be found at the knee point of the curve. The plot shown in Fig. 4(b) indicates that 3 or 4

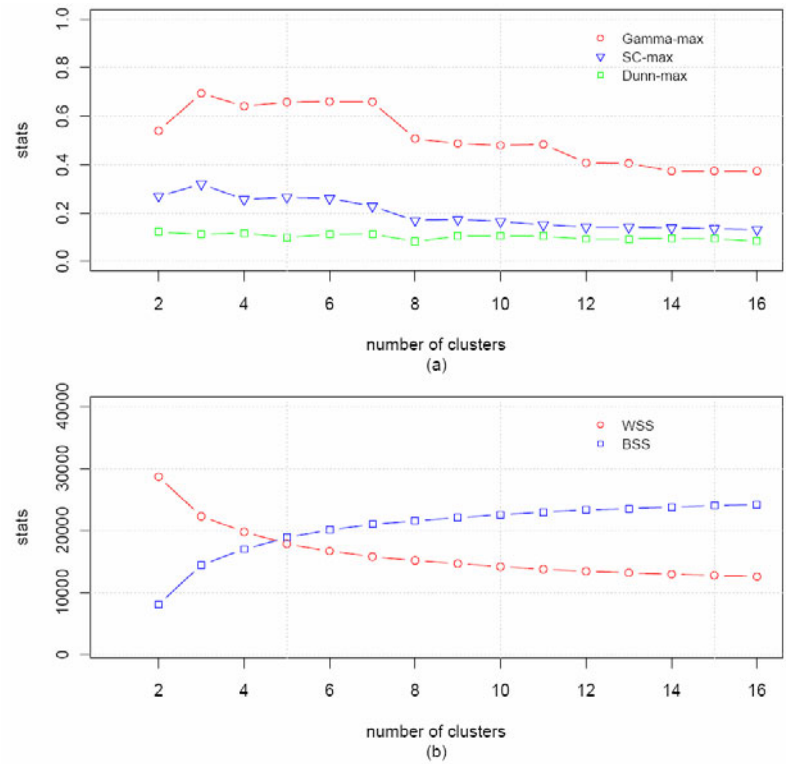


Fig. 4 Index values for solutions generated using k-means: (a) Common indexes; (b) Within-cluster and between-cluster sums of squares

clusters appear to be good knee points. Both the CHI and HI indexes imply that 6 or 7 clusters would be optimal, since the successive differences of both indexes are minimized there. The 3-cluster solution would produce relatively large and broad segments, which may not be useful within the business context. To obtain a finer-grained clustering result, the 6-cluster solution appears to be the next best choice.

For the fuzzy c-means algorithm, the partition coefficient (PC), partition entropy (PE), and Xie & Beni (XB) indexes were computed to help decide the optimal cluster number on the same data set. Fig. 5(a) shows the values of the PC and PE indexes on solutions with different number of clusters generated using the fuzzy c-means algorithm. The plot shows that 3 clusters appears to be the optimal solution, since the PC index is maximized and the PE index is minimized at that number. However, as with the crisp clustering, the 3-cluster solution would produce a very coarse-grained result. The 6-cluster solution appears to be the next best choice, as shown in Fig. 5(b).

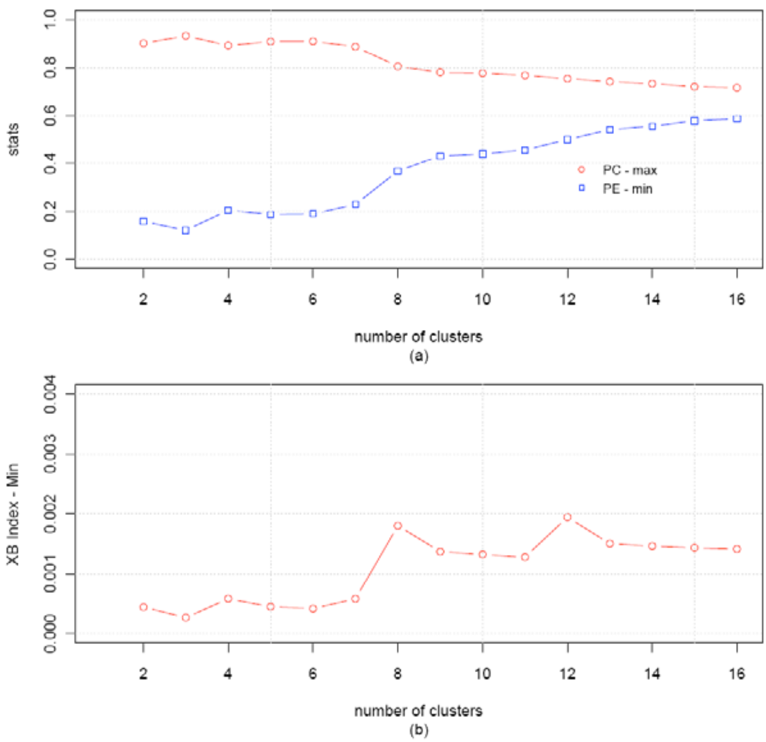


Fig. 5 Index values for solutions generated using fuzzy c-means: (a) Partition coefficient and partition; (b) Xie & Beni index

5.4 Validate Clustering Results

Various experiments were conducted to test the validity of the k-means, neural gas, genetic k-means, and fuzzy c-means clustering algorithms. First, the three crisp clustering algorithms and the fuzzy c-means algorithm were applied to both clean and noisy data sets to determine which algorithm was able to most efficiently generate stable and meaningful solutions. Second, the similarity of the clustering solutions was compared using the results from the crisp algorithms applied to multiple synthetic data sets that had same underlying structure and parameters. Ideally, since the synthetic data sets were constructed the same way, the features of the optimal clustering solutions generated for each of the different data sets should match one another. Third, the fuzzy c-means algorithm was applied to both clean and noisy data sets, and the stability of the solutions was analyzed. A detailed comparison of the k-means and the fuzzy c-means algorithm on both clean and noisy data sets is provided in Section 5.5.

To evaluate the stability of the three crisp algorithms, two trails for each algorithm were performed on the same data set and the results were compared using the Rand index metric discussed in Section 0. To avoid the problem of obtaining suboptimal clustering as a result of random initialization, the *k*-means algorithm was repeatedly run 3,000 times in each trial. In contrast, both neural gas and genetic k-means were run only once, but with a large number of iterations. The neural gas algorithm was run with the following learning rate parameters: 1,000 iterations and λ_i=10, λ_f=0.01, ε_i=0.5, ε_f=0.005. Increasing the number of iterations did not improve the stability of the results. For the genetic k-means algorithm, each clustering solution was represented by a vector of the cluster centres' coordinates, which has length 144 for a 6-cluster solution with 24 dimensions. The population size was defined as 100 and the mutation chance was chosen as 0.25%.

Table 5 shows the Rand index values for two 6-cluster solutions generated using each algorithm on several clean and noisy data sets. Repeated k-means produced the most stable performance across data sets with different levels of noise, followed by repeated fuzzy c-means. All the algorithms tended to become more volatile when the noise levels increased, especially the neural gas and genetic k-means algorithms. It was also observed that these algorithms' stability could be improved with repeated runs. However, the total execution time was much longer for neural gas and genetic k-means than k-means, even though they took fewer runs to reach a similar stability. For example, to generate a 6-cluster solution with 99% stability on the 15% noise data set, repeated k-means took less than 3 minutes to execute the Hartigan and Wong algorithm (Hartigan 1979) 3,000 times, whereas the neural gas algorithm took about 20 minutes and the genetic k-means algorithm took about 30 minutes to complete a single run. Given its overall stability under various conditions, repeated k-means was selected as the crisp clustering algorithm to be used as a baseline for compassion with the fuzzy c-means in the subsequent experiments.

Table 5 Rand index values on two runs of each algorithm for various data sets

Data Sets	*Clean*	*5% Noise*	*10% Noise*	*15% Noise*	*20% Noise*	*30% Noise*
Repeated k-means	1.000	0.998	0.982	0.991	0.969	0.994
Neural gas	0.997	0.930	0.922	0.821	0.707	0.735
Genetic k-means	0.883	0.845	0.828	0.769	0.698	0.717
Fuzzy c-means	0.998	0.995	0.983	0.986	0.966	0.991

Another observation was that there was significant variance in the clustering solutions obtained from the three algorithms, even at the 99% stability level. Table 6 shows a contingency table comparing two clustering solutions generated using k-means and neural gas algorithms on the clean data set. The table shows the number of points that fall in the same clusters between two clustering solutions. It was noticed that cluster 2 of neural gas was split into clusters 2 and 3 of k-means, while cluster 4 of k-means was split into cluster 3 and 5 of neural gas. This result was probably due to the fact that the algorithms can generate different suboptimal solutions that correspond to different local minima.

Table 6 Contingency table comparing two clustering solutions generated using k-means and neural gas algorithm

Cluster	*NG-1*	*NG-2*	*NG-3*	*NG-4*	*NG-5*	*NG-6*
k-means-1	170	0	3	1	1	0
k-means-2	0	248	0	0	0	0
k-means-4	0	0	331	0	523	0
k-means-6	0	0	0	310	10	0
k-means-3	0	162	0	0	0	1
k-means-5	0	0	0	0	0	167

To test whether the clustering procedure consistently identified clusters with the same features on different data sets, two different data sets were constructed with the same set of parameters were tested with the same process. To evaluate how close the two solutions were, a matching score, defined as the average percentage of matching features among the first six dominant features across all clusters, was computed. Table 7 shows the matching scores between the clustering solutions with different number of clusters on two different data sets. The matching scores between the clustering solutions using the selected set of variables show that more than 90% of features matched across all clusters for both k-means and neural gas. The reason for the low score with genetic k-means is largely due to the non-convergence of the algorithm in this case.

Table 7 Feature matching score between clustering solutions on different data sets

Algorithms	*Matching scores for different number of clusters*						
	3	*4*	*5*	*6*	*7*	*8*	*9*
Repeated k-means	1.000	0.958	1.000	1.000	0.929	1.000	0.833
Neural gas	1.000	0.958	0.967	1.000	0.912	1.000	0.907
Genetic k-means	1.000	0.958	0.733	0.750	0.690	0.792	0.704

To evaluate the stability of the fuzzy c-means algorithm, two runs of the algorithm were performed on both clean and noisy data sets. A contingency table was then produced that compared the number of points that fall into the same clusters by "hardening" the fuzzy clusters. Hardening the fuzzy clusters was achieved by uniquely assigning individual points to their closest cluster center. The Rand indexes for those solutions were calculated as 1.0 and 0.935, respectively.

5.5 Discussion and Analysis

As discussed in section 3.2, the synthetic data were generated from a population of eleven customer types. This allows a comparison matrix to be calculated from the derived clustering and the original data. Table 8 compares the result of running the k-means algorithm with six clusters on the clean data set with the original customer types. As can be seen, the original customer types, as defined in Table 8, largely fall cleanly into the generated clusters. The horizontal cluster IDs show how a six-cluster solution can be generated from the eleven customer types, and the vertical cluster IDs are those generated by the k-means algorithm. As there is no inherent meaning in a cluster ID, the rows have been reordered to show the correspondence between the two clustering more clearly.

Table 8. Contingency table comparing crisp 6-way clustering with customer types on clean data set

Customer Type	*RAN*	*RAF*	*RAP*	*SHN*	*SHW*	*SUN*	*SUW*	*WAD*	*WAF*	*WAP*	*WAU*
Cluster ID	*1*	*1*	*2*	*3*	*3*	*4*	*4*	*5*	*5*	*5*	*6*
2	167	82	5	0	0	0	0	0	0	0	0
3	0	0	162	0	0	0	0	0	0	0	1
5	0	0	0	67	100	0	0	0	0	0	0
6	0	0	0	0	0	130	80	0	5	5	0
4	0	0	0	0	0	1	3	112	578	160	0
1	0	2	0	0	0	3	1	0	1	2	166

Performing the same analysis while varying the number of clusters shows that all the statistically "good" clustering solutions, from three to seven clusters, correspond well to the original customer types. The three-cluster k-means solution distinguishes between high-school students, retired-age customers and approximately half of the unemployed adults, and the remaining population (university students and the other working-age adults). Moving to four clusters splits out the university students, and with five clusters the unemployed form a cluster on their own. The six-cluster solution shown above separates the retired-age customers who work part-time from the other retired-age customers, and with seven clusters, the university students are cleanly separated according to whether or not they have a job.

As the number of clusters increases, this clean separation would ideally continue until all of the eleven customer types form their own cluster. However, with eight clusters, the working population (employed or domestic) splits into two clusters, but not according to their customer type. As the number of clusters is increased further, the correspondence between customer types and clusters becomes still less clear. This demonstrates that there is a point at which the clustering algorithms become unable to distinguish between meaningful patterns in the data and random variations in customer behaviour. Although there are structural differences in the customers' interaction patterns, the statistical overlap between the different customer behaviour patterns is too great to allow the clustering to differentiate the underlying groups, using the interaction behaviour metrics chosen for analysis.

Had the variables used for clustering been finer grained, further cluster segmentation may have been possible. For example, the variables used to measure the time when interactions occurred for cluster analysis were defined as four, six-hour periods. This coarse categorization was surprisingly effective, given the more subtle timing differences that characterised the underlying data different customer groups. Had the analysis used hourly variables instead, better segmentation results for larger numbers of clusters may have been achievable. Because analysis effort was designed independently from the synthetic data construction, the benefit of using finer-grained time periods was not obvious *a priori*.

Table 9 Rand index values for different numbers of k-means clusters compared to customer types

Artificial Cluster-ing	*Number of clusters in k-means solution*							
	3	*4*	*5*	*6*	*7*	*8*	*10*	*12*
Same as k-means	0.862	0.872	0.964	0.966	0.945	0.669	0.590	0.525
11-cluster	0.645	0.799	0.852	0.874	0.884	0.881	0.900	0.900

Table 9 shows Rand index values comparing k-means clusterings to artificial clusterings based on the known customer type. Two types of artificial clusterings were generated: one by taking the same number of clusters as the k-means solution and assigning customer types to each cluster according to their weighting in the k-means solutions, and an 11-cluster solution with one cluster corresponding to each customer type. As can be seen, the Rand index values comparing the solutions to the same-sized artificial clustering drop sharply at eight and more clusters, meaning that the k-means clusterings do not closely correspond to the artificial clusterings. This agrees with the results of the visual inspection and statistical analysis obtained earlier to determine the optimal number of clusters. When comparing the k-means clusterings to the eleven predefined customer types, the higher numbers of clusters perform best, although significant agreement is reached from six clusters onwards.

Table 10 Comparison of fuzzy and crisp clusterings on clean data

First clustering	*Second clustering*	*Fuzzy Rand Index*
k-means	Known customer types	0.874
Fuzzy c-means	Known customer types	0.832
k-means	Fuzzy c-means	0.857

In order to compare the crisp and fuzzy clustering solutions, a fuzzy membership matrix was generated from the crisp clusterings, and the Fuzzy Rand Index was computed to compare the crisp, fuzzy, and customer type solutions on the clean data set.

Table 10 shows that the fuzzy c-means algorithm performs well on the clean data set, although not as well as the k-means algorithm. This is not surprising as the customer types constitute a crisp clustering, and the Fuzzy Rand index tends to give higher values when comparing crisp clusterings.

The effect of adding noise, as discussed in Section 3.3, to the data sets can be seen in Table 11. Both k-means and fuzzy c-means clusterings were computed, with six clusters, and then the clustering was compared to the 11-way known customer type clustering, by calculating the Fuzzy Rand index. Note that, while the Fuzzy Rand index gives the same result as the Rand index when comparing crisp clusterings, it always produces smaller values when applied to fuzzy clusterings. For example, whereas two identical crisp solutions always have a Fuzzy Rand Index of 1.0, using the Fuzzy Rand index to compare two fuzzy clusterings will always produce a value less than 1, even when they are identical.

Table 11 Fuzzy Rand index values for 6-cluster solutions compared with known customer types

Data set	*k-means*	*Fuzzy c-means*
Clean	0.874	0.832
5% noise	0.833	0.776
20% noise	0.783	0.761

Both crisp and fuzzy algorithms show some degradation in the presence of noise at the 5% level, and this effect is significantly more pronounced at the 20% level.

For the data set that included transitional customers, the crisp clustering represented the transitional customers as being split between the clusters corresponding to working and unemployed customers. In the fuzzy clustering, however, they had a fractional membership in each cluster. The fuzzy representation is more meaningful in this case and corresponds better to the actual business scenario. If a number of customers spend half of the time employed and half unemployed, it is more accurate to describe them all as having partial membership of both groups, rather than to arbitrarily categorise some of them as employed and others as unemployed. In order to compare clustering of data including transitional customers, a fuzzy 11-cluster solution was generated from the known customer types, allocating the transitional customers 50% weight in each of the WAF and WAU groups. Moreover, fuzzy clustering would also be expected to quickly detect customer transitions; small shifts in cluster weightings would become apparent soon after the transition occurred and then continue to increase over time.

Table 12 Comparison of fuzzy and crisp clusterings on transitional customers

First clustering	*Second clustering*	*Fuzzy Rand Index*
k-means	Known customer types	0.842
Fuzzy c-means	Known customer types	0.823
k-means	Fuzzy c-means	0. 917

Table 12 shows that both crisp and fuzzy clusterings produce very similar solutions. While the k-means has a slightly better Fuzzy Rand index score, this metric does not reflect the benefit of the fuzzy solution to describe partial membership across multiple clusters.

6 Conclusion

In summary, both crisp and fuzzy clustering algorithms appear to be useful for extracting meaningful groupings from customer interaction information. For the given data sets, the customers were accurately segmented according to their underlying types using common characteristics related to the time, channel, and transaction type of their interactions. These results suggest the techniques could be applied to customer channel interaction data with unknown characteristics with the expectation of drawing substantive conclusions about customer groupings based on their interaction behaviour patterns. For marketing purposes, these groupings – related to communication and lifestyle preferences – could be used instead of existing demographic and transactional-based customer segmentation models. Alternatively, they could be used in conjunction with existing segmentation models to provide new criteria for subdividing existing segments.

Fuzzy clustering was found to be superior for accurately describing customers whose underlying group membership was in flux. Hence, this technique may be better suited for applications where the customers migrate between groups. Using soft competitive learning and genetic algorithms to generate crisp clusters did not, in this particular case, appear to improve the efficiency of the clustering process. However, the soft competitive learning algorithm produced meaningful clustering results that were fairly close to the k-means results.

There are a number of potential business applications that could benefit from these techniques. The first, and most obvious, is to supplement existing customer marketing segmentation models with information derived from customer interaction information. Segments derived from demographic and transaction-based models could be compared with interaction-based clusters to gain additional insights into customer groupings. For example, whereas traditional segmentation tools might group all college students into one segment, clustering based on interaction behaviour could be used to further divide the group into sub-segments of “scholars”, “socialites” and “video gamers”. Specific products and services could then be marketed more effectively to each of these sub-groups.

Another potential application is to market products and services to customers solely based on the customers’ interaction behaviour groupings, ignoring any demographic disparities. For example, if a cluster analysis shows that there is a subset of retirees who behave in a way that is similar to university students, it could be beneficial to treat them as university students from a marketing perspective, even though they do not share the same demographic profile. The key point here is that that the demographic data may be insufficient to fully understand the desires and needs of some customers. Analysis of customer interaction information can provide an alternative angle from which to view customers and their behaviour.

Since fresh customer interaction data are generated continuously, it may be beneficial to perform cluster analysis on a regular basis. In particular, migration of customers between clusters over time could provide companies with a data-driven way of detecting new marketing opportunities. For example, if a customer's interaction pattern shifts from an "employed" cluster to a "retired" cluster, different products or services could be marketed towards them, accordingly. If they migrated to an "unemployed" cluster, collection of payments due could be pursued more aggressively. Alternatively, if cluster analysis showed that certain customers were moving to a cluster that was highly correlated with near-term customer attrition, retention efforts could be increased for those customers. Overall, fuzzy clustering techniques show the greatest potential for being able to identify the movement of customers between clusters.

While the results of this research are quite promising, further validation using real-world data sets is required. Real-world data are expected to differ from the synthetic data in two important ways. First, because companies may not be able to capture or easily aggregate multi-channel customer interaction data, fewer variables may be available for analysis. This consideration would have the most impact on the variable selection and data validation parts of the analysis process, and may affect the accuracy of the results. Second, the data variability is expected to be larger than was simulated in the synthetic data set. Higher variability could potentially lead to greater difficulty in identifying the optimal number of clusters and increased overlap between clusters where fuzzy clustering methods. Hence, interpreting the clustering results could become more challenging with real-world data.

Besides applying these techniques to live customer data, further investigation is also warranted in several areas. First, it would be beneficial to understand the sensitivity of the different clustering techniques to data set size, in order to understand the minimum amount of data required to achieve reasonably accurate results. Second, determining the effects of increasing the channel metrics' granularity – particularly time increments – would be of interest. Finally, determining the advantage provided by taking into account additional interaction details, such as the phone number of inbound calls or the Internet protocol (IP) address of the customer for web sessions, could also increase the strength of this approach.

References

Allred, C., Hite, K., Fonzone, S., Greenspan, J., Larew, J.: Modeling and data analysis in the credit card industry: bankruptcy, fraud, and collections. In: IEEE Systems and Information Design Symposium (2002)

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., Lewis, P.A.: A study of the classification capabilities of neural networks using unsupervised learning - A comparison with K-means clustering. Psychometrika 59(4), 509–525 (1994)

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., Lewis, P.A.: Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. European Journal of Operational Research 93(2), 346–357 (1996)

Bezdek, J.C.: Cluster validity with fuzzy set. Journal of Cybernet 3, 58–72 (1974)

Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. Communication in statistics 3, 1–27 (1974)

Campbell, D., Frei, F.: The persistence of customer profitability: empirical evidence and implications from a financial services firm. Journal of Service Research 7(2), 107–123 (2004)

Campello, R.J.G.B.: A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment. Pattern Recognition Letters 7(28), 833–841 (2007)

Dunn, J.C.: Well separated clusters and optimal fuzzy partitions. Journal of Cybernetica 4, 95–104 (1974)

Edelman, D.B.: An application of cluster analysis in credit control. IMA Journal of Mathematics Applied in Business and Industry 4, 81–87 (1992)

Fritzke, B. (1997), Some competitive learning methods, `http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/research/gsn/JavaPaper/` (accessed July 24, 2009)

Goldberg, R.: Proc. Factor: How to interpret the output of a real-world example. In: SESUG 1997 (1997)

Gordon, A.D.: Classification, 2nd edn. Chapman and Hall, Boca Raton (1999)

Halkidi, M., Batistakis, Y., Vazirgiannis, M.: On clustering validation techniques. Journal of Intelligent Information Systems 17, 107–145 (2001)

Halkidi, M., Batistakis, Y., Vazirgiannis, M.: Cluster validity methods. SIGMOD 31, 40–45 (2002)

Han, J.W., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann Publishers, San Francisco (2001)

Hartigan, J.A.: Clustering algorithms. Wiley, New York (1975)

Hartigan, J.A., Wong, M.A.: A K-means clustering algorithm. Applied Statistics 28, 100–108 (1979)

Hruschka, E.R., Campello, R.J.G.B., Freitas, A.A., De Carvalho, A.C.P.L.F.: A survey of evolutionary algorithms for clustering. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 39(2), 133–155 (2009)

Hitt, L., Frei, F.: Do better customers utilize electronic distribution channels? The case of PC banking. Management Science 48(6), 732–748 (2002)

Kaufman, L., Rousseeuw, P.J.: Finding groups in data. In: An Introduction to cluster analysis. Wiley, New York (1990)

Kohonen, T.: Self-organizing maps. Springer, New York (2001)

MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proc. of the Fifth Berkeley Symposium on Math., pp. 281–297 (1967)

Martinetz, T., Berkovich, S., Schulten, K.: 'Neural-Gas' network for vector quantization and its application to time-series prediction. IEEE Transactions on Neural Networks 4(4), 558–569 (1993)

Nargundkar, S., Olzer, T.J.: An application of cluster analysis in the financial services industry. Case Study (2000)

Nikhil, R.P., James, C.B., Richard, J.H.: Sequential competitive learning and the fuzzy c-means clustering algorithms. Neural Networks 9(5), 787–796 (1996)

Peng, Y., Kou, G., Shi, Y., Chen, Z.: Improving clustering analysis for credit card accounts classification. In: International Conference on Computational Science, pp. 548–553 (2005)

Pison, G., Struyf, A., Rousseeuw, P.J.: Displaying a clustering with CLUSPLOT. Computational Statistics & Data Analysis 30(4), 381–392 (1999)

Rho, J.J., Moon, B.J., Kim, Y.J., Yang, D.H.: Internet customer segmentation using web log data. Journal of Business & Economics Research 2(11), 59–74 (2004)

Sinisalo, J., Salo, J., Karjaluoto, H., Leppaniemi, M.: Mobile customer relationship management: underlying issues and challenges. Business Process Management Journal 13(6), 771–787 (2007)

Tan, P.N., Steinbach, M., Kumar, V.: Introduction to data mining. Addison Wesley, Reading (2005)

Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13(8), 841–847 (1991)

Zakrzewska, D.: On integrating unsupervised and supervised classification for credit risk evaluation. Information Technology and Control 36(1A), 98–102 (2007)

Marketing Intelligent System for Customer Segmentation

Brano Markic and Drazena Tomic

Faculty of Economics, University of Mostar, Bosnia and Herzegovina,
e-mail: brano.markic@tel.net.ba or brano.markic@sve-mo.ba
drazena.tomic@tel.net.ba or drazena.tomic@sve-mo.ba

Abstract. Marketing intelligent system consists of people, procedures, software, databases, and devices that are used in problem-specific decision-making and problem-solving. Marketing intelligent system is an interdisciplinary field that relates to databases, data warehouse, machine learning, expert systems (formalisms of knowledge representation), statistics and operational research and data visualization. The common goal of integrating these different fields is extracting knowledge from data stored in large databases and data warehouses.

Marketing intelligent system uses sophisticated software for satisfaction manager's quires. Software is designated so that its use is relatively simple. Top manager can very quickly receive the essential and key information about the basic economic indicators. Long running education of managers for implementation of marketing intelligent system is unnecessary. Information is short, condensed and visualized.

Marketing intelligent system for customers' segmentation performs useful tasks for marketing researches. They will make marketing researchers more productive allowing doing more work in less time and creating interesting information for marketing decision making. They comprise enough knowledge to react quickly.

In the paper is analyzing and building marketing intelligent system for customers segmentation based on crisp and fuzzy set clustering. Fuzzy logic is a well proven and well established logic of degrees and provides a framework for dealing quantitatively and logically with vague concepts. In fuzzy logic a data point's membership in a set is not crisp (crisp means either in or out of the set) but can be specified as a degree of membership. Fuzzy logic has a wide range of applicability (in clustering, machine learning, expert system, neural networks and decision trees). Marketing intelligent system built in the paper uses fuzzy clustering algorithm and assigns a set of multiple clusters with varying degrees of membership, unlike conventional cluster analysis that assigns a value to a single cluster. Data for customers clustering are stored in relational data warehouse that is temporarily loaded from transactional data bases.

Keywords: marketing intelligent system, fuzzy c-means clustering, market segmentation.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 79–111.
springerlink.com

1 Introduction

Marketing intelligent system for segmentation has the main task to discover the market segment relevance for market decision making. Market segmentation may be based on different criteria but the common goal is to satisfy the needs and desires of different kinds of customers. Market segmentation can be defined as it [2] follows: “market segmentation is the process of dividing up a market into more-or-less homogenous subset for which it is possible to create different value propositions”. In the process of market segmentation the first steps is identifying relevant segmentation variables and analyzing the market. We are concentrated on business to business context (B2B) in determining the customer segments. B2B context differs from business to customer context (B2C) in at least three main characteristics. First, the number of customers in B2B is fewer than in B2C context. Secondly, the relationships between business customers are closer then the relationships between the business and consumer customers (e.g. household). Thirdly, customers on business markets are much larger then customers in context B2C. The main idea of customer segmentation can be graphically illustrated (Figure. 1):

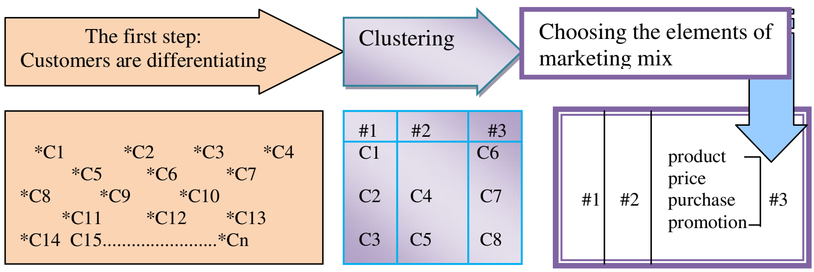


Fig. 1 The basic steps in business customer’s segmentation

Business customers’ (denoted as C1, C2,…,Cn) characteristics are called firmgraphics and may include size, income, number of employees, profitability, liquidity etc. The first step is defining the variables for customers’ segmentation followed by clustering algorithm which allocates the customers to one of segments #1, #2, #3…The customers possess the characteristics that are more closely associated with that segment than any other segment. Now it is possible to choose and design the adequate elements of marketing mix: price, product or service, purchase and promotion to each customer segment.

The main task and role of our marketing intelligent system is permanent surveillance of environment and clustering the customers on business market according to three attributes: amounts of purchasing, the profit and average number of days for payments of bills.

2 Components of Marketing Intelligent System

Marketing intelligent systems have two main characteristics. The first is ability to solve complex tasks at the level of human abilities and knowledge. The second is constant surveillance of environment with the goal to provide the necessary information for decision making in marketing. They make the people's job more productive (to do more in less time). In informatics sense intelligent system today has four interconnecting levels: level of operational data, level of derived data, level of data mining algorithms and visual display level. "To facilitate building market and customer intelligence, it is necessary to have integrated database systems that link together data from sales, marketing, customer, research, operations, and finance" [3]. "An intelligent information system in general is integrating artificial intelligence and database technology [19]. Artificial intelligence aspect concentrates on "intelligent reasoning over data" representing a fraction of the "world". Reasoning enables analyzing of data, check its quality and consistency and reacts to events in the environment. Database concentrates on data representation and storage.

The next Figure illustrates the conceptual model of marketing intelligent system:

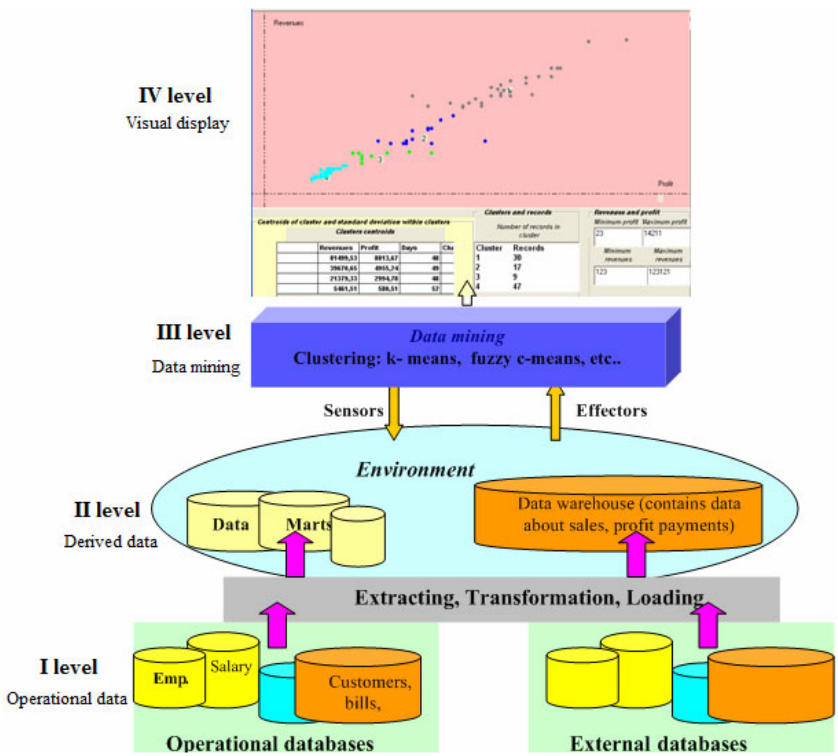


Fig. 2 Marketing intelligent system [15]

The first level represents operational data stored in operational databases. They store data about customers, orders, products, order details, costs etc. In the relation data model these data are represented among related tables.

2.1 Operational Data

Marketing intelligent system deals at the lowest level with operational data. Operational data are detailed data used to run the day-to-day operations of the organization. In most cases relational database is a set of relations (tables) connected with foreign keys. Thus, a data table consists of a number of columns (attributes). Every attribute describes a property of entity type. Several such tables are related by means of a common attribute found in two or more data tables. The common attribute must be defined with same domains and their names must be spelled alike. The entity relation model for our marketing intelligent system is shown in Figure 3.:

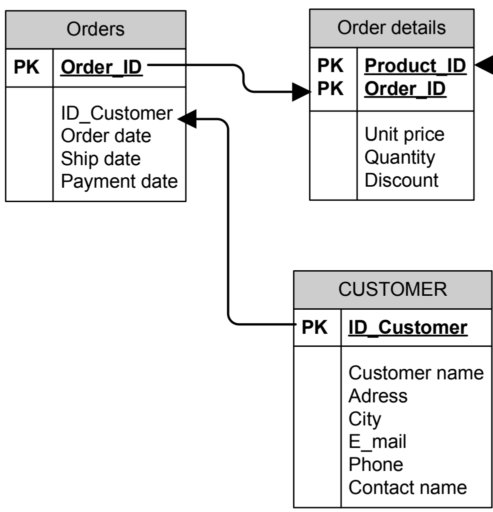


Fig. 3 Entity relationship model

In our model there are four tables:

Orders(OrderID, ID_Customer, Order date, Ship date, Payment date);
Order_details(Produc_ID, Order_ID, Unit price, Quantitiy, Discount);
Products(Product_ID, Product name, Unit price, Costs);
Customer(ID_Customer, Customer name, Address, City, E_mail, Phone, Contact name).

The difference between payment date and ship date is the interval within which the customer pays the bill. These data will be used later for customer segmentation.

2.2 Derived Data

Namely, for business operation and analyses it is necessary to differentiate two kinds of data:

- primitive or operational data
- derived or decision support (DSS) data.

Derived data are data that are summarized or otherwise calculated to meet the needs of the management of the organization.

Because there is a host of differences between primitive and derived data, prevail opinion is that both primitive and derived data would not fit in a single database. Foundation of the data warehouse concept is a separation of day-to-day operations production applications from operation of analyzing and reporting which are done by analysts or managers.

2.2.1. Data Warehouse – Foundation for Quality Data

Data warehouse concept is no revolutionary, but evolutionary one. A data warehouse serves as a central repository for recording historical data about the entire business. The data are pulled from many sources, including internal database and external databases.

Internal database may be information prepared by planning, sales, or marketing organizations that contain data such as budgets, forecasts, revenues, or sales quotas (see Figure 4.). Internal database must be treated like any other source system data. It must be transformed, documented in metadata, and mapped between the source and target databases. External database is useful for competitive analysis and marketing research. External database is important if one wants to compare the performance of its business against others.

The importance of quality data in the data warehouse cannot be over-emphasized. Kimball [11] states data staging process as the key part of the data warehouse project which includes set of processes that clean, transform, combine, de-duplicate, household, archive and prepare source data for use in the data warehouse. It is commonly called ETL (Extraction, Transformation, and Loading) describing the series of processes that, as it is obvious from Figure 4, do the following:

- Detect the changes made to source data required for the warehouse,
- Move data from source systems,
- Clean up and transform the data,
- Restructure keys,
- Index and summarize the data,
- Maintain the metadata,
- Load the data into the warehouse (analytical data).

These processes are absolutely fundamental in ensuring that the data resident in the warehouse are: required by and useful to the business users, good quality to ensure good information, accurate to ensure accurate information and easy to access so that the warehouse is used efficiently and effectively by the business users.

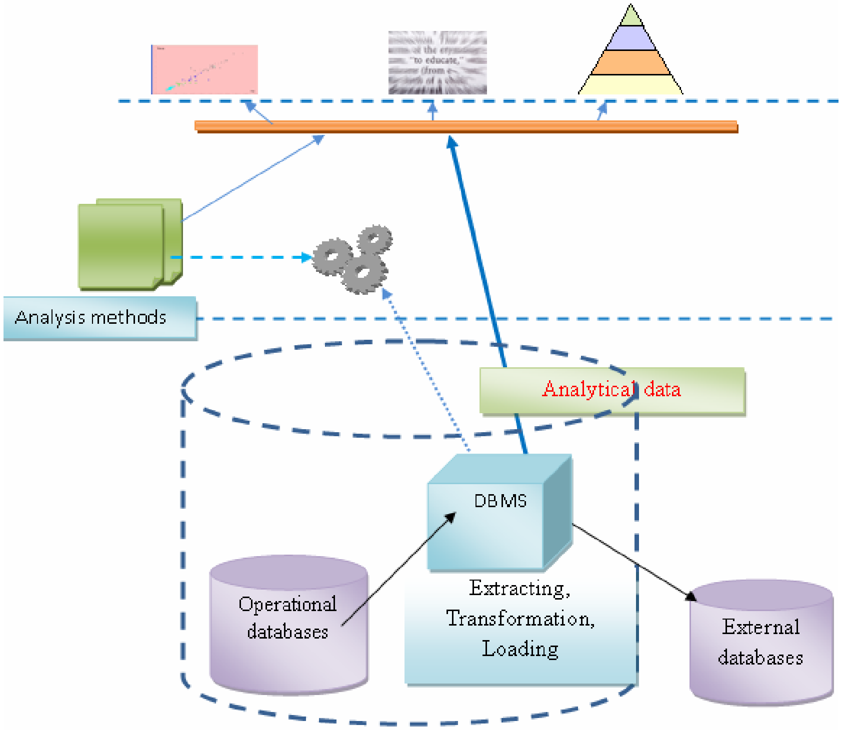


Fig. 4 Data staging (ETL) process

Former review of basic activities related to data staging (ETL) process stressed complexity, volume and importance of quality data providing. It is essential that these (ETL) activities are regular component of data warehouse building and maintenance. It means that they are performed on regular basis according to defined rules and procedures, in contrast to classical data extract processing in DSS (Decision Support System) and other systems, which were performed on irregular basis depending on particular decision makers needs.

Owing to ETL process the warehouse becomes a single distribution point of information for the enterprise and other levels of the organization, through feeds into data marts and desktop applications.

2.2.1.1 Data Warehouse for Market Segmentation

Data model for passing data from transaction database to data warehouse is represented by entity relationship model. Entity relationship model (Figure 3.) includes four relations tables:

Orders, Order details, Products and Customer.

Operational data are stored within these four tables and are updated every day. The process of developing marketing intelligent system inherently consists of several steps. The first step is *to understand the application domain and to formulate* *the problem.* Our application domain is segmentation of customers according to appropriate attributes. This research field is common for many disciplines in economy such as marketing, accounting, business informatics and customer relationship management. Knowledge from these disciplines must be integrated to successfully solve the research problem. Primarily the segmentation of customers may be assigned to customer relationship management because it faces the problem of customers clustering with the final goal to formulate adequate price discounts and payments as key element of the contracts between supplier and customer.

This step is clearly a prerequisite for extracting useful knowledge and for choosing appropriate data mining methods in the third step according to the application target and the nature of data.

The second step is *to collect and preprocess the data*, including selection of the data sources, removal of noise or outliers, treatment of missing data, the transformation (discretization if necessary) and reduction of data, etc.

a) Dimension Centroid

Physical Name	Data Type	Req'd	PK	Notes
ID_Cluster	BYTE	☑	☑	ID_Cluster identifies Centroid
Revenues	SINGLE	☐	☐	Revenues is of Centroid
Profit	SINGLE	☐	☐	Profit is of Centroid
Days_of_payment	INTEGER	☐	☐	Days_of_payment is of Centroid

b) *Dimension Customer*

Physical Name	Data Type	Req'd	PK	Notes
ID_Customer	INTEGER	☑	☑	ID_Customer identifies Customer
Customer name	CHAR(15)	☐	☐	Customer name is of Customer
Region	CHAR(20)	☐	☐	Region is of Customer
City	CHAR(15)	☐	☐	City is of Customer
Country	CHAR(15)	☐	☐	Country is of Customer

c) Dimension Time

Physical Name	Data Type	Req'd	PK	Notes
ID_Time	INTEGER	☑	☑	ID_Time identifies Time
Week	INTEGER	☐	☐	Week is of Time
Month	INTEGER	☐	☐	Month is of Time
Year	BYTE	☐	☐	Year is of Time

d) Fact table – Customer-cluster

Physical Name	Data Type	Req'd	PK	Notes
ID_Customer	INTEGER	☑	☑	ID_Customer identifies Customer_cluster
ID_Cluster	BYTE	☑	☑	ID_Cluster partly identifies
ID-Time	DATETIME	☑	☑	ID-Time partly identifies Customer_cluster
Amount_of_sale	SINGLE	☐	☐	Amount_of_sale is of Customer_cluster
Profit	SINGLE	☐	☐	Profit is of Customer_cluster
Days_of_payment	INTEGER	☐	☐	Days_of_payment is of Customer_cluster

Fig. 5 Definition of dimensions tables and fact table

Defining three dimensions and fact table with data warehouse is represented in details in the next Figure:

After definition schema of tables and data types we propose the star data warehouse architecture model. Figure 6. presents data mapping from transaction database to data warehouse model that makes base for example of cluster analysis.

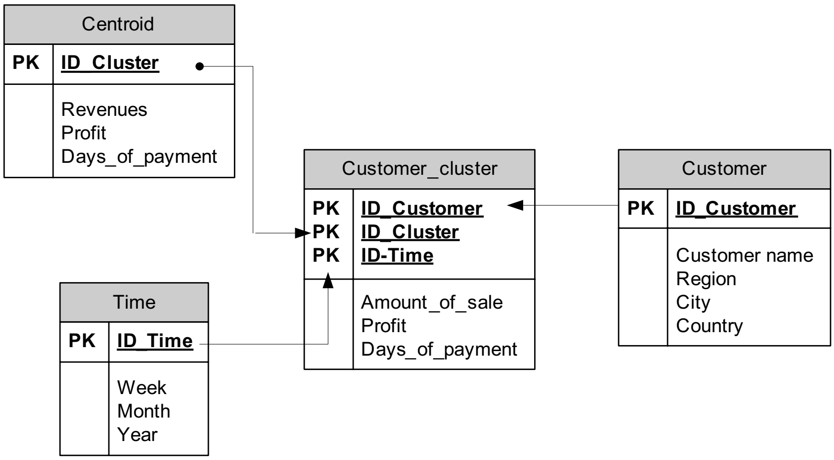


Fig. 6 Data warehouse model – base for cluster analysis

Data for customers' clustering are stored in relational data warehouse that is temporarily loaded from transactional data bases. The fact table in data warehouse named as Customers_cluster includes three measures: sales, profit and days of payments. Fact table is populated from data model using SQL *select* statement and corresponding aggregate function.

Data warehouse includes three dimensions: *Customer, Time* and *Centroid* (cluster). Dimension *Customer* is table Customer from transactional databases. Fact table can be populated by aggregate functions. Attribute "Sales" is sum of sold products $\sum_{i=1}^{n}(\text{Quantity} * \text{Unit price})$ for given customer for given time period (for example week, month or year), attribute "Profit" is sum $\sum_{i=1}^{n}(\text{Quantity} * (\text{Unit price - Costs})$ for all products sold to given customer in defined time period and attribute "Days_of_payment" is average number of days counted as quotient between the sum $\sum_{i=1}^{m}(\text{Payment date - Order date})$ and number of Orders m. This step usually takes the most time needed for the whole process of market segmentation.

3 Customers' Segmentation Using Partitioning Method

The third step in the process of market segmentation and building marketing intelligent information system is implementing adequate data mining algorithms [24] that extract patterns or models from data which are hidden in data warehouse.

A model can be viewed as "a global representation of a structure that summarizes the systematic component underlying the data or that describes how the data may have arisen" [7]. In contrast, "a pattern is a local structure, perhaps relating to just a handful of variables and a few cases". The major classes o*f data mining methods [6] are predictive modeling such as classification* and *regression, segmentation (clustering), dependency modeling such as* graphical models or density estimation, *summarization* such as finding the relations between fields, associations, visualization, *change and deviation detection/modeling* in data and knowledge.

Clustering is the process of grouping the data into classes (clusters) so that the data objects (examples) are similar to one another within the same cluster and dissimilar to the objects in other clusters. A *good clustering* method will produce high quality clusters with high *intra-class* similarity and low *inter-class* similarity.

Simple and very popular clustering algorithm is k-Means clustering algorithm [22].

It is an *iterative distance-based* clustering method where the number of clusters *k* has to be specified in advance. It can be implemented in four steps (Figure 7. is flowchart of k-means clustering):

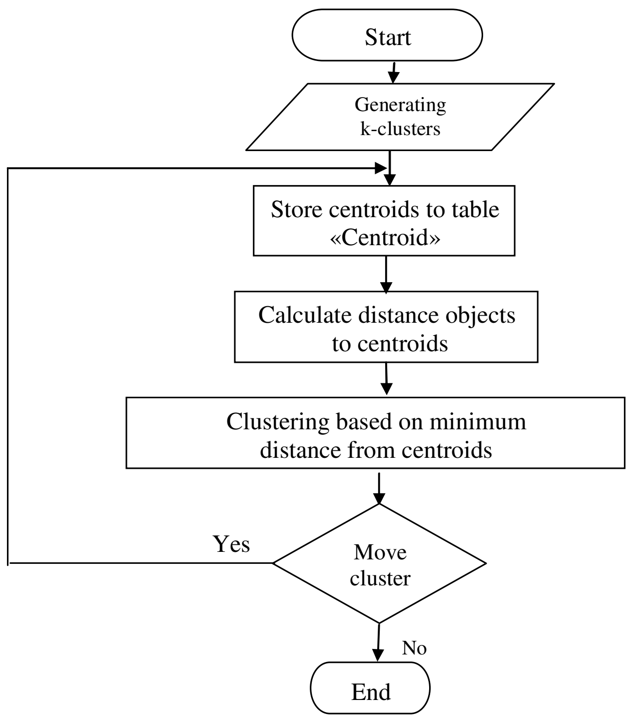


Fig. 7 Step by step k-means clustering algorithm

1. Generating k-clusters. Choose *k* seeds (vectors with the same dimensionality as the input.

examples. The first *k* examples are selected as seeds.

2. Apply an example, calculate the distance from it to all seeds and assign it to the cluster with the nearest seed point.

3. At the end of each epoch compute the centroid (means) of the clusters.

4. If the stopping criteria is satisfied (no changes in the assignment of the examples or max number of epochs reached), stop. Otherwise, repeat 2 and 3 with the new centroids taking the role of the seeds.

A good clustering method will produce high quality clusters with high intra-class similarity and low inter-class similarity. The similarity is measured using a distance function e.g. David-Bouldin index (DB) – a heuristic measure of the quality of clustering [4]:

$$DB = \frac{1}{c} * \sum_{i=1}^{c} \max_{j \neq i} = \frac{D(x_i) + D(x_j)}{D(x_i, x_j)}$$

DataKlaster (Customer_cluster)

ID_Customer	Amount_of_sale	Profit	Days_of_payment	ID_Cluster
1	234	24	30	1
2	23.456	3.451	45	2
3	1.341	45	56	1
4	1.245	89	23	1
5	123	23	45	1
6	5.644	458	30	1
7	14.351	2.340	55	2
8	755	70	65	1
9	894	99	75	1
10	7.777	698	45	1
11	34.432	7.890	45	3
12	24.251	4.352	45	2
……	………..	…….	……….	……….
……..	………..	…….	……….	……….

- c – number of clusters,
- $D(x_i)$ –mean-squared distance from the points in the cluster i to the center,
- $D(x_i,x_j)$ – distance between the centers of cluster i and j.

Data in fact table of data warehouse for customers clustering are stored in Access or in any other database management system. The fact table has the following schema and the part of data looks like this:

There is about 1500 customers stored in table Customer_cluster. The first step is randomly choosing *k* rows from Custome_cluster table as initial clusters (the working name for table Customer_cluster is DataKlaster).

The next code (Program 1.) describes one solution for choosing randomly *k* Centroids:

```
Dim Izbor As Integer
rsDataKlaster.MoveFirst // The first record in DataSet.
   For j = 1 To Val(Trim$(Text1.Text)) //Text box Text1.Text preserves
                                              // the number of cluster.
         DoEvents
         Randomize
         Izbor = Int(Rnd(j) * rc)
                  If  Izbor = 0 Then
                  Izbor = 1
                  End If
rsCentroid.AddNew
rsDataKlaster.Filter = "ID_Customer = " & Izbor // Record Izbor is choosen.
         If Izbor <> 0 Then
                  rsCentroid! ID_Customer = rsDataKlaster! ID_Customer
                  rsCentroid! Revenues = rsDataKlaster!Revenues
                  rsCentroid! Profit = rsDataKlaster!Profit
                  rsCentroid! Days = rsDataKlaster!Days
                   rsCentroid!Klaster = j
         End If
Next j
rsCentroid.Update
```

Program 1. Randomly choose k clusters from n tubles in Data set

After creating *k* clusters every customer will be clustered calculating the minimum Euclidian distance [10] from the center of clusters. The next code written in Visual Basic (Program 2.) reflects the idea if minimum distance[1]:

[1] Generally speaking Euclidian distance between x and y is calculated using this formula:

$$d(x,y) = \left(\left(x_1 - y_1\right)^2 + \left((x_2 - y_2\right)^2 + \ldots\left(x_d - y_d\right)^2\right)^{\frac{1}{2}}.$$

```
  min = 10 ^ 10 // Assigns a very huge number to variable
min.
      For DK = 1 To Val(Trim$(Text1.Text)) //Text1.Text preserves the
number of cluster.
    xc = rsCentroid!Revenues // Variable xc stores the values in column
Revenues in table Centroid.
            yc = rsCentroid!Profit // Variable yc stores the values in column
        Profit in table Centroid.
            zc= rsCentroid!Days //// Variable zc stores the values in column
        Days in table Centroid.
    d = Sqr( ((Xp - xc) ^2 + (Yp - yc) ^2+ (Zp – zc) ^2), 2)) // variable d
stores the distance from Centroid.
            If d < min Then
                  min = d
                  Kl = DK
            End If
            rsCentroid.MoveNext
Next DK
```

Program 2. Calculating the minimum distance from Centroid

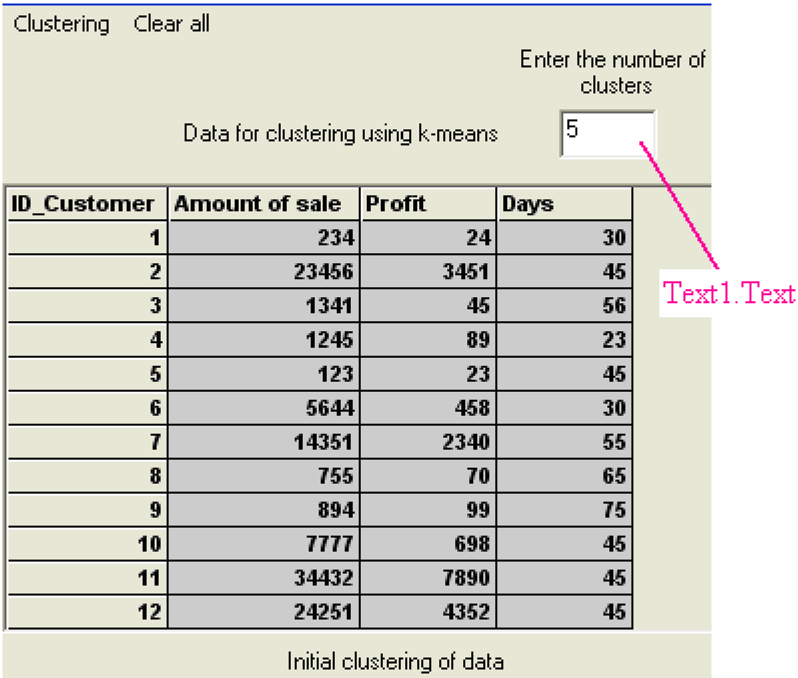
The next step is calculating new centroids and finding out if any of customers (tuples in DataKlaster table) leaves cluster. If there is moving a customer among the clusters then the process of calculating centroids and minimum distance has to continue what clearly presents the next code (Program 3.):

```
        Dim SK As Byte, a, m, DK, Kl
        Dim Xp As Single, PZbroj, RZbroj, Yp, xc, yc, min
        Dim rsD As Integer, rsC, P, podijeli, k
        Dim Komp As Boolean
        Dim rsDKlaster As adodb.Recordset
        Set rsDKlaster = New adodb.Recordset
             If cn.State <> adStateOpen Then
              cn.Open
            End If
            If rsDKlaster.State <> adStateOpen Then
                 rsDKlaster.Open "DataKlaster", cn, adOpenKeyset, adLockOptimistic
            End If
           If rsCentroid.State <> adStateOpen Then
               rsCentroid.Open "Centroid", cn, adOpenKeyset, adLockOptimistic,
              adCmdTable
           End If
        Komp = True
        k = 0
          Do While Komp = True
                 If rsDKlaster.AbsolutePosition <> rsDKlaster.BOF Then
                 rsDKlaster.MoveFirst
                 End If
           Komp = False
            For P = 1 To rsDKlaster.RecordCount
                 If rsDKlaster.AbsolutePosition = rsDKlaster.EOF Then
                     Exit For
                 End If
                  SK = 0,   Xp = 0,  Yp = 0
                  SK = rsDKlaster!Cluster
                  Xp = rsDKlaster!Revenues
                  Yp = rsDKlaster!Profit
                 If rsCentroid.AbsolutePosition <> rsCentroid.BOF Then
                          rsCentroid.MoveFirst
                 End If
min = 10 ^ 10
  For DK = 1 To Val(Trim$(Text1.Text))
        xc = rsCentroid!Revenues,  yc = rsCentroid!Profit  zc= rsCentroid!Days
        d = Sqr( ((Xp - xc) ^2 + (Yp - yc) ^2+ (Zp – zc) ^2), 2))
        If d < min Then
            min = d
            Kl = DK
        End If
      rsCentroid.MoveNext
```

```
                'MsgBox ("Is the cluster being changed")
                If Kl <> SK Then
                        rsDKlaster!Cluster = Kl
                        Komp = True
                        Else
                        'MsgBox ("Cluster does not change")
                End If
                rsDKlaster.Update
                rsDKlaster.MoveNext
            Next P
        rsCentroid.MoveFirst
                For m = 1 To Val(Trim$(Text1.Text))
                    PZbroj = 0
                   RZbroj = 0
                   podijeli = 0
                   If rsDKlaster.AbsolutePosition <> rsDKlaster.BOF Then
                   rsDKlaster.MoveFirst
                   End If
            Do While Not rsDKlaster.EOF
                     If rsDKlaster!Cluster = m Then
                       PZbroj = PZbroj + rsDKlaster!Revenues
                             'MsgBox ("Summing the revenues for cluster m")
                             RZbroj = RZbroj + rsDKlaster!Profit
                             podijeli = podijeli + 1
                      End If
                      rsDKlaster.MoveNext
             Loop
                   If podijeli <> 0 Then
                            rsCentroid!Revenues = PZbroj / podijeli
                            rsCentroid!Profit = RZbroj / podijeli
                   End If
                   If rsCentroid.AbsolutePosition <> rsCentroid.EOF Then
                               rsCentroid.MoveNext
                   End If
                Next m
        MsgBox ("I am at the end of program. Again from Begin")
        MsgBox ("The value for  komp is " & Komp)
        k = k + 1
    Loop
        Label5.Caption = "The number of pass thru loop is: " & k
```

Program 3. Moving a customer record among the clusters

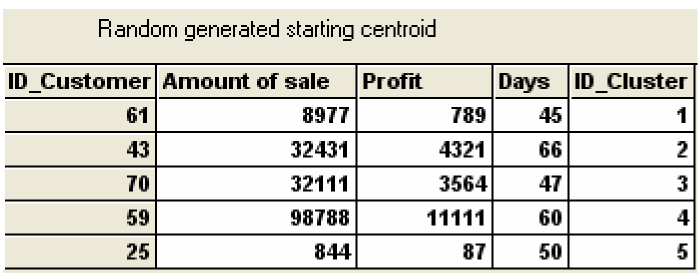
Program 1, program 2 and program 3 are the core of k-means clustering algorithm. Data in fact table includes four attributes: ID_Customer, Amount_of_sale, Profit and Days (average number of days necessary that one customer pays received bill). All customers must be cluster according to values of three attributes Amount_of_sale, Profit and Days (Figure 8 - initial clustering of data).



ID_Customer	Amount of sale	Profit	Days
1	234	24	30
2	23456	3451	45
3	1341	45	56
4	1245	89	23
5	123	23	45
6	5644	458	30
7	14351	2340	55
8	755	70	65
9	894	99	75
10	7777	698	45
11	34432	7890	45
12	24251	4352	45

The part of dataset (only 12 records) is presented. For example, the customer ID_Customer=8 has purchased 755 money units, the profit for supplier is 70 and the customer pays the bills in 65 days in average.

The user determines the number of clusters (in our session five clusters) from ***n*** tuples and each of five tuples initially represents a cluster mean or center (Figure 8 – randomly generated starting centroid).



ID_Customer	Amount of sale	Profit	Days	ID_Cluster
61	8977	789	45	1
43	32431	4321	66	2
70	32111	3564	47	3
59	98788	11111	60	4
25	844	87	50	5

For example, the tuple (record) with attributes values ID_Customer=70, Amount_of_sale=32111, Profit=3564 and Days=47 defines the cluster 3. The values of these attributes are the center of cluster 3 (ID_Cluster=3).

ID_Customer	Revenues	Profit	Days	ID_Cluster
1	234	24	30	5
2	23456	3451	45	3
3	1341	45	56	5
4	1245	89	23	5
5	123	23	45	5
6	5644	458	30	1
7	14351	2340	55	1
8	755	70	65	5
9	894	99	75	5
10	7777	698	45	1
11	34432	7890	45	2
12	24251	4352	45	3
13	969	100	45	5
14	1267	150	30	5

Each customer is assigned to one cluster.

Fig. 8 Initially generated starting centroid and customers assigned to cluster

It is necessary to assign each of the remaining tuples in dataset to one most similar cluster (Figure 8 – each customer is assigned to one cluster). This similarity is based on the Euclidian distance between the tuples and the cluster mean. After one execution of loop for assigning each customer to one cluster, it is necessary again to compute the new mean of each cluster (customer is presented as one row in dataset). Mapping between customers and clusters is N:1. That means one, two or more customers may belong to the same cluster and one cluster is assigned only to one. The end of the loop for assigning the customer to clusters is reached when no movement of customers to clusters exists. Such state is final clustering state or the result of k-means clustering (see Figure 9).

In Figure 9 different colors represent different customers' clusters. It is visible that in cluster 3 are there are 21 customers, in cluster 4 only 10 customers and so on. Now it is possible to make corresponding contracts with customers depending on cluster that is assigned to customer, choose and design the adequate elements of marketing mix: price, product or service, purchase and promotion to each customer segment (cluster).

The fourth step in building intelligent information systems is *to interpret (post-process) discovered knowledge*, the interpretation in terms of description and prediction. Experiment with k-means clustering shows that discovered patterns from data are not always of interest or direct use. From business computing point of view of it is possible to build expert systems for judgment of discovered knowledge.

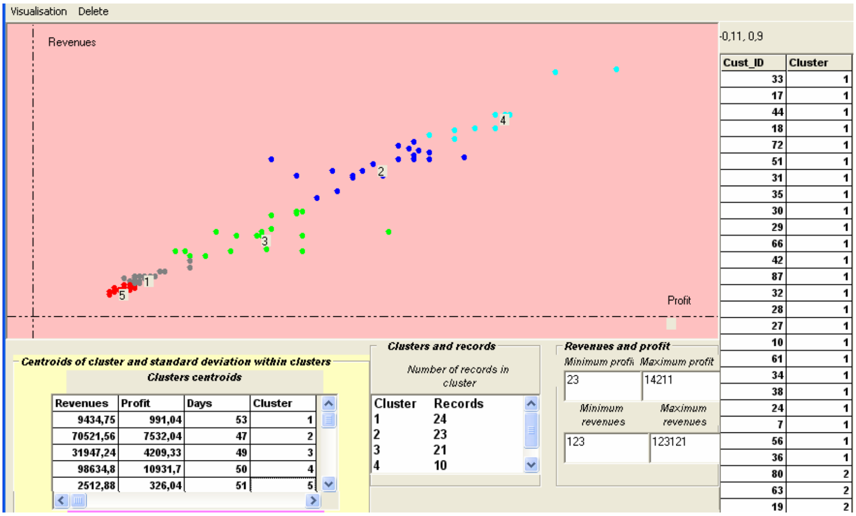


Fig. 9 The results of k-means clustering

3.1 Interpretation of Intelligent Systems Outputs by Expert Systems

An expert system is a computer program that contains stored knowledge and solves problems in a specific field in much the same way in which a human expert would [20]. One of the main problems and most difficult tasks in building rule based expert systems is representing the knowledge discovered by k-means clustering.

The knowledge typically comes from a series of conversations between the developer of the expert system and one or more experts. A peculiarity of expert systems for interpretation the results of k-means clustering is that the knowledge comes from two sources. The first source is dimensions of centroids in clusters (dimensions are revenues, profit and days of payment) and the second source is manager. After analyzing the clusters centroids (see Figure 9) the managers proposes that the distance between the centroids of neighbored clusters are the base for further customers grouping and preparing for signing delivering contracts of goods or services and defining elements of marketing mix.

The judgment of discovered knowledge (clustering each one customer) will be performed by building rule based expert systems (see 4. Collaboration knowledge based system and the fuzzy c-means clustering implementation results). The completed system applies the knowledge of customers clustering.

The format that a knowledge engineer uses to capture the knowledge is called a knowledge representation. The most popular knowledge representation is the production rule (also called the if-then rule). Production rules intend to reflect the "rules of thumb" that experts use in their day-to-day work. These rules of thumb are also referred to as heuristics. A knowledge base that consists of rules is sometimes called a rule base.

Except the executable code in the form of production rules, the knowledge could be represented by decision trees with four levels.

4 Marketing Intelligent Sytems for Customers Clustering Using Fuzzy C-Means Clustering

Clustering is a method that will produce high quality clusters with high *intra-class* similarity and low *inter-class* similarity.

Hard k-means algorithm executes a sharp clustering, in which each object is either assigned to a cluster or not [15].

The k-means algorithm partitions a set of N vector into **c** clusters (clusters G_i, i=1,..,c). The goal is finding cluster centers (centroids) for each cluster. The algorithm minimizes a dissimilarity (or distance) function which is given in Equation 1.

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{k, x_k \in G_i} d(x_k - c_i) \tag{1}$$

c_i is the centroid of cluster i;
$d(x_k - c_i)$ is the distance between i_{th} centroid(c_i) and k_{th} data point;
Overall dissimilarity function is expressed as in Equation 2

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k, x_k \in G_i} \| x_k - c_i \|^2 \right) \tag{2}$$

Partitioned groups can be defined by a binary membership matrix(u), where the element u_{ij} is 1 if the j_{th} data point x_j belongs to cluster i, and 0 otherwise (Equation 3)

$$u_{ij} = \begin{cases} 1 & if \| x_j - c_i \|^2 \leqslant \| x_j - c_k \|^2, for\ each\ k \neq i, \\ 0 & otherwise \end{cases} \tag{3}$$

Since a record can only be in a cluster, the membership matrix (U) has two conditions which are given in equation 4 and equation 5.

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1, \ldots, n \tag{4}$$

$$\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} = n \tag{5}$$

Centroids are computed as the mean of all vectors in group i:

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \tag{6}$$

$|G_i|$ is the size of G_i.

The software solution for k-means algorithm and all necessary steps [22] does not guarantee for an optimum solution steps with a recordset x_j, j=1,..n. The performance of the algorithm depends on the initial positions of centroids.

Fuzzy clustering allows that one tuple belongs at the same time to several clusters but with different degrees. This is an important feature for segmentation business markets to increase the sensitivity. Fuzzy c-means clustering was developed by Dunn [4] and improved by Bezdek [1] and is separated from hard k-means that employs hard partitioning. Fuzzy partitioning a tuple (fact table record in data warehouse) can belong to all groups with different membership grades between 0 and 1. Fuzzy c-means is an iterative algorithm. The aim of fuzzy c-means is to find cluster centers (centroids) that minimize a dissimilarity function. We will present fuzzy c-means clustering algorithm as a sequence of unambiguities and executable steps written in artificial programming language Visual Basic.

1. The first step is random generating of the clusters. The clusters are chosen again from fact table (table name is DataKlasterF). Seven seed clusters are chosen from fact table DataKlasterF. The next figure shows the random generated starting centroid. For example, Cluster 5 (ID_Cluster=5) is chosen from row 20 (ID_Customer=20) and the values of other three attributes are Amount_of_sale=74839, Profit=9899 and Days=75.

Random generated starting centroid

ID_Customer	Amount of sale	Profit	Days	ID_Cluster
30	7282	867	75	1
21	23145	2142	60	2
101	2314	546	45	3
39	3432	453	45	4
20	74839	9899	75	5
2	23456	3451	45	6
29	10234	1324	66	7

2. The second step is calculating the membership matrix(u) according to Equation 7:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,\dots,n \tag{7}$$

The dissimilarity function which is used in fuzzy c-means clustering is given in Equation 8

$$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}{}^m d_{ij}{}^2 \tag{8}$$

u_{ij} is between 0 and 1;
c_i is the centroid of cluster i;
d_{ij} is the Euclidian distance between i_{th} centroid(c_i) and j_{th} data point;
m ϵ $[1,\infty]$ is a weighting exponent.

To reach a minimum of dissimilarity function there are two conditions. These are given in Equation 9 and Equation 10.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}{}^m x_j}{\sum_{j=1}^{n} u_{ij}{}^m} \tag{9}$$

$$m(i, j) = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^2} \tag{10}$$

The next code (Program 4) written in Visual Basic implements the equation 10 (membership matrix m(i,j))[2]:

The membership matrix m(i,j) for the first twelve records is presented by Figure 10. For example, the customer with the value of attribute ID_Customer=2 only belongs to cluster 6 (membership value is 1 or 100%), while the customer with identification number ID_Customer=9 belongs to cluster 1 : 3,758%, cluster 2: 0,0312%, cluster 3: 70,164%, cluster 4: 23,686%, cluster 5: 0,032%, cluster 5: 0,028%, cluster 6: 0,299% and cluster 7: 1,753%. It is simple to conclude that customer with ID_Customer=9 belongs the most to cluster 3. Now it is easy to read the all values at Figure 10 and to make adequate conclusions.

The sum of all values of membership function for one customer is always 1. If we sum these values for our ID_Customer=7:

3,758% + 0,0312%+70,164% + 23,686% + 0,032%+0,028%+0,299%+1,753% = 100%;

The result is 100% or as coefficient the result is 1.

The membership matrix is calculated for seven randomly chosen clusters as centroids satisfy constrains

$$\sum_{j=1}^{n} m(i, j) = 1 \quad \forall j = 1,2,\ldots n. \tag{11}$$

[2] Visual Basic in our examples very often implements the array as data structure. Such m(i,j) is a two dimensional array with *i* rows and *j* columns.

```
rsDataKlasterF.MoveFirst  // rsDataKlasterF is a recordset
 ReDim Preserve m(rc, Val(Trim$(Text1.Text)))
 ReDim Preserve a(Val(Trim$(Text1.Text)))
 For i = 1 To rc
  X = rsDataKlasterF!Revenues,   Y = rsDataKlasterF!Profit,
  Z = rsDataKlasterF!Days
   rsCentroidF.MoveFirst
                For j = 1 To Val(Trim$(Text1.Text))
                        xc = rsCentroidF! Revenues
                        yc = rsCentroidF! Profit
                        zc = rsCentroidF!Days
    a(j) = Round(Round((X - xc) ^ 2, 0) + Round((Y - yc) ^ 2, 0) + (Z - zc) ^ 2, 0)
                      rsCentroidF.MoveNext
                Next j
     Dim S As Single
     Dim k As Integer
      For j = 1 To Val(Trim$(Text1.Text))
          S = 0
                    For k = 1 To Val(Trim$(Text1.Text))
                            If a(k) = 0 Then
                                     m(i, k) = 1
                                     S = 0
                                    Exit For
                             Else
                                    S = S + Round(a(j) / a(k), 5)
                            End If
                    Next k
                    If S <> 0 Then
                             m(i, j) = Round(1 / S, 5)
                   End If
    Next j
          rsDataKlasterF.MoveNext
 Next i
```

Program 4. Membership matrix for customers' records

	Fuzzy membership						
Record	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
1	0,05719	0,00544	0,62653	0,27676	0,00051	0,00523	0,02834
2	0	0	0	0	0	1	0
3	0,02528	0,0019	0,75921	0,20036	0,00017	0,00182	0,01127
4	0,02736	0,0021	0,74981	0,20623	0,00018	0,00201	0,01231
5	0,06021	0,00585	0,61665	0,28101	0,00055	0,00563	0,03011
6	0,5027	0,00464	0,12922	0,29304	0,00029	0,0044	0,06572
7	0,16711	0,11262	0,05883	0,07096	0,00234	0,10356	0,48456
8	0,04196	0,00359	0,68274	0,24809	0,00032	0,00344	0,01985
9	0,03758	0,00312	0,70164	0,23686	0,00028	0,00299	0,01753
10	0,93575	0,00108	0,0086	0,01356	0,00006	0,00101	0,03995
11	0,06783	0,33251	0,04914	0,05249	0,03259	0,38057	0,08486
12	0,00383	0,18838	0,00232	0,00256	0,00044	0,79686	0,00559

Fig. 10 The membership matrix m(i,j)

3. The third step is calculating the center of centroids (new centroids of clusters) using membership matrix m(i,j) and values for profits, sales and days of payments:

$$c_j = \frac{\sum_{i=1}^{n} m(i,j)^2 * X_i}{\sum_{i=1}^{n} m(i,j)^2}, \tag{12}$$

m(i,j) is between 0 and 1; c_j is the centroid of cluster j;

After calculating the new centroid it is necessary to calculate the distance between the cluster center and every record. Stop if its improvement over previous iteration is below a threshold. Stop condition in this example is defined by statement:

```
If Abs(pC(j, 1) - rsCentroidF!promet) < 60 And Abs(rucC(j, 2) –
    rsCentroidF!ruc) < 6 And   Abs(dopC(j, 3) - rsCentroidF!dop)
    < 1 Then
                nastavi = False
      Else
                nastavi = True
      End If
```

or generally do while $\sum_{j=1}^{c} \left\| c_j^{Previous} - c_j \right\| > \varepsilon$.

By iteratively updating the cluster centers and the membership matrix [9] for each record, fuzzy c-means iteratively moves the cluster centers to the "right center" within data records.

Cluster centers (centroids) initialized using randomly selected records and fuzzy c-means does not ensure that it converges to an optimal solution. Performance depends on initial centroids and may be improved in two ways:

1) Using an algorithm to determine all of the centroids (for example: arithmetic means of all records).

2) Run fuzzy c-means several times each starting with different initial centroids.

We preferred the second approach.

After calculating the new centroid it is necessary to calculate the distance between the cluster center and every record. Stop if its improvement over previous iteration is below a threshold ε.

Stop condition in our example is defined by statement:

```
If Abs(pC(j, 1) - rsCentroidF!Revenues) < 60 And Abs(rucC(j, 2) –
     rsCentroidF!Profit) < 6 And   Abs(dopC(j, 3) - rsCentroidF!Days)
     < 1
     Then
             nastavi = False
     Else
            nastavi = True
     End If
```

or generally

do while ($\sum_{j=1}^{c} \left\| c_j^{Previous} - c_j \right\| > \varepsilon$).

The next code (Program 5) implements calculating new centroids, test the conditions and if the stop condition is not satisfied calculate the new membership matrix.

```
Dim Brojnik() As Single, Nazivnik()
Dim nastavi As Boolean, Dim Xr() As Single
ReDim Xr(Val(Trim(Text1.Text)), 3)
ReDim Brojnik(Val(Trim(Text1.Text)), 3)
ReDim Nazivnik(Val(Trim(Text1.Text)), 3)
  Dim p As Single, ruc, dop, Dim pC() As Single,   Dim rucC() As Single,
 Dim dopC() As Single
nastavi = True
Do While nastavi
rsCentroidF.MoveFirst
For j = 1 To Val(Trim(Text1.Text))
                Brojnik(j, 1) = 0,  Nazivnik(j, 1) = 0, Brojnik(j, 2) = 0, Nazivnik(j, 2) = 0
                Brojnik(j, 3) = 0,  Nazivnik(j, 3) = 0
                rsDataKlasterF.MoveFirst
        For i = 1 To rc
                Xr(j, 1) = rsDataKlasterF!Revenues, Xr(j, 2) = rsDataKlasterF!Profit
                Xr(j, 3) = rsDataKlasterF!Days
                 Brojnik(j, 1) = Brojnik(j, 1) + Round(m(i, j) ^ 2 * Xr(j, 1), 2)
                 Nazivnik(j, 1) = Round(Nazivnik(j, 1) + m(i, j) ^ 2, 3)
                 Brojnik(j, 2) = Round(Brojnik(j, 2) + m(i, j) ^ 2 * Xr(j, 2), 3)
                 Nazivnik(j, 2) = Round(Nazivnik(j, 2) + m(i, j) ^ 2, 3)
                 Brojnik(j, 3) = Round(Brojnik(j, 3) + m(i, j) ^ 2 * Xr(j, 3), 3)
                 Nazivnik(j, 3) = Round(Nazivnik(j, 3) + m(i, j) ^ 2, 3)
                 rsDataKlasterF.MoveNext
         Next i
   Next j
```

```
ReDim Preserve pC(Val(Trim(Text1.Text)), 3)
                ReDim Preserve rucC(Val(Trim(Text1.Text)), 3)
                ReDim Preserve dopC(Val(Trim(Text1.Text)), 3)
                pC(j, 1) = rsCentroidF!Revenues,  rucC(j, 2) = rsCentroidF!Profit
                dopC(j, 3) = rsCentroidF!Days
                p = Round(Brojnik(j, 1) / Nazivnik(j, 1), 3)
                ruc = Round(Brojnik(j, 2) / Nazivnik(j, 2), 3)
                dop = Round(Brojnik(j, 3) / Nazivnik(j, 3), 2)
                rsCentroidF!Revenues = p, rsCentroidF!Profit = ruc,  rsCentroidF!Days = dop
                rsCentroidF.MoveNext
 rsCentroidF.MoveFirst
         For j = 1 To Val(Trim(Text1.Text))
        If Abs(pC(j, 1) - rsCentroidF!Revenues) < 60 And Abs(rucC(j, 2) –
       rsCentroidF!Profit) < 6 And   Abs(dopC(j, 3) - rsCentroidF!Days) < 1 Then
                    nastavi = False
                   Else
                             nastavi = True
                   End If
                   rsCentroidF.MoveNext
         Next j
 rsDataKlasterF.MoveFirst
 For i = 1 To rc
         X = rsDataKlasterF!Revenues
         Y = rsDataKlasterF!Profit
         Z = rsDataKlasterF!Days
         rsCentroidF.MoveFirst
        For j = 1 To Val(Trim$(Text1.Text))
        xc = rsCentroidF!Revenues,  yc = rsCentroidF!Profit,  zc = rsCentroidF!Days
        a(j) = Round(Round((X - xc) ^ 2, 0) + Round((Y - yc) ^ 2, 0) + (Z - zc) ^ 2, 0)
       rsCentroidF.MoveNext
         Next j
 For j = 1 To Val(Trim$(Text1.Text))
                   S = 0
                   For k = 1 To Val(Trim$(Text1.Text))
                             If a(k) = 0 Then
                             m(i, k) = 1
                              S = 0
                             Exit For
                             Else
                                       S = S + Round(a(j) / a(k), 5)
                    End If
                   Next k
                             If S <> 0 Then
                                       m(i, j) = Round(1 / S, 5)
                             End If
          Next j
          rsDataKlasterF.MoveNext
Next i
Loop
```

Program 5. Calculating new centroids

By iteratively updating the cluster centers and the membership matrix for each record, fuzzy c-means iteratively move the cluster centers to the "right center" within data records. Cluster centers (centroids) initialized using randomly selected records and fuzzy c-means do not ensure that it converges to an optimal solution. Performance depends on initial centroids and may be improved on two ways:

1) using an algorithm to determine all of the centroids. (for example: arithmetic means of all records)

2) Run fuzzy c-means several times each starting with different initial centroids.

In market segmentation we prefer the second approach.

4.1 Experimental Results

Marketing intelligent system for segmentation business markets (MISSEM) assigns the customers to different clusters, different market segments. After surveillance the environment (data warehouse) MISSEM calculates the centers for each cluster, assigns every customer to segment and calculates corresponding membership grades [15]. For example, the cluster 5 has the values: Revenues=102872, Profit=11431,39 and Days=52. The center for each of seven clusters and corresponding membership grades are represented in Figure 11.

New centroid

Revenues	Profit	Days	ID_Cluster
22729,6	3111,726	47	1
65579,61	6802,999	46	2
2138,963	287,817	50	3
8813,486	869,271	54	4
102872	11431,39	52	5
79418,45	8695,98	47	6
35477,07	4645,935	49	7

Fuzzy membership

Record	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
1	0,00676	0,00081	0,94192	0,04688	0,00033	0,00055	0,00276
2	0,99074	0,00036	0,00137	0,00288	0,0001	0,0002	0,00436
3	0,00147	0,00016	0,98547	0,01213	0,00007	0,00011	0,00058
4	0,00175	0,0002	0,9829	0,01425	0,00008	0,00013	0,00069
5	0,00744	0,00089	0,93685	0,0508	0,00036	0,00061	0,00305
6	0,01817	0,0015	0,44108	0,53172	0,00057	0,00099	0,00599
7	0,26128	0,007	0,12063	0,56348	0,00234	0,00433	0,04096
8	0,00385	0,00045	0,9648	0,02888	0,00018	0,0003	0,00154
9	0,00317	0,00036	0,97051	0,02431	0,00015	0,00025	0,00127
10	0,00462	0,00031	0,03317	0,96023	0,00012	0,0002	0,00135
11	0,06492	0,01068	0,00942	0,0147	0,00221	0,00512	0,89295
12	0,94547	0,00213	0,00721	0,01455	0,00058	0,00119	0,02889
13	0,00283	0,00032	0,97337	0,022	0,00013	0,00022	0,00113
14	0,00163	0,00018	0,98399	0,01335	0,00007	0,00012	0,00064

Fig. 11 Centroids of the clusters and corresponding membership grades

MISSEM shows that the first customer (record 1) belongs 94,18% to market segment where the average amount of sale is 2138,963 (Revenues=2138,963); average number of payment days is 50 and realized profit per customer 287,817.

MISSEM allows that one customer can belong to several market segments at the same time but with different degrees. This is an important feature for market segmentation because it increases the sensitivity of analysis.

If the membership degree is close to 0,5 then such case may be denoted as suspicious. We do not know where to assign the customer. Assigning the customer to one cluster could be wrong. Therefore MISSEM gives very reliable results because it finds all customers with membership grade close to 0,5. In our example MISSEM identifies all suspicious customers, their membership grade and assigned cluster (see Figure 12).

Suspicious records

ID_Customer	Cluster	Membership
6	3	0,44108
6	4	0,53172
7	4	0,56348
15	3	0,54289
15	4	0,43083
23	6	0,53251
50	3	0,40172
50	4	0,57109
53	5	0,43901
53	6	0,42327
68	7	0,59011
76	6	0,53519
77	5	0,44377
77	6	0,4187
97	2	0,50092
98	2	0,50209

Fig. 12 Membership degrees of suspicious samples

Now it is necessary that the expert judgment assign the customers 6, 7, 15, 23, 50, 53, 68, 76, 77, 97 and 98 to adequate market segment. Namely, the MISSEM extracts the customers for which the membership grade is between 0,4 and 0,65. This interval may be closer.

5 Collaboration of Knowledge Based System and the Fuzzy C-Means Clustering Implementation Results

A knowledge based system is a computer program that contains stored knowledge and solves problems in a specific field in much the same way that a human expert would. One of the main problems and most difficult tasks in building rule based knowledge systems is representing the knowledge discovered by c-means clustering.

The knowledge typically comes from a series of conversations between the developer of the expert system and one or more experts. A peculiarity of knowledge based systems for interpretation of the results of c-means clustering is that the knowledge comes from two sources. The first source is dimensions of centroids in clusters (dimensions are revenues, profit and days of payment) and the second source is manager. After analyzing the clusters centroids (see Figure 11) the managers propose that the distance between the centroids of neighbored clusters are the base for further customers grouping and preparing for signing delivering contracts of goods. The judgment of discovered knowledge (clustering every customer) will be performed by building rule based knowledge system. The completed system applies the knowledge of customers' clustering.

The format that a knowledge engineer uses to capture the knowledge is called a knowledge representation. The most popular knowledge representation is the production rule (also called the if-then rule). Production rules are intended to reflect the "rules of thumb" that experts use in their day-to-day work. These rules of thumb are also referred to as heuristics. A knowledge base that consists of rules is sometimes called a rule base.

Except the executable code in the form of production rules, the knowledge could be represented by decision trees with four levels [14]:

Level I: Revenues –clusters center (centroids).
Level II: Profit – clusters center (centroids).
Level III: Number of payment days – average of all clusters.
Level IV: Clusters as results of applying rule based knowledge system.

The key components of knowledge based system for interpretation of the results of customers' clustering will be transformed into the statements (clauses) of Visual Prolog.

Leafs of decision tree would represent the adequate customers' cluster. The Visual Prolog syntax ensures, by relatively simple and easy way, the knowledge representation about the objects properties as the relationships among objects and its properties. The knowledge is represented by production rules: IF (condition) THEN (action). The model of integrating the results of c-means clustering and manager's experience and knowledge is represented in Figure 13:

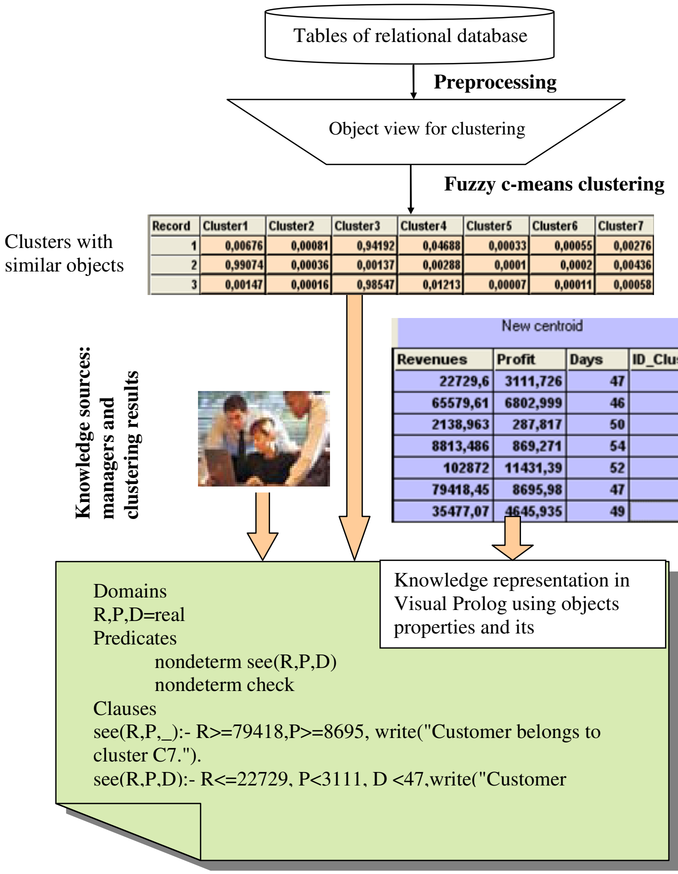


Fig. 13 The conceptual model of collaboration of knowledge based system and the fuzzy c-means clustering

Now it is possible for each new customer, by consulting with expert system, to find adequate cluster to which customer belongs. If the relationships among revenues, profit and days of payments are not acceptable for current market state then expert system will react and give managers adequate warning [14]. For example, one dialog with knowledge base system is presented in Figure 14.

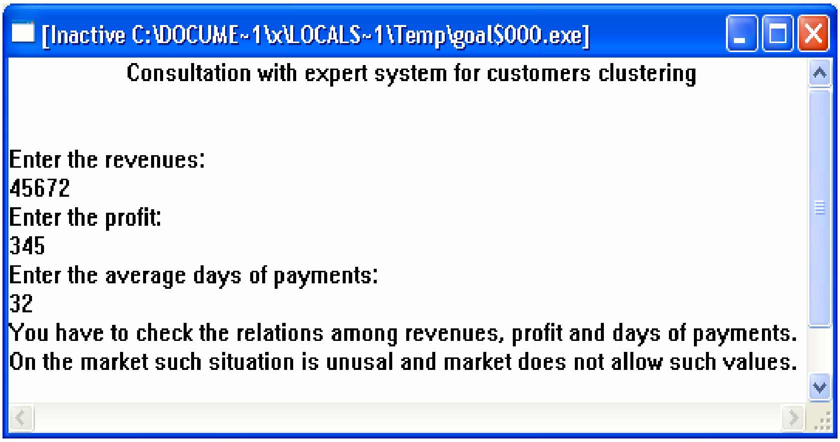


Fig. 14 Consultation with knowledge base system

If the amount of revenues is 45672 $, profit is only 345 $ and average number of payments days is 32, the knowledge based system advices the manager: "You have to check the relations among revenues, profit and days of payments".

Now clustering is fully integrated with technology of expert systems whereas clustering algorithm (k-means) helps to determine the vector dimensions (centroids) which are the input information for building knowledge base of expert system. The output of clustering algorithms is input to expert system. This integration shows enormous application power.

6 Conclusion

This paper theoretically and practically presents the components of marketing intelligent system. Marketing intelligent system may be build only as integration knowledge of databases, data warehouse, data mining and marketing research. The data mining component of the knowledge discovery process is mainly concerned with algorithms by which patterns are extracted from the data (fuzzy c-means clustering). Data for customers' clustering is stored in relational data warehouse that is temporarily loaded from transactional data bases, after running the program for fuzzy c-means clustering written in Visual Basic. Net development environment follows the results that are easy understand and explain. On the business market the firm may very easy define the required number of clusters (five, ten, twenty etc.) and the software will assign the customers to adequate cluster with membership degree. The sensitivity and broad applicability of the software and the concept of knowledge discovery are assured in this way.

In this study, marketing intelligent system (MISSEM) identifies the market segment and assigns every customer to adequate segment. MISSEM implements k-means and fuzzy c-means clustering. In market segmentation fuzzy c-means algorithm gives the better results than hard-k-means algorithm.

The paper clearly shows and realizes the collaboration among knowledge based systems and fuzzy c-means clustering. Fuzzy c-means clustering automatically clusters the database records into a number of groups and the results are the inputs into knowledge based system. Knowledge based system integrates managers knowledge and knowledge extracted by clustering. Clustering methodology is appropriate for the exploration of the interrelationships among samples (customers) and knowledge based system shows the strength and power for interpretation of received results.

7 Practical Utilities for Marketing Management

Marketing practitioners are under constant pressure to ensure adequate answers to market challenges. Building intelligent information systems will help to react quickly and provide information base for preparing efficient decisions. This paper presents the software solution for customer segmentation as a key stone in creating business police and shaping marketing actions to identified segments. A few elements must be especially stressed.

First, in tables in operational database warehouse is created as snowflake architecture by transforming, extracting and loading data. The reader can form clear image how to come from data in operational database to information and knowledge about customers.

Secondly, we built original model for customer segmentation. This model includes three attributes: sales, profit and average number of days during the year for the payment of bills by customer.

Thirdly, we are presented the part of source code for customer clustering by implementing k-means or fuzzy c-means clustering.

Fourthly, there are obvious differences between fuzzy and crisp clustering. Fuzzy clustering is more complex but the results are much realistic and applicable.

Fifthly, the model was tested in operational data in concrete firm and has shown satisfactory solution for defining business policies and formulation of optimal targeting strategies for each segment (customers that belong to corresponding cluster). The most attractive opportunities are segments with highest sales and profit and lowest average days for payment.

Sixthly, intelligent information system for customer clustering may be very easy extended by adding new criteria for segmentation.

Seventhly, the paper presents the possible collaboration of knowledge base systems and the results of clustering (using Visual Prolog developing environment).

References

[1] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York (1981)

[2] Buttle, F.: Customer Relationship management, Concepts and Technologies, 2nd edn. Elsevier, Amsterdam (2009)

[3] Chiu, S., Domingo, T.: Data Mining and Market Intelligence for Optimal Marketing Returns, Jordan Hill, Oxford OX28DP, Elsevier (2008)

[4] Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(2), 224–227 (1979)

[5] Dunn, J.C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3, 32–57 (1973)

[6] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, S., Uthurusamy, R.: Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge (1996)

[7] Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufman, San Francisco (2000)

[8] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2001)

[9] Bezdek, J.C.: Fuzzy Mathemathics in Pattern Classification. PhD Thesis, Applied Math. Center, Cornell University, Ithaca (1973)

[10] Kantardžić, M.: Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-Interscience, Piscataway (2003)

[11] Kimball, R.: The data warehouse toolkit. Wiley, Chichester (1998)

[12] Looney, C.G.: Interactive clustering and merging with a new fuzzy expected value. Pattern Recognition Lett. 35, 187–197 (2002)

[13] Markić, B., Tomić, D.: Executive information system for customers clustering, Društvo i tehnologija, Međunarodni simpozij, Hrvatska, Zadar (2005)

[14] Markić, B., Tomić, D.: Integrating cluster algorithms and expert systems. In: The 8th International Conference Modern Technologies in Manufacturing, Technical University of Cluj-Napoca, Romania (October 2005)

[15] Markić, B., Tomić, D.: Software solutions in marketing research for knowledge discovery in databases by fuzzy clustering. Informatologija, Zagreb 39(4), 240–244 (2006)

[16] Markić, B., Tomić, D.: Building software agents for market segementation, Baden Baden (2006)

[17] Joel, M.: Murach's Visual Basic 2008. Mike Murach & Associates, Inc. (2008)

[18] Joel, M.: Murach's SQL Server 2005 for Developers. Mike Murach & Associates, Inc. (2008)

[19] Rainer, M.: Intelligent Information Systems (SS 2009), `http://www.informatik.uni-bonn.de/~manthey/IIS09/`

[20] Shuliang, L., Barry, D., Edwards, J., Kinman, R., Duan, Y.: Integrating group Delphi, fuzzy logic and expert systems for marketing strategy development: the hybridisation and its effectiveness. Journal: Marketing Intelligence & Planning 20(5), 273–284 (2002)

[21] Ryu, T.-W., Eick, C.F.: A Database Clustering Methodology and Tool, Department of Computer Science University of Houston, Information Science in Spring (2005)

[22] Teknomo, K.: K-Means Clustering Tutorials, `http:\\people.revoledu.com\kardi\tutorial\kMean\`

[23] Turban, E., Aronson, J.E., Liang, T.-P.: Decision Support Systems and Intelligent Systems, 7th edn. Pearson, Prentice Hall (2005)

[24] Witten, I.H., Frank, E.: Data Mining. Academic Press, London (2000)

Information related to the authors

Markić Brano Ph.D. full time professor and
Tomić Draženа Ph.D. associated professor at
Faculty of Economics University of Mostar, Bosnia and Herzegovina.

Using Data Fusion to Enrich Customer Databases with Survey Data for Database Marketing

Peter van der Putten and Joost N. Kok

LIACS, Leiden University, P.O. Box 9512,
2300 RA Leiden, The Netherlands
e-mail: putten@liacs.nl, joost@liacs.nl

1 Introduction and Motivation

Many data mining papers start with claiming that the exponential growth in the amount of data provides great opportunities for data mining. Reality can be different though. In real world applications, the number of sources over which this information is fragmented can grow at an even faster rate, resulting in barriers to widespread application of data mining and missed business opportunities. Let us illustrate this paradox with a motivating example from database marketing.

In marketing, direct forms of communication are becoming increasingly popular. Instead of broadcasting a single message to all customers through traditional mass media such as television and print, the customers receive personalized offers through the most appropriate channels, inbound (the customer contacts the company) and outbound (the company contacts the customer), in batch and real time. So it becomes more important to gather information about media consumption, attitudes, product propensity etc. at an individual level [20]. Basic, company specific customer information resides in customer databases, but market survey data depicting a richer view of the customer are only available for a small sample of potentially anonymous customers. Collecting all this information for the whole customer database in a single source survey would certainly be valuable but prohibitively costly if not impossible because of privacy constraints. The common alternative within business to consumer marketing is to buy syndicated socio-demographic data that have been aggregated at a geographical level. All customers living in a particular geographic location, for instance in the same zip code area, are associated with the same characteristics. In reality customers from the same area may behave differently. Furthermore, regional indentifiers such as zip codes may be absent in company specific surveys because of privacy concerns.

The zip code based data enrichment procedure can be seen as a very crude example of data fusion: the combination of information from different sources. However more general and powerful fusion procedures are required that cater to any number and kind of 'linking' variables. Data mining algorithms can help to carry out such

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 113–130.
springerlink.com

generalized fusions and create rich data sets for further data mining for marketing and other applications.

In this chapter we position data fusion as both a enabling technology and an interesting research topic for data mining in database marketing. A fair amount of work has been done on data fusion over the past 30 years, but primarily outside the knowledge discovery and database marketing communities, as its application was primarily limited to media and socio-economic research. All published cases we are aware of focus on fusing survey samples. However, our application domain of interest is database marketing not market research. We are not so much interested in fusing surveys, but in enriching customer databases with market surveys to enable behavioral targeting for one to one marketing. To our knowledge we were the first to report on the added value of fusion for predictive analytics, by comparing models on data sets with and without fusion data [21], [22], [24], [23], [25].

Note that data fusion can act as an important enabler for data mining, but in return the data fusion problem can be seen as a data mining, intelligent systems or soft computing problem. In almost all published cases statistical matching is used which can be seen as a special case of *k*-nearest neighbor or fuzzy matching, but in principle any data mining technique could be applied, see section 2.2 for examples. In other words, data fusion is a fertile, new research area for data mining research.

We would like to share and summarize the main approaches taken so far from a data mining perspective (section 2). A case study from database marketing serves as a clarifying example and a proof of principle result (section 3). We then generalize from the case results by giving a high level overview of a process model for carrying out data fusion projects for the purpose of mining customer databases (section 4). In section 5 we provide a summary and conclusions.

2 Data Fusion

Valuable work has been done on data fusion in areas other than data mining. From the end of the sixties until now, the subject has been both popular and controversial, with a number of initial applications in social economic research primarily in the US and Germany (for instance [3], [32], [2], [30], [11], [16], [31], [10]; [27] provides an overview)) and later in the field of media research with a focus on Europe and Australia (for example [15], [29], [8], [35], [36]; [1] provides an overview; see also [28], [5] for statistical textbooks).

Data fusion has yet to be discovered by the traditional knowledge discovery and machine learning communities as a standard topic for research, though a relatively new area is developing around mining uncertain data – note fused data can be seen as a special case of uncertain data [17]. Data fusion is also known as micro data set merging, statistical record linkage, multi-source imputation and ascription. Data fusion is sometimes used as a data mining related term in multi-sensor information fusion, however in that context it refers to a different concept: combining information from different sources about a single entity, where as in our case we enrich data about instance a (a customer for example) with information from other instances $b, c, \ldots$ (other customers).

Until today, in marketing data fusion is often used to reduce the required number of respondents or questions in a survey. For instance, for the Belgian National Readership survey questions regarding media and questions regarding products are collected in 2 separate groups of 10,000 respondents each, and then fused into a single survey, thereby reducing costs and the required time for each respondent to complete a survey. However, it is not commonly used yet to enrich customer databases.

2.1 Data Fusion Concepts

Let us introduce some key data fusion concepts. We assume that we start from two data sets. These can be seen as two tables in a database that may refer to disjoint data sets, i.e. it is actually not required that any of the instances in table 1 also occur in table 2. The data set that is to be extended is called the recipient set A and the data set from which this extra information has to come is called the donor set B. We assume that the data sets share a number of variables. These variables are called the common variables X. The data fusion procedure will add a number of variables to the recipient set. These added variables are called the fusion variables Z. Unique variables are variables that only occur in one of the two sets: Y for A and Z for B. See figure 1 for a marketing example. In general, we will learn a model for the fusion using the donor B with the common variables X as input and the fusion variables Z as output and then apply it to the recipient A.

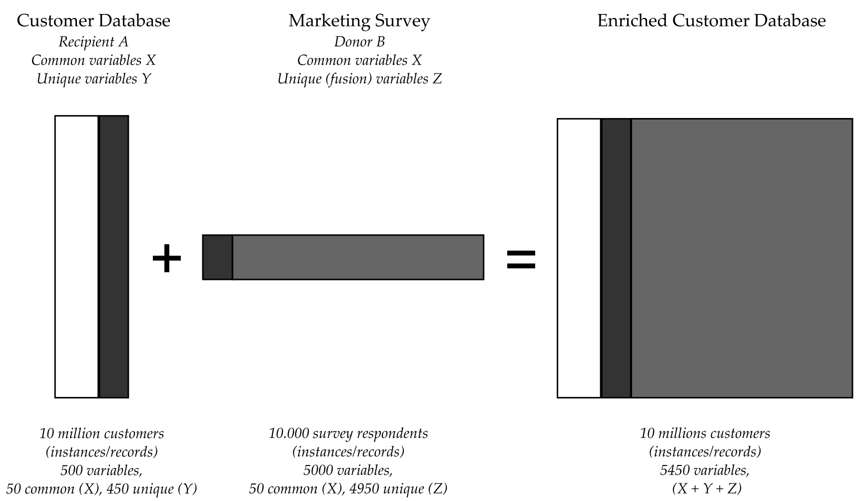


Fig. 1 Data fusion for database marketing: a customer database is enriched with market survey information for further data mining

2.2 Core Data Fusion Algorithms

In nearly all studies, statistical matching is used as the core fusion algorithm. The statistical matching approach can be compared to k-nearest neighbor prediction with the donor as training set and the recipient as a test or deployment set. The procedure consists of two steps. First, given some element from the recipient set, the set of k best matching donor elements is selected. The matching distance is calculated over some subset of the common variables.

Standard distance measures such as Euclidian distance can be used, but often more complex measures are designed to tune the fusion process. For instance, it may be desirable that men are never matched with women, to prevent that 'female' characteristics like 'pregnant last year' are predicted. In this case, the gender variable will become a so-called cell or critical variable; the match between recipient and donor must be 100% on the cell variable; otherwise these will not be matched at all. Weighting can be used to reflect the relative importance of the donor variables.

Another enhancement is called constrained matching. Experiments with statistical matching have shown that even if the donor and recipient are large samples of the same population, some donors are used more than others, which can result in a fusion that may not be representative, as the values for the fusion variables for these donors have a larger influence on predictions. Especially for donors with an average profile this can be the case; this is an artifact of the winner takes all character of nearest neighbor combined with the fact that the signal can get lost in high dimensional, noisy data. By taking into account how many times an element of the donor set has been used when calculating the distance, we can counter this effect [2], [1], [30], [18], [6].

It is interesting to note that within data fusion research this is seen as a generally accepted problem whereas within standard k-nearest neighbor research it is not identified as such. For instance whereas it is clear that overusing donors is a problem, it is not yet proven whether penalizing donors makes things better or worse, especially because this can be hard to evaluate. This is an area that warrants more theoretical debate in our opinion.

In the second step, the prediction for the fusion variables can be constructed using the set of best matching nearest neighbors, e.g. by calculating averages (numerical), modes (categorical) or distributions (categorical or numerical). In this step, the contribution of a neighbor is sometimes weighted inversely proportional to its distance from the recipient.

A number of constraints have to be satisfied by any fusion algorithm in order to produce valid results. Firstly, the donor must be representative for the recipient, or at least contain sub sets that are representative. This does not necessarily mean that the donor and recipient set need to be samples of the same population, although this would be preferable. For instance, in the case of statistical matching only the set of donors used needs to be representative of the recipient set. The recipients could be buyers of a specific product and the donor set could be very large sample of the general population that includes instances representative for these recipients. Methods that are not nearest neighbor based but that build a global, abstract model on the entire data set using donor data only, such as regression, may be more prone to errors in this example. This could be a possible explanation for the popularity of nearest neighbor based techniques for data fusion. The idea is that assuming the donor set is sufficiently large one can always find donors that are representative of the recipient, and predictions are made from these local recipient neighborhoods only ('product owners'). In contrast, a regression model to predict fusion variables would be developed on the donor data set, i.e. discover the relationships between common and fusion variables in the donor set alone ('the general population'), and the resulting global model would be applied to the recipient.

Secondly, the common variables must be good predictors for the fusion variables. In addition, the Conditional Independence Assumption must be satisfied: the commons X must explain all the relations that exist between unique variables Y and Z. In other words, we assume that $P(Y|X)$ is independent of $P(Z|X)$. This could be measured by the partial correlation $r(ZY,X)$, however if the recipient and donor data sets are disjoint there is no joint data available on X, Y and Z to compute this. As an intuitive explanation, consider there would be some other variable W that explains the relationship between Y and Z above and beyond what the commons X can explain; if it exists, finding out the exact relationship between Y and Z by predicting Z from X will not be possible. In most of our fusion projects we start with a small-scale fusion exercise to test the validity of the assumptions and to produce ballpark estimates of fusion performance.

In the wide majority of cases the standard statistical matching approach is being used, there are only very few examples of other approaches. In [2], constrained fusion is modeled as a large scale linear programming transportation model. The main idea was to minimize the match distance under the constraint that all donors should be used only once, given recipients and donors of equal size. This was recently extended to an approach that used genetic algorithms rather than classical optimization algorithms to solve the transportation problem [6]. Various methods derived from solutions to the well-known stable marriage problem [7] are briefly mentioned in [1]. In statistics extensive work has been done on handling missing data [11], including likelihood based methods based on explicit models such as linear and logistic regression. Some researchers have proposed to impute values for the fusion variables using multiple models to reflect the uncertainty in the correct values to impute [31]. In [9] a statistical clustering approach to fusion is described based on mixture models and the Expectation Maximization (EM) algorithm. In [34] also a clustering approach is taken, comparing k-means clustering with Self-Organizing Maps.

These examples of non nearest neighbor approaches are exceptions to the rule, and in most of the case above only a single technique is being used. To address this gap we have executed benchmarking fusion experiments comparing nearest neighbor based approaches with common data mining techniques such as naive Bayes, logistic regression, decision stumps, decision trees and feed forward neural networks [12].

2.3 Data Fusion Evaluation and Deployment

An important issue in data fusion is to how to measure the quality of the fusion; this is not a trivial problem [8]. We distinguish between internal evaluation and external evaluation. This refers to the different steps in the data mining process. If one considers data fusion to be part of the data step and evaluates the quality of the fused data set only within this stage then this is an internal evaluation. However, if the quality is measured using the results within the other steps in the data mining process, then we call this an external evaluation (see figure 2).

Assume for instance that one wants to improve the response on mailings for a certain set of products, and this is the reason why the fusion variables would be added in the first place. In this case, one way to evaluate the external quality is to check whether an improved mail response prediction model can be built when fused data is included in the input.

Ideally, the fusion algorithm is tuned towards the kinds of analysis that is expected to be performed on the enriched data set. In practice the external evaluation will provide the bottom line evaluation, but an enriched data set could be used for multiple purposes unknown at the time of the fusion, and the internal evaluation will provide smoke test results about the fusion quality. In other words, a fusion that passes internal evaluation can still deliver bad external evaluation results, but a fusion with bad internal fusion results will likely not deliver good external test results.

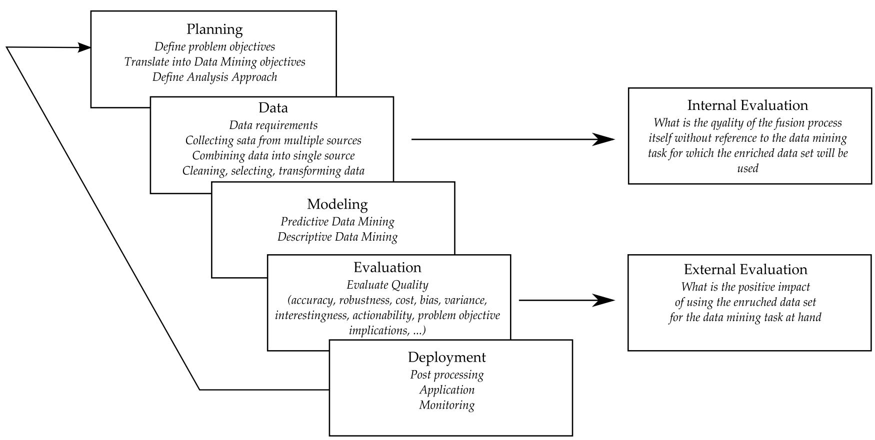


Fig. 2 Internal and external evaluation of data fusion quality within the overall data mining process

3 Case Study: Cross Selling Credit Cards

As a case example of using data fusion for predictive data mining, assume the following example. A bank wants to learn more about its credit card customers and expand the market for this product. Unfortunately, there is no survey data available that includes credit card ownership; this variable is only known for customers in the customer base. Data fusion is used to enrich a customer database with survey data. The resulting data set serves as a starting point for further data mining. The goal is to find out whether the enriched data has added value for the task at hand, i.e. predict who has a high probability to take up a credit card, and profile prospects in terms of survey variables, both of which can't be achieved using single source data only.

To simulate the bank case we do not use a separate donor; instead we draw a sample from an existing proprietary real world market survey (the Dutch SUMMO national readership survey) and split the sample into a disjoint donor set and recipient set, i.e. no donor instance can act as recipient and vice versa. The original survey contains over a 1000 variables and over 5000 possible variable values and covers a wide variety of consumer products and media. Whilst this is a simulation, it can be seen as representative for situations when the data sets to be fused are sufficiently large random samples from the same underlying population, which is a common use case especially in marketing.

Exceptions would be situations when samples differ by design or are poor samples of a population. An example of a difference by design is a customer database for a young and trendy mobile telecom provider versus survey on calling behavior for the general population in a given country. Note that some of the fusion methods presented in the previous section do not apply if samples are not meant to be representative, such as constrained matching. An example of poor representativeness could be various small data sets on cancer patients for hospitals with different overall life expectancy rates.

The recipient set representing a small sample from the customer database, contains 2000 records with a cell variable for gender, common variables for age, marital status, region, number of persons in the household and income. Furthermore, the recipient set contains a unique variable for credit card ownership. One of the goals is to predict this variable for future customers. The donor set representing the survey contains the remaining 4880 records, with 36 variables for which we expect that there may be a relationship to the credit card ownership: general household demographics, holiday and leisure activities, financial product usage and personal attitudes. These 36 variables are either numerical or Boolean.

First we discuss the specific kind of matching between the databases and then the way the matching is transformed into values of the fusion variables. The matching is done on all common variables. Given an element of the recipient set we search for elements in the donor set that are similar. Elements of the donor set need to agree on the cell variable gender. All the common variables are transformed to numerical values and simple Euclidean distance on the commons is used as the distance measure. We select the k best matching elements from the donor. For the values of the fusion variables, we take the average of the corresponding values of the k best matching elements from the donor set.

3.1 Internal Evaluation

As a baseline analysis we first compared averages for all common variables between the donor and the recipient. As could be expected from the donor and recipient sizes and the fact that the split was done randomly, there were not many significant differences between donor set and recipient set for the common variables. Within the recipient 'not married' was over represented (30.0% instead of 26.6%), 'married and living together' was under represented (56.1% versus 60.0%) and the countryside and larger families were slightly over represented. This provides a baseline expectation of magnitude of differences that could be caused by sampling error only (or lack of representativeness by design if that would apply).

Then we compared the average values between the values of the fusion variables and the corresponding average values in the donor. Only the averages of 'Way Of Spending The Night during Summer Holiday' and 'Number Of Savings Accounts' differed significantly, respectively by 2.6% and 1.5%. Compared to the differences between the common variables, which were entirely due to sampling errors, this was a good result.

Next, we evaluated the preservation of relations between variables, for which we used the following measures. For each common variable, we listed the correlation with all fusion variables, both for the fused data set and for the donor. The mean difference between common-fusion correlations in the donor versus the fused data set was 0.12 ± 0.028. In other words, these correlations were well preserved. A similar procedure was carried out for correlations between the fusion variables with a similar result.

3.2 External Evaluation

The external evaluation concerns the value of data fusion for further analysis. Typically only the enriched recipient database is available for this purpose. We first performed some descriptive data mining to discover relations between the target variable, credit card ownership, and the fusion variables using straightforward univariate techniques. We selected the top 10 fusion variables with the highest absolute correlations with the target (see Table 1).

Note that this analysis was not possible without the fusion, because the credit card ownership variable was only available in the recipient. If other new variables become available for the recipient customer base, e.g. product ownership of some new product, their estimated relationships with the donor survey variables can directly be calculated, without the need to carry out a new survey.

Next we investigated whether different predictive modeling methods would be able to exploit the added information in the fusion variables. The specific goal of

Table 1 Top ten fusion variables in recipient most strongly correlated with credit card ownership

Variable
Welfare class
Income household above average
Is a manager
Manages which number of people
Time per day of watching television
Eating out (privately): money per person
Frequency usage credit card
Frequency usage regular customer card
Statement current income
Spend more money on investments

the models was to predict a response score for credit card ownership for each recipient, so that they could be ranked from top prospects to suspects. We compared models trained only on values of common variables to models trained on values of common variables plus either all or a selection of correlated fusion variables. We used feed forward neural networks, linear regression, k nearest neighbor search and naive Bayes classification.

The feed forward neural networks had a fixed architecture of one hidden layer with 20 hidden nodes using a tanh activation function and an output layer with linear activation functions. The weights were initialized by Nguyen-Widrow initialization to enforce that the active regions of the layer's neurons were distributed roughly evenly over the input space [14] . The inputs were linearly scaled between -1 and 1. The networks were trained using scaled conjugate gradient learning as provided within Matlab [13] . The training was stopped after the error on the validation set increased during five consecutive iterations. For the regression models we used standard least squares linear regression modeling. For the k nearest neighbor algorithm, we used the same simple approach as in the fusion procedure, so without normalization and variable weighting, with k=75. We used our own implementation of the standard Naive Bayes algorithm. The core fusion algorithm was implemented in C++ using a object oriented library we originally developed for codebook based algorithms (codebooks, SOMs, LVQ etc. [19]); the algorithms to build the prediction models were developed using MATLAB [33].

We report results over ten runs with train and test sets of equal size. Error criteria such as the root mean squared error or accuracy do not always suffice to evaluate a ranking task. Take for instance a situation where there are few positive cases, say people that own a credit card. A model that predicts that no one is interested in credit cards has a low rmse, but is useless for the ranking and the selection of prospects. In fact, one has to take the costs and gains per mail piece into account. If we do not have this information, we can consider rank based tests that measure the concordance between the ordered lists of real and predicted cardholders.

We use a measure we call the c-index, which is a test related to Kendall's Tau [33]. The c-index is a rank based test statistic that can be used to measure how concordant two series of values are, assuming that one series is real valued and the other series is binary valued.

We use the following procedure to calculate the c-index. Assume that all records are sorted ascending on rank scores. Records can be positive or negative (for example, if these are credit card holders or not). We assign points to all positive records: in fact we give $k-0.5$ points to the k-th ranked positive record and records with equal scores share their points. These points are summed and scaled to obtain the c-index , so that an optimal predictor results in a c-index of 1 and a random predictor results in a c-index of 0.5. Under these assumptions, the c-index is equivalent (but not equal) to Kendalls Tau.

The scaling works as follows. Assume that l is the total number of points that we have assigned, and that we have a total of n records with s positive records. If the s positives all have a score higher than the other $n-s$ records, then the ranking is perfect and $l = s*(n-s/2)$. If the s positives all have a score that is lower than the $n-s$ others, then we have used a worst case model and $l = s^2/2$. The c-index is thus calculated by:

$$c - \text{index} = \frac{l - \frac{s^2}{2}}{s(n-\frac{s}{2}) - \frac{s^2}{2}} = \frac{l - \frac{s^2}{2}}{s(n-s)} \tag{1}$$

See Table 2 for some examples. Note that by definition $c = 0.5$ corresponds to random prediction and $c = 1$ corresponds to perfect prediction.

The results of our experiments can be found in Table 3. We provide the average c-value and standard deviation over all runs. We also measure the statistical significance of improvements by fusion through a one tailed two sample T test. The p-value intuitively relates to the probability that the improvement gained by using fusion is coincidental.

The results show that for the data set under consideration most models that are allowed to take the fusion variables into account outperform the models without the fusion variables. Assuming variable selection, three results are significant at the

Table 2 C-index calculation examples for the target list (0,0,0,1,1)

Score list	**Corresponding c-index**
(0.1, 0.2, 0.3, 0.4, 0.5)	$\frac{1}{6}*((3\frac{1}{2}+4\frac{1}{2})-2)=1$
(0.1, 0.2, 0.4, 0.3, 0.5)	$\frac{1}{6}*((2\frac{1}{2}+4\frac{1}{2})-2)=\frac{5}{6}$
(0.1, 0.2, 0.4, 0.4, 0.5)	$\frac{1}{6}*((3+4\frac{1}{2})-2)=\frac{11}{12}$

Table 3 External evaluation results: using enriched data generally leads to improved performance

	Only common variables	**Common and correlated fusion variables**	**Common and all fusion variables**
SCG neural network	c=0.692 ± 0.012	c=0.703 ± 0.015 p=0.041	c=0.694 ± 0.019 p=0.38
Linear regression	c=0.692 ± 0.014	c=0.724 ± 0.012 p=0.000	c=0.713 ± 0.013 p=0.002
Naive Bayes Gaussian	c=0.701 ± 0.015	c=0.720 ± 0.012 p=0.003	c=0.719 ± 0.012 p=0.005
Naive Bayes multinomial	c=0.707 ± 0.015	c=0.720 ± 0.011 p=0.200	c=0.704 ± 0.009 p not relevant
k-nearest neighbor	c=0.702 ± 0.012	c=0.716 ± 0.013 p=0.0093	c=0.720 ± 0.012 p=0.0023

0.01 level, one at the 0.05 level and one is not significant (p=0.2). Without variable selection three results are significant at the 0.01 level, one is not significant (p=0.38) and one result is not better. So without significance testing, nine out of ten results are better, seven out of ten are better at the 0.05 level, and six out of ten results are better at the 0.01 level.

From an algorithm perspective, the best results are achieved with logistic regression using commons and correlated fusion variables. A possible explanation for this is that regression is a high bias method that can only model linear relationships. Fusion in this case may have added additional variables to the model that make the problem more linear([26]. The results are significant with $p < 0.01$ for logistic regression, naive Bayes Gaussian and k-nearest neighbor. The neural network results are significant as well at the 0.05 level, provided variable selection has taken place, otherwise the results are not significant. This could be due to the fact that the number of degrees of freedom in a neural network, the network weights, is severely impacted by an increase in the number of inputs, so variable selection is even more important. The results for naive Bayes multinomial are actually worse if no variable selection has taken place, and with variable selection the improvement is not significant. This may be due to the fact that variables added are violating the naive Bayes assumption of independency, coupled with the issue of the multinomial over the Gaussian approach of having potentially too many unique values in the fusion variables to allow for proper estimation of model parameters. This demonstrates it is important to use a variety of modeling algorithms to find out which works best.

For four out of five algorithms, using variable selection on the enriched data set improves the performance. Fusion variables are derived information, not measured and even if the fusion process were perfect, specific fusion variables may not be relevant for the prediction task at hand. Variable selection can successfully be used to counter this effect. Assuming variable selection, linear regression seems to benefit

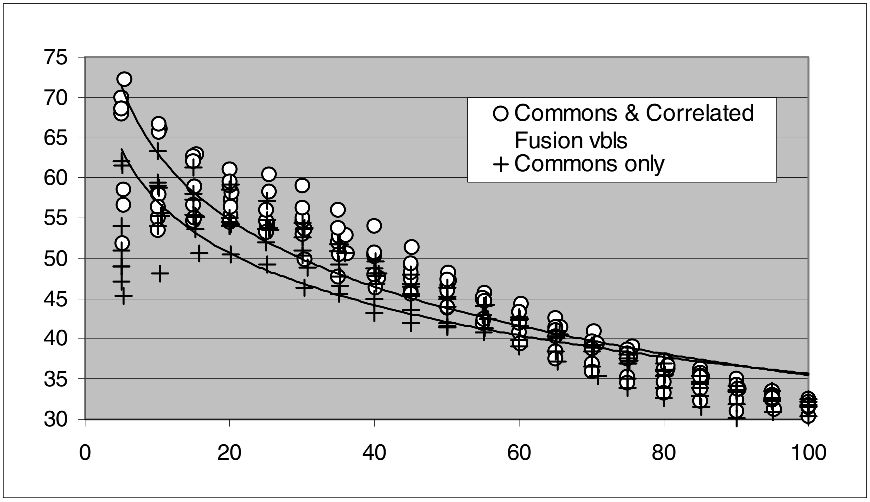


Fig. 3 Cumulative response curves linear regression models for predicting credit card ownership (seven random runs) with and without fusion variables. The x-axis corresponds to the cumulative volume of top scoring instances selected by the model, the y-axis corresponds to the percentage of positives (cardholders) in the selection.

most from enrichment through fusion: a difference of 0.032 versus 0.019 (naive Bayes Gaussian), 0.013 (naive Bayes Multinomial) and 0.011 (neural networks).

In figure 3, cumulative response curves are shown for the linear regression models, for commons only and commons plus fusion variables. A response curve displays the probability of positive, in this case percentage of card holders (y-axis) for model selections of increasing size, ordered from top to bottom model score (x-axis). Response curves are often used in database marketing, for instance to compare model quality at a specific volume cut off of customers to be contacted. Curves for all the runs are displayed and logistic trend lines are fitted to the series for commons only and enriched data.

As can be seen from the graph at the 100% cut off, the overall percentage of credit card holders is 32.5% cardholders. In general credit card ownership can be predicted quite well: the top 10% of cardholder prospects according to the model contains around 50-65% cardholders, the top 20% contains 50-60% card holders still. The spread of results for smaller volumes is larger, this is common and due to a smaller sample size and hence a less robust estimation of the true percentage of cardholders in smaller selections. The added logarithmic trend lines clearly indicate that the models which include fusion variables are better in 'creaming the crop', i.e. selecting the top prospects. At 10% the difference between trend lines is 6.0% (from 57.0% to 63.0% card owners), at 20% it is 4.1% (50.7% versus 54.8%), which is quite substantial and can translate to high impact on campaign ROI. For model selections over 40% the differences become a lot smaller. Again this is a common pattern; if the selection volume gets larger, the pool of cardholders to ‘fish’ from becomes substantially smaller, the overall percentage of cardholders drops, and the prediction task to select medium prospects is substantially noisier, so the various models will converge. From a business and customer centricity perspective these customers are less rewarding segments to contact in outbound campaigns, and in an inbound scenario medium or low propensity propositions will likely not ’win’ over other propositions, so this section of the curve is generally of less interest.

3.3 Case Discussion

Data fusion can be a valuable, practical tool. For descriptive data mining tasks, the additional fusion variables and the derived patterns may be more understandable and easier to interpret. This is not restricted to relations between commons and fusion variables, also relationships between variables that only appear in the recipient and donor can be studied, which can’t be achieved if donor and recipient are disjoint data sets. An example would be profiling the users of a particular new product as indicated by the customer database in terms of answers to an older survey, without requiring that information about the product was actually in the survey.

For predictive data mining, enriching a data set using fusion may make sense, notwithstanding the fact that the fusion variables are derived from information already contained in the donor variables. Fusion may make it easier for high bias algorithms such as linear regression to discover complex non-linear relations between commons and target variables by exploiting the information in the fusion variables. Of course, it is recommended to use appropriate variable selection techniques to remove the noise that is added by ’irrelevant’ fusion variables and counter the ’curse of dimensionality’, as demonstrated by the experiments [26].

There is also a practical dimension to this. Even if certain relationships could be studied by looking at single source data only for subsets of customers one would need to have access and knowledge of these data sets, or knowledge how to combine the results of mining exercises on separate data sets into a single result. In many cases it can be more practical to let a core expert team fuse a variety of data sources into a single set on a periodical basis, and make this available to a wider community of customer insight analysts. This is valid not just for database marketing, but for instance also in the case of providing public integrated mulit source data sets for scientific research, for instance in the medical domain. In media research this is already common practice, as many national readerschip surveys are based on fused surveys.

The fusion algorithms itself provide an interesting opportunity for further data mining research. There is no fundamental reason why the fusion algorithm should be based on k-nearest neighbor prediction instead of clustering methods, decision trees, regression, the expectation-maximization (EM) algorithm or other data mining algorithms, whereas examples are still rare. In addition, it is to be expected that future applications will require massive scalability. For instance, in the past the focus on fusion for marketing was on fusing surveys with surveys, each containing up to tens of thousands of respondents and hundreds of questions or more. In contrast, customer databases typically contain millions of customers. This requires scalable fusion algorithms, as well as scalable algorithms to mine the fused data, which also need to be able to deal with the uncertainty in this data.

It goes without saying that evaluating the quality of data fusion is also crucial. We hope to have demonstrated that this is not straightforward and that it ultimately depends on the type of data mining that will be performed on the enriched data set. As discussed, recently a new research area is developing around algorithms that are specifically adapted to mine uncertain data [17]. Fused data sets can be seen as a special case of such data and the fusion process can actually generate metadata that provide an indication of the degree of uncertainty in the fused data.

4 A Process Model for a Fusion Factory

In the previous sections we provided a example in which an enriched customer base provides an improved source for data mining, in this case better input to predict the propensity for a credit card. This provides proof of concept evidence for the feasibility of using data fusion for database marketing. However to take the step towards wide scale real world applications more is needed. This research project was carried out in the context of setting up a commercial data fusion service, a factory to carry out fusions on an ongoing, repeatable basis. So as a next step after proving the idea in principle, the decision was made to develop a model of the end to end fusion process, for which we will provide summary highlights here.

There is no guarantee that fusion will always deliver added value. Data fusion projects are complex, with many steps and pitfalls. Instead of a single data set, several heterogeneous data sources are involved in the procedure that need to be mapped onto each other. Source data sets with hundreds to thousands of variables in a wide range of logical and physical formats are not uncommon. The fusion process itself consists of many intertwined phases and steps, and a lot of choices have to be made. What the right choices are is predominantly determined by factors outside the core fusion procedure, namely the business and data mining goals for which the enriched data set will be used.

Despite these challenges, we envision a streamlined fusion procedure where the core steps can be carried out in less than a working week instead of weeks or months (the current best practice in media research), using a predictable, reproducible process. To standardize and structure fusion projects we decided to develop a data fusion process model, borrowing some key concepts from data mining process models like CRISP-DM [4]. The end goal of the fusion process model is to rationalize the process and automate it where possible, ultimately to the extent that end users of the fusion service could parameterize, control and execute large parts of it themselves. The development of the process model took place in parallel with three major data fusion projects carried out by a commercial data mining research company, Sentient Machine Research, for a financial services company, a charity and a marketing data

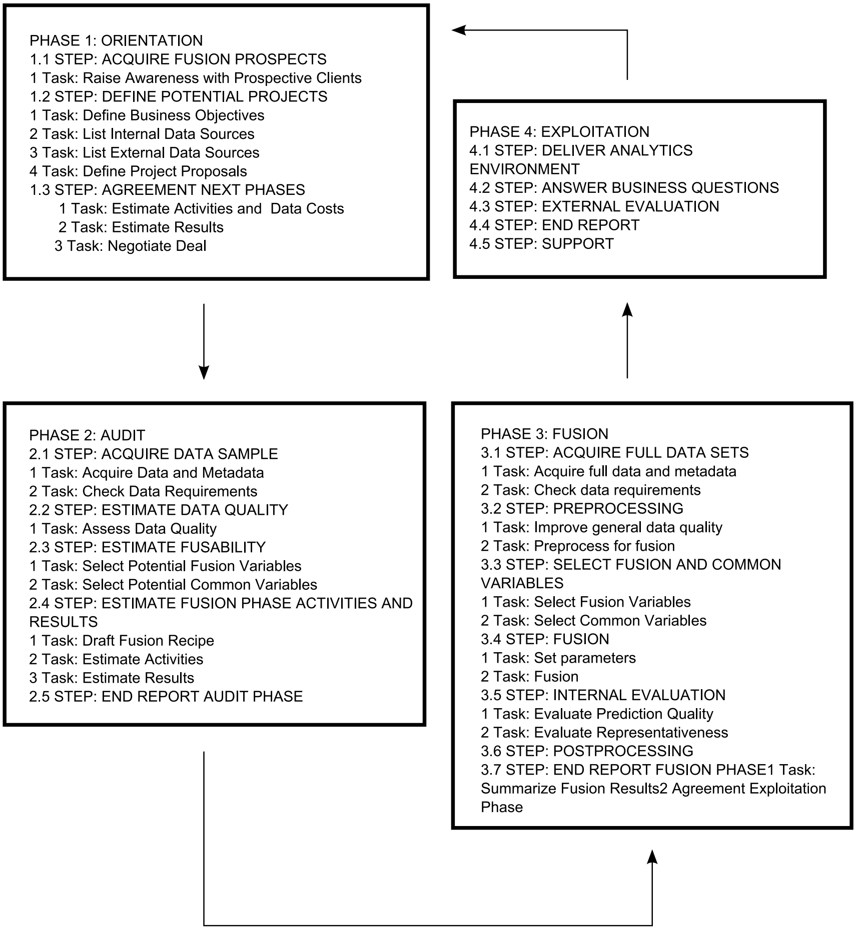


Fig. 4 Fusion Factory Process Model

provider. Further input was provided by previous experimentation on a variety of data sets and some 25 data fusion cases from research literature.

The high-level structure of the process model can be found in figure 4 and is described in detail in [25]. Four main phases have been identified, each of which will terminate in a go/no go decision. The first phase covers the scoping and definition of the project, including the data mining tasks for which the enriched data set will be used and a description of the donor and recipient data. Then an audit step takes place, in this phase the available data sets are analyzed separately and data quality and 'fusability' is assessed, through a variety of methods. On a go decision the actual fusion takes place including all internal evaluation activities. In the final phase, the enriched data set is being as an input to the regular data mining process to assess external quality. Ideally this then leads to further iterations of the overall process.

The process model could be used by any marketing analyst to follow a structured approach towards carrying out fusion projects. It applies to database marketing but is generic enough to be extended to other domains. Alternatively, as a blueprint for the overall process it can be used to analyze where bottlenecks arise and to provide end to end process automation support, or identify sub problems to be covered by data mining research.

5 Conclusion

In this chapter we started by discussing how the information explosion provides barriers to the application of data mining and positioned data fusion as a possible solution to the data availability problem. We presented an overview of the main approaches adopted by researchers from outside the data mining and database marketing communities and described a database marketing case, for which a data set that was enriched by data fusion was used to predict propensity for credit card ownership. Our work is to our knowledge the first published case that discusses the value added by data fusion for predictive data mining for behavioral targeting.

We hope to have shown that, despite its difficulties and pitfalls, the application of data fusion increases the value of data mining, because there is more integrated data to mine. Data mining algorithms can also be used to perform fusions, but publications on methods other than the standard statistical matching approach are rare. Therefore we think that data fusion is an interesting research topic for knowledge discovery and data mining research.

From a database marketing and managerial point of view it will allow marketeers to bring information together from all kinds of sources in the organization, no matter how small the sample for which the information was gathered. This resulting data can be used for one to one marketing at individual customer level, rather than aggregate market research analysis, as if one could have extensive interviews with each of its millions of customers, at a fraction of the cost of real surveys. That said, there is no such thing as a free lunch, further research will be required to avoid overestimating the validity of the fused data and develop mining algorithms that appropriately deal with uncertain data.

References

1. Baker, K., Harris, P., O'Brien, J.: Data fusion: An appraisal and experimental evaluation. Journal of the Market Research Society 31(2), 152–212 (1989)
2. Barr, R., Turner, J.: A new, linear programming approach to microdata file merging. In: 1978 Compendium of Tax Research. Office of Tax Analysis (1978)
3. Budd, E.: The creation of a microdata file for estimating the size distribution of income. Review of Income and Wealth 17, 317–333 (1971)

4. Chapman, P., Clinton, J., Khabaza, T., Reinartz, T., Wirth, R.: The crisp-dm process model. Tech. rep., Crisp Consortium (1999), http://www.crisp-dm.org/
5. D'Orazio, M., Di Zio, M., Scanu, M.: Statistical Matching: Theory and Practice. Wiley, Chichester (2006)
6. Flores, G.A., Albacea, E.A.: A genetic algorithm for constrained statistical matching. In: 10th National Convention on Statistics (NCS), Manila, Phillipines (2007)
7. Gusfield, D., Irving, R.W.: The stable marriage problem: structure and algorithms. MIT Press, Cambridge (1989)
8. Jephcott, J., Bock, T.: The application and validation of data fusion. Journal of the Market Research Society 40(3), 185–205 (1998)
9. Kamakura, W., Wedel, M.: Statistical data fusion for cross-tabulation. Journal of Marketing Research 34(4), 485–498 (1997)
10. Kum, H., Masterson, T.: Statistical matching using propensity scores: Theory and application to the levy institute measure of economic well-being. Working paper no. 535, The Levy Economics Institute of Bard College (2008)
11. Little, R., Rubin, D.: Statistical analysis with missing data. John Wiley and Sons, Chichester (1986)
12. Maat, B.: The need for fusing head and neck cancer data. can more data provide a better data mining model for predicting survivability of head and neck cancer patients? Master's thesis, ICT in Business, Leiden Institute of Advanced Computer Science. Leiden University, The Netherlands (2006)
13. Moller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6(4), 525–533 (1993)
14. Nguyen, D.H., Widrow, B.: Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In: IJCNN International Joint Conference on Neural Networks, vol. 3, pp. 21–26 (1990)
15. O'Brien, S.: The role of data fusion in actionable media targeting in the 1990's. Marketing and Research Today 19, 15–22 (1991)
16. Paass, G.: Statistical match: Evaluation of existing procedures and improvements by using additional information. In: Orcutt, G., Merz, K. (eds.) Microanalytic Simulation Models to Support Social and Financial Policy, pp. 401–422. Elsevier Science, Amsterdam (1986)
17. Pei, J., Getoor, L., de Keijzer, A. (eds.): First ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data, Paris, France, June 28. ACM, New York (2009)
18. van Pelt, X.: The fusion factory: A constrained data fusion approach. Master's thesis, Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands (2001)
19. van der Putten, P.: Utilizing the topology preserving property of self-organizing maps for classification. Master's thesis, Cognitive Artificial Intelligence, Utrecht University, The Netherlands (1996)
20. van der Putten, P.: Data mining in direct marketing databases. In: Baets, W. (ed.) Complexity and Management: A Collection of Essays, World Scientific Publishers, Singapore (1999)
21. van der Putten, P.: Data fusion: A way to provide more data to mine in? In: Proceedings 12th Belgian-Dutch Artificial Intelligence (2000)
22. van der Putten, P.: Data fusion for data mining: a problem statement. In: Coil Seminar 2000, Chios, Greece, June 22-23 (2000)
23. van der Putten, P., Kok, J.N., Gupta, A.: Data fusion through statistical matching. Tech. Rep. Working Paper No. 4342-02, MIT Sloan School of Management, Cambridge, MA (2002)

24. van der Putten, P., Kok, J.N., Gupta, A.: Why the information explosion can be bad for data mining, and how data fusion provides a way out. In: Grossman, R.L., Han, J., Kumar, V., Mannila, H., Motwani, R. (eds.) SDM, SIAM, Philadelphia (2002)
25. van der Putten, P., Ramaekers, M., den Uyl, M., Kok, J.N.: A process model for a data fusion factory. In: Proceedings of the 14th Belgium/Netherlands Conference on Artificial Intelligence (BNAIC 2002), Leuven, Belgium (2002)
26. van der Putten, P., van Someren, M.: A Bias-Variance Analysis of a Real World Learning Problem: The CoIL Challenge 2000. Machine Learning 57(1-2), 177–195 (2004)
27. Radner, D., Rich, A., Gonzalez, M., Jabine, T., Muller, H.: Report on exact and statistical matching techniques. statistical working paper 5. Tech. rep., Office of Federal Statistical Policy and Standards US DoC (1980)
28. Raessler, S.: Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches. Springer, Heidelberg (2002)
29. Roberts, A.: Media exposure and consumer purchasing: An improved data fusion technique. Marketing And Research Today 22, 159–172 (1994)
30. Rodgers, W.L.: An evaluation of statistical matching. Journal of Business & Economic Statistics 2(1), 91–102 (1984)
31. Rubin, D.B.: Statistical matching using file concatenation with adjusted weights and multiple imputations. Journal of Business & Economic Statistics 4(1), 87–94 (1986)
32. Ruggles, N., Ruggles, R.: A strategy for merging and matching microdata sets. Annals Of Social And Economic Measurement 3(2), 353–371 (1974)
33. de Ruiter, M.: Bayesian classification in data mining: theory and practice. Master's thesis, BWI, Free University of Amsterdam, The Netherlands (1999)
34. Smith, K.A., Chuan, S., van der Putten, P.: Determining the validity of clustering for data fusion. In: Proceedings of Hybrid Information Systems, Adelaide, Australia, December 11-12 (2001)
35. Soong, R., de Montigny, M.: Does fusion-on-the-fly really fly? In: ARF/ESOMAR Week of Audience Measurement (2003)
36. Soong, R., de Montigny, M.: No free lunch in data integration. In: ARF/ESOMAR Week of Audience Measurement (2004)

Collective Intelligence in Marketing

Tilmann Bruckhaus

eBay Inc.

Abstract. As marketing professionals communicate value and manage customer relationships, they must target changing markets, and personalize offers to individual customers. With the recent adoption of large-scale, Internet-based information systems, marketing professionals now face large volumes of complex data, including detailed purchase and service transactions, social network links, click streams, blogs, comments and inquiries. While traditional marketing methodologies struggled to produce actionable insights from such information quickly, emerging collective intelligence techniques enable marketing professionals to understand and act on the observed behaviors, preferences and ideas of groups of people. Marketing professionals apply collective intelligence technology to create behavioral models and apply them for targeting and personalization. As they analyze preferences, match products to customers, discover groups of similar consumers, and construct pricing models, they generate significant competitive advantage. In this chapter, we highlight publications of interest, describe analytic processes, review techniques, and present a case study of matching products to customers.

1 Introduction

As marketing organizations create, communicate and deliver value to customers and manage customer relationships, they must target offerings to changing markets, and they must personalize offers for individual customers. Marketing professionals require timely and accurate information about customers and markets to target and personalize successfully, and they need technology to process this information effectively. With the adoption of large-scale information systems that leverage the Internet, marketing professionals face large volumes of complex data that have previously not been available.

For example, organizations collect massive amounts of data about customer purchase transactions across multiple channels, membership information, customer service interactions, blogs, customer comments and inquiries, as well as external data sources available from third-party providers. The large size of the collected data and diversity of data sources pose difficult challenges to marketing professionals. However, other new challenges with modern sources of marketing-relevant data may present even greater dilemmas. These new challenges originate from the growing trend that each specific piece of information may be available

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 131–154.
springerlink.com

only for a fraction of all customers. Available data may be sparse, making it difficult to analyze relationships with other pieces of information. Similarly, data quality may be poor and many pieces of information may be incorrect due to the prohibitive cost of validating and correcting massive amounts of information. Traditional marketing methodologies have struggled to produce actionable insights from such data quickly, and new soft-computing techniques have emerged to address these new realities.

One promising field in soft computing is collective intelligence, a term that refers to combining of behaviors, preferences and ideas of a group of people to create novel insights. As Figure 1 illustrates, electronic commerce offers a variety of data source, which marketing professionals can mine to extract behaviors, preferences and ideas of consumers. The boundaries between behaviors, preferences and ideas may be somewhat fluid. Nonetheless, shoppers exhibit specific behaviors in their views of web pages, purchases and decisions to share items they find, among other behaviors. Customers similarly express preferences when they enter ratings and reviews, or mark favorites. Marketing professionals may also mine comments, suggestions, complaints and inquiries for ideas shoppers express.

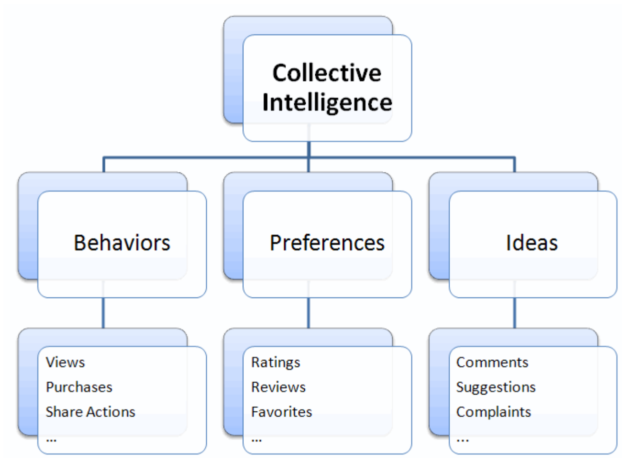


Fig. 1 Dimensions of collective intelligence and sources of expressed behaviors, preferences and ideas in electronic commerce

This chapter reviews a sample of some of the most promising collective intelligence techniques and their applications in marketing. For instance, collaborative filtering is a technique that allows collecting customer preferences, finding similar customers or consumers, matching them to products, and making personalized recommendations. Marketing professionals can also discover groups of similar consumers with a technique known as hierarchical clustering. Customers may find products of interest effectively through query-based search technology and document filtering. They may further construct pricing models with estimation techniques, such as k-nearest neighbors.

This chapter will review such collective intelligence techniques, provide examples and describe the advantages and challenges associated with these techniques. We begin by describing a simple example of data mining to illustrate the technology, and describe several broad classes of data mining algorithms that professionals frequently apply to marketing problems. We then highlight a number of popular business books that have recently advocated the use of data mining and analytics in business settings. Next, we describe how collective intelligence technology applies more specifically to marketing and proceed to review the process of applying data mining to marketing, along with its challenges. A case study then applies these concepts to the problem of matching customers to products.

2 Data Mining Technology in Marketing

One of the promising technologies to emerge recently in marketing is behavioral targeting based on collective intelligence. The essence of this analytic approach is to use data to understand past customer behavior, identify patterns, and make predictions for targeting.

Several concepts closely related to collective intelligence are analytics, data mining, mathematical modeling, and statistical learning. At the core of these methods is the realization that if patterns of customer behavior exist we should be able to discover them in historical data. To accomplish this feat, analytic approaches extract, prepare and analyze historical data and use it to construct mathematical models of customer behavior. For example, online retailers are able to make personalized product purchase recommendations to customers, based on their observed preferences and based on the observed preferences of similar customers. Collective intelligence technology complements other, more traditional technologies, such as expert-rules and heuristics, which marketing professionals derive from the knowledge of human experts and their insights into customer behaviors. This chapter introduces collective intelligence technology and shows how leading organizations use analytic approaches for behavioral targeting.

Wikipedia defines analytics as the study of

> "how an entity (i.e., business) arrives at an optimal or realistic decision based on existing data. Business managers may choose to make decisions based on past experiences or rules of thumb, or there might be other qualitative aspects to decision making; but unless there are data involved in the process, it would not be considered analytics. Common applications of Analytics include the study of business data using statistical analysis in order to discover and understand historical patterns with an eye to predicting and improving business performance in the future."

In the domain of behavioral targeting in marketing, this means to identify and examine attributes of customer behavior and expressed preferences, and to construct models that marketing professionals can use for targeting offers and messages to match observed behaviors.

For example, a widely used illustration of data mining is the analysis of data relating to the decision to play tennis, based on weather conditions. In this example, the observed behavior is past decisions about playing tennis, and a marketing professional may use behavioral patterns she might to extract from historical data for targeting. The following table shows how a tennis player decided to play tennis ("yes" vs. "no") on 14 days, based on weather outlook, temperature, humidity and windiness. This example is adapted from Witten and Frank (2005).

Day	Outlook	Temperature	Humidity	Windy	Play
1	sunny	85	85	not windy	No
2	sunny	80	90	windy	No
3	overcast	83	86	not windy	Yes
4	rainy	70	96	not windy	Yes
5	rainy	68	80	not windy	Yes
6	rainy	65	70	windy	No
7	overcast	64	65	windy	Yes
8	sunny	72	95	not windy	No
9	sunny	69	70	not windy	Yes
10	rainy	75	80	not windy	Yes
11	sunny	75	70	windy	Yes
12	overcast	72	90	windy	Yes
13	overcast	81	75	not windy	Yes
14	rainy	71	91	windy	No

Fig. 2 Example Historical Data for Tennis Problem

Analysts may then use mathematical modeling techniques to process the above data and build predictive models from the historical data, such as the following "decision tree" model:

The decision tree extracts several patterns from the tabular training data. For example, when the outlook is overcast the decision is to play tennis, independent of other conditions. However, when the outlook is sunny a second data point is

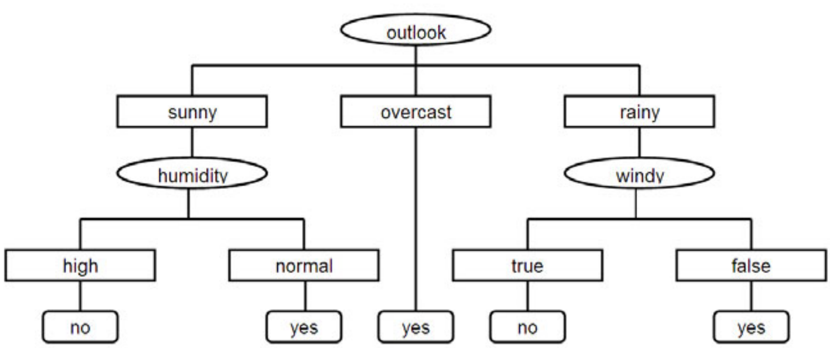


Fig. 3 Example Decision-Tree Model for Tennis Problem

considered, and the tennis player decides to play when humidity is normal, but not to play when humidity is high. Lastly, when the outlook is rainy, the player will prefer to play tennis when it is not windy but forego a match when it is windy. Data mining algorithms, such as decision tree algorithms can construct predictive models such as this one for millions of rows of data, representing "observations" and hundreds or thousands of columns, representing inputs, predictors or attributes. In this example, the algorithm was able to select some inputs for inclusion in the model and discard others, such as temperature. Similarly, the algorithm decided which variables to check first, in this case outlook, and which variables to check later, in this case humidity and windiness.

It is easy to see that mathematical modeling techniques that can predict whether today's weather allows for an enjoyable tennis match might also help predict whether a given customer might want to purchase a specific product, or whether a contact might respond to a specific offer. Many such techniques are available for constructing models of customer behavior and preferences. The academic data mining community has created many powerful mathematical modeling techniques, and individuals who use or evaluate mathematical models frequently use specific terminology to describe a model or a modeling technology. We can classify many of these techniques as linear vs. non-linear, and deterministic vs. probabilistic (stochastic). For example, a model of customer behavior might indicate propensity to respond to an offer increases with the number of products purchased in the past 30 days. With a linear model, propensity increases uniformly, with each increase in number of products purchased previously. That is, a linear model might estimate propensity to respond for customers who purchased zero products in the past 30 days as ten percent, 15 percent propensity for one product purchased, 20 percent for two products, and so forth. In contrast, a non-linear modeling technique might indicate a rapid increase in propensity from zero to three products purchased and a more modest increase beyond that point. Non-linear models are more flexible than linear models in their ability to reproduce patterns in historical data and therefore behavioral targeting often uses non-linear modeling techniques. Models for behavioral targeting are also typically stochastic or probabilistic models, because most observers agree that human behavior does not strictly follow deterministic rules.

Decision trees represent only one of many data mining algorithms that marketing professionals can apply to behavioral targeting. According to a survey paper, among the top 10 mathematical algorithms that have been among the most influential data mining algorithms in the research community are C4.5 and CART (two decision-tree classification algorithms), *k*NN, SVM, and Naive Bayes (classification algorithms not based on decision trees), *k*-Means (a clustering algorithm), and Apriori (an association algorithm). See Wu et al. (2008). Figure 4 lists Wu's top 10 algorithms and other prominent data mining algorithms, along with their type, sample use and prevalence in marketing applications.

Classification algorithms predict into which class an observation falls. For example, classification algorithms can predict which out of a million contacts will respond to an offer. Classification algorithms can also make predictions for problems with more than two classes. For instance marketing professionals can use

Algorithm	Type	Sample Use	Marketing Use
Apriori	Association	Cross Selling	Common
AdaBoost	Classification	Select Contacts	Common
C4.5	Classification	Select Contacts	Common
CART	Classification	Select Contacts	Common
kNN	Classification	Select Contacts	Common
Naive Bayes	Classification	Select Contacts	Common
SVM	Classification	Select Contacts	Less common
Neural Network	Classification Clustering	Select Contacts Market Segmentation	Common
EM	Clustering	Market Segmentation	Common
k-Means	Clustering	Market Segmentation	Common
Genetic Algorithm	Optimization	Resource Allocation	Less common
Simulated Annealing	Optimization	Resource Allocation	Less common
PageRank	Search Ranking	Search Ranking	Less common

Fig. 4 Data mining algorithms and their use in marketing

classification to predict for each visitor to a web site whether the visitor will not sign up for membership, sign up for a free membership, sign up for basic membership, or sign up for premium membership. One useful property of decision tree-based classification algorithms is that they are easy to interpret, as we saw for the play-tennis example above. Other classification algorithms are typically not easy or even very difficult to interpret, although they may generate predictions of similar accuracy.

Marketing professionals can also use classification algorithms to find natural groups of items that are similar to other items in the same group, and different from items in other groups. A common marketing application of clustering algorithms, such as neural networks, EM and k-Means, is to identify clusters of similar customers for market segmentation. The input for such a task might be a table containing one million customer records, each having a number of customer attributes, such as age, gender, income, address, occupation, marital status, number of purchases in the last 90 days, and so forth. Even though every one of the one million customers might be different from every other customer, and even though humans might find it impossibly difficult to identify patterns in such a large data set, clustering algorithms can segment such data sets into multiple clusters. With additional analysis, marketing professionals can then also summarize and characterize each cluster as well as differences among the clusters. Similarly, it is possible to identify specific customers who are most typical of each cluster, as well as other customers who are anomalous. Such anomalies are items that belong to one cluster, but exhibit characteristics that are less similar to other items in the same cluster than most.

Association algorithms are similarly useful in marketing applications. Amazon.com shows customers who purchase or browse for products, some other different but similar products in which the customer might also have an interest.

Amazon.com accomplishes this by searching for associations in sales transactions. For this application, association algorithms search for other customers who purchased or viewed the same items, and the algorithm then identifies other items that these other customers frequently purchased or viewed. Two metrics allow the marketing professional to fine-tune results from the association algorithm, the support and confidence metrics. For each pair of a currently viewed or purchased item together with a candidate similar item, support measures the occurrences where these items occurred together as a proportion of all items viewed or purchased. Confidence measures the occurrences where the currently viewed and the candidate similar item appeared together, as a proportion of all instances where users viewed or purchased the current item. Marketing professionals can interpret support as a measure of how popular a pair or set of items is in general. In contrast, we can interpret confidence as a measure of the conditional probability of the candidate item, given that a user is viewing or purchasing the current item.

While its use is increasing, marketing professionals use the SVM algorithm less commonly, because relatively fewer software tools offer this more recently developed algorithm to date. Although neural networks, simulated-annealing and genetic algorithms do not appear in Wu's top 10 list of data mining algorithms, they have received extensive attention in the data mining literature. They are also among the most powerful tools available to professionals who engage in data mining. Optimization algorithms, such as simulated annealing and genetic algorithms, solve problems that involve searching for the best solution when the value of each solution depends on complex, interrelated factors, as is the case in resource allocation problems. For example, optimization algorithms find near-optimal flight schedules for individuals in different locations who want to meet in one location, while minimizing ticket cost, layovers, travel times and waiting times, see Segaran (2007). Similar optimization problems in marketing involve allocating budget to a set of marketing initiatives, channels or campaigns. The PageRank algorithm originates from Google founders Sergey Brin and Larry Page, who devised this algorithm for search ranking using hyperlinks on the Web. Search plays an important role in all fields of business and its use is less specific to marketing applications.

A number of textbooks provide further information on the technical aspects of data mining. For example, the following are some of the more popular data mining textbooks: Berry and Linoff (1997), Han and Kamber (2005), Mitchell (1997), Quinlan (1993), Soukup and Davidson (2002), and Witten and Frank (2005).

3 Business Applications of Data Mining

Several recent, non-technical business books have also advertised the power of mathematical modeling techniques for maximizing profits in the commercial sector.

In "The Power to Predict," Vivek Ranadivé describes an ongoing "Predictive Business" revolution, and shows how to prepare a business for this new technology. Ranadivé reports how predictive businesses, such as Harrah's, Pirelli, E & J Gallo, and other leading-edge firms are making the transition from an event-driven real-time business model to a predictive one. See Ranadivé (2006). In another book entitled "Competing on Analytics," Thomas H. Davenport and Jeanne G. Harris argue that the frontier for using data to make decisions has shifted dramatically. Certain high-performing enterprises are now building their competitive strategies around data-driven insights that in turn generate impressive business results. Their "secret weapon" is analytics with its sophisticated techniques of quantitative and statistical analysis and predictive modeling. Davenport describes businesses that use analytics to identify their most profitable customers and offer them the right price, accelerate product innovation, optimize supply chains, and identify the true drivers of financial performance. Examples include organizations as diverse as Amazon, Barclay's, Capital One, Harrah's, Procter & Gamble, Wachovia, and the Boston Red Sox. See Davenport and Harris (2007).

In "Why Thinking-by-Numbers Is the New Way to Be Smart", Yale Law School professor and econometrician Ian Ayres argues that the recent creation of huge data sets allows knowledgeable individuals to make previously impossible predictions. He discusses the changes they are making to industries like medical diagnostics, air travel pricing, screenwriting and online-dating services. See Ayres (2007). In his book "The Long Tail," Chris Anderson shows how mathematical modeling tools can help increase revenue and sales by helping customers find items of interest. In The Long Tail, Chris Anderson offers a look at the future of business and common culture. The long-tail phenomenon, he argues, will affect industries, such as entertainment, and "re-shape our understanding of what people actually want to watch." See Anderson (2006). In his book "The Wisdom of Crowds," James Surowiecki shows how businesses can aggregate the collective wisdom of groups of people, such as users or customers, to make better decisions. See Surowiecki (2004). In "Freakonomics," Steven Levitt describes how statistical analysis can uncover hidden patterns in almost any economic activity imaginable, from betting on Sumo wrestling matches to drug trafficking and even prostitution. See Levitt (2006). Lastly, in "Moneyball," Michael Lewis recounts the tremendous success of Billie Beane's Oakland A's baseball team through the systematic use of mathematical modeling for selecting players and strategies. See Lewis (2003).

Examples of the power of mathematical modeling in the real world abound. From the Oakland A's in baseball, to Google's ad placement, many of today's leading enterprises take advantage of analytics technology for increased revenue and profits. Harrah optimizes customer care by analyzing historical records of customer behavior and spending. Verizon prevents cancellations of service contracts, by predicting "churn risk" and taking preventive measures such as offering incentives to high-risk customers. MetLife and other insurance companies use actuarial analysis and mathematical modeling to offer insurance at optimized premiums, and to minimize the risk that other insurances will offer lower rates. By taking the profitable business away from competitors through optimal pricing, MetLife looks to maximize profits. The Progressive insurance company goes as far as expecting that their competitors will lose money on the policy if a competitor offers a lower premium. Insurance companies are not only using analytics to price policies but they also use the technology to identify which of the insurance claims they receive are associated with the highest risk of fraud. Similar to the adoption of analytics in the insurance industry, banks have long analyzed credit applications to help decide which ones to approve and which ones to reject, and the IRS uses analytics to identify questionable tax returns for in-depth analysis and auditing.

3.1 Predicting Customer Preferences

When marketing professionals seek to target individual customers for offers they want to determine, among other things, to which of a set of offers each individual customer is most likely to respond. That is, marketing professionals want to understand individual customer preferences. Recent advances in electronic commerce and data mining technology make such improved targeting possible. E-commerce web sites often allow customers to rate products and services. It is possible to analyze these ratings and similarities between ratings from different users to infer which products or services each user might prefer. Similarly, web sites may track which customers visit which product pages. The data collected by these and other similar IT capabilities allow marketing professionals to target offers to individual preferences. Data about preferences, pages visited and similar activities provide information about user behavior. Therefore, marketers refer to practices that exploit such data for targeting as “behavioral targeting.”

Wikipedia defines behavioral targeting as follows:

> Behavioral targeting or behavioural targeting is a technique used by online publishers and advertisers to increase the effectiveness of their campaigns. Behavioral targeting uses information collected on an individual's web-browsing behavior, such as the pages they have visited or the searches they have made, to select which advertisements to display to that individual. Practitioners believe this helps them deliver their online advertisements to the users who are most likely to be influenced by them. Behavioral marketing can be used on its own or in conjunction with other forms of targeting based on factors like geography, demographics or the surrounding content. […] Behavioral Targeting allows site owners or ad networks to display content more relevant to the interests of the individual viewing the page.

For example, when a marketing campaign manager wants to run an email campaign that would offer a discount for one of three different digital home entertainment products, the campaign manager might take advantage of profile data members provide, including information such as age, gender, and interests. Similarly, when members log into a web site and rate products the retailer can associate each rating with a unique member. In situations where members do not log in or where unregistered users rate products, the retailer is often able to associate their rating with unique identifiers via web cookies that the web site stores on the visitor’s computer. Cookie-based associations are less reliable than member identification-based associations because visitors may access the site from different systems, which would lead at times to different cookies being associated with a single visitor. Similarly, visitors may from time to time erase cookies on their systems and later obtain a new cookie with a different identifier. Although these technological limitations reduce the retailer’s ability to correlate preferences to members and visitors, retailers can draw significant value from such behavioral data. This is possible because data mining technology does not require complete and accurate information to provide useful results. Although complete and accurate data would improve the performance of data mining-based capabilities, these technologies have an advantageous property of "graceful degradation." That is, small flaws in input data cause small degradation in data mining results, rather than inability to generate results or entirely incorrect results. Therefore, every piece of information available to a retailer can improve the retailer's ability to employ behavioral targeting technology.

For instance, one individual visitor may visit two different product pages for a digital camera and a digital music player in a web browser session on her laptop computer. Later that day, the same person may visit two different product pages for digital camera and a handheld game console in a different session on her desktop computer. The retailer can then gain valuable insights from information about both of these sessions, even though the retailer may not be able to associate these two sessions to the same visitor because both computers would contain different web cookies. By analyzing behavioral information across thousands of page visits, the retailer might learn that users who visit digital camera pages frequently also visit digital music player pages. The retailer may also learn that the association between digital cameras and digital music players might be stronger than the association between digital cameras and flat-screen TVs. To apply behavioral targeting technology for email campaigns, the campaign manager would prepare a list of contacts. She would then gather data about each contact, including information about page visits, profile information, and other similar information she can associate with each contact. She would next proceed to analyze which products correlate to each contacts behavior and select one or more products to market to the contacts.

The correlation of products to a given contact may consider products that are associated directly with that contact or associated with similar contacts. For example, Mary may have visited pages about a specific digital cameras and she may have subsequently purchased that camera model. Because Mary already purchased the digital camera, the retailer may not want to market the same product again to Mary. In addition, the retailer may not have any information about other products associated with Mary. However, because the retailer knows that users who visited pages about digital cameras also frequently visited pages about game consoles, and because the retailer may not have any information indicating that Mary would have already purchased a game console, the retailer might decide to offer a loyalty discount on a digital game console to her. The retailer could further use information about product ratings to target offers to contacts. Mary may be highly satisfied with the digital camera she purchased, and she may rate the camera favorably on the retailer's web site. Although Mary might not have rated any other product, the retailer may know that favorable ratings on digital cameras are strongly associated with high ratings on a specific book about digital photography. With this knowledge in hand, the retailer may decide to market an offer for the digital photography book to Mary.

The retailer may analyze such information manually and select individual offers for specific customers one by one, but to scale the process to thousands of contacts, the retailer might automate the process by deploying a commercial campaign-management software solution, or building a custom IT capability in-house. The components of such a solution include capabilities for collecting behavioral data, extracting and preparing the data, analyzing prepared data with mathematical modeling algorithms, and applying the constructed models to contact lists to match contacts to offers. The campaign manager can then target offers to individual customers based on their behavior and based on the behavior of other similar users.

3.2 Finding Similar Customers or Consumers

One critical step in behavioral targeting is finding similar customers. Data mining algorithms identify groups of similar items using a family of algorithms known as clustering or segmentation techniques. When the objective is to find groups of users with similar preferences, such algorithms analyze preferences for each user for a set of items. Each instance of expressing a preference might be choosing a rating for a specific product or visiting a page on a web site that describes a particular product. Clustering algorithms can then analyze preferences for each contact across thousands of products. Customers who fall into the same cluster share similar preferences and behaviors, and marketing professionals can use this fact for targeting, by selecting customers from one or more clusters for email campaigns. They might then recommend specific items, which many members of a cluster have purchased, to those cluster members who have not purchased these items. Other similar marketing strategies can also take advantage of this type of clustering analysis.

In many markets, different subsets of the larger market are dissimilar and follow their own rules and behaviors. To account for this, marketing professionals might identify distinct populations in their market and characterize differences. Marketing professionals can use the Hierarchical Clustering data-mining algorithm, among other similar techniques, to matching customers to products, make personalized recommendations, and discover groups of similar consumers. The Hierarchical Clustering algorithm begins by placing each item in its own cluster. That is, when searching for clusters of similar customers, each customer initially forms their own small cluster. Then, the algorithm finds the two most similar clusters and joins them into a new higher-level cluster. The algorithm repeats this last step until all items are contained in a single cluster. Two methods for computing measures of similarity between two customers or two clusters of customers are the Euclidian distance and Pearson correlation. While a description of these methods goes beyond the scope of this chapter, readers can find more information on these techniques in standard statistics textbooks, such as McClave et al. (2008).

Web-based business can take advantage of collective intelligence-based knowledge about similarity of customers and products by incorporating similarity into search functionality. When a user searches for a product, web sites can rank search results to favor products that similar customers prefer, and to rank products higher that are similar to products the customer prefers. If a customer has provided text in the form of comments, evaluations, feedback or similar methods, the marketing professional can also utilize query-based search technology and document filtering to personalize the customer's web experience. Search algorithms are able to match queries to specific relevant documents by learning which documents contain words that have strong links to the words appearing in the customer's query. A popular data-mining algorithm for performing this analysis is the Naïve Bayes algorithm. Marketing professionals can use such algorithms to match customer queries to product attributes and words that occur in product descriptions and other product-related texts.

Collective intelligence methods also apply to the development of pricing models, where marketing professionals use estimation techniques, such as k-nearest neighbors to find prices for products that are most consistent with a set of products for which pricing is available. For example, when a marketing professional wants to determine a price for a digital camera that is consistent with a set of prices for other cameras, she can use k-nearest neighbors to learn how camera features affect pricing, such as megapixels, zoom, battery type, and LCD screen size. The k-nearest neighbors-algorithm uses a measure of similarity between any two products, to find a small subset of products that are most similar in their features to the product the marketing professional is looking to price. The algorithm then averages the prices of the most similar products to produce a price that is consistent with the pricing of the other products.

4 Applying Collective Intelligence in Marketing

Many authors and businesses describe success with data mining, but how can a business go about applying analytics in their business context to improve targeting and personalization in marketing? The data mining industry has developed a

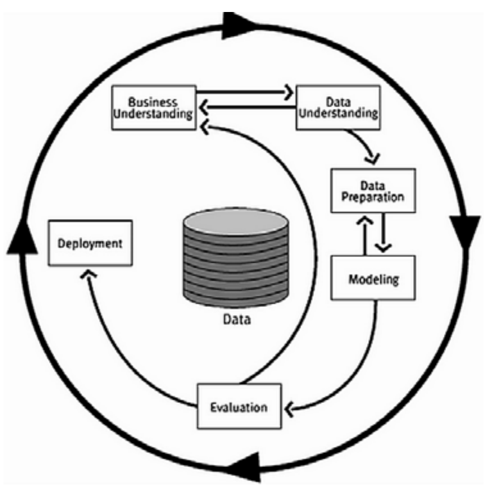


Fig. 5 CRISP-DM Data Mining Process

prescription for how to implement data mining, called the "CRoss-Industry Standard Process for Data Mining (CRISP-DM)", see Chapman et al. (2000). The CRISP-DM standard identifies a cycle of multiple process steps that comprise business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Figure 5 illustrates the CRISP-DM process.

The process breaks down into a series of steps with backtracking to earlier steps as required, and the process begins with business understanding. In this step, the organization learns about the business context, business process, project business objectives, the business systems that generate data, and the manner in which results from the data-mining project will generate business benefits. The marketing professional should begin the data mining process with an understanding of business needs and success criteria. If it is not clear what would constitute success of the project, she might document specific performance criteria, such as improving the "Click-Through Rate" (CTR) for an email campaign or reducing "Cost per Acquisition" (CPA), also known as "Cost per Action."

After the business-understanding step follows a data review to gain an initial understanding of the data. Here, the marketing professional identifies required and available data for executing the analytic project. For example, she might want to use twelve months of click-stream data to investigate which product pages customers tend to visit in the same session to understand associations among products. If some of the required data is not currently available, she must reduce the scope of the project or obtain additional data. For instance, the IT organization may have only seven months of click stream data available. If the project is to produce predictions rather than descriptive results, such as predictions of which customers will respond to which offers, then the marketing professional must also obtain sample output data for "training" the predictive model. For the example of predicting customer responses to offers, the marketing professional would collect information about which customers responded to past offers, or which customers purchased a product by clicking through an email campaign.

Following this review, the organization performs a data load that brings all the data together into a format that is convenient for expeditious and sophisticated analysis and mining. When the augmented data set is available, in-depth analysis and modeling begins, and the organization applies modeling algorithms to the prepared data to generate one or more mathematical models of the prepared data. Such a mathematical model expresses patterns of relationships between input and output data. For a give marketing campaign, the inputs might consist of various customer attributes in a customer database, and the output might be the propensity to purchase a specific product. In order to model such patterns, a data-mining algorithm must extract them from a data set that contains historical data, known as a training data set. For example, a training data set might contain records that represent purchase transactions, including attributes of the customer who purchased and which products they selected. Often, marketing professionals can find such data in multiple relational databases, web transaction logs and other similar sources. Now the analyst must prepare input data by arranging all input data in a single table. For predicting contact responses to an email campaign, each row corresponds to a contact who received the campaign and each input column corresponds to a contact attribute that might help predict the contact's response, such as age, number of pages visited, total revenue from purchases, and so forth. The analyst should then analyze input and output data characteristics to understand the nature of the specific data set she prepared, and to detect any problems and issues the data set might contain. For instance, if the marketing professional expects revenue in the order of hundreds of dollars, and the data in the revenue column is in the order of millions, then the analyst must investigate and resolve the discrepancy.

Having generated one or more models, the organization would then analyze model behavior to understand how each model reacts to its inputs and how it behaves under various distinct conditions. With the best performing model identified the marketing professional proceeds to deploy that model. The deployment includes two distinct facets: the technical deployment into the client's IT infrastructure, and the deployment into the client's business process. Once the organization deploys and integrates the predictive model into its business process, tangible benefits can emerge, such as increased revenue and profits, and improved customer satisfaction.

5 Challenges of Applying Collective Intelligence in Marketing

As with many new technologies, to succeed with analytics, one has to overcome challenges and limitations. Presently, the most significant challenge is probably the lack of skilled professionals. Few universities produce graduates that specialize in analytics, and few university programs incorporate analytics training into their curricula, see Baker and Leak (2006). Moreover, few organizations that do not already practice analytics would have the tools in place that enable the sophisticated analytic techniques that make data mining possible. The complexity of the subject matter aggravates the issue of scarcity of skilled professionals. The math underlying statistical modeling technology is notoriously impenetrable for those who do not have the required mathematical training and predisposition. In addition to skilled professionals, marketing organizations also require access to technology, and analytics software offered by leading vendors is currently costly, although prices have been declining somewhat. We might expect that the number of tools available will continue to increase as analytics becomes more popular, and their prices might continue to fall as the market for these tools grows and vendors compete more aggressively. Yet, selecting and putting in place the required tooling will likely remain a significant challenge to organizations that do not yet practice analytics.

Fortunately, lack of data is becoming less and less of a problem in the modern economy. Instead of suffering from a lack of data, organizations more frequently report that they "drown in data." Numerous information technology solutions provide seemingly unlimited streams of data. Customer Relationship Management (CRM) solutions provide data from sales-force automation, marketing, and customer support. Issue tracking systems complement CRM and configuration data with defect reports and complaints about products. Web site access logs contain click-stream data that many organizations leverage to understand how customers navigate web sites and purchase products. See Tancer (2008). Typically, all that information is available from relational data sources that can serve as a source for mathematical analysis. When relational data is not available, IT professionals can often extract information from logs and collections and make it available for analysis, although this may require substantial effort. In situations where the quantity or quality of required data is unavailable, the marketing professional may have to postpone the construction of collective intelligence models until she can create or enhance data collection processes. Here, the Goal-Question-Metric (GQM) paradigm can help marketing professionals implement such data collection improvements systematically. GQM begins by documenting the goals one wishes to achieve, and the next step is to determine which questions we must answer in order to accomplish these goals. Finally, questions break down into a series of metrics. See Basili et al. (1994).

Independent of the source, data quality is usually a major concern. When humans manually enter information into an IT system, such as when call center agents enter details on customer support cases into a CRM system, they introduce errors of all imaginable kinds into the data. Users may misspell, pick wrong items, miss important entries, and pick random answers when they lack information. Obviously, this source of poor data quality is not limited to CRM data but present everywhere humans enter data. Another and more pernicious source of poor data arises when humans providing data have incentives that relate to the content they enter. For example, when employees have goals to reduce the severity of required time to close reported problems, they may provide skewed data. In many situations, it is necessary to identify data for rejection by removing any records containing values the marketing professional believes to be incorrect. A more sophisticated alternative to rejecting and removing incorrect data is to correct data defects. However, it is not always possible to correct bad data because of the amount of additional effort and time required to make such corrections, and because it may not be possible to determine correct values to replace for the incorrect values. The marketing professional usually bases the decision of whether to remove or correct faulty data on the overall amount of data available, the ratio of good to bad data, the resources available for the project and the required precision of the analysis. Since data-mining projects can be complex, it might be a reasonable choice to reject poor-quality data, and consider coming back to repair faulty data later, once the marketing professional has developed an understanding of the quality and utility of the results she can obtain without correcting faulty data.

A more subtle pitfall in analytics concerns the behavior of the mathematical model. One of the questions that analysts might consider before deploying a model are whether the model can help with what-if scenario analysis. If a model is to support what-if scenarios, then one may expect that the model should react to changes in inputs with predictions that change in the expected direction. For example, if a model predicts propensity to purchase a product based on the customer's income, it might be surprising if a model were to predict the propensity to purchase as repeatedly rising and falling, and fluctuating up and down, as income increases. Yet, mathematical models can display such behavior if analysts do not carefully construct them for a specific, desired behavior. Similarly, the analyst might consider whether a model can extrapolate. Consider a model that predicts the total purchase volume per year based in part on customer income. If the historical training data that the analyst used to build the model contained customer income in the range of zero to $100,000, what will the model predict when it must predict annual purchase volume for a customer with annual income of $200,000? Depending on how the model functions internally, such a model may generate predictions of trillions of dollars.

Figure 6 is a simple flowchart depicting the process of what-if model evaluation. The marketing professional might begin with a base-case data record for which the model provides a score. For instance, this base case might represent a contact earning $60,000. The marketing professional then develops an expectation of how scores should change when income changes in either direction. The marketing professional might expect propensity scores to change smoothly with income, and peak at one of the extremes of the income range or somewhere inside this range. To investigate whether model behavior conforms to expectations, the marketing professional generates one or more what-if cases and obtains model scores for the new cases. If the model generates scores that fluctuate repeatedly up and down as income increases, or if the model produces scores of trillions of dollars for particularly input values, the marketing professional might be reluctant to accept that model. A more demanding expectation may involve a particular shape of model response with peaks and lows at specific points, and having steeper or shallower slopes in specified locations.

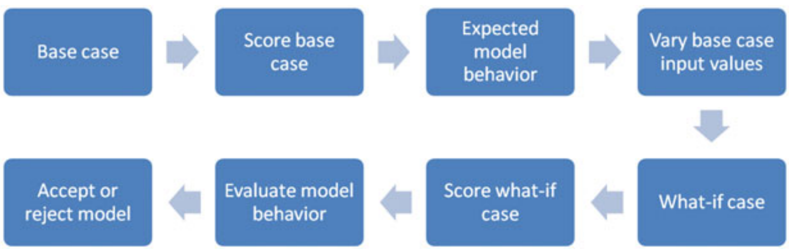


Fig. 6 Using what-if analysis to evaluate a model against expected behavior

Once a marketing professional has constructed one or more candidate models, she will want to evaluate model accuracy and test the model to explore what performance she can expect for her specific need. She may then review the behavior and accuracy of the model to assess its utility and capability. If the model meets her needs, she will have to train staff in the use of the new capability that her model enables. If the new capability is autonomous and does not require staff to act on the results the model generates, then required training may be limited to IT staff or other personnel who monitor model performance, and those who would retrain or refresh the model when its performance degrades below a predetermined threshold.

It is usually advisable to track model performance and to refresh model periodically to avoid silent deterioration of business performance that might go unnoticed. After refreshing a collective intelligence model, the updated model might be able to capture newly emerging trends in the market that may not have existed when the previous model became available. As marketing professionals consider deploying mathematical models to leverage collective intelligence, they might also ask themselves how long they can expect a model to perform well before they must update the model. Business processes change over time and data-mining professionals think of this phenomenon as “concept drift”, because they aim to cast a concept as a mathematical model and as the process changes, this concept drifts away from the model. To avoid working with an out-of-date model it is advisable to monitor model performance over time and to retrain and refresh predictive models as the underlying concepts drift away from them.

6 Case Study: Keyword-Based Product Suggestions

In this case study, we describe how a marketing professional can personalize offers using collective intelligence. The data presented in this case study originates from a business offering products and services relating to computers, entertainment, wireless and other types of products. The marketing professional has collected texts that customers have entered into web site dialogs and forms, along with information on which type of product, “wireless,” “entertainment” or “computer,” each customer selected. The marketing professional has then created a model of how keywords in these texts predict which type of product customers chose. For example, the model might incorporate a pattern that suggests that customers or prospects who mention the word “TV” in a comment might be interested in products and services relating to “entertainment,” rather than “computer” or “wireless.”

The marketing professional can use such a collective intelligence model in various ways. For example, the professional might purchase keywords with a search engine provider, such as purchasing keywords in Google’s AdWords, and the professional could then direct visitors to a landing page that provides different specific information for visitors who are more likely to be interested in entertainment, wireless or computers. Similarly, when customers interact with a call center, the Customer Relationship Management (CRM) software the call center uses might instruct customer services representatives to make specific personalized offers, based on keywords customers use in chat messages, or using voice recognition software. Other similar uses of such keyword-based product suggestion models are also possible.

The first step in this case study is to extract raw data from the data source. Our data source is a relational database that constitutes the backend of the web site into which customers entered text. In this case, we extract three attributes from the data source, the customer text identifier, “CUSTOMER_TEXT_ID”, the text the customer entered, “TEXT”, and the type of product the customer selected, “PRODUCT_TYPE”. The following table shows ten sample rows of raw data to illustrate the raw data we use in this case study.

CUSTOMER_TEXT_ID	TEXT	PRODUCT_TYPE
1067102	Fibre Channel Controller	Computers
1154801	Audio Interface	Wireless Communications
1222223	4 bands GSM/GPRS	Wireless Communications
1232202	HD-DVD/Bru-ray	Entertainment Products
1248206	Digital Still Camera	Entertainment Products
1287981	Video and Graphics controller	Computers
1478315	Video Co-processor	Entertainment Products
1608812	Personal Computer	Computers
1640722	Digital cellular phone	Wireless Communications
1992838	Embedded Graphics Controller	Computers

Fig. 7 Extracted Raw Data

We next review the distribution of the product type in the raw data and observe that the distribution is imbalanced as it contains more records for entertainment than for the other types, and more records wireless than for computers. Figure 8 illustrates the imbalance. Such imbalances in the target attribute can lead to undesirable model behavior. For illustration, consider a data set in which 99% of the records pertain to entertainment, while only one percent belongs to the other two categories. Given such a highly imbalanced dataset, many data mining algorithms produce models that always predict "entertainment" irrespective of attribute values of input cases. Such a result is understandable, considering that the data set contains much evidence for the outcome "entertainment" and little evidence for either of the other two possible outcomes. Moreover, in this scenario, a model that always predicts "entertainment" would be 99% accurate when tested against a similarly imbalance data set. To avoid these issues with imbalanced outcomes, data mining professionals typically "balance" training data before constructing models, or they may use algorithms that balance data automatically. For more information on mining imbalance data, see Chawla et al. (2004).

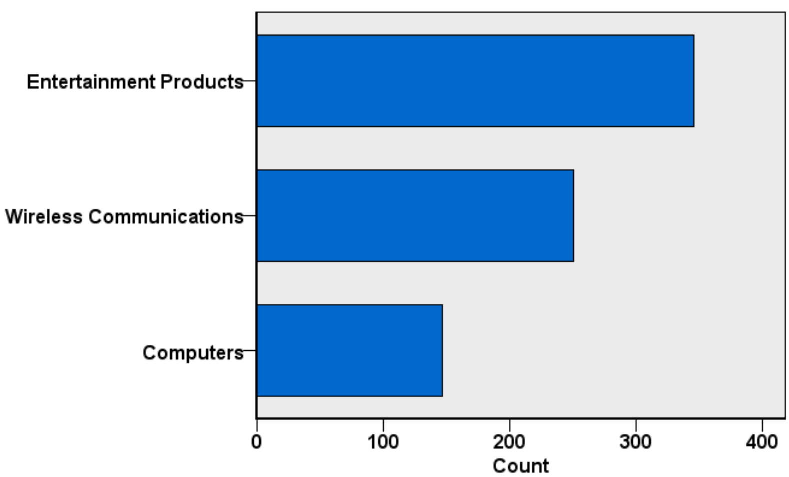


Fig. 8 Imbalanced Raw Data

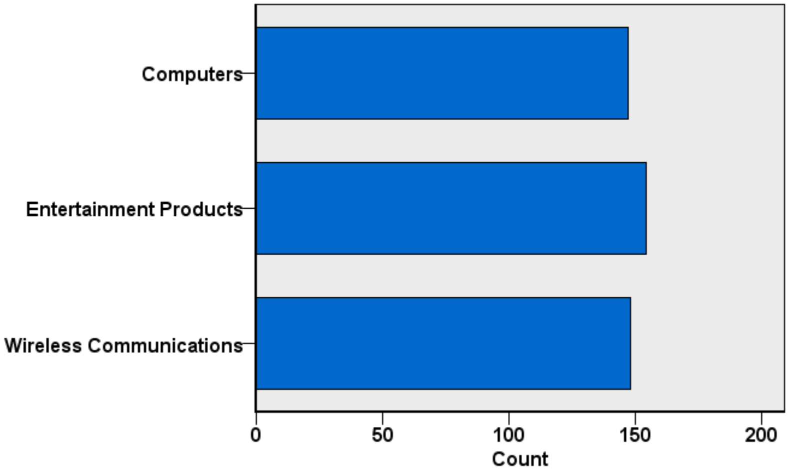


Fig. 9 Raw Data after Balancing

The marketing professional now balances the data by sampling the larger categories. This process ensures that data mining algorithms can best extract patterns from this data that would allow predicting the product type. Figure 9 shows the distribution after balancing the raw data.

Although we balanced the data set, we obtained slightly unequal numbers of cases for each product choice. This remaining imbalance occurred because we applied single-pass, random sampling to balance the data, which results in an only approximately balanced data set. Although this issue would likely not affect the analysis significantly, the marketing professional could further revise the analysis to avoid this issue. After balancing the raw data, the marketing professional prepares the data to generate a new flag column for every keyword of interest that occurs in the comments. For this case study, we generated 353 such flag columns. Figure 10 illustrates these flags for six of the extracted keywords. Because there

CUSTOMER_TEXT_ID	PRODUCT_TYPE	analog	bit	bridge	camera	chipset	clock
1330210	Computers	0	0	1	0	0	0
1683793	Entertainment Products	0	0	0	0	0	0
1298110	Computers	0	0	0	0	0	0
1190691	Computers	0	0	0	0	1	0
1813767	Computers	0	1	0	0	0	1
1085539	Entertainment Products	1	0	0	0	0	0
1484331	Entertainment Products	0	0	0	0	0	0
1828620	Entertainment Products	1	0	0	0	0	0
1706757	Entertainment Products	0	0	0	1	0	0
1211690	Entertainment Products	0	0	0	1	0	0

Fig. 10 Keyword Flags

are numerous keywords across all comments, and because the comments are relatively short, the resulting table will be sparse in the sense that the overwhelming majority of cells contain the value zero and few cells contain the value one. For illustration, Figure 10, below, shows a selection of columns that contain at least one entry of one for the sample rows shown. A more typical portion of the flag table would show fewer entries of one and more zeros.

The marketing professional now uses this table as input to a decision-tree algorithm. The column "PRODUCT_TYPE" serves as output and all keyword flag columns serve as inputs. Figure 11, below, shows a portion of such a decision tree model.

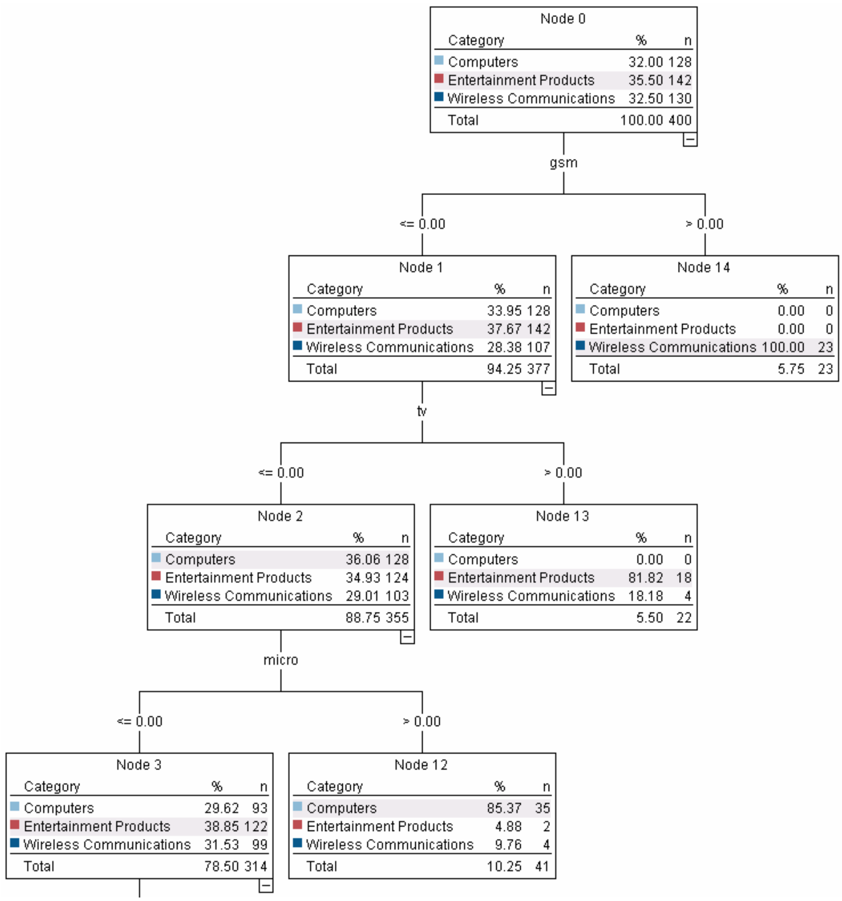


Fig. 11 Decision Tree Model (partial)

This model indicates that when the keyword "gsm" is present the model predicts the product type Wireless. If the keyword "gsm" is not present, the model checks for the keyword "tv." If "tv" is present, the model predicts the product type Entertainment. Otherwise, the model checks for the keyword "micro" and predicts the product type Computers. The model then continues to check for other keywords, including “pc” (Computers), “dma” (Wireless), “cd” (Entertainment), and “controller” (Computers).

The marketing professional may now evaluate the performance of this model as Figure 12 below shows.

Product		Computers	Entertainment	Wireless	Total
Computers	Count	59	2	67	**128**
	Row %	46.1	1.6	52.3	**100**
	Column %	84.3	2.7	26.3	**32**
Entertainment	Count	2	67	73	**142**
	Row %	1.4	47.2	51.4	**100**
	Column %	2.9	89.3	28.6	**35.5**
Wireless	Count	9	6	115	**130**
	Row %	6.9	4.6	88.5	**100**
	Column %	12.9	8.0	45.1	**32.5**
Total	Count	**70**	**75**	**255**	**400**
	Row %	**17.5**	**18.8**	**63.8**	**100**
	Column %	**100.0**	**100.0**	**100.0**	**100**

Fig. 12 Evaluation Matrix for Decision Tree Model

The evaluation matrix tabulates the frequency of correct and incorrect predictions. To do so, the table shows actual product choices across the rows and predicted product choices in the columns. For example, the value 59 near the top left of the table indicates that the model correctly predicted “Computers” in 59 cases. The values to the right show that, for the other cases where the customer chose “Computers,” the model incorrectly predicted “Entertainment” in two cases and Wireless in 67 cases. Summing these three values, we see that the total number of cases with an actual choice of “Computers” is 128. The next row, labeled “Row %,” expresses these values as percentages of the total 128 cases. The following row, labeled “Column %,” allows the marketing professional to evaluate all predictions where the model made a specific prediction. For example, the value 84.3% shows that when the model predicted “Computer” the prediction was correct in 84.3% of those predictions. The Totals in the three bottom-most rows of the table tally all predictions, right or wrong, which the model made for “Computers” (70 or 17.5% of all predictions), “Entertainment” (75 or 18.8% of all predictions), and “Wireless” (255 or 63.8% of all predictions).

Reviewing the table data by columns, we already saw that the model correctly predicted “Computers” for 59 cases. We observe further that the model also predicted “Computers” when the correct choice would have been “Entertainment” in two cases (2.9%), and “Wireless” in nine cases (12.9%). The right-most column indicates that the full data set contained 128 cases where the actual product choice was “Computers”, which represents 32%, 142 cases for “Entertainment” (35.5%), and 130 “Wireless (32.5%), for a total of 400 cases analyzed. We observe a smaller number of observations than we obtained after balancing the data set (see Figure 9). This reduction occurred because some of the cases we selected did not provide any text, that is, the data field was NULL. Although these issues did not likely affect the analysis significantly, the marketing professional could revise the analysis to avoid these issues, for example by replacing NULLs with empty strings, or other similar techniques.

Depending on the needs of the marketing professional, there are various insights and conclusions to draw from this evaluation, and we will consider a subset of these. For a more comprehensive analysis, see Bruckhaus (2007). Turning our attention to the performance of the predictive model, we observe that this model predicted "Wireless" more frequently (63.8%) than Computers (17.5%) or Entertainment (18.8%). This occurred because the keywords that are most indicative of Wireless also occur frequently in comments from customers who selected Computer and Entertainment. In contrast, keywords that are indicative of Computers and Entertainment appear to be more specific. Therefore, only 45% of the "Wireless" predictions were correct, while 84.3% of the "Computers" predictions were correct, and 89.3% of the "Entertainment" predictions were correct. However, the tendency to predict "Wireless" leads to correctly classifying 88.5% of all cases where the customer chose "Wireless," whereas the model only classifies 46.1% of the Computers cases and 47.2% of the Entertainment cases correctly.

Because of the model bias toward predicting "Wireless," the model would perform better in situations where it is important to achieve one or both of the following goals:

- Identify the majority of customers who are interested in Wireless products. This objective might be important if Wireless products provide a greater profit margin, or if a business wants to grow revenue for this type of product for other reasons. The model achieves a recall rate of 88.5% for Wireless; however, precision for wireless is lower at 45.1%.
- Prevent false positives when predicting that customers might be interested in Computers or Entertainment. This objective might be important if presenting offers for Computers or Entertainment to customers who would not select those products entails significant cost, might dissatisfy the customer, or might have other undesirable effects. The model achieves precision rates of 84.3% (Computers) and 89.3% (Entertainment); however recall for these product types is lower at 46.1% (Computers) and 47.2% (Entertainment).

In the author's experience with introducing collective intelligence technology in real-life situations it has often been more important initially to focus on preventing false positives, and to put less emphasis on preventing false negatives. When businesses introduce new methods and tools, they are typically careful to avoid or minimize false positives because false positives can lead the organization to take action inappropriately. For example, when targeting offers, it can be a good strategy to target initially only a small group of individuals with the highest estimated propensity to respond. After gaining some experience with such an approach, marketing professionals may be able to develop procedures for handling false positives. For instance, when service and support call center agents use behavioral targeting for cross-selling and present offers to callers, it may be important to develop processes for those false positives where the caller reacts negatively to an offer. After gaining experience with handling false positives, the marketing professional may then shift focus to improve recall and accept an increased incidence of false positives.

For a discussion of these and other model evaluation metrics, see Caruana and Niculescu-Mizil (2004), and Caruana and Niculescu-Mizil (2006).

7 Summary and Conclusions

In this chapter, we reviewed how marketing professionals apply novel collective intelligence technology to solve marketing problems. Marketing professionals can apply data mining technology to analyze data they collect from their customers and the market in general. They can use this data to build sophisticated models that express the patterns inherent in this data, and then apply those models for behavioral targeting. For example, marketing professionals can target and personalize offers for individual customers and identify customers with similar behavior. We presented a case study that demonstrated how marketing professionals could use customer comments to personalize offers with behavioral targeting to leverage the collective intelligence of all customers who commented.

For the marketing professional, the practical implications and utilities of the collective intelligence techniques presented here are of primary importance. Of particular interest are the people, processes and tools required to use collective intelligence technology successfully for practical applications. The lack of skilled professionals a key challenge as the successful application of collective intelligence analysis requires substantial training and experience. To aid in working through the process of applying collective intelligence technology to practical marketing tasks, marketing professionals may employ the “CRoss-Industry Standard Process for Data Mining (CRISP-DM)”, which guides the practitioner through business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Although it is possible to implement collective intelligence software from scratch, only the most resourceful marketing professionals would consider that route. Many established vendors offer marketing software that supports advanced analytics and modeling, and open source tools are increasingly becoming available as a viable alternative.

If we can believe the literature, rich rewards await those organizations that adopt analytics. Most observers would agree that, when applied appropriately, analytic capabilities could present a significant competitive advantage over those that do not employ the technology. The recent boom in products, books and publications on analytics is evidence of a tremendous business trend. Many of those that have adopted analytics are already enjoying the fruits of their labor, and their competitive advantage over late adopters may continue to grow with further advances in technology.

References

Anderson, C.: The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion (2006)

Ayres, I.: Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart. Bantam Dell (2007)

Baker, S., Leak, B.: Math Will Rock Your World. Business Week, 54–62 (January 23, 2006)

Berry, M., Linoff, G.: Mastering Data Mining: The Art and Science of Customer Relationship Management. John Wiley & Sons, Chichester (1999)

Basili, V.R., Caldiera, G., Rombach, H.D.: Goal Question Metric Paradigm. In: Marciniak, J.J. (ed.) Encyclopedia of Software Engineering, pp. 528–532. John Wiley & Sons, Chichester (1994)

Bruckhaus, T.: The Business Impact of Predictive Analytics. In: Zhu, Q., Davidson, I. (eds.) Knowledge Discovery and Data Mining: Challenges and Realities with Real World Data, Idea Group Publishing, USA (2007)

Caruana, R., Niculescu-Mizil, A.: Data mining in metric space: an empirical analysis of supervised learning performance criteria. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 69–78. ACM, New York (2004)

Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd international conference on Machine learning, pp. 161–168. ACM, New York (2006)

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0: Cross Industry Standard Process for Data Mining CRISP-DM Consortium (2000)

Chawla, N.V., Japkowicz, N., Kolcz, A. (eds.): Special Issue on Learning from Imbalanced Datasets. SIGKDD, 6(1) (2004)

Davenport, T.H., Harris, J.: Competing on analytics: the new science of winning. Harvard Business School Press, Boston (2007)

Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. The Morgan Kaufman Series in Data Management Systems. Morgan Kaufman, San Francisco (2005)

Levitt, S.D., Dubner, S.J.: Freakonomics: A Rogue Economist Explores the Hidden Side of Everything. William Morrow (2005)

Lewis, M.: Moneyball: The Art of Winning an Unfair Game. WW Norton & Company, New York (2003)

McClave, J.T., Benson, P.G., Sincich, T.: Statistics for Business and Economics: International Edition. Pearson Prentice Hall, London (2008)

Mitchell, T.: Machine Learning, 1st edn. McGraw Hill, New York (1997)

Quinlan, J.R.: C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco (1993)

Ranadive, V.: The Power to Predict: How Real Time Businesses Anticipate Customer Needs, Create Opportunities and Beat the Competition. McGraw-Hill Professional, New York (2006)

Segaran, T.: Programming Collective Intelligence: Building Smart Web 2.0. Applications. O'Reilly, Sebastopol (2007)

Soukup, T., Davidson, I.: Visual data mining: Techniques and tools for data visualization and mining. Wiley & Sons, Chichester (2002)

Surowiecki, J.: The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. Doubleday (2004)

Tancer, B.: Click: What Millions of People Are Doing Online and Why It Matters. Hyperion (2008)

Witten, I.H., Frank, E.: Data mining: practical machine learning tools and techniques with Java implementations, 2nd edn. Morgan Kaufman, San Francisco (2005)

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S.: Top 10 algorithms in data mining. Knowledge and Information Systems 14(1), 1–37 (2008)

Predictive Modeling on Multiple Marketing Objectives Using Evolutionary Computation

Siddhartha Bhattacharyya

Information and Decision Sciences,
College of Business Administration,
University of Illinois,
Chicago, 312-996-8794
e-mail: sidb@uic.edu

Abstract. Predictive models find wide use in marketing for customer segmentation, targeting, etc. Models can be developed to different objectives, as defined through the dependent variable of interest. While standard modeling approaches embody single performance objectives, actual marketing decisions often need consideration of multiple performance criteria. Multiple objective problems typically characterize a range of solutions, none of which dominate the others with respect to the different objectives - these specify the Pareto-frontier of non-dominated solutions, each offering a different level of tradeoff. This chapter examines the use of evolutionary computation to obtain a set of such non-dominated models. An application using a real-life problem and data-set is presented, with results highlighting how such multi-objective models can yield advantages over traditional approaches.

1 Introduction

Predictive models find wide use in marketing for various customer segmentation and targeting applications, customer acquisition and retention, cross sell and upsell, lifetime value modeling and others (Berry and Linoff 2004). Statistical regression based models typically form the basis for such models, and techniques like decision trees and neural networks have also gained acceptance in practice. Models can be developed to different objectives, as defined through dependent variable of interest. In customer targeting, for example, response models may be built from data identifying individuals as responders/non-responders. Or models may seek to identify individuals with the highest response frequency in previous solicitations, or those that have generated most revenue in earlier purchases.

While standard modeling approaches embody single performance objectives, actual marketing decisions often need consideration of multiple performance criteria. For example, marketers may look for individuals likely to respond to a solicitation and also generate high purchase revenues; or, a cellular carrier may seek to identify customers most likely to churn and who also generate high usage revenues, in order to minimize potential losses likely from these individuals leaving

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 155–179.
springerlink.com

for a competitor. While separate models optimized on the different criteria can be combined to obtain a joint measure of expected performance, handling multiple objectives separately in this manner seldom yields adequate solutions to the overall problem. Further, different performance objectives sought can often run counter to each other – for example, high revenue potential can run counter to churn likelihood, or customers most likely to respond to a solicitation may not the ones with high purchase revenues. Given conflicting objectives, high performance from a model on one objective can correspond to poor performance on the others; a suitable solution here will involve obtaining an acceptable *tradeoff* amongst the multiple objectives.

Multi-criteria optimization problems are often reformulated as single objective problems. Aggregation functions based on domain knowledge and decision-maker preference may be used, and linear weighted averages are often considered, with weights for the different objectives specified according to desired tradeoffs. Where the nature of such tradeoffs is not well understood – as in the case of most complex data-mining scenarios - a precise articulation of preferences becomes difficult. Usually, multiple solutions incorporating varying tradeoffs amongst the objectives need to be obtained, and the most satisfactory amongst these chosen.

Multi-criteria problems, especially when considering conflicting objectives, do not carry a single best solution, but are instead, characterized by range of solutions, none of which dominate the others with respect to the different objectives. These specify the *Pareto-frontier* of non-dominated solutions, where each solution offers a different level of tradeoff, and can be the decision model of choice. When multiple performance criteria are of importance, the effectiveness of models from across the Pareto-frontier thus needs to be considered. An exploration of such solutions through ad-hoc manipulation of a weighted objective function is inefficient and tiresome. A preferred approach is to obtain a set of Pareto-optimal solutions in a single invocation of the model development procedure. This chapter examines the use of evolutionary computation to obtain a set of such non-dominated models.

Evolutionary computation (EC) techniques like genetic algorithms (Goldberg 1989, Michalewicz 1994) and genetic programming (Koza 1993) offer a search approach based loosely on principles of natural selection and biological evolution. They provide a powerful, general purpose search mechanism that has found application in problems ranging from the scheduling problems, development of financial trading rules and portfolio management, design of engines and aerospace structures, to modeling of varied economic phenomena, mechanisms for adaptive behavior in autonomous agents, and others. They have been found useful, in general, for obtaining solutions for hard optimization problems that are not amenable to solution using traditional approaches.

EC approaches have also been applied for classification and data mining. A unique advantage stems from the representational flexibility on the model structure. Appropriate model structure in data mining problems can vary, based on the nature of the problem and available data and solution characteristics desired. Model structure usually arises from the modeling technique used. For example, logistic regression yields a model for the dependent variable that is functionally linear in the set of predictor variables, while CHAID or CART models take the form of decision trees or restricted rule sets. Model representation is crucial since it largely determines the nature of patterns that are discernable from the data. Evolutionary search can be usefully applied with a range of representational forms. They have been applied to learn linear discriminant functions, condition-action rules, decision tree models, association rules, as well as for learning neural network and support vector machine models. Genetic programming can represent general program structures, and have been used in data mining to obtain varied non-linear models on the predictor variables.

A second key advantage of EC models arises from the flexibility in formulation of the search objective (fitness function). The search objective determines the nature of obtained models and performance characteristics. Traditional approaches like logistic and least-squares regression models seek to maximize likelihood or minimize sum-of-squares of errors; decision tree models seek to minimize some measure of 'impurity' at the nodes; others may seek to minimize classification error rates. Model performance is then assessed on a range of measures, like accuracy, true/false positives and negatives, lift, AUC (area under the ROC curve), etc. Performance of models in application also needs to consider the business context, and the business objective may not correspond well to the precise objective function of the model development procedure. Given the business requirements in many direct marketing problems of specific targeting depths given certain budget constraints, genetic algorithms have been proposed to obtain models optimized for specified targeting depths (Bhattacharyya 1999). EC, with its ability to tailor the search to specific performance and business needs through a flexible fitness function formulation, offers notable advantages in this regard.

Evolutionary computation techniques like genetic algorithms have been noted to hold advantages for multi-objective optimization problems (Coello et al. 2006, Deb 2001, Fonseca and Fleming 1995). Various papers in recent years report on the use of evolutionary algorithms for data mining (Bhattacharyya 1999, Kim and Street 2004, Sikora and Piramuthu 2005, Zhang and Bhattacharyya 2004). Their use for multi-objective data mining problems has been suggested (Bhattacharyya 2000, Casillas and Martinez-Lopez 2009, Dehuri et al. 2006, Dehuri and Mall 2006, Freitas et al. 2002, Handl and Knowles 2004, Ishibuchi and Yamamoto 2003, Kaya 2006, Murty et al. 2008, Pappa and Freitas 2009, Thilagam and Ananthanarayana 2008); see Dehuri et al. (2008) for a recent review. A discussion on multi-objective problems in data-mining is given in Freitas (2004).

We first examine the concept of non-dominated solutions and the Pareto frontier, provide a brief introduction to essential concepts in genetic search and examine the literature in multi-objective evolutionary computation and its application to data mining. The following section then describes how non-dominated solutions on multiple objectives are obtained using genetic search, model representation, search operators, the fitness functions to embody the marketing objectives considered, and how performance is evaluated. Next, an application using a real-life problem and data-set is presented, with results highlighting how such multi-objective models yield advantages over traditional approaches.

2 Background

This section introduces non-dominated solutions and the Pareto frontier in multi-objective problems, and provides an introduction to genetic search. Next, an overview of multi-objective evolutionary computation and its application to data mining is given.

2.1 Non-dominated Solutions

Problems with multiple objectives typically carry multiple solutions, each offering different tradeoffs amongst the objectives. This is shown in Figure 1, with π_1 and π_2 as two objectives. The solutions of interest are those displaying better performance than others on at least one objective. This is the set of non-dominated solutions, which shows strictly better performance on the two objectives than the dominated solutions.

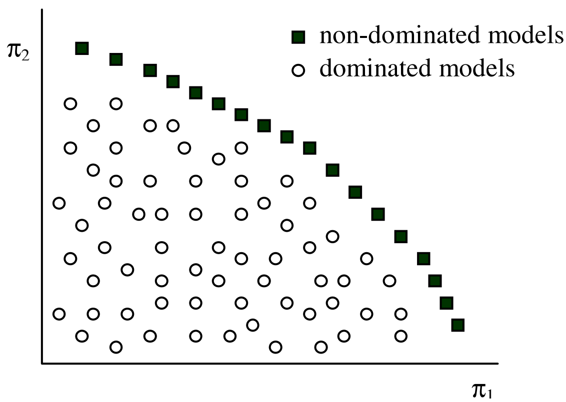


Fig. 1 Multiple objectives and non-dominated solutions

Non-dominance of solutions with respect to the objectives π_i can be formalized as follows: consider n objectives $\pi_i(f(\mathbf{x})),\ i=1,..,n$ where $f(\mathbf{x})$ is a model defined on the vector $\mathbf{x}$ of predictors. Assuming, without loss of generality, the goal of maximization on all objectives, a model $f^a(\mathbf{x})$ is said to *dominate* another model $f^b(\mathbf{x})$ iff:

$$\forall i : \pi_i^d(f^a(\mathbf{x})) \geq \pi_i^d(f^b(\mathbf{x})) \text{ and } \exists j : \pi_j^d(f^a(\mathbf{x})) > \pi_j^d(f^b(\mathbf{x})).$$

Otherwise the models $f^a(\mathbf{x})$and $f^b(\mathbf{x})$ are *non-dominated* with respect to each other. The set of models that are non-dominated by other models forms the non-dominated or *Pareto-optimal* set of models.

Several non-dominated models typically exist for multi-objective problems, especially when considering conflicting objectives. Here, high performance on one objective corresponds to poor performance on the other. The set of non-dominated models along the Pareto-frontier represent different levels of tradeoff amongst the objectives. Solutions from traditional methods that optimize single objectives will typically be towards the extremities of the frontier.

Adequately addressing multi-objectives requires consideration of solutions along the entire Pareto frontier, so that decision makers can examine different tradeoffs. Traditional methods attempt to find solutions distributed across the non-dominated frontier by, for example, optimizing on an aggregated function of the objectives (Zeleny 1982) and varying the parameters to obtain individual solutions. Weighted combination of objectives into a single function to optimize will foster search towards a specific part of the tradeoff frontier. Linear weighted averages are often considered, with weights on the objectives based on desired tradeoffs or other domain knowledge. Without adequate prior understanding of the nature of tradeoffs and different solutions obtainable, considering such weighted combinations of objectives presents an unreliable and ad-hoc approach (Freitas 2004). This, however, remains the common approach to addressing multiple objectives in data mining. Other traditional approaches like hierarchical regressions also yield only a single solution, with the model-builder having little control over the tradeoff manifest in this solution.

Evolutionary computation based approaches, with their population based search process, present effective mechanisms for searching along multiple objectives in parallel (Coello 2000, Coello et al. 2006, Deb 2001). Here, solutions along the entire Pareto frontier are simultaneously obtained, without any need for preference weighting on objectives. It thus readily provides decision-makers with a range of models exhibiting varying levels of tradeoff. Decision on a model to implement may then be taken after consideration of performance tradeoffs that different models along the Pareto frontier reveal.

2.2 Genetic Search

Genetic algorithms provide a stochastic search procedure based on principles of natural genetics and survival of the fittest. They operate through a simulated evolution process on a population of structures that represent candidate solutions in the search space. Evolution occurs through (1) a selection mechanism that implements a survival of the fittest strategy, and (2) genetic recombination of the selected strings to produce ‘offspring’ for the next generation.

The basic operation of a simple GA is illustrated in Figure 2, where each population carries N solutions. Each solution is evaluated against a fitness function (the search objective) that assigns a numeric fitness f_i. The selection operation probabilistically chooses high fitness solutions into a ‘mating pool’ – solutions with higher than average fitness have a higher occurrence in the mating pool, while low fitness solutions may be eliminated from further consideration. Next, pairs of solutions from the mating pool are recombined to form new solutions (‘offspring’) for the next generation population. Crossover is a recombination operator where offspring are formed by combining parts of the ‘parent’ solutions.

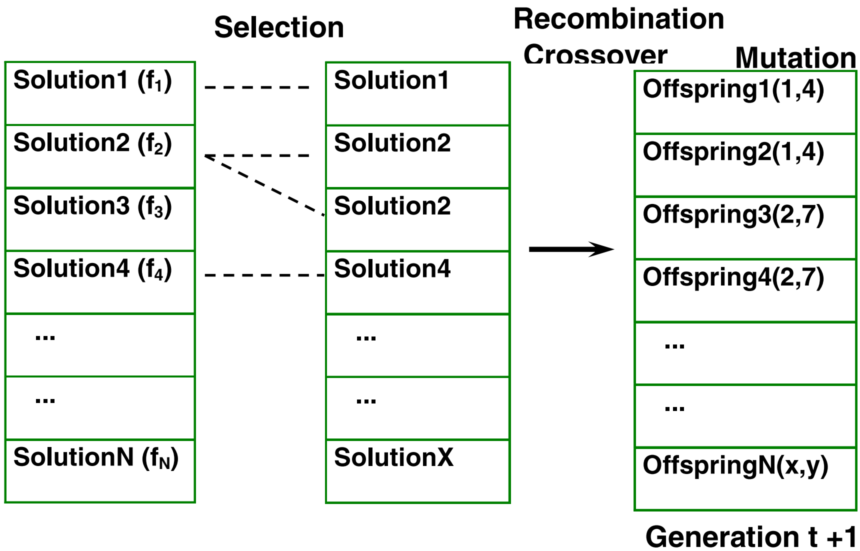


Fig. 2 Genetic search –basic operation

For example, in Figure 2, crossover applied to Solution1 and Solution4 yields Offspring 1 and Offspring2. The mutation operator makes random changes to theoffspring and is applied with low probability. The population of new solutions is then again evaluated against the search objective in an iterative search procedure.

Genetic search is known to be effective because of its ability to process and obtain good 'building blocks' – sub-structures in the solution -- that progressively yield better solutions, and from the implicit parallelism that arises from its simultaneous consideration of multiple solutions (see Goldberg (1989) for a detailed discussion.) GAs are considered suitable for application to complex search spaces not easily amenable to traditional techniques, and are noted to provide an effective tradeoff between exploitation of currently known solutions and a robust exploration of the entire search space. The selection scheme operationalizes exploitation and recombination effects exploration.

2.3 Multi-objective Evolutionary Computation

Various multi-objective evolutionary algorithms (MOEA) have been proposed (Coello et al. 2006, Deb 2001). Key differences among these are in terms of selection and Pareto ranking, diversity preservation approaches, and use of secondary populations. In the vector-evaluated GA (Schaffer 1985), sub-populations are selected separately based on fitness along each of the different objectives; reproduction operators are then applied after shuffling all these sub-populations In Pareto-based selection schemes, the selection of members for the new generation is based on some non-dominance criterion. Non-dominated solutions may be assigned equal selective pressure or population members can be ranked by the number of solutions in the population that they are dominated by. Various approaches have been suggested for ranking population members based on non-dominance (see Coello et al. (2006), Deb (2001) for a full discussion).

Genetic search typically converges to a single solution due to stochastic errors in selection. Fitness sharing (Goldberg and Richardson 1987), whereby population members in the same neighborhood have their fitness reduced through a sharing function, is used to foster search around multiple peaks in the fitness landscape and thus maintain diversity among population members. A sharing parameter determines the neighborhood distance within which such fitness adjustment occurs. In the multi-objective context, such techniques help maintain population members from across the Pareto frontier. Various sharing/niching techniques have been proposed to enhance Pareto-GAs by fostering wider sampling along the non-dominated frontier (Coello 2000, Deb 2001, Veldhuizen and Lamont 2000). Performance of sharing is sensitive to sharing parameters, and guidelines for appropriate values have been suggested for some of the techniques.

Among recent MOEA algorithms with noted strong performance are the non-dominated sorting GA (Deb et al. 2002), the strength Pareto approach (Zitzler and Theile 1999), the Pareto archived evolutionary strategy (Knowles and Corne 2000), and the evolutionary local search algorithm (Menczer et al. 2000). Studies have reported comparisons between different MOEAs (Kollat and Reed 2005, Shaw et al. 1999, Veldhuizen and Lamont 1999), though as noted in Veldhuizen and Lamont (2000), there is no clear evidence favoring a specific ranking method or sharing approach.

Another important aspect to MOEAs in practice is the use of a secondary population to store the non-dominated solutions found as genetic search progresses. This is necessary since non-dominated solutions in one generation can be lost in the stochastic search process. A basic approach can store all non-dominated solutions from each generation in the second population, which can be updated to remove dominated solutions. Alternately, solutions from this second population can be inserted periodically into the regular population to participate in the search (Veldhuizen and Lamont 2000). This is similar in concept to elitist selection in genetic search which preserves the current best solution into the next population. For Pareto-optimal solutions, the current non-dominated set can take up much of the next population; care thus needs to be taken to ensure adequate search.

A number of papers in recent years report on application of MOEA to data mining problems. For classification, MOEAs have been suggested to simultaneously optimize accuracy as well as model compactness (Freitas et al. 2002; Kim, 2004; Dehuri and Mall, 2006; Pappa and Freitas, 2009). While most reported work considers objectives arising from traditional measures of data-mining performance, MOEAs can also directly try to optimize business objectives (Bhattacharyya 2000). MOEA has been applied to association rules mining considering support, interestingness and comprehensibility as different objectives (Ghosh and Nath, 2004; Kaya, 2006; Thilagam and Ananthanarayana, 2008). MOEAs have also been found useful for clustering (Handl and Knowles 2004, Kim et al. 2000, Murty et al. 2008), where different objectives can consider cluster cohesiveness, separation between clusters, minimal number of clusters, and minimal attributes used to describe clusters (Dehuri et al. 2008). A recent paper (Casillas and Martinez-Lopez 2009) obtains fuzzy rules predicting consumer behavior using MOEA, with error, rule set size and a measure of rule set interpretability as three objectives. Becerra et al. (2008) gives an overview on incorporating knowledge to facilitate search in MOEA through fitness function, search operators and initialization. Such mechanisms can be useful in improving the efficiency and efficacy of MOEAs in real-world applications.

3 Multi-objective Models Using Genetic Search

This section describes the algorithm used to obtain multi-objective predictive models in this chapter, the fitness function formulation to incorporate business objectives for the marketing problem and dataset considered, and how performance of obtained models is evaluated.

3.1 Model Representation

Solutions in a population can take a variety of representational forms, and in a data-mining context, the representation determines the nature of patterns that can be discerned from the data. Each population member can specify a weight vector on the predictor variables as in a linear regression; solutions then represent models of the form $y = 1.67x_1 + 11.6x_2 + \ldots$ Solutions can also represent a model expressed in symbolic rule form as in, for example,

[(50K < Income <= 82K) AND (Account Balance >=5.5K) OR (…) …] => Buyer.

Such representations can capture varied non-linear relationships in data.

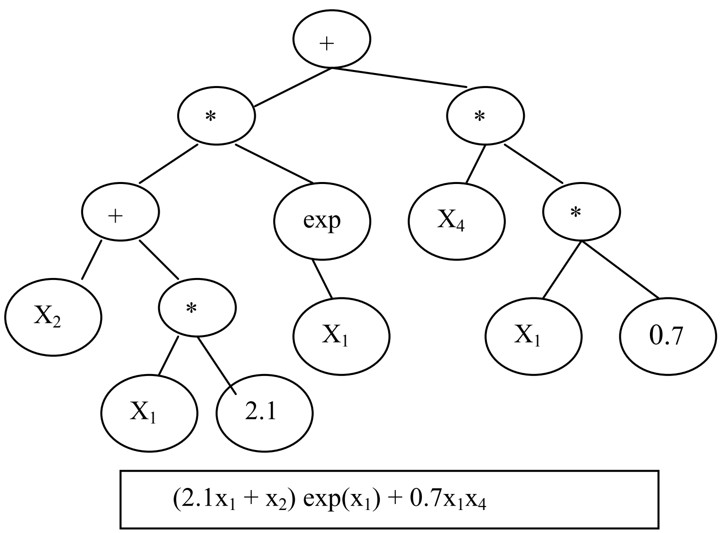


Fig. 3 Non-linear representation of GP

The tree-structured representation of genetic programming (Koza 1993) allows arbitrary functional forms based on a set of functional primitives. Models here specify a function $f(\boldsymbol{x})$ of the predictor variables that can be depicted as a parse tree, thus allowing arbitrarily complex functions based on a defined set of primitives. The functional primitives usable at different internal (non-terminal) nodes are defined through the Function Set, and terminal nodes are obtained from the Terminal Set. For the models reported in this paper, the *function-set* F = {+, -, *, /, *exp, log*} is used, and the *terminal-set* is T = $\{\Re, x_1, x_2, ..., x_n\}$, where $\Re$ denotes the set of real numbers (in a specified range) and x_i the predictor variables in the data. Figure 3 provides an example of a GP tree-based model.

3.2 Genetic Search Operators

Crossover and mutation form the two basic recombination operators. Crossover implements a mating scheme between pairs of "parents" to produce "offspring" that carry characteristics of both parents. Mutation is a random operator applied to insure against premature convergence of the population; mutation also maintains the possibility that any population representative can be ultimately generated.

Appropriate implementations of the crossover and mutation operators are used based on model representation, and are detailed in various standard texts. For the tree-structured models used for the experiments reported here, regular GP crossover and mutation operators are used. Crossover exchanges randomly chosen sub-trees of two parents to create two new offspring; Figure 4 shows an example.

The tree mutation operator randomly changes a sub-tree. Regular GP attempts to learn numeric constants through the combination of numeric values (terminals in the tree) by function sets operators; as detailed in Evett and Fernandez (1998), this is often inadequate for obtaining accurate constants. Non-uniform mutation where the search gets focused with increasing generations is used here for learning numeric constants.

<u>Non-uniform mutation</u> (Michalewicz, 1994): A numeric value s_k at a terminal node is replaced by

$$s_k' = \begin{cases} s_k + \Delta(t, u - s_k) \\ s_k - \Delta(t, s_k - l) \end{cases}.$$

Here, [*u, l*] represents the legal range of values for each element and a uniform random choice determines whether the increment or the decrement be applied. The mutation value $\Delta(t,x)$ returns a value in [0,x] that decreases with increasing t. Thus, with t as the number of generations of search, this operator seeks used wider search in the initial stages, but gradually focuses the search as the generations progress. The following implementation is used:

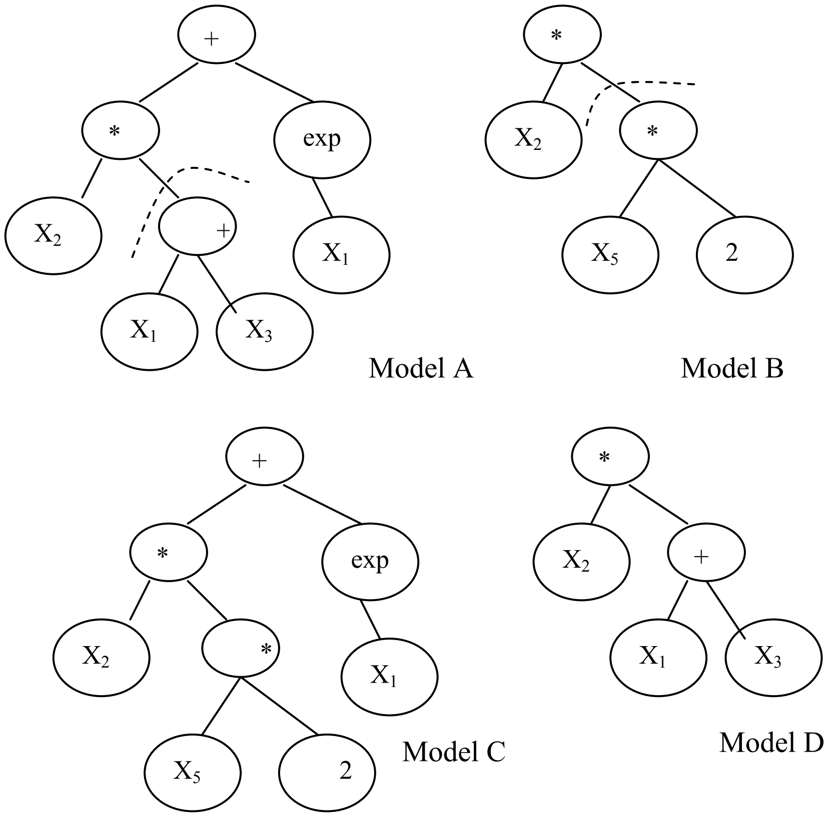


Fig. 4 Crossover of models A and B to give two new models C and D

$$\Delta(t,x) = x\,(1 - r^{(1-\frac{t}{T})^{b}})$$

where r is uniformly generated in [0,1], T gives the total number of generations of search, and b is a parameter determining degree of non-uniformity (a value of b=2 was used for the experiments).

3.3 Multi-objective Models Using Pareto-Selection

This study adopts the simple and elegant Pareto-based scheme of Louis and Rawlins (1993) to obtain the set of non-dominated solutions. This is a variant of binary tournament selection and operates as follows: a pair of solutions (parents) is randomly selected from the current population, and the recombination operators (crossover and mutation) applied in the usual manner to generate two new solutions (offspring). Then the Pareto-optimal set of parents and offspring is produced, and two solutions from this set are randomly selected for the new population. This procedure is repeated to fill the entire population for the next generation. The process in general can be applied with tournament sizes greater than two also. This manner of selection is noted to naturally foster the development of niches exploring different regions of fitness tradeoffs (Louis and Rawlins 1993).

We incorporate elitism into the Pareto selection process. Elitism, where the best solution is retained intact into the next generation, has been noted to be crucial for effective search. In a Pareto-GA context, a population will usually contain several non-dominated solutions. Elitism is incorporated by retaining the current population's non-dominated solutions into the next generation. Note that elitist selection here reduces the number of population members that actively participate in the search using the genetic recombination operators, and can thus impose a burden.

The genetic learning procedure begins with a population of randomly generated models, and can be summarized as:

```
While (not terminating-condition) {
    Evaluate-fitness of population members
    Determine the non-dominated set in the population
    Insert non-dominated set into the next generation
    While next-generation population is not full {
        Select two parents randomly from current population
        With probability pcross
            perform crossover on two parents to get two new offspring,
        With probability pmutate
            perform mutation on each offspring
        With probability pnumutate
            perform non-uniform mutation on each offspring
        Obtain the Pareto-optimal set of parents and offspring
        Select two solutions randomly from the Pareto-optimal set
        Insert selected solutions into next generation
    }
}
```

The search is terminated after a fixed number of iterations.

3.4 Fitness Function

The fitness function specifies the search objective and provides a numerical figure-of-merit or utility measure for a solution in the population. A key advantage of GA/GP arises from the flexibility allowed in formulation of the fitness function. Unlike in many other techniques, there are no constraints of smoothness, continuity or linearity in the function – the only requirement is that the fitness function provide a numerical value indicating desirability of a solution; it may even be specified as a rule-set for model performance assessment.

The flexibility in fitness function formulation allows the development of models tailored to specific business objectives. For predictive modeling tasks, the search objective is often framed around the dependent variable. For example, with a binary dependent variable measuring a buy/no-buy decision, customer churn, response to a solicitation, etc., the fitness function can be specified along traditional performance measures to maximize overall accuracy, correctly identify 'responders', etc. Considering the way models will be implemented and given budgetary or other constraints, the fitness function can also be defined to identify, say, 30% of individuals most likely to buy (Bhattacharyya 1999). Traditional objectives (and models developed using conventional approaches that seek to maximize overall likelihood, minimize errors, etc.) may not yield models that are best suited with respect to specific implementation considerations like these. For continuous dependent variables, too, the fitness function can be set to obtain a 'profit' model that seeks, for example, the most profitable customers for specific targeting-depths, or those with the highest frequency of response, etc.

In the decile-maximization approach (Bhattacharyya 1999) that obtains models to maximize performance at specific targeting depths, fitness of models is estimated at a specified depth-of-file d (d is specified as a fraction of the total data - e.g. d=0.1 for the top 10 percent or first decile, etc.). Given multiple objectives π_i, fitness may be evaluated along each of the objectives as follows:
Consider a data set D containing N observations:

$$D = \{(\boldsymbol{x}, z)_k,\ k=1,..N\},$$

where $\boldsymbol{x}$ denotes the vector of predictors, and $\boldsymbol{z} = \{z_i,\ i=1,..n\}$ gives the variables corresponding to n objectives. Then, considering a specific model f, model evaluation scores the observations as: $\hat{y}_k = f(x_k)$. Let the data ranked in descending order of model scores be denoted as $\hat{y}_k^s$, and the ranked observations up to the specified depth d be given by

$$D^d = \{ (\boldsymbol{x}, z)_k\!:\ \hat{y}_k^s,\ k=1,..,N_d\},$$

where $N_d = d.N$ gives the number of observations up to the depth d. Then, the model's fitness for the i-th objective is obtained as

$$\pi_i^d = \sum_{k \in D^d} (z_i)_k .$$

Evaluating the fitness of a model involves obtaining values for each of the multiple objectives defining the problem.

Alternately, to generally maximize lifts across all deciles, rather than focus on specific deciles, fitness functions can be defined to seek an optimal ordering of observations based on the dependent variables. This can be obtained by considering differences between the dependent variable values in score-ranked observation and in the optimal ordering of observations from high to low values of the dependent variables. This is the approach taken for the results presented in this chapter.

The fitness evaluation may also be defined to guard against over-fit by, for example, utilizing resampling techniques, as found useful in Bhattacharyya (1999). It can also incorporate a preference for simpler models carrying fewer variables, or for models exhibiting desired tradeoffs amongst conflicting characteristics.

3.5 Performance Measures

Given the context of the direct marketing dataset, we assess model performance on the cumulative lifts at different file-depths. This is preferred over traditional measures of performance like overall accuracy, error rate, etc. in direct marketing where models are often used to identify a subset of the total customers expected to maximize response to a solicitation, revenue generated, or other performance criterion. A decile analysis is typically used to evaluate model performance across file-depths. Here, customers are ranked in descending order of their respective model scores – higher scores indicating better performance – and separated into 10 equal groups. Table 1 shows a typical decile analysis, where performance is assessed on response. The first row indicates performance for the top 10% of individuals as identified by the model. The Cumulative Lifts at specific depths of file provide a measure of improvement over a random mailing, and are calculated as:

$$\text{Cumulative Lift}_{\text{decile}} = \frac{\text{cumulative average performance}_{\text{decile}}}{\text{overall average performance}} * 100.$$

Thus, in Table 1, a cumulative lift of 3.7 in the top decile indicates that the model in question is expected to provide a mailing response that is 3.7 times the response expected from a random mailing to 10% of the file. Where a dependent variable gives the revenue generated, a similar decile analysis can be used to evaluate the performance on cumulative revenue at different deciles.

Performance of a model on the Response and Revenue objectives is indicated by the Response-Lift and Revenue-Lift at a considered decile - a model that captures more responders/revenue at the top deciles thus shows superior performance on the response/revenue objective. Note that individual lift values indicate performance on a single objective only, without regard for the other objective. As mentioned earlier, where the two objectives do not relate well, high performance of a model on one objective will correspond to poor performance on the other. In such situations, different levels of performance tradeoffs exist and are captured by the models along the Pareto frontier.

Performance of a model on the Response and Revenue objectives is indicated by the Response-Lift and Revenue-Lift at a considered decile - a model that captures more responders/revenue at the top deciles thus shows superior performance on the response/revenue objective. Note that individual lift values indicate performance on a single objective only, without regard for the other objective. As mentioned earlier, where the two objectives do not relate well, high performance of a model on one objective will correspond to poor performance on the other. In such situations, different levels of performance tradeoffs exist and are captured by the models along the Pareto frontier.

The determination of a specific model to implement can be based on various factors considered by a decision-maker - for instance, it may be desirable that performance on both objectives be above some minimal threshold level, and judgments may consider individual, subjective factors too. Given the application

Table 1 Sample Decile Analysis

Decile	Number of Customers	Number of Responses	Cumulative Responses	Cumulative Response Rate (%)	Cumulative Response Lift
top	9000	1332	2179	14.8%	3.70
2	9000	936	3932	12.6%	3.15
3	9000	648	4328	10.8%	2.70
4	9000	324	4439	9.0%	2.25
5	9000	144	4549	7.52%	1.88
6	9000	72	4634	6.4%	1.60
7	9000	72	4701	5.6%	1.40
8	9000	36	4770	4.95%	1.24
9	9000	20	4819	4.43%	1.11
bottom	9000	16	4874	4.0%	1.00
Total	90,000	3600			

considered here, the overall modeling objective is taken as the maximization of the expected revenue that can be realized through identification of high-revenue customers. This can be estimated at a specific decile or file-depth d as follows (Bhattacharyya 2000):

Let V denote the total revenue over all individuals in the data, and R the total number of responders. Consider V_d and R_d the cumulative total revenue and cumulative total number of responders respectively at decile d. Then, if N denotes the overall total customers in the data and N_d is the total customers up to the decile level d, the cumulative response and revenue lifts are:

$$Response\ Lift = (R_d/N_d)/(R/N) \text{ and } Revenue\ Lift = (V_d/N_d)/(V/N).$$

The expected revenue up to the file-depth d is given by:

$$\begin{aligned}&(\text{Average-response per customer})_d * (\text{Average revenue per customer})_d\\&= (R_d/N_d)*(V_d/N_d)\\&= (ResponseLift * RevenueLift) * [(R/N) * (V/N)].\end{aligned}$$

The product of Response Lift and Revenue Lift values then gives the *cumulative lift on the expected-revenue* as:

$$[(R_d/N_d)*(V_d/N_d)] / [(R/N) * (V/N)].$$

This Product of Lifts thus provides a useful measure to evaluate the performance of models on expected revenue. This measure depends on both objectives, and models with high performance on only one objective may not perform well on Product of Lifts which indicates expected-revenue from a model at specified file-depths.

4 Data and Multi-objective Model Performance

We examine the effectiveness of the multi-objective genetic search approach in handling multi-objective data-mining problems using a real-life dataset. This section describes the dataset used and presents the performance of different models along the Pareto frontier. For comparison, we also show the performance of least squares regression and logistic regression models on the same data.

4.1 Data

The dataset pertains to direct mail solicitations from past donors for a non-profit organization[1]. It carries historical data on solicitations and donations from 10/86 through 6/95. All donors in the data received at least one solicitation in early 10/95 and the data carries the number and amount donations in the subsequent Fall, 1995 period. We consider the development of models to p redict response (at least one donation) and dollars in the later time period, given the history of prior solicitations and contributions.

The dataset carries 99,200 cases, with each being defined through 77 attributes. The list of data attributes are given in Appendix A. Various transformations were conducted to obtain a modeling dataset.

Solicitation codes in the data are specific to program type and date of solicitation. The data specifies these codes as one of four program types - A, B, C, or a miscellaneous group. The codes are resolved to obtain the total number of solicitations of different types for each observation in the data. Similarly, for contribution codes, we obtain the total number of contributions to different types. Contribution dollars for the different codes are also aggregated to get total contributions made for the different contribution types. Binary variables were created for each type to depict whether a solicitation and contribution for that type was ever made. Thus the following variables were derived for each of the four types:

Number of solicitations of Type x
Number of contributions of Type x
Dollars of contribution to Type x
Solicitation Type x (yes/no)
Contribution Type x (yes/no)

Date fields were converted to months-since in relation to a baseline of 10/1/1995.The following date related fields were retained for modeling: dates of first contribution, largest contribution, latest solicitation and latest contribution, change of address date.

The State variable was transformed to retain only the 9 states found to have large concentration of customers, and these were converted to binary indicator variables. After preliminary examination, the Reinstatement Code, Rental Exclusion Code, 2nd Address Indicator variables were also retained in the modeling

[1] The dataset is provided for academic use by the Direct Marketing Educational Foundation.

dataset. Where more than two categories were specified, multiple binary variables were created. Certain additional fields were also created:

Longevity: time period between first and latest contributions
Ratio of Lifetime Contributions to Solicitations
Average Lifetime Contribution
Average Contribution per Solicitation

The modeling dataset had 58 predictor variables, and two dependent variables for Response and Dollars in the later Fall 1995 period.

4.2 Models along the Pareto-Frontier

Here we examine the performance of a set of GP models obtained using the genetic search procedure with Pareto-selection as described above[2]. For comparison, we also show the performance of an ordinary least squares regression model and a logistic regression model on the same data. Models were developed on a training dataset of 30,000 cases and performance of models is considered on the separate validation dataset comprising the remaining cases.

For the genetic search, multiple initial runs with different random seeds were conducted and the non-dominated solutions from each were saved. A final run with initial population seeded using the saved non-dominated solutions was used to obtain the non-dominated solutions shown below. The fitness function was defined to optimize the ordering of dataset observations on the dependent variables so as to maximize lifts at the upper deciles.

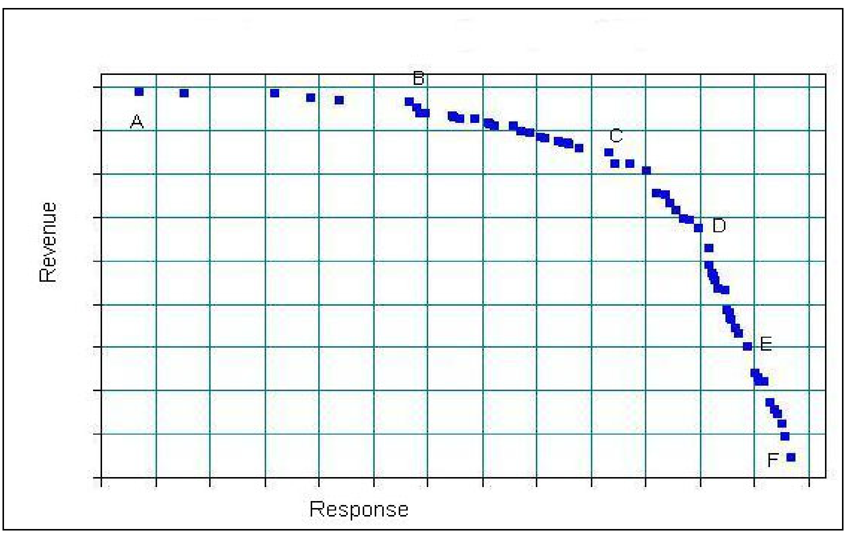


Fig. 5 Non-dominated models

[2] The models were obtained using the evolveDM^{TM} software for data mining using evolutionary computation techniques. Details are available from the author.

The non-dominated models obtained are shown in Figure 5. Different models incorporating varying levels of tradeoff among the two objectives are seen. Models toward the upper-left show good performance on the Revenue objective, but perform poorly on Response. Conversely, models at the lower-right have good performance on Response but low Revenue performance. We consider different models along the Pareto frontier, as indicated by A, B… F in the figure.

Since performance of models in the upper deciles is typically of interest, we examine the cumulative lifts of different models at the first, second and third deciles, corresponding to 10%, 20% and 30% file-depths. Tables 2a-2c give the performance of the different models at these deciles. Here, OLS represents the ordinary least squares regression on the continuous dependent variable for revenue, and LR is for the logistic regression model on the binary dependent variable for response. The graphs in Figure 6 plot the lifts of the different models on the two individual objectives, at different file-depths.

The single objective OLS and LR models perform well on the respective revenue and response objectives that they are built to optimize on. As can be expected, performance of these models on the other objective is lower. The multi-objective models from the upper-left region of the Pareto-frontier are seen to perform better on Revenue Lift than the OLS model, and those from the lower-right perform better on Response Lift than the LR model. This is not surprising and shows that evolutionary search, using the nonlinear representation and seeking to maximize a fitness function that is more related to lift performance, is able to obtain better solutions than the traditional models.

The multi-objective models from the middle region of the Pareto-frontier, exhibiting tradeoffs amongst the two objectives, are seen to perform well on the Product of Lifts measure. This arises from the conflicting objectives in the data, as evident in the Pareto-frontier obtained. All these models do not show better expected-revenue lifts than the OLS and LR models. At the top decile, the OLS model has higher performance than the multi-objective models A and B in the upper-left region of the Pareto-frontier, and the LR model does better than the models E and F which are from the lower-right region. It is interesting to observe that for this top decile, the LR model is not dominated on both objectives by any other model, and only model C dominates the OLS model. On product-of- lifts, the LR model does better than OLS, but models C and D exhibit the highest performance. At the second decile, it is the OLS model whose performance is non-dominated on both objectives by any of the other models. The LR model, however, displays a higher product-of-lifts. Performance of both OLS and LR is surpassed by models B, C and D from the middle region of the Pareto frontier. At the third decile, too, these three models outperform the others on product-of-lifts.

Table 2a Lifts for top decile

10% depth	Response Lift	Revenue Lift	Prod Lifts
A	1.42	2.67	3.79
B	1.49	2.70	4.02
C	1.97	2.58	5.09
D	2.21	2.43	5.38
E	2.29	2.04	4.68
F	2.42	1.85	4.47
OLS	1.61	2.46	3.97
LR	2.34	2.10	4.91

Table 2b Lifts for second decile

20% depth	Response Lift	Revenue Lift	Prod Lifts
A	1.37	2.08	2.85
B	1.42	2.09	2.97
C	1.82	2.01	3.67
D	1.92	1.94	3.73
E	1.99	1.81	3.61
F	2.07	1.49	3.08
OLS	1.49	2.04	3.04
LR	1.91	1.72	3.29

Table 2c Lifts for third decile

30% depth	Response Lift	Revenue Lift	Prod Lifts
A	1.33	1.82	2.42
B	1.38	1.80	2.50
C	1.61	1.77	2.83
D	1.78	1.71	3.04
E	1.82	1.62	2.95
F	1.91	1.29	2.47
OLS	1.38	1.79	2.48
LR	1.69	1.46	2.48

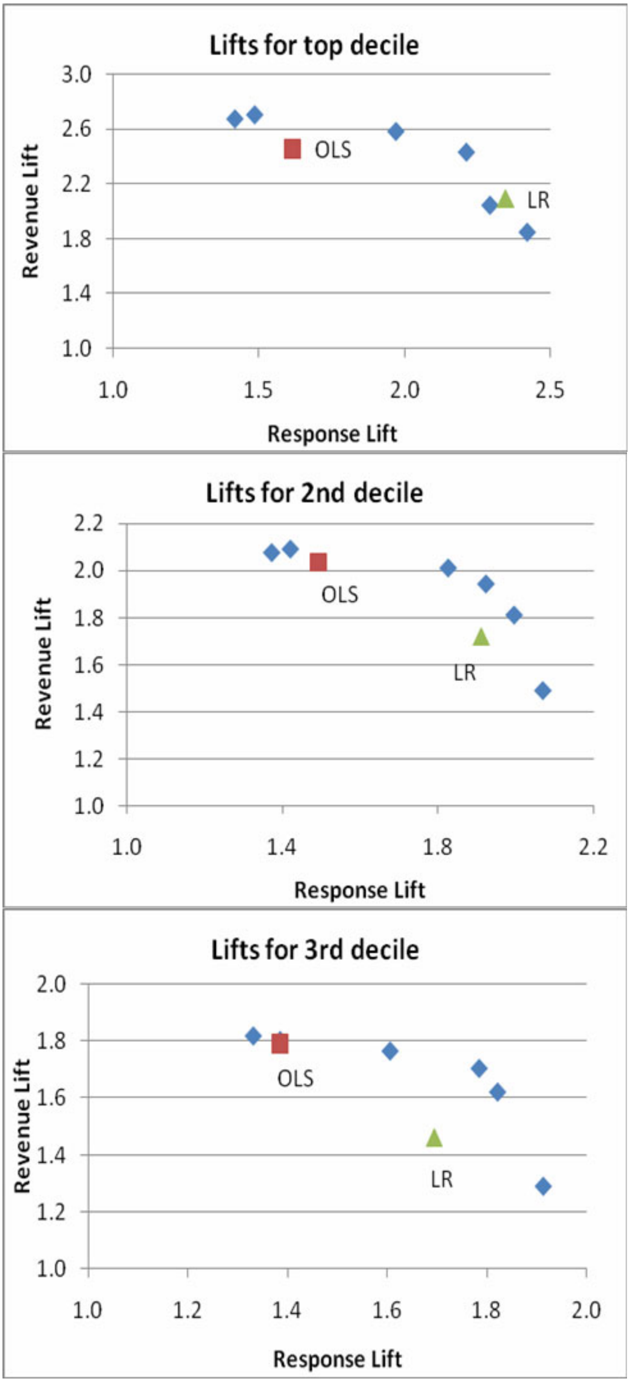


Fig. 6 Lifts at different deciles

5 Conclusions

The Pareto-genetic search scheme used is seen to be effective at obtaining models with varying levels of tradeoff on multiple data-mining objectives. Models from the extremes of the tradeoff frontier are found to perform favorably in comparison with OLS and logistic regression models on the respective single objectives. The traditional OLS and logistic regression models - popular in industry use - perform well on the single criteria that they model. In the context of multiple objectives, however, they do not provide a decision-maker with different tradeoffs that the data may admit. With multiple models as obtained from the genetic search based approach here, selection of a specific model to implement, from amongst the non-dominated set, can be made in consideration of a variety of factors of possible concern to a decision-maker. For the application and data in this study, considering expected-revenue as a criterion for judging overall model performance, the best of the multiple-objective models were seen to yield superior performance over the logistic regression and OLS models at different file-depths.

With multiple objectives to consider, and a set of models with performance ranging along the Pareto frontier, the choice of a model to implement is an important next step. Where decision makers have a clear understanding of desired tradeoffs among objectives, the selection of 'best' model can be straightforward, especially where models are developed to optimize business objectives as in the case presented in this paper. For situations where the modeling objectives may not directly correspond with business criteria, observed tradeoffs on the modeling objectives may not present adequate information for a decision-makers choice of a model to implement. For such situations, further analyses on the Pareto optimal set of models may be required to provide decision-makers adequate insight into application performance tradeoffs of alternate models. The need for additional analyses also occurs where three or more objectives are considered, and visualization tools have been investigated to aid in decision-making (Kollat and Reed 2007).

In marketing applications like the one presented in this paper, a combination of models can be useful to obtain the 'best' overall solution. With response and revenue/profit maximization, for example, as the two modeling objectives, a decision-maker may prefer a model with a certain tradeoff in profit and response likelihood at the upper deciles. From the potential customers identified through this model, one may also want to distinguish those indicated as high response potential by another model which exhibits high performance in the response objective; these individuals may be of interest for a different marketing treatment. In a similar manner, consideration of multiple models from the Pareto set can also be useful for identifying groups of customers that are best targeted with varying marketing approaches. Consideration of multiple models from the Pareto set in this way is under investigation.

While traditional data-mining approaches usually obtain models built to objectives like maximum likelihood, minimal classification error, etc., the business problem can consider alternate criteria. In a direct marketing context, for example, managers may be concerned about the targeting depth that yields optimal returns, maximizing the number of responders within a certain budget, or identifying responders that are also likely to generate high purchase revenue. As noted earlier, evolutionary computation allows the incorporation of such criteria into the fitness function and can thereby help build models that directly optimize business objectives of interest (Bhattacharyya 1999, 2000). Most EC based data mining work to date, however, have not taken advantage of this, and models are developed to general performance criteria. Incorporation of varied business criteria in fitness functions and their evaluation presents an important research opportunity. Consideration of multiple managerial objectives across different business problems using MOEA is also a promising area for future research.

Many real-world problems addressed through marketing analytics and data mining can profitably utilize the multi-objective evolutionary search based approach presented here. In catalogue and retail sales, for example, models identifying potential buyers that also will not return purchased goods are useful; similarly, models that identify potential responders to mailings who are also likely to buy some specific product are often sought. Multiple and often conflicting objectives are also seen in the context of many cross-selling marketing campaigns. Problems in the telecommunications industry often seek to model customers' tenure in combination with usage - identifying people who have long tenure and high usage of services. Further application examples occur in the financial services industry, where models, for example, can seek customers who are likely to be approved for credit and who can also be expected not to make late payments or default on loans.

References

Berry, M.J.A., Linoff, G.S.: Data Mining Techniques for Marketing, Sales and Customer Relationship Management. John Wiley & Sons, Chichester (2004)

Becerra, R.L., Santana-Quintero, L.V., Coello, C.C.: Knowledge Incorporation in Multi-objective Evolutionary Algorithms. In: Ghosh, A., Dehuri, S., Ghosh, S. (eds.) Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Databases. Studies in Computational Intelligence, vol. 98, pp. 23–46. Springer, Heidelberg (2008)

Bhattachryya, S.: Direct Marketing Performance Modeling using Genetic Algorithms. INFORMS Journal of Computing 11(13), 248–257 (1999)

Bhattacharyya, S.: Evolutionary algorithms in data mining: Multi-objective performance modeling for direct marketing. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, Massachusetts, pp. 465–473 (2000)

Casillas, J., Martínez-López, F.J.: Mining Uncertain Data with Multiobjective Genetic Fuzzy Systems to be Applied in Consumer Behaviour Modeling. Expert Systems with Applications 36(2), 1645–1659 (2009)

Coello, C.C.: An Updated Survey of GA-Based Multiobjective Optimization Techniques. ACM Computing Surveys 32(2), 109–143 (2000)

Coello, C.C., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation). Springer, New York (2006)

De La Iglesia, B., Richards, G., Philpott, M.S., Rayward-Smith, V.J.: The Application and Effectiveness of a Multi-objective Metaheuristic Algorithm for Partial Classification. European Journal of Operational Research 169(3), 898–917 (2006)

Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Inc., New York (2001)

Deb, K., Pratap, A., Agrawal, S., Meyarivan, T.: A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197 (2002)

Dehuri, S., Ghosh, S., Ghosh, A.: Genetic Algorithm for Optimization of Multiple Objectives in Knowledge Discovery from Large Databases. In: Ghosh, A., Dehuri, S., Ghosh, S. (eds.) Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Databases. Studies in Computational Intelligence, vol. 98, pp. 1–22. Springer, Heidelberg (2008)

Dehuri, S., Jagadev, A.K., Ghosh, A., Mall, R.: Multi-objective Genetic Algorithm for Association Rule Mining using a Homogeneous Dedicated Cluster of Workstations. American Journal of Applied Sciences 88, 2086–2095 (2006)

Dehuri, S., Mall, R.: Predictive and Comprehensible Rule Discovery using a Multi-objective Genetic Algorithm. Knowledge-Based Systems 19(6), 413–421 (2006)

Evett, M., Fernandez, T.: Numeric Mutation Improves the Discovery of Numeric Constants in Genetic Program. In: Koza, J.R., et al. (eds.) Proceedings of the Third Annual Genetic Programming Conference, Wisconsin, Madison, Morgan Kaufmann, San Francisco (1998)

Fonseca, C.M., Fleming, P.J.: An Overview of Evolutionary Algorithms in Multi-Objective Optimization. Evolutionary Computation 3(1), 1–16 (1995)

Freitas, A.A.: A Critical Review of Multi-objective Optimization in Data Mining: a Position Paper. SIGKDD Explorations. Newsletter 6(2), 77–86 (2004)

Freitas, A.A., Pappa, G.L., Kaestner, C.A.A.: Attribute Selection with a Multi-objective Genetic Algorithm. In: Proceedings of the 16th Brazilian Symposium on Artificial Intelligence, pp. 280–290. Springer, Heidelberg (2002)

Ghosh, A., Nath, B.: Multi-objective Rule mining using Genetic Algorithms. Information Sciences 163(1-3), 123–133 (2004)

Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading (1989)

Goldberg, D.E., Richardson, K.: Genetic algorithms with Sharing for Multi-modal Function Optimization. In: Proceedings of the 2nd International Conference on Genetic Algorithm, pp. 41–49 (1987)

Hand, D.J.: Construction and Assessment of Classification Rules. John Wiley and Sons, Chichester (1997)

Handl, J., Knowles, J.: Multiobjective Clustering with Automatic Determination of the Number of Clusters, Technical Report No. TR-COMPSYSBIO-2004-02, UMIST, Department of Chemistry (August 2004)

Kaya, M.: Multi-objective Genetic Algorithm based Approaches for Mining Optimized Fuzzy Association Rules. Soft Computing: A Fusion of Foundations, Methodologies and Applications 10(7), 578–586 (2006)

Kim, D.: Structural Risk Minimization on Decision Trees using an Evolutionary Multiobjective Algorithm. In: Keijzer, M., O'Reilly, U.-M., Lucas, S., Costa, E., Soule, T. (eds.) EuroGP 2004. LNCS, vol. 3003, pp. 338–348. Springer, Heidelberg (2004)

Kim, Y., Street, N.W.: An Intelligent System for Customer Targeting: a Data Mining Approach. Decision Support Systems 37(2), 215–228 (2004)

Kim, Y., Street, W.N., Menczer, F.: Feature Selection in Unsupervised Learning via Evolutionary Search. In: Proc. 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2000), pp. 365–369 (2000)

Knowles, J.D., Corne, D.W.: Approximating the Non-dominated Front using the Pareto Archived Evolution Strategy. Evolutionary Computation 8(2), 49–172 (2000)

Kollat, J.B., Reed, P.M.: The value of online adaptive search: A performance comparison of NSGAII, ε-NSGAII and εMOEA. In: Coello, C.C., Aguirre, A.H., Zitzler, E. (eds.) EMO 2005. LNCS, vol. 3410, pp. 386–398. Springer, Heidelberg (2005)

Kollat, J.B., Reed, P.M.: A framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VIDEO). Environmental Modelling & Software 22(12), 1691–1704 (2007)

Koza, J.R.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge (1993)

Louis, S.J., Rawlins, G.J.E.: Pareto-Optimality, GA-Easiness and Deception. In: Forrest, S. (ed.) Proceedings of the Fifth International Conference on Genetic Algorithms, pp. 118–123 (1993)

Massand, B., Piatetsky-Shapiro, G.: A Comparison of Different Approaches for Maximizing the Business Payoffs of Prediction Models. In: Simoudis, E., Han, J.W., Fayyad, U. (eds.) Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 195–201 (1996)

Menczer, F., Degeratu, M., Street, N.W.: Efficient and Scalable Pareto Optimization by Evolutionary Local Selection Algorithms. Evolutionary Computation 8(2), 223–247 (2000)

Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs, 2nd edn. Springer, Heidelberg (1994)

Murty, M.N., Babaria, R., Bhattacharyya, C.: Clustering Based on Genetic Algorithms. In: Ghosh, A., Dehuri, S., Ghosh, S. (eds.) Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Databases. Studies in Computational Intelligence, vol. 98, pp. 137–159. Springer, Heidelberg (2008)

Pappa, G.L., Freitas, A.A.: Evolving Rule Induction algorithms with Multi-objective Grammar-based Genetic Programming. Knowledge and Information Systems 19(3), 283–309 (2009)

Richardson, J.T., Palmer, M.R., Liepins, G., Hilliard, M.: Some Guidelines for Genetic Algorithms with Penalty Functions. In: Schaffer, J.D. (ed.) Proceedings of the Third International Conference on genetic Algorithms, pp. 191–197 (1989)

Schaffer, J.D.: Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. In: Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms, pp. 93–100. Lawrence Erlbaum, Mahwah (1985)

Shaw, K.J., Nortcliffe, A.L., Thompson, M., Love, J., Fonseca, C.M., Fleming, P.J.: Assessing the Performance of Multiobjective Genetic Algorithms for Optimization of a Batch Process Scheduling Problem. In: Angeline, P. (ed.) Congress on Evolutionary Computation, pp. 37–45. IEEE Press, Piscataway (1999)

Sikora, R., Piramuthu, S.: Efficient Genetic Algorithm Based Data Mining Using Feature Selection with Hausdorff Distance. Information Technology and Management 6(4), 315–331 (2005)

Thilagam, P.S., Ananthanarayana, V.S.: Extraction and Optimization of Fuzzy Association Rules using Multi-objective Genetic Algorithm. Pattern Analysis and Applications 11(2), 159–168 (2008)

Van Veldhuizen, D.A., Lamont, G.B.: Multiobjective Evolutionary Algorithm Test Suites. In: Carroll, J., Haddad, H., Oppenheim, D., Bryant, B., Lamont, G.B. (eds.) Proceedings of the 1999 ACM Symposium on Applied Computing, New York, pp. 351–357 (1999)

Van Veldhuizen, D.A., Lamont, G.B.: Multiobjective Evolutionary Algorithms: Analyzing the State-of-Art. Evolutionary Computation 8(2), 125–147 (2000)

Zhang, Y., Bhattacharyya, S.: Genetic Programming in Classifying Large-scale Data: an Ensemble Method. Information Sciences 163(1-3), 85–101 (2004)

Zeleny, M.: Multiple Criteria Decision Making. McGraw-Hill, New York (1982)

Zitzler, E., Thiele, L.: Multi-objective Evolutionary Algorithms: a Comparative Case study and Strength Pareto Approach. IEEE Transactions on Evolutionary Computation 3, 257–271 (1999)

Appendix

Appendix A – Original Data set attributes

ACCNTNMB, Donor ID

TARGDOL Dollars of Fall 1995 Donations
TARGRES Number of Fall 1995 Donations

Contributions history:
CNCOD1 to CNCOD10 Latest to 10th Latest Contribution Code
CNDAT1 to CNDAT10 Latest to 10th Latest Contribution Date
CNDOL1 to CNDOL10 Latest to 10th Latest Contribution
CNTMLIF Times Contributed Lifetime
CNTRLIF Dollars Contribution Lifetime
CONLARG Largest Contribution
CONTRFST First Contribution
DATEFST First Contribution Date
DATELRG Largest Contribution Date

Solicitation history:
SLCOD1 to SLCOD11 Latest to 11th Latest Solicitation Code
SLDAT1 to SLDAT11 Latest to 11th Latest Solicitation Date
SLTMLIF Times Solicitated Lifetime

FIRMCOD Firm/Head HH code
MEMBCODE Membership Code
NOCLBCOD No Club Contact Code
NONPRCOD No Premium Contact Code
NORETCOD No Return Postage Code
NOSUSCOD No Sustain Fund Code
PREFCODE Preferred Contributor Code
REINCODE Reinstatement Code
REINDATE Reinstatement Date
RENTCODE Rental Exclusion Code
CHNGDATE Change of Address Date
SECADRIN 2nd Address Indicator
SEX Gender
STATCODE State
ZIPCODE ZIP Code

Automatic Discovery of Potential Causal Structures in Marketing Databases Based on Fuzzy Association Rules

Albert Orriols-Puig[1], Jorge Casillas[2], and Francisco J. Martínez-López[3]

[1] Grup de Recerca en Sistemes Intelligents
Enginyeria i Arquitectura La Salle (URL), 08022 Barcelona, Spain
e-mail: aorriols@salle.url.edu

[2] CITIC-UGR (Research Center on Communication and Information Technology)
Department of Computer Science and Artificial Intelligence
University of Granada, 18071 Granada, Spain
e-mail: casillas@decsai.ugr.es

[3] Department of Marketing
University of Granada, 18071 Granada, Spain and
Universitat Oberta de Catalunya, 08035 Barcelona, Spain
e-mail: fjmlopez@ugr.es

Abstract. Marketing-oriented firms are especially concerned with modeling consumer behavior in order to improve their information and aid their decision processes on markets. For this purpose, marketing experts use complex models and apply statistical methodologies to infer conclusions from data. In the recent years, the application of machine learning has been identified as a promising approach to complement these classical techniques of analysis. In this chapter, we review some of the first approaches that undertake this idea. More specifically, we review the application of Fuzzy-CSar, a machine learning technique that evolves fuzzy association rules online, to a certain consumption problem analyzed. As a differentiating sign of identity from other methods, Fuzzy-CSar does not assume any aprioristic causality (so model) within the variables forming the consumer database. Instead, the system is responsible for extracting the strongest associations among variables, and so, the structure of the problem. Fuzzy-CSar is applied to the real-world marketing problem of modeling web consumers, with the aim of identifying interesting relationships among the variables of the model. In addition, the system is compared with a supervised learning technique, which is able to extract associations between a set of input variables and a pre-fixed output variable, expressly designed for this marketing problem. The results show that Fuzzy-CSar can provide interesting information for marketing experts that was not detected by the classical approach, and that the extraction of fuzzy association rules is an appealing alternative, in general, to refine or complement the modeling results obtained with the use of traditional methods of analysis applied for these purposes; in particular, we focus on, and take as a reference, the structural equation modeling.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 181–206.
springerlink.com

1 Introduction

Companies are constantly searching for suitable marketing opportunities to survive in increasingly turbulent and volatile markets. For this purpose, marketing experts are especially concerned with the creation and management of key information about the market [6]. In management and marketing disciplines, the use of models has been usual to drive the database analysis. Model-based analytical processes imply that a structure of relations among the elements (i.e., variables) of this previously known model be used to, by means of analytical methods of study, describe or predict the behavior of those relations. This analytical approach matches the procedure classically set by the scientific method; i.e., a researcher works with a set of hypotheses of expected relationships among variables, those hypotheses are empirically tested and, finally, some conclusions are extracted (e.g., see [20]). Basically, these are the core questions in marketing modeling, which are usually followed to drive the information search process in marketing databases with the aim of supporting marketing decisions. But, would it be plausible to work without models? Doubtless, models are very necessary, especially in the academic field, where the arsenal of statistical and, in general, analytical tools are usually applied with a theory-driven approach. However, mostly from the practitioners' perspective, their usage may limit the added-value extracted from the data when applied to certain kind of decision problems in marketing. In particular, in non- or ill-structured problems, analysis based on the *a priori* information offered by a model, which may disregard important relationships due to the weak structure of the problem, may not be as effective as a decision maker would expect.

Hence, though the support of models is helpful to address the search of information in marketing databases, there are situations, both in the practitioners' and scholars' arena, where the use of other non model-based solutions, either on their own or as a complementary tool to a information search process based on models, might produce profitable results. For instance, from an academic perspective, when analyzing the validity of a theoretical model, an additional approach to the traditional would be to adjust all the possible causal structures (models), reasonable or unreasonable, and then, theoretically analyze those configurations with better fitness. However, as causal (theoretic) structures of reference increase in complexity, the number of possible configurations is considerably higher [3], so the development of the said approach would be more difficult to accomplish. In this case, powerful analytical methods are necessary to undertake this task with efficiency. Time ago, some authors [7] pointed out that a process of search and analysis for all the possible configurations of causal models, in a certain marketing database, could be automatized using some Computation Science-based method. However, these authors also recognized that years of evolution would be necessary to be able to work with suitable procedures.

Nowadays, the so-called knowledge-based marketing support systems offer an excellent framework to develop methods with this purpose (see [3]). In this regard, several authors have proposed to apply supervised machine learning methods, which are informed with little prior knowledge about the problem, resulting in the extraction of key knowledge that was not detected by the classical analysis methodology (e.g., see [4, 17]). Continuing with these efforts, the application of unsupervised learning techniques which have no knowledge about the problem structure—letting the machine extract interesting, useful, and unknown knowledge about the market—appears as an appealing approach to these problems.

The purpose of this chapter is to review the work done on the extraction of fuzzy association rules to discover new interesting knowledge from marketing databases. Specifically, we focus on a database that contains information about the consumer behavior. To achieve this, we apply Fuzzy-CSar, a learning classifier system (LCS) [11] that assumes no structure about the problem and evolves a diverse set of fuzzy association rules that describe interesting associations among problem variables. Fuzzy-CSar uses a fuzzy representation that enables the system to deal with the imprecision of the marketing data. The system is compared with an evolutionary multi-objective (EMO) approach that extracts fuzzy rules that define a particular prefixed output variable [17]. The results highlight that fuzzy association rules permit extracting key knowledge that was discovered neither by the classical approach nor by the EMO approach.

The chapter is organized as follows. Section 2 describes the type of data usually found in marketing databases, with especial attention to the particularities of the kind of variables (i.e. constructs) forming complex causal models in marketing, explains the classical marketing analysis approach in more detail, and motivates the use of machine learning to tackle these problems. Section 3 provides the basic concepts of association rules, and Sect. 4 describes Fuzzy-CSar. Section 5 presents the experimental methodology, and Sect. 6 analyzes the results. Finally, Sect. 7 concludes and presents future work lines.

2 Previous Considerations on the Adaptation of Marketing Data

A common practice in marketing modeling, and consumer behavior modeling in particular (field where the method proposed here is applied to), when working with complex models (i.e., with multiple relations of dependent and independent variables), is specifying such models to be empirically analyzed by structural equation modeling [17]; other types of causal models, so statistical estimation methods, are also used, though we focus our research on the most difficult case to solve of the complex models. These models are compounded by elements (constructs) which are inferred from imprecise data, i.e., the indicators or variables related to every element of the model. As follows, we explicate these types of problems, specifically focusing on the type of data that is made available for analysis. Then, we outline some significant aspects related to this structural modeling methodology when applied to a consumer behavior model and motivate the use of machine learning techniques to obtain new interesting information. Then, we explain how marketing data can be transformed into fuzzy semantics, and finally, we discuss different strategies to let machine learning techniques deal with the particularities of the marketing data.

2.1 Data Collection in Marketing

Generally, when working with complex models for consumer behavior analysis, so with structural models, the elements of the model are divided into two categories: (1) *unobserved/latent* variables, also known as *constructs*, which are conceptually those whose measurement cannot be made directly with a single measure; and (2) observed variables or indicators, those related to every single measure (i.e., an item in a multi-item measurement scale) developed to be related to a construct. The underlying idea is that an observed variable is an imperfect measure of a construct, but a set of indicators related to a construct, considered altogether, may lead to a reliable measurement of said construct. Therefore, every construct in a model is usually related to a set of observed variables. This is currently the predominant measurement approach, known as the partial-interpretation philosophy [22].

Finally, there is an especial category of constructs known as second-order constructs. These are characterized by not having direct association with indicators in the measurement model, as an ordinary/first order construct has, but by being defined by a combination of first-order constructs related to them. Note that the overall structure of these data is unconventional. Thus, machine learning techniques need to be adapted to deal with them.

2.2 The Classical Approach to Deal with Marketing Data

To extract key knowledge from a database (usually generated after a survey that administered questionnaires to a sample of the target population), marketing experts use the following approach, addressed as the *classical approach* of analysis in the rest of this chapter. First, the expert establishes a theoretical model, which denotes the relationships—and directions of these relationships—among the variables of the problem. Marketing experts base such models on diverse sources, where we highlight the theoretical basis, the *a priori* information of the market, and their own experience. Then, the models are used to establish a set of hypotheses that explain the relationship among constructs that have been connected in the structural model. Thereafter, a measurement model is set and statistical methods based on structural modeling methodologies are used to contrast these hypotheses. The conclusions extracted from the analysis may cause the marketing expert to refine the structural model and to apply again the same analysis procedure.

While it has been shown that the classical approach may provide key knowledge of the consumer behavior analyzed, which may be used to support decision making [20], based on a conceptual/structural model to drive the search of information in the database, it may hamper the discovery of some key knowledge. To extract further interesting information, several authors have successfully applied machine learning techniques to these types of problems. For example, in [4], the authors used supervised learning techniques to model the consumer behavior in the Internet, resulting in new interesting knowledge not detected by the classical approach. This approach permitted extracting fuzzy rules that always predicted the same variable in the consequent. In the present chapter, we take some of the ideas presented in [4] as starting point and extend them to build a system that extracts fuzzy association rules from consumer behavior databases, but with a different approach. In particular, we do not consider any *a priori* information about the system and expect that the system provides us with any relevant association among variables. Before proceeding with the description of this approach, the next subsections briefly present key questions related with the transformation of the original data (i.e. marketing measurement scales) into fuzzy semantic and finally discuss how a general learning system can be adapted to deal with the particularities of the marketing data.

2.3 Transformation of Marketing Scales into Fuzzy Semantic

The machine learning stage could not work without transforming the original marketing data into fuzzy terms. Some notes are deserved to be commented in this regard. The transformation process differs depending on the type of marketing scale, subjacent to every variable of the marketing database. In order to simplify the problem, let us focus on the following traditional classification of measurement scales [23, 24]: nominal, ordinal, interval, and ratio. The transformation of these basic measurement scales into fuzzy variables is useful for all those cases where a measurement scale entails, as minimum, certain order. This premise would involve all the types of measurement scales, with the exception of the nominal. Next, we offer some general reflections for each of the four scales.

Nominal scales. The characteristics of this scale (e.g., consumer's gender, nationality, etc.) just allow identifying and classifying into some of the categories of the scale. But, there is no relation of order or grade between the categories. Consequently, it does not have any sense applying the fuzzy reasoning, as nature of the scale's categories is purely deterministic. This fact involves that these scales are considered as singleton fuzzy sets, a particular case of fuzzy sets; i.e., if certain consumer belongs to certain category, he/she has a membership degree of one to the fuzzy set related to that category, and zero to the others.

Ordinal scales. When the transformation of these types of scales is tackled, there is a main inconvenient: as they are non-metric/qualitative scales, there is just information about the consumer's membership or non-membership to one of the categories in which the marketing variable was structured. This fact limits the possibilities to determine the extreme and central values of the fuzzy sets defining the linguistic variable associated with that marketing variable.

Likewise, regardless the previous question, the marketing expert should solve the following dilemma: should or should not the linguistic variable explicitly consider the structure of categories defining the original marketing variable? In general, it is widely accepted the convenience of a linguistic variable synthesizes the information provided by the original scale, in order to improve the interpretation of relations among the elements of the model, as well as to draw on the potentials of fuzzy inference. However, a subsequent aggregation of original categories is difficult to implement due to the lack of information provided by an ordinal scale; i.e., references are needed, for instance the extremes and central points of the fuzzy sets obtained after the aggregation. On the other hand, there are studies which require the original categories of certain ordinal scale to be maintained, with the aim of analyzing, for instance, an eventual research question. Though, it is also true that there are other situations where the aggregation of categories, classes of the variable, can be done without any inconvenient for the research purposes.

Therefore, based on the above reflections, there are two possibilities when transforming an ordinal scale into a linguistic variable: (1) maintaining the categories of the original scale or (2) aggregating such categories, so to obtain a linguistic variable with fewer terms than categories had the original ordinal scale. Diverse questions, mainly the problems that would have to be faced with the latter, make the first option to be more convenient. In Fig. 1 we show an example for the variable "Weekly use of the Web" (ordinal scale extracted from the database used in [20]), structured as follows: (1) $x \leq 1$ hour; (2) $1 < x \leq 5$ hours; (3) $5 < x \leq 10$ hours; (4) $10 < x \leq 20$ hours; (5) $20 < x \leq 40$ hours; and (6) $x > 40$ hours.

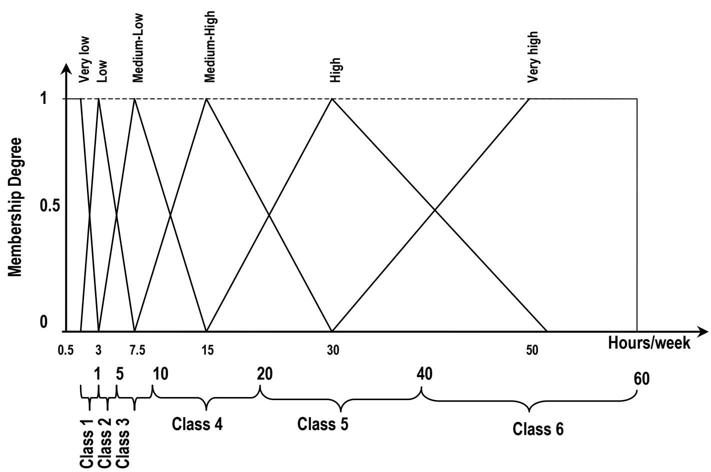


Fig. 1 Example of transformation of a marketing ordinal scale into a linguistic variable (classes of the original scale are maintained)

Interval scales. These scales are metric/quantitative, so they allow more possibilities when being transformed into linguistic variables. Notwithstanding, now we are going to focus our reflections on the particular case of the rating scales, as they are the scales habitually used to measure the items related to constructs. Two main questions should be tackled: the number of linguistic terms to use and the type of membership function more convenient to represent the behavior of the fuzzy sets. With respect to the former, though the particularities of each research has to be taken into account, considering the number of points commonly used by these types of scales (i.e. between five and eleven), it is convenient to work with a number of fuzzy set between three and five. Likewise, as these scales generally measure the consumer's opinion intensity on variables of interest for the certain research, we propose using, in general, the following labels or terms: *low*, *medium/indifferent*, and *high* when working with three fuzzy sets; and *very low*, *low*, *medium/indifferent*, *high*, and *very high* when working with five fuzzy sets. With respect to the second question, the membership function type, it is more convenient the transformation of scales using a triangular function. In particular, triangular functions must be necessarily used for the case of the extreme fuzzy sets defining a fuzzy variable related to certain rating scale. The main argument to support this is based on the characteristics of these scales.

For instance, let us consider a seven-point semantic differential scale (1: Bad - 7: Good), used to measure the consumer's attitude toward a brand. We know that when consumer has a totally negative attitude, his/her valuation will be 1. However, if his/her valuation were 2, that would mean a low attitude, though not the lowest level linked to a valuation of 1. Therefore, the fuzzy set *low* should show a membership degree of 1 when the marketing scale value is 1, decreasing with a linear tendency to zero for the rest of numeric values associated with said set. This reasoning would be equally valid for the case of the fuzzy set *high*, though it would be a fuzzy set with a membership function linearly increasing till the highest value of 7 in the marketing scale. Finally, as it is logic, the superior limit of the fuzzy set *low*, as well as the inferior limit of the fuzzy set *high*, would match with that value of the marketing scale in which the fuzzy set *medium* takes a membership degree of 1. Therefore, it is also necessary to define the domain of such central fuzzy set of the variable. The procedure to define this set differs depending on whether we are dealing with forced or non-forced scales. For the case of non-forced scales, the solution is intuitive and straight. As the central point of the scale represents an intermediate or indifferent position, said point would be the central point of the fuzzy set, with membership degree of 1. However, the solution for the case of forced-scales would imply another solution, as there is no central point. This fact makes necessary the use of a trapezoidal functions to define the behavior of the central fuzzy set of the variable, in such a way that the central points would be always formed by an even number of points in the scale, i.e., the central points. Figure 2 illustrates a graphic representation of two fuzzy variables with three linguistic terms each, associated with two semantic differential (non-forced and forced) scales.

Ratio scales. These present less restrictions and inconvenient to be transformed into fuzzy variables than any of the other types already described. As these scales are truly continuous, with zero as the lowest value, the numbers of linguistic terms, the determination of the domains for the established set of fuzzy sets, and the membership functions to use are completely flexible. The only inconvenient for the marketing expert is how to fix the maximum value for the scales in order to define the domain of the last fuzzy set of the variable.

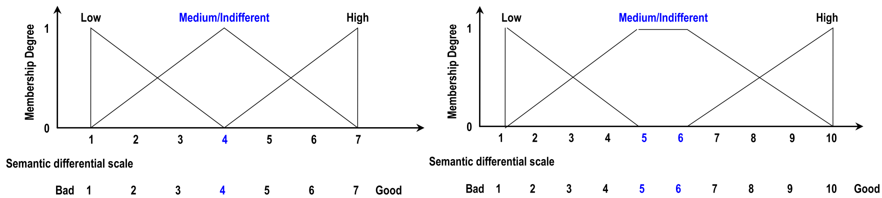


Fig. 2 Examples of membership functions for non-forced (seven-point) and forced (ten-points) rating scales

2.4 Application of Machine Learning to the Marketing Data

In general, two strategies could be used to let learners deal with the marketing data: (1) preprocessing the input data to render them tractable with a general learner or (2) adapting the learning technique to the particularities of the data. The former approach implies transforming the data into a simpler format. An intuitive approach would be to reduce the different items of a specific first-order construct to a single value (e.g., by averaging the values); a similar approach should be used to get an average value for second-order constructs. Another approach would be to expand any variable measured by multiple items to multiple variables measured by a single item and do not consider the existence of second-order constructs; then, the data set could be reduced by means of instance selection.

Nevertheless, the underlying problem of data preprocessing is that relevant information may be lost in the transformation process. For this purpose, Casillas and Martínez-López [4] proposed a modification of the inference process of fuzzy rule-based systems to deal with this especial type of data, which was addressed as *multi-item fuzzification*. The idea of this approach is to use fuzzy operators to (1) aggregate by fuzzy unions (T-conorms) the information provided by the multiple items that define a single variable and (2) intersect (with T-norms) the partial information provided by the first-order variables that describe second-order variables. This mechanism, included in Fuzzy-CSar, is detailed in Sect. 4.2.

3 Mining Association Rules

Association rule mining (ARM) [1] consists in extracting interesting patterns, associations, correlations, or causal structures among the variables in sets of usually unlabeled data. There has been and increasing interest in ARM in the recent years due to the existence of real-world applications in industry that generate large volumes of unlabeled data that have to be processed in order to extract novel and useful information for the company, which in turn may help guide the decision process of the business. This section briefly introduces ARM by first reviewing the initial approaches applied to data described by boolean variables and by then going to more recent approaches that can deal with numeric variables.

3.1 Association Rules: The Beginning

Initial research on ARM was mainly motivated by the analysis of market basket data, which enabled companies to get a better understanding of the purchasing behavior of their customers. Therefore, association rules were first applied to problems featured by binary or boolean variables. The problem of ARM can be described as follows.

Let $T = \{t_1, t_2, ..., t_n\}$ be a set of transactions, where each transaction consists of a set of items $I = \{i_1, i_2, \ldots, i_k\}$. Let an *itemset* X be a collection of items $I = \{i_1, i_2, ..., i_m\}$. A frequent itemset is an itemset whose support (*supp(X)*) is greater than a threshold specified by the user (this threshold is typically addressed as *minsupp* in the literature). The support of the rules is computed as

$$supp(X) = \frac{|X(T)|}{|T|}. \tag{1}$$

That is, the support is the number of transactions in the database which have the itemset X divided by the total number of transactions in the database.

Then, an *association rule* R is an implication of the form $X \rightarrow Y$, where both X and Y are itemsets and $X \cap Y = \emptyset$. As previously mentioned, ARM aims at extracting interesting association rules. Although many different measures have been developed to measure the interest of association rules so far [15], there are two basic indicators of the quality of the rules: support ($supp$) and confidence ($conf$). The support of a rule is defined as the ratio of the support of the union of antecedent and consequent to the number of transactions in the database, i.e.,

$$supp(R) = \frac{supp(X \cup Y)}{|T|}. \tag{2}$$

The confidence is computed as the ratio of the support of the union of antecedent and consequent to the support of the antecedent, i.e.,

$$conf(R) = \frac{supp(X \cup Y)}{supp(X)}. \tag{3}$$

Therefore, support indicates the frequency of occurring patterns, and confidence evaluates the strength of the implication denoted in the association rule.

Many algorithms have been proposed to extract association rules since the first proposal in [1]. Agrawal et al. [2] presented the Apriori algorithm, one of the most influential algorithms that set the basis for further research in association rule mining. Apriori uses two different phases to extract all the possible association rules with minimum support and confidence: (1) identification of all frequent itemsets and (2) generation of association rules from these large itemsets. The first phase is based on an iterative process that builds k-length itemsets by combining all the (k-1)-length itemsets whose support is greater than or equal to the minimum support fixed by the used. The support of each new itemset is computed by scanning the database. The second phase takes each frequent itemset and generates rules that contain some of the items in the antecedent of the rule and the remaining ones in the consequent of the rule. The confidence of each rule is computed by scanning the database for each rule, and only those with a minimum confidence are returned. As this process is time consuming, especially in large databases, new approaches that try to reduce the number of scans of the database have been proposed (e.g., see [9]).

3.2 Association Rule Mining with Continuous Variables

The first approaches to ARM only focused on analyzing whether an item was present in a transaction or not, describing the problem with boolean variables. Nonetheless, real-world problems are typically featured by continuous attributes, and these attributes can contain many distinct values. While the support of particular values for these attributes tends to be low, the support of interval of values is much higher. This created the need for building algorithms that could deal with intervals of values, yielding two approaches to the problem: quantitative association rules and fuzzy association rules.

Several authors proposed algorithms to deal with interval-based rules, which are typically addressed as quantitative association rules. In these algorithms, the aim was shifted to extracting rules in which variables are defined by intervals, such as "if *experience* $\in$ [5-10] years then *income* $\in$ [30 000 - 40 000]\$." One of the first methods that falls under this category can be found in [21], which, previously to extracting frequent itemsets, uses an equi-depth partitioning to define a set of intervals for each continuous attribute. The method creates a new variable for each interval, transforming therefore the problem into a binary problem. Then, an Apriori-like algorithm is applied to extract association rules from the transformed data. Although this approach and similar ones could deal with continuous variables, it was detected that these types of algorithms could either ignore or over-emphasize the items that lay near the boundary of intervals if the attributes were not properly partitioned. This was addressed as the *sharp boundary* problem. Two main approaches have been followed to tackle this problem. On the one hand, some authors have applied different clustering mechanisms to extract the best possible intervals from the data [16, 19, 25]. On the other hand, there have been some proposals that adjust these intervals during learning [18].

In parallel to these approaches, some authors faced the problem of sharp boundaries by incorporating fuzzy logics into ARM. In this case, variables were defined by fuzzy sets, allowing the system to extract rules such as "if *experience* is large then *income* is high," where *large* and *high* are two linguistic terms represented by fuzzy sets. As variables were represented by fuzzy sets, the problem of the sharp boundary was overcome. Among others, one of the most significant approaches under this category was proposed in [12, 13, 14], which redefined support and confidence for fuzzy association rules and designed an algorithm that combined ideas of Apriori with concepts of fuzzy sets to extract association rules described by variables represented by linguistic terms.

Since the consumer behavior modeling problem addressed in this chapter is featured by continuous attributes, we employ a system that falls under this last category. Therefore, Fuzzy-CSar is a system that creates fuzzy association rules from the database and utilizes the typical definitions of support and confidence defined for fuzzy systems to evaluate the interestingness of rules. The details of the algorithm are further explained in the following section.

4 Description of Fuzzy-CSar

Fuzzy-CSar is an ARM algorithm that follows a Michigan-style learning classifier system architecture [11] to extract fuzzy association rules from databases. Differently from most of the state-of-the-art algorithms in fuzzy ARM, Fuzzy-CSar (1) uses a fixed-size population to search for the most promising associations among variables, and so, does not necessarily create all the association rules with minimum support and confidence, (2) extracts association rules from streams of examples instead of from static databases, and, as a consequence, (3) does not scan repetitively the data base but incrementally learns from the stream of examples. The system uses an apportionment of credit technique to incrementally adjust the parameters of association rules and a genetic algorithm [8, 10] to discover new promising rules online. In addition, the system is provided with the multi-item fuzzification in order to deal with the particularities of the marketing data. In what follows, the system is described in some detail by first presenting the knowledge representation and the multi-item fuzzification and then explaining the learning organization.

4.1 Knowledge Representation

Fuzzy-CSar evolves a *population* [P] of *classifiers*, where each classifier individually denotes an association among problem variables. Therefore, the solution to the problem is the whole population. Note thus that the population size fixes an upper bound on the number of interesting associations that can be found; that is, at maximum, the system will be able to discover as many interesting relationships as number of classifiers in the population.

Each classifier consists of a *fuzzy association rule* and a set of parameters. The fuzzy association rule is represented as

$$\textbf{if } x_i \text{ is } \widetilde{A}_i \text{ and } \cdots \text{ and } x_j \text{ is } \widetilde{A}_j \textbf{ then } x_c \text{ is } \widetilde{A}_c,$$

in which the antecedent contains a set of ℓ_a input variables $x_i, \ldots, x_j$ ($0 < \ell_a < \ell$, where ℓ is the number of variables of the problem) and the consequent consists of a single variable x_c which is not present in the antecedent. Thus, we allow rules to have an arbitrary number of variables in the antecedent, but we require that rules have always one variable in the consequent.

Each variable is represented by a disjunction of *linguistic terms* or *labels* $\widetilde{A}_i = \{A_{i1} \vee \ldots \vee A_{in_i}\}$. However, the number of linguistic terms per variable is limited in order to avoid the creation of largely general rules that may provide poor information about the problem. That is, if no restriction were required, the system would tend to generate rules whose variables in the antecedent and consequent had all the possible linguistic terms, since they would cause the rule to match any possible input, and so, its support and confidence would be very high. To prevent the system from creating these rules, we allow the configuration of the maximum number of linguistic terms permitted per input variable (maxLabIn) and output variable (maxLabOut).

In addition to the rule itself, each classifier has also six main parameters: (1) the support $supp$, an indicator of the occurring frequency of the rule; (2) the confidence $conf$, which denotes the strength of the implication; (3) the fitness F, which is computed as a power of the confidence, so reflecting the quality of the rule; (4) the experience exp, which counts the number of times that the antecedent of the rule has matched an input instance; (5) the numerosity num, which reckons the number of copies of the classifier in the population; and (6) the average size of the association sets as in which the classifier has participated. The function of the different parameters, as well as the process followed to create and evolve these rules, is further explained with the process organization of Fuzzy-CSar in Section 4.3. But before that, next section introduces the multi-item fuzzification included in the system to deal with the marketing data.

4.2 Multi-item Fuzzification

In [17], the authors proposed the concept of multi-item fuzzification in order to deal with problems featured by unobserved variables described by multiple items and second-order constructs partially defined by first-order constructs. This procedure, which was incorporated into Fuzzy-CSar to deal with this kind of marketing data, considers both (1) how to compute the matching degree of a set of items with a variable and (2) how to calculate the matching of several first-order variables with a second-order variable.

The first idea of the method is that each individual item provides partial information about the corresponding unobserved variable or first-order variable. Therefore, the authors proposed to compute the matching degree as the aggregation (T-conorm) of the information given by each item. Thence, the matching degree of a variable i with the vector of items $\mathbf{x_i} = (x_1^i, x_2^i, \ldots, x_{p_i}^i)$ is

$$\mu_{\widetilde{A}_i}(\mathbf{x_i}) = max_{h_i=1}^{p_i} \mu_{\widetilde{A}_i}(x_{h_i}^i). \tag{4}$$

In our experiments, we considered the maximum as the union operator.

On the other hand, second-order variables are those defined by the intersection of the information provided by the corresponding first-order variables. For this reason, multi-item fuzzification calculates the matching degree of second-order variables as the T-norm of the matching degrees of each corresponding first-order variable. In our implementation, we used the minimum as T-norm.

4.3 Process Organization

After explaining the classifier representation and the mechanism to compute the matching degree in the marketing data, now we are in position to review the learning organization of Fuzzy-CSar. Fuzzy-CSar incrementally tunes the parameters of the classifiers as new examples are received and periodically applies the GA to *niches* of classifiers in order to create new rules that denote promising associations. The process is explained as follows. At each learning iteration, Fuzzy-CSar receives an input example (e_1, e_2, ..., e_ℓ) and takes the following actions. First, the system creates the *match set* [M] with all the classifiers in the population that match the input example with a degree larger than 0. If [M] contains less that θ_{mna} classifiers, the *covering operator* is triggered to create as many new matching classifiers as required to have θ_{mna} classifiers in [M]. Then, classifiers in [M] are organized in *association set candidates*.

Each association set candidate is given a probability to be selected that is proportional to the average confidence of the classifiers that belong to this association set. The selected *association set* [A] goes through a *subsumption* process which aims at diminishing the number of rules that express similar associations among variables. Then, the parameters of all the classifiers in [M] are updated. At the end of the iteration, a GA is applied to [A] if the average time since its last application is greater than θ_{GA}. This process is repeatedly applied, therefore, updating the parameters of existing classifiers and creating new promising rules online.

To fully comprehend the system process, five elements need further explanation: (1) the covering operator, (2) the procedure to create association set candidates, (3) the association set subsumption mechanism, (4) the parameter update procedure, and (5) the rule discovery by means of a GA. The subsequent subsections explicate each one of these elements in more detail.

4.3.1 Covering Operator

The covering operator is the responsible for providing the population with the initial classifiers which will be latter evaluated as new examples are received and evolved by the genetic algorithm. In order to create coherent rules, the operator generates rules that denote associations that are actually strong in the sampled example e from which covering has been activated. For this purpose, the covering operator uses the following procedure. Given the sampled input example e, covering creates a new classifier that contains some variables of the problem in the antecedent and the consequent of the rule and that matches e with maximum degree. That is, for each variable, the operator randomly decides (with probability $1 - P_\#$) whether the variable has to be in the antecedent of the rule, with the constraints (1) that, at least, a variable has to be selected and (2) that, at most, $\ell - 1$ variables can be included in the antecedent. Then, one of the remaining variables is selected to be in the rule consequent. Each of these variables is initialized with the linguistic label that maximizes the matching degree with the corresponding input value. In addition, we introduce generalization by permitting the addition of any other linguistic term with probability $P_{\#}$, with the restrictions that each variable in the antecedent and consequent respectively contains *maxLabIn* and *maxLabOut* linguistic terms at maximum.

4.3.2 Creation of Association Set Candidates

The system organizes the population rules in different niches that individually contain rules with similar associations with the aim of establishing a collaboration among niches and a competition of rules inside each niche. That is, the collaboration/competition scheme is produced by the niche-based genetic algorithm and the population-based deletion scheme, which are explained in subsection 4.3.5. The following explains the heuristic process employed to create these niches.

The system relies on the idea that rules that have the same variable with the same or similar linguistic terms in the consequent must belong to the same niche, since probably they would denote similar associations among variables. Therefore, in order to create the different association set candidates, Fuzzy-CSar first sorts the rules of [M] ascendantly depending on the variable of the consequent. Given two rules r_1 and r_2 that have the same variable in the consequent, the system considers that r_1 is smaller than r_2 if $\ell_1 < \ell_2$ or ($\ell_1 = \ell_2$ and $u_1 > u_2$), where ℓ_1, u_1, ℓ_2, and u_2 are the position of first and the last linguistic term of the output variable of each rule respectively.

Once [M] has been sorted, the association set candidates are built as follows. At the beginning, an association set candidate [A] is created and the first classifier in [M] is added to this association set candidate. Then, the following classifier k is added if it has the same variable in the consequent, and ℓ_k is smaller than the minimum u_i among all the classifiers in the current [A]. This process is repeated until finding the first classifier that does not satisfy this condition. In this case, a new association set candidate is created, and the same process is applied to add new classifiers to this association set. At the end, this process creates a non-fixed number of niches and distributes the rules through these niches.

4.3.3 Association Set Subsumption

The system explained thus far may generate similar rules that would coexist in the population. In order to avoid the maintenance of similar rules in the population, which would consume resources that may be useful to discover rules that denote different associations, Fuzzy-CSar incorporates a subsumption mechanism that searches for similar rules and only maintains the most general one.

The subsumption procedure works as follows. Each rule in [A] is checked for subsumption with each other rule in [A]. A rule r_i is a candidate subsumer of r_j if it satisfies the following four conditions: (1) r_i has higher confidence and it is experienced enough (that is, $conf^i > conf_0$ and $exp^i > \theta_{exp}$, where $conf_0$ and θ_{exp} are user-set parameters); (2) all the variables in the antecedent of r_i are also present in the antecedent of r_j (r_j can have more variables in the antecedent than r_i); (3) both rules have the same variable in the consequent; and (4) r_i is more general than

r_j. A rule r_i is more general than r_j if all the input and the output variables of r_i are also defined in r_j, and r_i has, at least, the same linguistic terms as r_j for each one of its variables.

4.3.4 Parameter Update

At the end of each learning iteration, the parameters of all the classifiers that belong to the match set are adjusted according to the information provided by the sampled instance. First, the experience of the classifier is incremented. Second, the support of each rule is updated as

$$supp_{t+1} = \frac{supp_t \cdot (\ell time - 1) + \mu_{\widetilde{A}}(x^{(e)}) \cdot \mu_{\widetilde{B}}(y^{(e)})}{\ell time}, \tag{5}$$

where $\ell time$ is the life time of the classifier, that is, the number of iterations that the classifier has been in the population, and $\mu_{\widetilde{A}}(x^{(e)})$ and $\mu_{\widetilde{B}}(y^{(e)})$ are the matching degrees of the antecedent and the consequent with $x^{(e)}$ and $y^{(e)}$ respectively. Note that this formula computes the support considering all the examples sampled to the system since the rule was created.

Thereafter, the confidence is computed as $conf_{t+1} = sum\_imp_{t+1}/sum\_mat_{t+1}$, where

$$sum\_imp_{t+1} = sum\_imp_t + \mu_{\widetilde{A}}(x^{(e)}) \cdot max\{1 - \mu_{\widetilde{A}}(x^{(e)}), \mu_{\widetilde{B}}(y^{(e)})\}, \text{ and} \tag{6}$$

$$sum\_mat_{t+1} = sum\_mat_t + \mu_{\widetilde{A}}(x^{(e)}). \tag{7}$$

Initially, $sum\_imp_{t+1} = sum\_mat_{t+1} = 0$. That is, *sum_imp* maintains the addition of the matching degree of each example sampled so far with the implication of the rule, and *sum_mat* keeps the addition of the matching degrees of the antecedent of the rule with each example sampled since the rule creation.

Next, the fitness of each rule in [M] is computed as a function of the confidence, i.e., $F = conf^{\nu}$, where ν permits controlling the pressure toward highly fit classifiers. Finally, the association set size estimate of all rules that belong to [A] is updated. Each rule maintains the average size of all the association sets in which it has participated.

4.3.5 Discovery Component

Fuzzy-CSar uses a niche-based GA to create new promising classifiers. The GA is triggered on [A] when the average time from its last application upon the classifiers in [A] exceeds the threshold θ_{GA}. The time elapsed between GA applications enables the system to adjust the parameters of the new classifiers before the next application of the GA.

Once triggered, the GA selects two parents p_1 and p_2 from [A], where each classifier has a probability of being selected proportional to its fitness. The two parents are crossed with probability P_χ, generating two offspring ch_1 and ch_2. Fuzzy-CSar uses a uniform crossover operator that contemplates the restriction that any offspring has to have, at least, a variable in the rule's antecedent. If crossover is not applied, the children are exact copies of the parents. The resulting offspring may go through three different types of mutation: (1) mutation of antecedent variables (with probability $P_{I/R}$), which randomly chooses whether a new antecedent variable has to be added to or one of the antecedent variables has to be removed from the rule; (2) mutation of the linguistic terms of the variable (with probability P_{μ}), which selects one of the existing variables of the rule and mutates its value; and (3) mutation of the consequent variable (with probability P_C), which selects one of the variables of the antecedent and exchanges it with the variable of the consequent. Thereafter, the new offspring are introduced into the population. If the population is full, excess classifiers are deleted from [P] with probability directly proportional to their association set size estimate and inversely proportional its fitness.

To sum up, Fuzzy-CSar is a population-based ARM that evaluates rules online as new examples are sampled to the system and that periodically applies a GA to create new promising rules. Note that the system does not require the user to determine the minimum support and minimum confidence of the rules. Instead of this, the system evolves a set of rules with maximum support and confidence, and the number of rules is limited by the population size. The rule activation based on matching prioritizes rules that match a larger number of training examples with respect to those that match a lower number of training examples. In addition, the confidence-based selection of [A] and the inside-niche competition established by the GA pressure toward the creation of rules with progressively higher confidence.

5 Problem Description and Methodology

After motivating the use of ARM for modeling the user behavior and presenting a technique that is able to extract fuzzy association rules without specifying the minimum support and the minimum confidence of the rules, we now move on to the experimentation. This section first explains the details of the marketing problem analyzed in this chapter and presents previous structural models extracted from this problem by using classical marketing analysis techniques. Then, we detail the experimental methodology.

5.1 Problem Description

The present work addresses the problem of modeling web consumers to extract key knowledge that enable marketing experts to create a compelling online environment for these users with the final goal of using this information to create a competitive advantage on the Internet. To tackle this problem, several authors have proposed causal models of the consumer experience on the Internet [5]. These models have mainly focused on the description of the state of *flow* during consumer navigation of the Web, that is, the cognitive state experienced during online navigation. Reaching

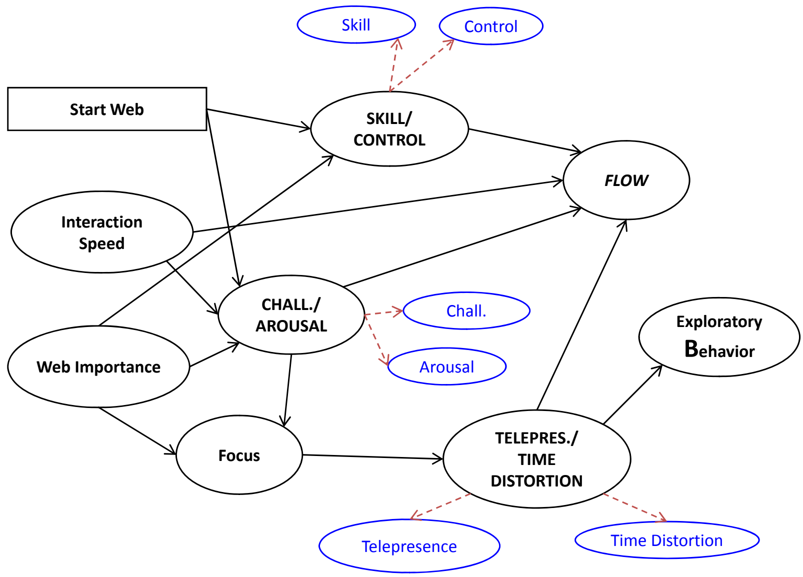


Fig. 3 Theoretical model of the user experience on the Internet

the state of flow comprises a "complete involvement of the user with his activity." Therefore, marketing experts are especially concerned with the identification of the factors that lead the user to the state of maximum flow.

In this chapter, we consider one of the most influential structural models as starting point, which was analytically developed by Novak et al. [20]. The structural model, illustrated in Fig. 3, consisted of nine *first-order constructs*: skill, control, interactive speed, importance, challenge, arousal, telepresence, time distortion, and exploratory behavior. In addition to these first-order variables, the model also contained three *second-order constructs*: skill/control, chall/arousal, and telepresence/time distortion, which are partially defined by first-order constructs. The model also considered the variable *startWeb*, which indicated for how long the user had been using the web.

The data were obtained from a large sample Web-based consumer survey conducted in [20]. These surveys posed a set of questions or *items* that partially described each one of the nine first-order constructs. The user was asked to grade these questions with Likert nine-point rating scales that ranged from "strongly disagree" to "strongly agree." The startWeb variable was measured with a six-ordinal rating scale that comprised different options of usage time.

The analysis conducted in [20] identified that the following four constructs were the most important ones to determine the state of *flow*: (1) skill and control, (2) challenge and arousal, (3) telepresence and time distortion, and (4) interactive speed. The other constructs were found to be meaningless to define *flow*. However, it is worth noting that the conclusions extracted by the classical approach depended on the initial causal model. Therefore, some key relationships may had not been discovered. In the following, we use the model-free system Fuzzy-CSar in order to identify associations among the variables of the problem with the aim of detecting any further relationship not captured by the causal model of Novak et al. The approach is not proposed as an alternative to the classical marketing analysis tools, but as a complement to these techniques. The next subsection details the experiments conducted.

5.2 Experimental Methodology

The aim of the experiments was to study whether the application of machine learning techniques could result in the identification of not only the same but new important associations between variables with respect to those detected in the causal model of Novak et al. In addition, we also analyzed the benefits of association rule mining over other machine learning techniques for data prediction, i.e., techniques in which the target variable is predetermined. For this purpose, we included an EMO approach expressly designed for creating rules with a fixed variable in the consequent for the marketing problem [4]. The approach used a genetic cooperative competitive scheme to evolve a Pareto set of rules with maximum support and confidence. For more details on the algorithm, the reader is referred to [4]. Herein, we explain the three experiments conducted to analyze the added value provided by ARM.

Experiment 1. The first experiment aimed at studying whether Fuzzy-CSar could capture the same knowledge represented in the structural model of Novak et al. This model focused on predicting the variable *flow*. The analytical study detected that there were four relevant variables to determine the state of *flow*: (1) skill and control, (2) challenge and arousal, (3) telepresence and time distortion, and (4) interactive speed. The remaining variables were considered irrelevant. Thus, we applied Fuzzy-CSar to the data extracted from the questionnaires, but only considering these four variables and fixing the variable *flow* as the output variable of the association rules. As the output variable was fixed, we could also apply the aforementioned EMO approach in order to analyze whether Fuzzy-CSar could obtain similar rules to those created by a system specialized to optimize the support and confidence of the rules.

Experiment 2. Since the first experiment did not consider all the input variables, some important knowledge could be overlooked by no considering important interactions between these missing variables. To study this aspect, we ran Fuzzy-CSar on the data described by all the input variables and compared the results with those obtained from the first experiment. Again, the variable *flow* was fixed as the target variable of the association rules. In addition, we also run the EMO approach on these data, extending the comparison of Fuzzy-CSar and the EMO approach started in the previous experiment.

Experiment 3. The two first experiments permitted the analysis of whether the machine learning techniques could extract similar knowledge to that provided by the structural model of Novak et al. and whether new important knowledge was discovered. Nevertheless, the two first experiments did not test the added value provided by extracting association rules online. Therefore, in the third experiment, we ran Fuzzy-CSar on the input data without forcing any variable in the consequent. Thus, the system was expected to evolve rules with different variables in the consequent, and so, to evolve the rules with maximum support and confidence. The aim of the experiment was to examine whether new interesting relationships, not captured by the structural model, could be discovered by Fuzzy-CSar. Note that, since the output variable was not fixed in this experiment, the EMO approach could not be ran.

In all the experiments, Fuzzy-CSar was configured with a population size of 6 400 rules and the following parameters: $P_{\#} = 0.5$, $P_{\chi} = 0.8$, $\{P_{I/B}, P_{\mu}, P_C\} = 0.1$, $\theta_{GA} = 50$, $\theta_{exp} = 1\,000$, $conf_0 = 0.95$, $\nu = 1$, $\delta = 0.1$. All the variables, except for *startWeb*, used Ruspini's strong fuzzy partitions with three linguistic terms. *startWeb* used six membership functions, each centered in one of the values that the variable could take. In all cases, $maxLabIn = 2$ and $maxLabOut = 1$. For the EMO approach, we employed the same configuration used by the authors [4]. That is, the system was configured to evolve a population of 100 individuals during 100 iterations, with crossover and mutation probabilities of 0.7 and 0.1 respectively. The variables used the same semantics as Fuzzy-CSar ones.

Before proceeding to the analysis of the results, it is worth highlighting the underlying differences between Fuzzy-CSar and the EMO approach. As aforementioned, the first important difference is in the knowledge representation: Fuzzy-CSar creates fuzzy association rules where the output variable is not fixed and the EMO approach creates rules with a prefixed target variable. Therefore, Fuzzy-CSar, and ARM algorithms in general, could create rules that denote important associations among variables in a single run; on the other hand, the EMO approach has to fix the output variable at each run. The second important difference is the process organization and the goal of the method. Fuzzy-CSar aims at learning a set of association rules distributed through different niches according to the genotype of the rule; in addition, the learning is done online. The fitness-based inside-niche selection and population-based deletion pressure toward obtaining rules with maximum confidence and support. Conversely, the EMO approach explicitly optimizes the rules with respect to their support and confidence, that is, it optimizes the Pareto front. Therefore, the EMO approach is more likely to evolve rules that maximize support and confidence, since it is specifically designed with this objective, while Fuzzy-CSar is more focused on evolving a diverse set of rules that have maximum confidence. Notwithstanding, we are interested in analyzing how our approach performs in comparison with a system which is specialized in optimizing the Pareto front.

6 Analysis of the Results

This section examines the results of the three experiments from two perspectives. First, we study the results on the objective space by analyzing the rules of the Pareto set, that is, those rules for which there do not exist any other rule in the population that has both a higher support and a higher confidence than the given rule. With this analysis we consider the ability of Fuzzy-CSar to create different rules with high support and confidence that are distributed through the solution space and compare it with the EMO approach, but we do not study the utility of the rules from the point of view of the marketing expert. This analysis is conducted afterwards, where several rules that provide new interesting knowledge about the problem, not captured by the structural model, are illustrated.

6.1 Analysis of the Rules in the Objective Space

Figure 4 shows the shape of the Pareto fronts evolved by Fuzzy-CSar and by the EMO approach in the first experiment, which considers only the four most relevant variables in the antecedent and forces *flow* to be in the consequent. The first row of Table 1 complements this information by reporting the average number of rules in the population of Fuzzy-CSar and the average number of rules in the Pareto set of Fuzzy-CSar and the EMO approach. In addition, to indicate the distribution of solutions in the Pareto set, the sum of the distance crowding between consecutive solutions in the Pareto front are also provided in parentheses.

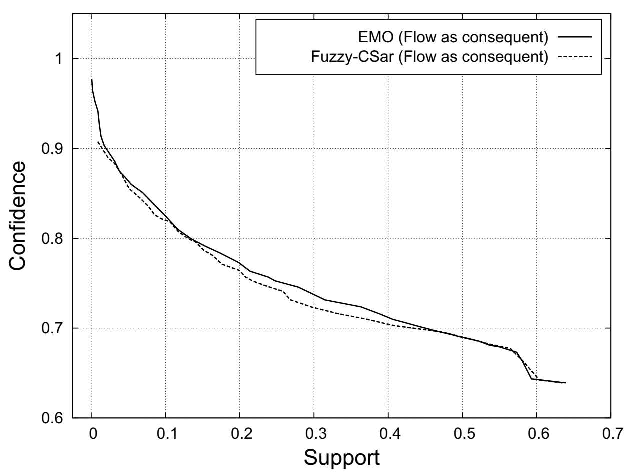


Fig. 4 Average Pareto front obtained by Fuzzy-CSar and the EMO approach considering the 4 variables of the marketing data identified as the most important variables by the structural model and fixing the variable *flow* as target of the rules

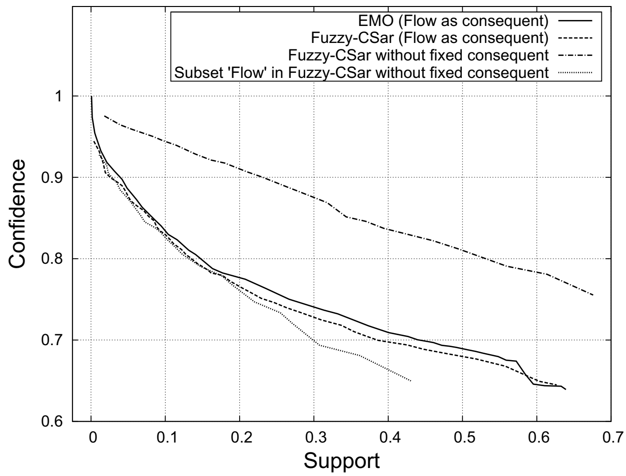


Fig. 5 Average Pareto front obtained by Fuzzy-CSar and the EMO approach considering the 9 variables of the marketing data

Table 1 Average number of rules evolved by Fuzzy-CSar, average number of these rules that are in the Pareto set, and average number of rules in the Pareto sets obtained by the EMO approach. For the Pareto sets, the average crowding distance of the population is provided in parentheses.

	FCSar All	FCSar Pareto	EMO Pareto
Experiment 1	479.2	76.3 $(1.53 \cdot 10^{-2})$	82.6 $(1.49 \cdot 10^{-2})$
Experiment 2	1259.7	105.9 $(1.07 \cdot 10^{-2})$	84.4 $(1.49 \cdot 10^{-2})$
Experiment 3	1752.5	468.3 $(2.58 \cdot 10^{-3})$	—

These results show the similarity of the results obtained with both methods. That is, both techniques discovered a similar number of solutions in the Pareto set, and these solutions were distributed uniformly through the objective space. The similarity of the results highlight the robustness of Fuzzy-CSar, which was able to generate Pareto fronts that were very close to those created by a competent technique which specifically optimized the Pareto front. In addition, the strongest rules obtained denote the same relationships provided by the structural model. Nonetheless, we put aside further discussion about the utility of the rules until the next section.

Figure 5 together with the second row of Table 1 show the same information but for the second experiment, which considers any of the nine variables in the antecedent and forces *flow* to be in the consequent. These Pareto fronts are very similar to those obtained in the first experiment. Actually, the EMO approach can discover practically the same rules than those obtained with the first experiment. On the other hand, Fuzzy-CSar obtains a significantly larger number of rules in the Pareto set; as a consequence, the average crowding distance decreases, since solutions in the Pareto set are closer to each other. Nonetheless, the shape of the Pareto sets is almost the same in both cases, which supports the hypothesis that the four variables identified as the most important ones by the models in [20] are indeed the most relevant ones to describe the *flow* construct.

Finally, Fig. 5 together with the third row of Table 1 supply the results of Fuzzy-CSar for the third experiment, where any variable can be in the antecedent or in the consequent of the association rules. These results show the potential of our approach. In a single run, Fuzzy-CSar was able to evolve a set of rules with large confidence and support, resulting in a Pareto front that was clearly better than those of Fuzzy-CSar and the EMO approach when the *flow* construct was fixed in the rule consequent.

To complement the results of the third experiment, the same figure plots the objective values of the rules of the Pareto front evolved by Fuzzy-CSar that predict the *flow* construct. Notice that, for large confidence, this Pareto front is close to the one evolved by the EMO approach and Fuzzy-CSar in previous experiments where *flow* was fixed in the consequent. On the other hand, the solutions in the Pareto front degrade as the confidence of the rules decreases. This behavior can be easily explained as follows. As the number of possible variables in the consequent increases, Fuzzy-CSar needs to maintain a larger number of rules that belong to different niches. In this case, the implicit niching system together with the niche-based GA and population-wise deletion operator of Fuzzy-CSar make pressure toward maintaining a diverse set of solutions. On the other hand, the GA also pressures toward evolving rules with maximum confidence. Therefore, the system maintains a diverse set of solutions with maximum confidence, which goes in detriment of solutions with smaller confidence, but larger support.

Similar results could be obtained by the EMO approach by running nine different experiments, each one fixing a different variable in the consequent. This would yield nine Pareto sets, each one with rules that predict one of the nine variables. Then, these Pareto sets could be joined and processed to get the final Pareto set. Nevertheless, it is worth noting that Fuzzy-CSar provides a natural support for the extraction of interesting association rules with different variables in the consequent, evolving a set of distributed solutions in parallel, and maintaining only those with maximum confidence.

6.2 Analysis of the Utility of the Rules from the Marketing Expert Perspective

After showing the competitiveness of Fuzzy-CSar with respect to the EMO approach, this section analyzes the importance of the knowledge provided by some of the rules discovered by Fuzzy-CSar. For this purpose, we show two particular examples of rules that provide key knowledge considered neither by the structural model [20] nor by the EMO approach [4].

Firstly, we selected a rule that predicted *exploratory behavior*, that is,

R_1: **IF** *importance* **is** Medium **and** *skill/control* **is** {Small or Medium} **and** *focusedAttention* **is** {Small or Medium} **and** *flow* **is** {Small or Medium} **THEN** *exploratoryBehavior* **is** Medium [Supp.: 0.22; Conf.: 0.87].

The model proposed by Novak et al. considered that *exploratory behavior* was related to only *telepresence/time distortion*, that is, the degree of telepresence and the effect of losing the notion of time while browsing the web. However, rule R_1 does not consider this relationship. Instead, it denotes that *exploratory behavior* is determined by *importance*, perceived *skill/control*, *focused attention* in the browsing process, and the state of *flow*. Thence, this rule indicates that intermediate values of the variables of the antecedent explicate, with confidence 0.87, states of moderate exploratory behaviors in the Web. The knowledge denoted by the rule may cause the marketing expert to consider other associations among variables that were not considered in the initial model. In particular, this relationship was initially considered in the causal model built in [20], but it was further discarded after a process of model refinement. Nonetheless, R_1 is alerting of the importance and strength of this association.

Secondly, we chose the following rule, which described *focused attention*:

R_2: **IF** *importance* **is** {Small or Medium} **and** *chall/arousal* **is** {Small or Medium} **and** *telepres/time distortion* **is** Medium **and** *exploratoryBehavior* **is** {Medium or Large} **THEN** *focused attention* **is** Medium [Supp.: 0.21; Conf.: 0.84]

In Novak's et al. model, *focused attention* was related to neither *importance* nor *chall/arousal*. However, rule R_2 indicates that these two variables together with *telepres/time distortion* and *exploratory behavior* may determine moderate degrees of attention in the Web browsing. This information is especially interesting since it contradicts the causal model. This contradiction is reasonable if we consider the following. Differently from [20], Fuzzy-CSar does not assume any type of problem structure. Thence, Fuzzy-CSar can discover new relations among variables that may appear to be very useful and interesting. This may be the case of R_2, which implies that increasing the experience in the navigation process may influence, together with the other variables, the capacity of users to focus their attention on the Web. In summary, R_2 proposes a new scenario that was not considered before, and marketing experts may analyze whether this new knowledge needs to be included in further revisions of the causal model.

In addition to these particular examples, it is worth emphasizing that, in general, unsupervised learning techniques such as Fuzzy-CSar may be relevant tools in problems for which *a priori* information is unknown. In these cases, association rules may discover interesting, useful, and hidden associations among the variables forming a database that help marketing experts better understand a certain problem they are approaching to.

7 Conclusions and Further Work

This chapter started by discussing the importance of the use of machine learning techniques to give support to classical methodologies for marketing analysis. Among the different techniques in machine learning, we identified ARM as one of the most appealing approaches since it enables the automatic identification of associations or relationships among variables from a data set. That is, differently from the classical approach, which requires that marketing experts work with a theoretical model, ARM does not require any type of *a priori* information about the problem.

In order to show the added value that ARM could provide to marketing experts, we reviewed the application of Fuzzy-CSar, a general-purpose ARM technique that evolves a set of association rules online and that uses adapted inference mechanisms to deal with the particularities of the marketing data. Then, we applied it to the problem of modeling the user navigational process in online (the Web) environments; in particular, we were based on the Novak et al. [20] data and *flow* model to develop the experimental stage. Additionally, the system was compared to a predictive EMO approach that needed to fix the target variable of the evolved rules. The empirical results highlighted the added value of applying machine learning techniques to the marketing problem and, more specifically, of extracting association rules. That is, Fuzzy-CSar was able not only to generate rules that expressed the same knowledge as that contained in the theoretical (structural) marketing model of reference, but also to capture additional relationships among variables not previously considered in the theoretical background. We have shown how some of such uncovered relationships are very interesting from the analyzed marketing problem perspective. To sum up, these results suggest the suitability of ARM for marketing databases analysis. In particular, it has demonstrated to be helpful in consumer behavior modeling, especially as a complementary analytical tool to the traditional methods applied there. Anyhow, marketing researchers and practitioners, especially the formers, must not forget that the outcomes of these new, less orthodox, analytical methods are desirable to be interpreted and assimilated without forgetting to connect with the subjacent theoretical frameworks of the marketing issues they face.

Acknowledgments. The authors would like to thank *Ministerio de Educación y Ciencia* for its support under projects TIN2008-06681-CO6-01 and TIN2008-06681-CO6-05, *Generalitat de Catalunya* for its support under grant 2005SGR-00302, and *Junta de Andalucía* for its support under project P07-TIC-3185.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington D.C., May 1993, pp. 207–216 (1993)

2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, September 1994, pp. 487–499 (1994)
3. Bollen, K.A.: Structural equations with latent variables. Wiley-Interscience, Hoboken (1989) (A division of John Wiley & Sons, Inc.)
4. Casillas, J., Martínez-López, F.J.: Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. Expert Systems With Applications 36(2), 1645–1659 (2009)
5. Csikszentmihalyi, M.: Finding flow: The psychology of engagement with everyday life (1997)
6. Drejer, A.: Back to basics an beyond: Strategic management – an area where practice and theory are poorly related. Management Decision, 42(3/4), 508–520
7. Glymour, C., Scheines, R., Spirtes, P., Kelly, K.: Discovering causal structure. Academic Press, Orlando (1987)
8. Goldberg, D.E.: Genetic algorithms in search, optimization & machine learning, 1st edn. Addison Wesley, Reading (1989)
9. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery 8(1), 53–87 (2004)
10. Holland, J.H.: Adaptation in natural and artificial systems. The University of Michigan Press, Ann Arbor (1975)
11. Holland, J.H., Reitman, J.S.: Cognitive systems based on adaptive algorithms. In: Waterman, D.A., Hayes-Roth, F. (eds.) Pattern-directed inference systems, pp. 313–329. Academic Press, San Diego (1978)
12. Hong, T.P., Kuo, C.S., Chi, S.C.: A fuzzy data mining algorithm for quantitative values. In: Proceedings International Conference on Knowledge-Based Intelligent Information Engineering Systems, pp. 480–483 (1999)
13. Hong, T.P., Kuo, C.S., Chi, S.C.: Mining association rules from quantitative data. Intelligent Data Analysis 3, 363–376 (1999)
14. Hong, T.P., Kuo, C.S., Chi, S.C.: Trade-off between computation time and number of rules for fuzzy mining from quantitative data. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 9(5), 587–604 (2001)
15. Lenca, P., Meyer, P., Vaillant, B., Lallich, S.: On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research 184, 610–626 (2008)
16. Lent, B., Swami, A.N., Widom, J.: Clustering association rules. In: Procedings of the IEEE International Conference on Data Engineering, pp. 220–231 (1997)
17. Martínez-López, F.J., Casillas, J.: Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic fuzzy systems. Industrial Marketing Management (2009), doi:10.1016/j.indmarman.2008.02.003
18. Mata, J., Alvarez, J.L., Riquelme, J.C.: An evolutionary algorithm to discover numeric association rules. In: SAC 2002, pp. 590–594. ACM, New York (2002)
19. Miller, R.J., Yang, Y.: Association rules over interval data. In: SIGMOD 1997: Proceedings of the 1997 ACM SIGMOD international conference on Management of data, pp. 452–461. ACM, New York (1997)
20. Novak, T., Hoffman, D., Yung, Y.: Measuring the customer experience in online environments: A structural modelling approach. Marketing Science 19(1), 22–42 (2000)

21. Srikant, R., Agrawal, R.: Mining quantitative association rules in large relational tables. In: Jagadish, H.V., Mumick, I.S. (eds.) Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, pp. 1–12 (1996)
22. Steenkamp, J., Baumgartner, H.: On the use of structural equation models for marketing modelling. International Journal of Research in Marketing 17, 195–202 (2000)
23. Stevens, S.S.: On the theory of scales of measurement. Science, 677–680 (1946)
24. Stevens, S.S.: Measurement, psychophysics and utility. John Wiley, New York (1959)
25. Wang, K., Tay, S.H.W., Liu, B.: Interestingness-based interval merger for numeric association rules. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, KDD, pp. 121–128. AAAI Press, Menlo Park (1998)

Fuzzy–Evolutionary Modeling of Customer Behavior for Business Intelligence

Célia da Costa Pereira and Andrea G.B. Tettamanzi

Università degli Studi di Milano
Dipartimento di Tecnologie dell'Informazione
via Bramante 65, I-26013 Crema, Italy
e-mail: {celia.pereira,andrea.tettamanzi}@unimi.it

Abstract. This chapter describes the application of evolutionary algorithms to induce predictive models of customer behavior in a business environment. Predictive models are expressed as fuzzy rule bases, which have the interesting property of being easy to interpret for a human expert, while providing satisfactory accuracy. The details of an island-based distributed evolutionary algorithm for fuzzy model induction are presented and a case study is used to illustrate the effectiveness of the approach.

Keywords: Business Intelligence, Data Mining, Modeling, Strategic Marketing, Forecast, Evolutionary Algorithms.

1 Introduction

Companies face everyday problems related to uncertainty in organizational planning activities: accurate and timely knowledge means improved business performance. In this framework, business intelligence applications represent instruments for improving the decision making process within the company, by achieving a deeper understanding of market dynamics and customers' behaviour.

Particularly, in the fields of business and finance, the executives can improve their insight of market scenarios by foreseeing customers' behaviour. This information allows to maximize revenues and manage costs through an increase in the effectiveness and efficiency of all the strategies and processes which involve the customers.

Predictions about customers' intentions to purchase a product, about their loyalty rating, the gross operating margins or revenue they will generate, are fundamental for two reasons. Firstly, they are instrumental to an effective planning of production volumes and specific promotional activities. Secondly, the comparison of projections to actual results allows to spot meaningful indicators, useful for improving performance.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 207–225.
springerlink.com

This chapter describes a general approach to business intelligence, which exploits an evolutionary algorithm to design and optimize fuzzy-rule-based predictive models of various types of customer behavior.

The chapter is organized as follows: Section 2 discusses the scenarios where evolutionary predictive modeling may be employed, and Section 3 gives an outline of the approach. The next sections introduce the main ingredients of the approach: Section 4 provides an introduction to fuzzy rule-based systems and Section 5 to evolutionary algorithms. Section 6 gives a detailed description of the approach, and Section 7 illustrates its effectiveness by means of a real-world case study. Section 8 concludes.

2 The Context

Traditional methods of customer analysis, like segmentation and market research, provide static knowledge about customers, which may become unreliable in time. A competitive advantage can be gained by adopting a data-mining approach whereby predictive models of customer behaviour are learned from historical data. Such knowledge is more fine-grained, in that it allows to reason about an individual customer, not a segment; furthermore, by re-running the learning algorithm as newer data become available, such an approach may be made to take a continuous picture of the current situation, thus providing dynamic knowledge about customers.

The described approach uses evolutionary algorithms (EAs) for model learning, and expresses models as fuzzy rule bases. EAs are known to be well-suited to tracking optima in dynamic optimization problems [5]. Fuzzy rule bases have the desirable characteristic of being intelligible, as they are expressed in a language typically used by human experts to express their knowledge.

2.1 Application Scenarios

The system has been applied to a wide range of situations to achieve different goals, including:

- credit scoring in the banking sector [20];
- estimating the lifetime value of customers in the insurance sector [23];
- debt collection, i.e., predicting the probability of success of each of several alternative collection strategies in order to minimize cost and maximize effectiveness of collection;
- predicting customer response to new products, as a help to target advertisement campaigns or promotional activities in general;
- predicting customer response to pricing and other marketing actions;
- modeling of financial time series for automated single-position day trading [7, 8].

- predicting the revenue and/or gross operating margins for each individual customer and each product, as an aid to optimizing production and sales force planning [24].

A case study on the last scenario is presented in Section 7.

Once the range of revenue (or gross operating margin) in which a customer will be located in the next quarter has been foreseen, the manager can evaluate the capability of the selling force: for example, if all the customers followed by a group of sellers will generate the minimum hypothetical revenue, the considered sellers are not so effective. In this case, the manager can compare different groups on the basis of their respective expected results, even if their specific targets and environments are heterogeneous. Moreover, the obtained projections enable the company to target strategic marketing actions aimed at proposing new products to the customers that really have a predisposition for a specific kind of products and at increasing customers loyalty. Comparing the expected and actual results, a manager can have a detailed picture of the business process in order to promptly manage critical variables.

2.2 Related Work

The idea of using fuzzy logic for expressing customer models or classifications and exploiting such information for personalizing marketing actions targeted to individual (potential) customers has been explored for at least a decade, with the first proposals of fuzzy market segmentation [25] dating back to almost twenty years ago.

For instance, the fuzzy modeling of client preferences has been applied to selecting the targets of direct marketing actions. This has been shown to provide advantages over the traditional practice of using statistical tools for target selection [21].

Other researchers have applied fuzzy logic to marketing by inducing a fuzzy classification of individual customers that can be exploited to plan marketing campaigns [13]. This idea finds its natural application in e-commerce [26, 27].

A similar approach has been proposed for personalized advertisement [29], to determine which ad to display on a Web site, on the basis of the viewer's characteristics.

Some researchers have taken the further step of combining fuzzy logic with other soft computing methods, like neural networks, to build decision support systems that assist the user in developing marketing strategies [14].

A natural combination with other soft computing methods consists of using evolutionary algorithms to induce fuzzy classification rules. This is the approach we describe in this chapter and other researchers have pursued with slightly different techniques, notably Herrera and colleagues [10].

3 Data Mining

In the area of business intelligence, data mining is a process aimed at discovering meaningful correlations, patterns, and trends between large amounts of data collected in a dataset. Once an objective of strategic marketing has been established, the system needs a wide dataset including as many data as possible not only to describe customers, but also to characterize their behaviour and tracing their actions.

The model is determined by observing past behaviour of customers and extracting the relevant variables and correlations between data and rating (dependent variable) and it provides the company with projections based on the characteristics of each customer: a good knowledge of customers is the key for a successful marketing strategy.

The described approach is based on the use of an EA which recognizes patterns within the dataset, by learning classifiers represented by sets of fuzzy rules. Using fuzzy rules makes it possible to get homogeneous predictions for different clusters without imposing a traditional partition based on crisp thresholds, that often do not fit the data, particularly in business applications. Fuzzy decision rules are useful in approximating non-linear functions because they have a good interpolative power and are intuitive and easily intelligible at the same time. Their characteristics allow the model to give an effective representation of the reality and simultaneously avoid the "black-box" effect of, e.g., neural networks.

The output of the application is a set of rules written in plain consequential sentences. The intelligibility of the model and the high explanatory power of the obtained rules are useful for the firm, in fact the rules are easy to be interpreted and explained, so that an expert of the firm can clearly read and understand them. An easy understanding of a forecasting method is a fundamental characteristic, since otherwise the managers are reluctant to use forecasts [1]. Moreover, the proposed approach provides the managers with an information that is more transparent for the stakeholders and can easily be shared with them.

4 Fuzzy Rule-Based Systems

This section provides a gentle introduction to fuzzy rule-based systems, with particular emphasis on the flavor employed by the described approach.

Fuzzy logic was initiated by Lotfi Zadeh with his seminal work on fuzzy sets [31]. Fuzzy set theory provides a mathematical framework for representing and treating vagueness, imprecision, lack of information, and partial truth.

Very often, we lack complete information in solving real world problems. This can be due to several causes. First of all, human expertise is of a qualitative type, hard to translate into exact numbers and formulas. Our understanding of any process is largely based on imprecise, "approximate" reasoning.

However, imprecision does not prevent us from performing successfully very hard tasks, such as driving cars, improvising on a chord progression, or trading financial instruments. Furthermore, the main vehicle of human expertise is natural language, which is in its own right ambiguous and vague, while at the same time being the most powerful communication tool ever invented.

4.1 Fuzzy Sets

Fuzzy sets are a generalization of classical sets obtained by replacing the characteristic function of a set A, χ_A which takes up values in $\{0, 1\}$ ($\chi_A(x) = 1$ iff $x \in A$, $\chi_A(x) = 0$ otherwise) with a *membership function* μ_A, which can take up any value in $[0, 1]$. The value $\mu_A(x)$ is the membership degree of element x in A, i.e., the degree to which x belongs in A.

A fuzzy set is completely defined by its membership function. Therefore, it is useful to define a few terms describing various features of this function, summarized in Figure 1. Given a fuzzy set A, its *core* is the (conventional) set of all elements x such that $\mu_A(x) = 1$; its *support* is the set of all x such that $\mu_A(x) > 0$. A fuzzy set is *normal* if its core is nonempty. The set of all elements x of A such that $\mu_A(x) \geq \alpha$, for a given $\alpha \in (0, 1]$, is called the α-cut of A, denoted A_α.

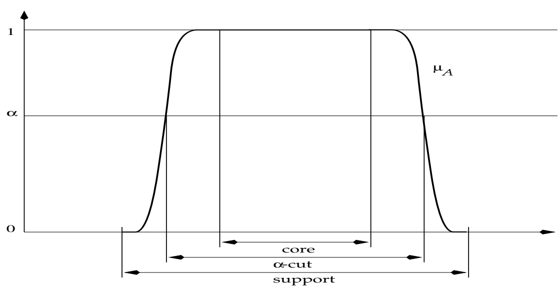


Fig. 1 Core, support, and α-cuts of a set A of the real line, having membership function μ_A

If a fuzzy set is completely defined by its membership function, the question arises of how the shape of this function is determined. From an engineering point of view, the definition of the ranges, quantities, and entities relevant to a system is a crucial design step. In fuzzy systems all entities that come into play are defined in terms of fuzzy sets, that is, of their membership functions. The determination of membership functions is then correctly viewed as a problem of design. As such, it can be left to the sensibility of a human expert or more objective techniques can be employed. Alternatively, optimal membership function assignment, of course relative to a number of design goals that have to be clearly stated, such as robustness, system performance, etc., can be estimated by means of a machine learning or optimization method. In particular, evolutionary algorithms have been employed with success to this aim. This is the approach we follow in this chapter.

4.2 Operations on Fuzzy Sets

The usual set-theoretic operations of union, intersection, and complement can be defined as a generalization of their counterparts on classical sets by introducing two families of operators, called triangular norms and triangular co-norms. In practice, it is usual to employ the min norm for intersection and the max co-norm for union. Given two fuzzy sets A and B, and an element x,

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}; \tag{1}$$
$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}; \tag{2}$$
$$\mu_{\bar{A}}(x) = 1 - \mu_A(x). \tag{3}$$

4.3 Fuzzy Propositions and Predicates

In classical logic, a given proposition can fall in either of two sets: the set of all true propositions and the set of all false propositions, which is the complement of the former. In fuzzy logic, the set of true proposition and its complement, the set of false propositions, are fuzzy. The degree to which a given proposition P belongs to the set of true propositions is its degree of truth, $T(P)$.

The logical connectives of negation, disjunction, and conjunction can be defined for fuzzy logic based on its set-theoretic foundation, as follows:

$$\text{Negation} \quad T(\neg P) = 1 - T(P); \tag{4}$$
$$\text{Disjunction} \quad T(P \vee Q) = \max\{T(P), T(Q)\}; \tag{5}$$
$$\text{Conjunction} \quad T(P \wedge Q) = \min\{T(P), T(Q)\}. \tag{6}$$

Much in the same way, a one-to-one mapping can be established as well between fuzzy sets and fuzzy predicates. In classical logic, a predicate of an element of the universe of discourse defines the set of elements for which that predicate is true and its complement, the set of elements for which that predicate is not true. Once again, in fuzzy logic, these sets are fuzzy and the degree of truth of a predicate of an element is given by the degree to which that element is in the set associated with that predicate.

4.4 Fuzzy Rulebases

A prominent role in the application of fuzzy logic to real-world problems is played by fuzzy rule-based systems. Fuzzy rule-based systems are systems of fuzzy rules that embody expert knowledge about a problem, and can be used to solve it by performing fuzzy inferences. The ingredients of a fuzzy rule-based systems are *linguistic variables*, *fuzzy rules*, and *defuzzification* methods.

4.4.1 Linguistic Variables

A linguistic variable [32] is defined on a numerical interval and has linguistic values, whose semantics is defined by their membership function. For example, a linguistic variable *temperature* might be defined over the interval $[-20°C, 50°C]$; it could have linguistic values like *cold, warm,* and *hot,* whose meanings would be defined by appropriate membership functions.

4.4.2 Fuzzy Rules

A fuzzy rule is a syntactic structure of the form

$$\text{IF } \textit{antecedent} \text{ THEN } \textit{consequent}, \tag{7}$$

where each *antecedent* and *consequent* are formulas in fuzzy logic.

Fuzzy rules provide an alternative, compact, and powerful way of expressing functional dependencies between various elements of a system in a modular and, most importantly, intuitive fashion. As such, they have found broad application in practice, for example in the field of control and diagnostic systems [19].

4.4.3 Inference

The semantics of a fuzzy rule-based system is governed by the calculus of fuzzy rules [33]. In summary, all rules in a fuzzy rule base take part simultaneously in the inference process, each to an extent proportionate to the truth value associated with its antecedent. The result of an inference is represented by a fuzzy set for each of the dependent variables. The degree of membership for a value of a dependent variable in the associated fuzzy set gives a measure of its compatibility with the observed values of the independent variables.

Given a system with n independent variables $x_1, \dots, x_n$ and m dependent variables $y_1, \dots, y_m$, let R be a base of r fuzzy rules

$$\begin{array}{ll} \text{IF } P_1(x_1, \dots, x_n) & \text{THEN } Q_1(y_1, \dots, y_m), \\ \quad\vdots & \quad\vdots \\ \text{IF } P_r(x_1, \dots, x_n) & \text{THEN } Q_r(y_1, \dots, y_m), \end{array} \tag{8}$$

where $P_1, \ldots, P_r$ and $Q_1, \ldots Q_r$ represent fuzzy predicates respectively on independent and dependent variables, and let τ_P denote the truth value of predicate P. Then the membership function describing the fuzzy set of values taken up by dependent variables $y_1, \ldots, y_m$ of system R is given by

$$\begin{aligned} &\tau_R(y_1, \ldots, y_m; x_1, \ldots, x_n) \\ &\quad = \sup_{1 \le i \le r} \min\{\tau_{Q_i}(y_1, \ldots, y_m), \tau_{P_i}(x_1, \ldots, x_n)\}. \end{aligned} \tag{9}$$

4.4.4 The Mamdani Model

The type of fuzzy rule-based system just described, making use of the min and max as the triangular norm and co-norm, is called the Mamdani model. A Mamdani system [15] has rules of the form

$$\text{IF } x_1 \text{ is } A_1 \text{ AND } \ldots \text{ AND } x_n \text{ is } A_n \text{ THEN } y \text{ is } B, \tag{10}$$

where the A_is and B are linguistic values (i.e., fuzzy sets) and each clause of the form "x is A" has the meaning that the value of variable x is in fuzzy set A.

4.4.5 Defuzzification Methods

There may be situations in which the output of a fuzzy inference needs to be a crisp number y^* instead of a fuzzy set R. Defuzzification is the conversion of a fuzzy quantity into a precise quantity.

At least seven methods in the literature are popular for defuzzifying fuzzy outputs [12], which are appropriate for different application contexts. The *centroid method* is the most prominent and physically appealing of all the defuzzification methods. It results in a crisp value

$$y^* = \frac{\int y \mu_R(y) dy}{\int \mu_R(y) dy}, \tag{11}$$

where the integration can be replaced by summation in discrete cases.

The next section introduces evolutionary algorithms, a biologically inspired technique which we use to learn and optimize fuzzy rule bases.

5 Evolutionary Algorithms

EAs are a broad class of stochastic optimization algorithms, inspired by biology and in particular by those biological processes that allow populations of organisms to adapt to their surrounding environment: genetic inheritance and survival of the fittest.

An EA maintains a population of candidate solutions for the problem at hand, and makes it evolve by iteratively applying a (usually quite small) set of stochastic operators, known as *mutation*, *recombination*, and *selection*.

Mutation randomly perturbs a candidate solution; recombination decomposes two distinct solutions and then randomly mixes their parts to form novel solutions; and selection replicates the most successful solutions found in a population at a rate proportional to their relative quality.

The initial population may be either a random sample of the solution space or may be seeded with solutions found by simple local search procedures, if these are available.

The resulting process tends to find, given enough time, globally optimal solutions to the problem much in the same way as in nature populations of organisms tend to adapt to their surrounding environment.

Books of reference and synthesis in the field of EAs are [9, 3, 2]; recent advances are surveyed in [30].

Evolutionary algorithms have enjoyed an increasing popularity as reliable stochastic optimization, search and rule-discovering methods in the last few years. The original formulation by Holland and others in the seventies was a sequential one. That approach made it easier to reason about mathematical properties of the algorithms and was justified at the time by the lack of adequate software and hardware. However, it is clear that EAs offer many natural opportunities for parallel implementation [17]. There are several possible parallel EA models, the most popular being the fine-grained or *grid* [16], the coarse-grain or *island* [28], and the master-slave or *fitness parallelization* [6] models. In the grid model, large populations of individuals are spatially distributed on a low-dimensional grid and individuals interact locally within a small neighborhood. In the master-slave model, a sequential EA is executed on what is called the *master* computer. The master is connected to several *slave* computers to which it sends individuals when they require evaluation. The slaves evaluate the individuals (fitness evaluation makes up most of the computing time of an EA) and send the result back to the master.

In the island model, the population is divided into smaller subpopulations which evolve independently and simultaneously according to a sequential EA. Periodic migrations of some selected individuals between islands allow to inject new diversity into converging subpopulations. Microprocessor-based multicomputers and workstation clusters are well suited for the implementation of this kind of parallel EA. Being coarse-grained, the island model is less demanding in terms of communication speed and bandwidth, which makes it a good candidate for a cluster implementation.

6 An Island-Based Evolutionary Algorithm for Fuzzy Rule-Base Optimization

This section describes an island-based distributed evolutionary algorithm for the optimization of fuzzy rule bases. In particular, the discussion will focus on the specialized mutation and crossover operation, as well as on the fitness function and ways to prevent overfitting.

The described approach incorporates an EA for the design and optimization of fuzzy rule-based systems that was originally developed to automatically learn fuzzy controllers [22, 18], then was adapted for data mining, [4] and is at the basis of MOLE, a general-purpose distributed engine for modeling and data mining based on EAs and fuzzy logic [24].

Each classifier is described through a set of fuzzy rules. A rule is made by one or more antecedent clauses ("IF ...") and a consequent clause ("THEN ..."). Clauses are represented by a pair of indices referring respectively to a variable and to one of its fuzzy sub-domains, i.e., a membership function.

A MOLE classifier is a rule base, whose rules comprise up to four antecedent and one consequent clause each. Input and output variables are partitioned into up to 16 distinct linguistic values each, described by as many membership functions. Membership functions for input variables are trapezoidal, while membership functions for the output variable are triangular.

Classifiers are encoded in three main blocks:

1. a set of trapezoidal membership functions for each input variable; a trapezoid is represented by four fixed-point numbers, each fitting into a byte;
2. a set of symmetric triangular membership functions, represented as an area-center pair, for the output variable;
3. a set of rules, where a rule is represented as a list of up to four antecedent clauses (the IF part) and one consequent clause (the THEN part); a clause is represented by a pair of indices, referring respectively to a variable and to one of its membership functions.

An island-based distributed EA is used to evolve classifiers. The sequential algorithm executed on every island is a standard generational replacement, elitist EA. Crossover and mutation are never applied to the best individual in the population.

6.1 Genetic Operators

The recombination operator is designed to preserve the syntactic legality of classifiers. A new classifier is obtained by combining the pieces of two parent classifiers. Each rule of the offspring classifier can be inherited from one of the parent programs with probability 1/2. When inherited, a rule takes with it to the offspring classifier all the referred domains with their membership functions. Other domains can be inherited from the parents, even if they are not used in the rule set of the child classifier, to increase the size of the offspring so that their size is roughly the average of its parents' sizes.

Like recombination, mutation produces only legal models, by applying small changes to the various syntactic parts of a fuzzy rulebase.

Migration is responsible for the diffusion of genetic material between populations residing on different islands. At each generation, with a small probability (the migration rate), a copy of the best individual of an island is sent to all connected islands and as many of the worst individuals as the number of connected islands are replaced with an equal number of immigrants.

A detailed description of the evolutionary algorithm and of its genetic operators can be found in [18].

6.2 Fitness

Modeling can be thought of as an optimization problem, where we wish to find the model M^* which maximizes some criterion which measure its accuracy in predicting $y_i = x_{im}$ for all records $i = 1, \ldots, N$ in the training dataset. The most natural criteria for measuring model accuracy are:

- the mean absolute error,

$$\mathrm{err}(M) = \frac{1}{N} \sum_{i=1}^{N} |y_i - M(x_{i1}, \ldots, x_{i,m-1})|; \tag{12}$$

- the mean square error,

$$\mathrm{mse}(M) = \frac{1}{N} \sum_{i=1}^{N} (y_i - M(x_{i1}, \ldots, x_{i,m-1}))^2. \tag{13}$$

One big problem with using such criteria is that the dataset must be *balanced*, i.e., an equal number of representative for each possible value of the predictive attribute y_i must be present, otherwise the underrepresented classes will end up being modeled with lesser accuracy. In other words, the optimal model would be very good at predicting representatives of highly represented classes, and quite poor at predicting individuals from other classes.

To solve this problem, MOLE divides the range $[y_{\min}, y_{\max}]$ of the predictive variable into 256 bins. The bth bin, X_b, contains all the indices i such that

$$1 + \lfloor 255 \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \rfloor = b. \tag{14}$$

For each bin $b = 1, \ldots, 256$, it computes the mean absolute error for that bin

$$\mathrm{err}_b(M) = \frac{1}{\|X_b\|} \sum_{i \in X_b} |y_i - M(x_{i1}, \ldots, x_{i,m-1})|, \tag{15}$$

then the total absolute error as an integral of the histogram of the absolute errors for all the bins, $\mathrm{tae}(M) = \sum_{b:\|X_b\| \neq 0} \mathrm{err}_b(M)$. Now, the mean absolute error for every bin in the above summation counts just the same no matter how many records in the dataset belong to that bin. In other words, the level of representation of each bin (which, roughly speaking, corresponds to a class) has been factored out by the calculation of $\mathrm{err}_b(M)$. What we want from a model is that it is accurate in predicting all classes, independently of their cardinality.

The fitness used by the EA is given by $f(M) = \frac{1}{\mathrm{tae}(M)+1}$, in such a way that a greater fitness corresponds to a more accurate model.

6.3 Selection and Overfitting Control

In order to avoid overfitting, the following mechanism is applied: the dataset is split into two subsets, namely the training set and the test set. The training set is used to compute the fitness considered by selection, whereas the test set is used to compute a test fitness. Now, for each island, the best model so far, M^*, is stored aside; at every generation, the best individual with respect to fitness is obtained,

$$M_{\text{best}} = \mathrm{argmax}_i f(M_i).$$

The test fitness of M_{best}, $f_{\text{test}}(M_{\text{best}})$, is computed and, together with $f(M_{\text{best}})$, it is used to determine an optimistic and a pessimistic estimate of the real quality of a model: for all model M, $f_{\text{opt}}(M) = \max\{f(M), f_{\text{test}}(M)\}$, and $f_{\text{pess}}(M) = \min\{f(M), f_{\text{test}}(M)\}$. Now, M_{best} replaces M^* if and only if $f_{\text{pess}}(M_{\text{best}}) > f_{\text{pess}}(M^*)$, or, in case $f_{\text{pess}}(M_{\text{best}}) = f_{\text{pess}}(M^*)$, if $f_{\text{opt}}(M_{\text{best}}) > f_{\text{opt}}(M^*)$.

Elitist linear ranking selection, with an adjustable selection pressure, is responsible for improvements from one generation to the next. Overall, the algorithm is elitist, in the sense that the best individual in the population is always passed on unchanged to the next generation, without undergoing crossover or mutation.

7 A Case Study on Customer Revenue Modeling

The system has been applied to the predictive modeling of the revenue generated by customers of an Italian medium-sized manufacturing corporation operating in the field of timber and its end products.

7.1 The Company

Ever since 1927, the corporation we have targeted has been working proficiently and professionally in the industry of timber and its by-products both on the domestic and on the international market, becoming renowned for innovation, guarantee of quality, and careful customer care. Understanding the market has been their winning strategy to be dynamic and always proactive towards their clients, by offering customized solutions in every sector and often anticipating their needs. Indeed, every market sector has distinctive features and specific trends: the corporation's products serve multiple purposes and are suited to different uses.

Their studies and continuous research on high technology and specialist materials, together with the expertise and the energy of their sales network have afforded the ideal answer to many major firms which have chosen the corporation as their strategic partner.

The company is a leader in manufacturing special plywood, assembled with pioneering and technological materials—all certified and guaranteed by the ISO 9001:2000 Company Quality System—and specializes in selling timber and semi-finished products coming from all over the world.

The company manufactures two broad types of products, namely technical plywood and dimensional lumber. The company mainly sells its products to construction companies, shipyards (especially those building yachts, where high-quality timbers are demanded and appreciated), and retailers.

For the purposes of this case study, the products marketed by the company may be divided into four homogeneous classes, by combining two orthogonal classification criteria:

- type of product: plywood vs. dimensional lumber;
- distribution channel: direct sale to construction and shipbuilding companies vs. sale to distributors.

The four homogeneous product classes are thus the following:

1. production lumber—dimensional lumber sold directly to manufacturing companies;
2. production plywood—plywood sold directly to manufacturing companies;
3. commercial lumber—dimensional lumber sold to distributors for the retail market;
4. commercial plywood—plywood sold to distributors for the retail market.

The rationale for this four-way classification is that each of the four resulting classes has its own specific customer requirements and channel of distribution, which is reflected by the internal organization of the marketing department and of the sales network of the corporation.

7.2 Aim of the Study

In the Fall of 2005, a pilot test was performed to demonstrate the feasibility of an innovative approach to customers modeling in revenue segments. In order to reduce time and costs, the traditional statistical analysis of data was skipped.

Classifying customers into revenue segments can be useful not only to plan the activities and forecast the overall returns for the next period, but also to identify characteristics which describe different patterns of customers, to recognize strategic and occasional customers, to target commercial/marketing activities and so on.

7.3 The Data

The described approach was used to develop a predictive model to foresee the customers' revenue segments for a quarter using historical data of the year before the analysis. Customers were classified into three quarterly revenue segments:

1st segment: revenue >50,000 euro/quarter;
2nd segment: revenue between 10,000 and 50,000 euro/quarter;
3rd segment: revenue <10,000 euro/quarter.

Historical data on revenue generated by each customer c were available in the form of monthly revenues m_{ip}^{c}, for the ith month ($i = 1, \dots, 24$) and four homogeneous classes of products, $p = 1, \dots, 4$. These data have been aggregated on a quarterly basis, giving a vector of quarterly revenues q_{jp}^{c}, to be used in order to perform an analysis of the 12-months trend-cycle. Data were adjusted seasonally, since the observations relating to the month of August, when most businesses shut down for vacations, were supposed to be not significant.

The dataset given to MOLE as input is extracted from the deseasonalized quarterly historical data q_{jp}^{c} as follows: a sliding window of one year (i.e., four quarters) plus the revenue segment y_{j+4} for the forward quarter (based on total customer revenue) is used to extract a staggered set of records in the form

$$q_{j1}^{c}, \dots, q_{j4}^{c}, \dots, q_{j+3,1}^{c}, \dots, q_{j+3,4}^{c}, y_{j+4}, \tag{16}$$

for $j = 1, \dots, 19$. Such a record provides a summary of a year of activity by a customer, along with the associate value of the predictive variable y_{j+4}.

With reference to the three selected revenue segments for this pilot test, MOLE was also given a customer segment, calculated by aggregating partial revenues related to every single product during the following period. For

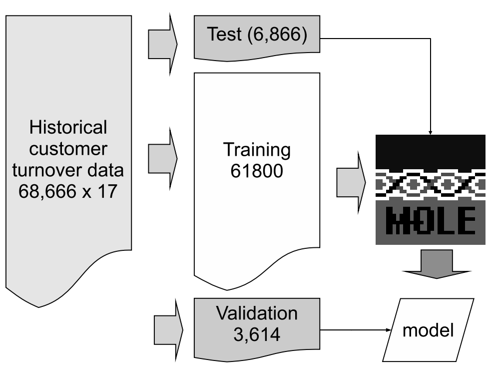


Fig. 2 A block diagram of the experimental setup

example, if we consider quarterly data for Q1, Q2, Q3, Q4, then the revenue segment is calculated based on the sum of four single products revenues in the first quarter of the following year.

Finally, data concerning customer's industrial sector, geographical location, and average quarterly revenue during the previous year were also added to each record.

The resulting dataset, consisting of 19 distinct records for each customer, i.e., 68,666 records overall, was fed to the MOLE evolutionary engine for model learning, with 4 islands of 100 individuals each connected according to a ring topology and a migration rate of 0.1; for each island, the mutation rate was set to 0.1 and the crossover rate to 0.5.

Figure 2 summarizes the setup of the experiment.

7.4 Discussion of the Results

In order to simplify the procedures for this preliminary test, the authors assigned meaningful labels (e.g., low, medium, high, etc.) to the membership functions generated by MOLE, although such labels should normally be established jointly with the customer.

The y_{j+4} variable (the "rating") represents the revenue segment for the following quarter.

The algorithm underlying the model evaluates all the rules at the same time and provides forecasts by calculating the average between values assigned to the rating in the consequent of every rule, weighted by the degree of satisfaction of the antecedent.

Model output is usually generated as result of the interaction between more than one fuzzy rule weighted using the corresponding satisfaction degree.

Following some of the rules representing the selected model are proposed:

IF $q_{j+1,3}$ **is** medium **AND** $q_{j+2,3}$ **is** very high
THEN y_{j+4} **is** first segment

IF $q_{j+3,3}$ **is** medium-low **AND** $q_{j+1,4}$ **is** very high
THEN y_{j+4} **is** first segment

IF $q_{j+1,3}$ **is** very low **AND**
$q_{j+2,2}$ **is** medium-high **AND**
$q_{j+3,3}$ **is** medium-high **AND**
$q_{j+1,1}$ **is** any
THEN y_{j+4} **is** first segment

The managers of the company could easily evaluate the correlations suggested by the rules. For example, by analyzing the presented rules, it is possible to recognize the trend of purchasing for product class 3: purchases of this product class have a high frequency and the customer which has given the company a revenue, even if medium, in a recent period probably will generate an high revenue in the next period.

7.5 Validation of the Model

The model thus determined using data up to September 2005 has been employed to predict future revenue for every customer in the fourth quarter 2005; then, at the beginning of 2006, the predictions have been compared with actual data referring to the same period: the model correctly hits 2,672 records out of 3,614 and the fitness of the model to the data is 0.44, while the total error is 1.27.

The error has been compared to the one obtained using simple forecasts normally used by the company and based exclusively on the average revenue during the previous year.

The error of the models in predicting revenue resulted significantly lower for segments 1 and 2, which are the most strategic for the company, as they comprise the customers that make up for the greatest part (almost 70%) of the total revenue. Table 1 shows the distribution of the error with the evolved model (concentrated in the third segment) and the difference between the results obtained using the evolved model versus the estimation based on the average revenue during the previous year.

Table 1 Comparison between error in predictions obtained on the basis of average revenue during the previous year on one hand and using the evolved model on the other hand

Segment	Error using Average	Error using Model
1 (>50,000 EUR/Q)	0.71	0.5
2 (10,000–50,000 EUR/Q)	0.54	0.5
3 (<10,000 EUR/Q)	0.03	0.26

8 Conclusions

This chapter has described an innovative soft-computing approach to customer modeling for strategic marketing based on evolutionary algorithms and fuzzy logic. The main theoretical contribution of this chapter lies in the unique combination of fuzzy rule-based models and of an evolutionary algorithm specifically designed and refined over the years to deal with the optimization of fuzzy rule-based models. In particular, the mutation operator, the fitness function, and the original overfitting control mechanism built into the selection operator make the described approach effective in producing accurate models with remarkable generalization capabilities.

On the practical side, the pilot test was implemented without a preliminary elaboration of the data. The positive results allow the authors to expect further improvements through an analysis of the data, to identify other potential explicative variables or useful information about customer behaviour, i.e., for example, frequency of purchases, etc.

Fuzzy-evolutionary forecasts provide the company with new qualitative knowledge of customers, also due to the flexibility of the instrument that can be used in many different applications. Moreover, these additional information are always up to date, since the system automatically considers new data as soon as they are added to the dataset.

Dynamic predictive knowledge of customers is the most important factor to develop an effective marketing strategy. In fact, the aim of business intelligence is improving the comprehension of market dynamics and actors, and these forecasts allow the firm to understand the development of customer preferences and complete the framework of internal knowledge of the company in order to be able to answer the continuous new customer needs. The company has thus the opportunity to effectively plan production volumes and financial flows for the following period and can target specific promotional and cross-selling activities. Moreover, the executives have an instrument to evaluate the policies adopted by sellers in different regions: the gap between forecasts and real data can be the term of comparison between the performance of different sellers.

The decision-making process within the company takes an important advantage in terms of invested time and resources, and finally the effectiveness of marketing activities can considerably increase.

References

1. Armstrong, J.S., Brodie, R.J., McIntyre, S.H.: Forecasting methods for marketing: Review of empirical research. International Journal of Forecasting 3, 335–376 (1987)
2. Bäck, T.: Evolutionary algorithms in theory and practice. Oxford University Press, Oxford (1996)
3. Bäck, T., Fogel, D., Michalewicz, Z.: Evolutionary Computation. IoP Publishing, Bristol (2000)
4. Beretta, M., Tettamanzi, A.: Learning fuzzy classifiers with evolutionary algorithms. In: Bonarini, G.P.A., Masulli, F. (eds.) Soft Computing Applications, pp. 1–10. Physica Verlag, Heidelberg (2003)
5. Branke, J.: Evolutionary Optimization in Dynamic Environments. Kluwer Academic Publishers, Dordrecht (2001)
6. Cantú-Paz, E.: A survey of parallel genetic algorithms. Tech. Rep. IlliGAL 97003, University of Illinois at Urbana-Champaign (1997), `http://citeseer.ist.psu.edu/article/cantu-paz97survey.html`

7. da Costa Pereira, C., Tettamanzi, A.: Fuzzy-evolutionary modeling for single-position day trading. In: Brabazon, A., O'Neill, M. (eds.) Natural Computing in Computational Finance. Studies in Computational Intelligence, vol. 100, pp. 131–159. Springer, Berlin (2008)
8. da Costa Pereira, C., Tettamanzi, A.G.B.: Horizontal generalization properties of fuzzy rule-based trading models. In: Giacobini, M., Brabazon, A., Cagnoni, S., Di Caro, G.A., Drechsler, R., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Fink, A., McCormack, J., O'Neill, M., Romero, J., Rothlauf, F., Squillero, G., Uyar, A.Ş., Yang, S. (eds.) EvoWorkshops 2008. LNCS, vol. 4974, pp. 93–102. Springer, Heidelberg (2008)
9. DeJong, K.A.: Evolutionary Computation: A unified approach. MIT Press, Cambridge (2002)
10. del Jesus, M.J., Gonzalez, P., Herrera, F., Mesonero, M.: Evolutionary fuzzy rule induction process for subgroup discovery: A case study in marketing. IEEE Transactions on Fuzzy Systems 15(4), 578–592 (2007)
11. Eiben, A.E., Smith, J.E.: Introduction to Evolutionary Computing. Springer, Berlin (2003)
12. Hellendoorn, H., Thomas, C.: Defuzzification in fuzzy controllers. Intelligent and Fuzzy Systems 1, 109–123 (1993)
13. Kaufmann, M., Meier, A.: An inductive fuzzy classification approach applied to individual marketing. In: NAFIPS 2009, Cincinnati, Ohio (2009)
14. Li, S.: The development of a hybrid intelligent system for developing marketing strategy. Decision Support Systems 27(4), 395–409 (2000)
15. Mamdani, E.H.: Advances in linguistic synthesis of fuzzy controllers. International Journal of Man Machine Studies 8, 669–678 (1976)
16. Manderick, B., Spiessens, P.: Fine-grained parallel genetic algorithms. In: Schaffer, J.D. (ed.) Proceedings of the Third International Conference on Genetic Algorithms. Morgan Kaufmann, San Mateo (1989)
17. Mühlenbein, H.: Parallel genetic algorithms, population genetics and combinatorial optimization. In: Schaffer, J.D. (ed.) Proceedings of the Third International Conference on Genetic Algorithms, pp. 416–421. Morgan Kaufmann, San Mateo (1989)
18. Poluzzi, R., Rizzotto, G.G.: An evolutionary algorithm for fuzzy controller synthesis and optimization based on SGS-Thomson's W.A.R.P. fuzzy processor. In: Sanchez, L.A.Z.E., Shibata, T. (eds.) Genetic algorithms and fuzzy logic systems: Soft computing perspectives. World Scientific, Singapore (1996)
19. Ross, T.: Fuzzy Logic with Engineering Applications. McGraw-Hill, New York (1995)
20. Sammartino, L., Simonov, M., Soroldoni, M., Tettamanzi, A.: Gamut: A system for customer modeling based on evolutionary algorithms. In: Whitley, L.D., Goldberg, D.E., Cantú-Paz, E., Spector, L., Parmee, I.C., Beyer, H.G. (eds.) Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2000), Las Vegas, Nevada, USA, July 8-12, p. 758. Morgan Kaufmann, San Francisco (2000)
21. Setnes, M., Kaymak, U.: Fuzzy modeling of client preference from large data sets: An application to target selection in direct marketing. IEEE Transactions on Fuzzy Systems 9(1), 153–163 (2001)

22. Tettamanzi, A.: An evolutionary algorithm for fuzzy controller synthesis and optimization. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 5/5, pp. 4021–4026 (1995) IEEE Systems, Man, and Cybernetics Society
23. Tettamanzi, A., Sammartino, L., Simonov, M., Soroldoni, M., Beretta, M.: Learning environment for life time value calculation of customers in insurance domain. In: Deb, K., et al. (eds.) GECCO 2004. LNCS, vol. 3103, pp. 1251–1262. Springer, Heidelberg (2004)
24. Tettamanzi, A.G.B., Carlesi, M., Pannese, L., Santalmasi, M.: Business intelligence for strategic marketing: Predictive modelling of customer behaviour using fuzzy logic and evolutionary algorithms. In: Giacobini, M. (ed.) EvoWorkshops 2007. LNCS, vol. 4448, pp. 233–240. Springer, Heidelberg (2007)
25. Wedel, M., Steenkamp, J.B.E.M.: A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. Journal of Marketing Research XXVIII, 385–396 (1991)
26. Werro, N., Stormer, H., Meier, A.: Personalized discount—a fuzzy logic approach. In: Funabashi, M., Grzech, A. (eds.) Challenges of Expanding Internet: E-Commerce, E-Business, and E-Government. 5th IFIP Conference e-Commerce, e-Business, and e-Government (I3E 2005), Poznań, Poland. IFIP, vol. 189, pp. 375–387 (2005)
27. Werro, N., Stormer, H., Meier, A.: A hierarchical fuzzy classification of online customers. In: IEEE International Conference on e-Business Engineering (ICEBE 2006). Shanghai (2006)
28. Whitley, D., Rana, S., Heckendorn, R.B.: The island model genetic algorithm: On separability, population size and convergence. Journal of Computing and Information Technology 7(1), 33–47 (1999)
29. Yager, R.R.: Targeted e-commerce marketing using fuzzy intelligent agents. IEEE Intelligent Systems 15(6), 42–45 (2000)
30. Yao, X., Xu, Y.: Recent advances in evolutionary computation. Computer Science and Technology 21(1), 1–18 (2006)
31. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
32. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, i–ii. Information Science 8, 199–249, 301–357 (1975)
33. Zadeh, L.A.: The calculus of fuzzy if-then rules. AI Expert 7(3), 22–27 (1992)

An Evaluation Model for Selecting Integrated Marketing Communication Strategies for Customer Relationship Management

Tsuen-Ho Hsu[1,*], Yen-Ting Helena Chiu[2], and Jia-Wei Tang[3]

[1] Department of Marketing and Distribution Management,
National Kaohsiung First University of Science and Technology
Tel.: 886-7-6011000 ext. 4217.
e-mail: thhsu@ccms.nkfust.edu.tw
2 Jhuoyue Road, Nanzih District, Kaohsiung City, 811, Taiwan
[2] Department of Marketing and Distribution Management,
National Kaohsiung First University of Science and Technology.
e-mail: helena@ccms.nkfust.edu.tw
[3] Graduate School of Management, National Kaohsiung
First University of Science and Technology
e-mail: u9528911@ccms.nkfust.edu.tw

Abstract. Since the early 1990s, integrated marketing communication (IMC) has become the accepted practice in the marketing field. An increasing number of researchers consider the marketing communication strategies of IMC as offering key competitive advantages associated with customer relationship management. This paper develops an evaluating model for selecting strategies of IMC to solidify relationships with existing customers based on the quality function deployment (QFD) approach incorporating with the fuzzy analytic hierarchy process (FAHP) method. IMC is a concept by which a company systematically coordinates its multiple messages and different communication channels and integrates them into a cohesive and consistent marketing communications mix. Furthermore, fostering long-term customer relationships constitutes an essential part of IMC from a strategic perspective. The QFD approach is not only able to incorporate the voice of customer (VOC) into the marketing communication strategies of a company but also provides a systematic planning tool for incorporating information of elements to make appropriate decisions effectively and efficiently. In addition, the FAHP method can reduce imprecision and improve judgment when determining the relative importance of marketing decisions for customer relationship benefits. The proposed model has proven useful in evaluating value for the department store by presenting the results of an empirical study. Managers could apply this model to re-examine their own strategies of IMC and ensure that their strategies can satisfy or maintain the voice of customer for the purpose of relationship benefits in order to continually facilitate the evolution of marketing communication activities.

* Corresponding author.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 227–254.
springerlink.com

Keywords: Integrated Marketing Communication, Customer Relationship Management, Quality Function Deployment, Fuzzy Analytic Hierarchy Process.

1 Introduction

Integrated marketing communication (IMC) is a concept that entails a company's systematically coordinating its multiple messages and many communications channels, and integrating them into a cohesive and consistent marketing communications mix with the aim of sending the target market a clear, consistent message and image about it and its offerings (Lee and Park 2007). An increasing number of researchers and practitioners consider marketing communication strategies of IMC as offering key competitive advantages associated with customer relationship management. First, they have developed targeted marketing communication strategies of IMC which are designed to build closer relationships with customers in more narrowly segmented markets (e.g., Kotler and Armstrong 2005; Tauder 2005; Reynolds 2006; Martensen et al. 2007). At the same time, efforts to improve marketing communication's return on investment have intensified as demands for accountability have increased (Reid 2005). Second, evolutions in information technology have sped the movement toward segmented marketing and have enabled the marketers to adopt new fragmented communications channels (e.g., Sun 2006; Belch and Belch 2007; Mort and Drennan 2007). However, when customers obtain information about a brand, product or company from an increasing array of communications channels, they often get confused with disparate or inconsistent messages about the same subject (Puntoni and Tavassoli 2007). This is a result of marketers' tendency to neglect integrating and coordinating these various messages and communications strategies. In the customer's mind, all information from different media channels becomes part of the message about a company. Conflicting messages from these different sources can create a confusing company image in the customer's mind, and hinder efforts to build closer relationships with existing customers. Therefore, the needed model for systematic integration and coordination of IMC strategies to build customer relationships is not merely one of theoretical concern, but has become one of practical necessity.

The billions invested in customer relationship management (CRM) over the last decade are testimony to the desire companies have to find better ways to interact with customers and build long-term relationships. Unfortunately, because CRM systems have been driven by hardware and software companies, the communications dimension has been missing and the majority of CRM initiatives have been considered failures (Duncan 2005). CRM's unfortunate failures have probably hurt investment in IMC, which is ironic because IMC is a process that not only incorporates the data-driven qualities of CRM, but also has a strong focus on interactive communication—adding the human dimension to CRM (Kerr et al. 2008). In addition, it seems too much of existing research focuses on "how to do IMC" from a marketer's or company's view. The real question is what do customers get from IMC strategy (customer's view), and that seems to be the area where research focus is lacking (Schultz 2005). Thus, research should start with the customer's need and work back through the process or approach of IMC strategy, rather than starting with how and in what way we want to develop marketing programs. In other words, identifying appropriate IMC strategies based on consumer preference, while staying within the segmented markets and satisfying the desired marketing performance is indeed necessary.

Accordingly, the present article has two goals, that is, to contribute to the above incremental endeavor and to overcome the above challenges. The first goal, this study develops an evaluating model for selecting appropriate strategies of IMC based on the quality function deployment (QFD) approach and incorporates the fuzzy analytic hierarchy process (FAHP) method to solidify relationships with existing customers. The second goal, this study presents an empirical example and demonstrates the applicability of the proposed model to customer relationships of a department store, which same model can be applied to other service organizations. The QFD approach is able not only to translate the real needs of customers when building customer relationships into the strategies of IMC, but also provides a systematic planning tool for incorporating and coordinating information to make appropriate decisions about IMC strategy effectively and efficiently. In addition, the FAHP method can reduce imprecision and improve judgment when evaluating the preference of customer opinions about customer relationship management.

Marketing managers could anticipate the current and expected performance of customer relationship management and the effects of IMC strategy activities by employing the results of analytic evaluation. Thus, managers could apply the evaluating model described in this study to re-examine their IMC strategies and ensure that these strategies satisfy their customers' view in terms of customer relational benefit as well as the continual facilitation of strategic evolution. Moreover, this evaluating model may assist SC researchers in the comprehensive inspection of IMC strategies and further to compute appropriate activities of IMC strategy for improving the customer relational benefit.

2 Literature Review

The main purpose of the literature review in this study is to comprehend relative knowledge concerning CRM, customer relational benefit, and IMC strategy which are foundation of constructing computational hierarchy for later analysis. In addition, the QFD approach and FAHP method were introduced as followings.

2.1 Customer Relationship Management (CRM)

CRM is a strategic approach that is concerned with creating improved customer value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of relationship marketing strategies and IT to create profitable, long-term relationships with customers. CRM provides enhanced opportunities to use data and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications (Payne and Frow 2005). The main body of literature discusses concepts that are relevant to CRM, such as the influence of prior experience on future customer expectations, the different treatment of each customer, and the value of long-term relationships. Concurrently, marketing scholars turned their attention to the core capabilities of the firm that were necessary to develop and maintain good customer relationships (e.g., Vorhies and Neil 2005; Javalgi et al. 2006; Osarenkhoe and Bennani 2007; Wu et al. 2008). These studies emphasized the establishment of good information processes and capabilities within the firm to understand the needs and wants of customers, thus making firms more efficient and effective in managing customer relationships. In addition, companies began to focus on acquiring new customers; retaining their current customers (i.e., building long-term relationships); and enhancing these relationships through such activities as customized communications, cross-selling, and the segmentation of customers, depending on their value to the company (Payne and Frow 2005). Implementation of CRM solutions also requires firms to have a customer relational orientation (Jayachandran et al. 2005; Srinivasan and Moorman 2005) and to have processes in place to collect, analyze, and apply the acquired customer information (Jayachandran et al. 2005).

The value of the needs and wants of a customer that are fulfilled by the company draws on the concept of the benefits that enhance the customer offer (Levitt 1969; Lovelock 1995). These benefits can be integrated in the form of a value proposition that explains the relationship among the performance of the product, the fulfillment of the customer's needs, and the total cost to the customer over the customer relationship life cycle (Lanning 1998). In recent years, the studies concerning customer benefits received have gradually begun to focus on the consumer relational benefits and became the key issue for CRM. The relevant issues focus on investigating the interaction between customer relational benefit, customer satisfaction, commitment, trust, and customer loyalty (Hennig-Thurau et al. 2002; Marzo-Navarro et al. 2004; Colgate et al. 2005; Simon et al. 2005; Kinard and Capella 2006; Palmatier et al. 2007).

2.2 Customer Relational Benefit

The Customer relational benefit is defined as those benefits which customers receive from long-term relationships above and beyond the core service performance (Gwinner et al. 1998; Liljander and Roos 2002). Specifically, Gwinner et al. (1998) indicated that when consumers develop a long-term relational exchange relationship with a service provider, they perceive the heightening of three benefit types from maintaining that relationship, including confidence, social, and special treatment benefits. Additionally, their research determined that confidence benefits are consistently more important to consumers across various service typologies than social and special treatment benefits. According to several studies, confidence benefits will reduce anxiety levels associated with a service offering, increase perceived trust in the provider, diminish the perception of risk, and enhance knowledge of service expectations (Berry 1995; Bitner 1995; Hennig-Thurau et al. 2002). Consumers perceive social benefits from forging a long-term relationship with a service provider, such as personal recognition with employees, customer familiarity, and the development of a friendship with the service provider (Berry 1995; Gremler and Gwinner 2000). In addition, consumers may attain special treatment benefits from prolonged relationships, such as economic and customization benefits that they do not receive from other service providers (Gwinner et al. 1998; Reynolds and Beatty 1999). Kamakura et al. (2005) stated the customer relationship management research process using the customer life-cycle framework, and described the issues and methodological challenges unique to each stage. Grégoire and Fisher (2006) examined the effects of relationship quality on customers' desire to retaliate after service failures. According to the results of previous studies, three customer relational benefits included the social benefits, confidence benefits, and special treatment benefits are more targeted value to customers. Thus, this study utilized these three distinct benefit types proposed by Gwinner et al. (1998) to construct the relational benefit dimensions that represent the perceived value that customers attempt to acquire and, moreover, to develop several attributes of customer relational benefit under this dimension.

To determine whether the customer value proposition is likely to result in a superior customer experience, a company should undertake a value assessment to quantify the relative importance that customers place on the various relational benefit attributes. However, analytical tools such as conjoint analysis may also reveal substantial market segments with service needs that are not fully catered to by the attributes of existing offers (Payne and Frow 2005). In addition, Boulding et al. (2005) indicated that further studies concerning customer value should directly examine the link between CRM activities in a variety of other literature streams (e.g., IMC strategy) and customer value.

2.3 Integrated Marketing Communication Strategy

Literature on IMC reveals that some researchers recognized the notion of building relationships with customers in general terms. For example, Cathey and Schumanri (1996) stated that building a strong customer relationship is an important part of the "customer orientation" of IMC. Duncan (2002) suggested that all messages should create profitable customer relationships in the practice of IMC. Meanwhile, marketing communications practitioners and scholars in the field increasingly recognize the importance of building relationships with customers and retaining existing customers. Nowak and Siraj (1996) reported that the managers who practice IMC directed their communications efforts more toward existing customers in narrowly defined target markets, whereas those who do not practice IMC spent more communications efforts in attracting new customers in broadly defined targets. Lindberg-Repo and Grönroos (2004) present a framework that represents a strategic approach to managing service organizations' communication processes in order to maximize value generation in the relationship between the organization and the customer. Because the retention of existing customers is strategically important for the company's profitability, the marketing communication strategies should help the company to attain this goal by solidifying relationships with existing customers through communications activities. The company should not only use our proposed model to design its IMC strategies, but also implement specific activities to achieve that goal. Therefore, it is our view that an incremental, systematic and scientific model should be developed for determining IMC strategy toward customer relationship management. That is, the measurement of IMC strategy that researchers and practitioners consider relevant and important should be identified in the growing literature about IMC strategy and customer relationship management; at the same time, new or emerging views should also be identified and examined as the measurement and practice of IMC strategy continually evolves (Lee and Park 2007).

This study applies four major marketing activities often used in promotion programs as IMC strategies including advertising, personal sales, sales promotions, and direct marketing. (Kotler and Keller 2006): (1) Advertising is paid for, shows the sponsor's name, and allows for a non-personal presentation of ideas, goods, or services. Messages are usually conveyed through television, the Internet, magazines and other media to the target market (Sissors and Baron 1996; Rossiter and Bellman 1999); (2) Sales promotions utilize diverse short-term techniques to induce customer awareness, with the goal of interesting customers to purchase products or services. For short-term retail marketing, sales promotion is a powerful tool, tempting customers to make impulse purchases (Laroche et al. 2003; Honea and Dahl 2005); (3) In direct marketing, products/services are launched to the target market directly, through which there may be timely buying, selective contacts, savings of time, and an increase in convenience (Reardon and McCorkle 2002); (4) Personal selling is where the salespeople communicate with the customers in the target market. It has the advantages of two-way communication, sending sales messages to the customers, and ultimately, decreasing customer resistance. However, personal selling has small message coverage, and sometimes, the sales message may be inconsistent (Belch and Belch 2007). The use of IMC strategy combinations is usually based on marketing strategy. Using different IMC strategies for delivering messages to customers can result in varied responses.

2.4 Quality Function Deployment

Quality function deployment (QFD) is an integrated decision-making methodology that can ensure the elements of design and construction processes have all the requirements of a construction procedure and can improve on them as well (Yang et al. 2003). QFD can translate customers' values (also referred to as the voice of the customer, or VOC) into technical requirements, and this can lead to component characteristics, process steps and operational steps. Moreover, QFD utilizes a matrix to represent each of the translations, and four matrices can construct a House of Quality (HOQ) (Griffin 1992). A HOQ is made up of complex matrices and can provide the means for inter-functional planning and communications, as well as offer the specifications for product or service design through the relative-importance judgments of the voice of the customer (Hauser and Don 1988; Cohen 1995). The functions of QFD have been applied to diverse areas such as product development, quality management, design, decision-making, and strategic planning (Chan and Wu 2002). Three aspects unique to the QFD planning tool are as follows: (1) the complex interrelationships of inputs and outputs can be more easily understood through a planning matrix; (2) subsequent levels of requirements can be traced back to the VOC; (3) the competitive evaluation of a product or a service is based on quantitative analysis (Partovi 2001). For these reasons, QFD was gradually introduced into the service sector to develop quality service types or the appropriate strategies of services including professional services (Adiano 1998), engineering services (Pun et al. 2000), and government services (Lewis and Hartley 2001).

Various quantitative methods such as the analytic hierarchy process and analytic network process are combined with QFD and provide a more objective approach to evaluation. Fuzzy set theory may be used to improve the quality of the responsiveness for the requirements in the QFD process so as to mitigate the effects of the linguistic variables involving vagueness and imprecision (Kwong and Bai 2003). Further, Krcmar et al. (2001) proved that the QFD's application with fuzzy set theory is much more flexible than probability statistics with regard to vague and imprecise decisions. Partovi (2001) presented an analytical method for quantifying Heskett's strategic service vision based on QFD and used the analytic hierarchy process (AHP) to determine the intensity of the relationship between variables of each matrix. Hsu and Lin (2006a) presented a model that considers the attributes of customer value by means-end chain analysis and the utilization of fuzzy QFD and the entropy method to help structure the amount of information about customers' cognition. QFD requirements at each level of deployment can be tied back and coordinated with any decision regarding customers' opinions, and it is more easily understood by using a matrix format to show the complex interrelationships of requirement inputs and outputs (Griffin and Hauser 1993; Partovi 2001). However, the expression of customer relational benefits and relationship marketing strategies in the QFD approach with linguistic variables involves the quality of, vagueness of human cognition, and perception (Hsu and Lin 2006a). Therefore, it is inappropriate to use precise and numerical data when utilizing the QFD approach as a basis for developing an evaluation model of a relationship marketing strategy. To overcome this problem, the fuzzy analytic hierarchy process (FAHP) was integrated into the procedure of QFD for evaluating the relative weighting of customer relational benefits and IMC strategies. The FAHP method offers a systematic procedure for identifying and justifying the alternatives by applying the concept of fuzzy logic and multi-attribute decision-making (MADM) inherited from the traditional AHP method.

2.5 Fuzzy Analytic Hierarchy Process

The traditional AHP method decomposes a complex multi-criterion decision problem into a hierarchy and applies quantified judgments to permit decision makers to clearly analyze the problem, thus providing sufficient information to select the most suitable alternative. This method has been used in most applications related to different areas, including evaluation, allocation, selection, cost-benefit analysis, planning and development, priority and ranking, and decision-making (Crary et al.

2002; Badri 2001; Beynon 2002; Wei et al. 2005; Chandran et al. 2005). The AHP also provides a methodology to calibrate the numeric scale for the measurement of quantitative as well as qualitative performance (Omkarprasad and Sushil 2006). During the past decades, the AHP has been utilized to select, rank, evaluate, optimize, predict and benchmark decision alternatives (Chandran et al. 2005). Simultaneously, applications of this technique have been utilized in various areas, including Internet access technology selection (Malladi and Min 2005), production and distribution (Chan et al. 2005), evaluation of transport investment (Caliskan 2006), and facility layout design in manufacturing systems (Ertay et al. 2006). However, the traditional AHP seems inadequate for capturing customer values with linguistic expressions and accurately determining the relative importance of customers' needs (Kahraman et al. 2003). The AHP method is often criticized because of its use of an unbalanced scale of judgments and its inability to adequately handle inherent uncertainty and imprecision in the pairwise comparison process (Deng 1999). Hence, the fuzzy analytic hierarchy process (FAHP) was developed to overcome these deficiencies since decision-makers are usually more confident about giving interval judgments than fixed value judgments.

The fuzzy AHP method provides a systematic procedure for selecting and justifying the alternatives by using the concepts of fuzzy logic, fuzzy set theory, and hierarchical structure inherited from the traditional AHP method. The overall process of the fuzzy AHP is shown in Figure 1. Fuzzy logic is a system which uses approximate reasoning or estimation via numerical computations and symbolic manipulation to qualitatively tackle problems involving imprecision or uncertainty. Fuzzy set theory is primarily concerned with vagueness such as tends to characterize human judgments and perceptions (Beskese et al. 2004). Hsu and Lin (2006b) utilized fuzzy set theoretic techniques to analyze travel risk and develop the travel perceived risk averse strategy matrix. Hsu et al. (2009) integrate a fuzzy linguistic decision model with a genetic algorithm to extract the optimum promotion mix of a variety of tools for satisfying expected marketing performance within budget limitations. Moreover, the fuzzy AHP method is a popular approach for multiple criteria decision-making and has been widely used in the relevant literature. Chang (1996) applied triangular fuzzy numbers to construct the fuzzy pairwise comparison matrix in the AHP and used the extent analysis method to obtain the synthetic values of the pairwise comparisons. Sheu (2004) combined the fuzzy AHP with the fuzzy multi-attribute decision-making approach for identifying global logistics strategies. Kahraman et al. (2004) applied the fuzzy AHP to the comparison of catering firms via customer satisfaction. Chang's extent analysis method (Chang 1996) provides an easier way to construct a fuzzy reciprocal comparison matrix as well as derive the weight vectors for individual levels of the hierarchical requirements without weight overlapping, than do the other fuzzy AHP and traditional AHP approaches. In this study, Chang's extent analysis method (Chang 1996) is applied to evaluate the relative weight of customer relational benefit attributes for seeking appropriate IMC strategies.

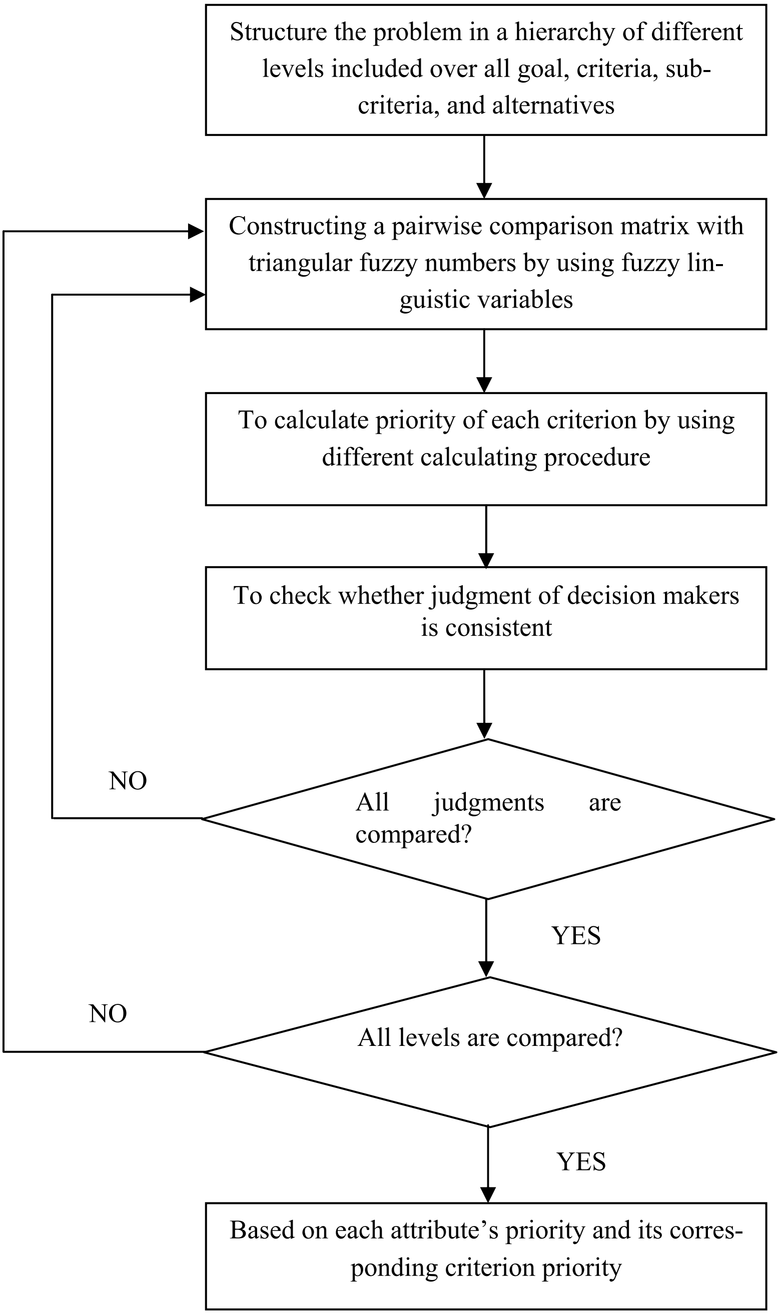


Fig. 1 The flowchart of the fuzzy AHP

3 Construction of an Evaluation Model for Selecting IMC Strategy on CRM

This study constructs the model through two phases in the deployment of the QFD system. In the first phase, market segmentation was implemented in order to express the voice of the customer in designing relational benefits in different customer segments. This study then applies the fuzzy AHP with the extent analysis method to systematically evaluate the relative-importance weighting among attributes of customer relational benefit in different segments (***Step 1*** to ***Step 4***). In the second phase, the relationship between customer relational benefit attributes and IMC strategies were estimated (including ***Step 5*** and ***Step 6***). The steps for developing the proposed model are shown in the following:

Step 1: Constructing the hierarchy of customer relational benefit

In this step, we first extract the attributes of customer relational benefit from early studies including Berry (1995), Gwinner et al. (1998), Gremler and Gwinner (2000), and Hennig-Thurau et al. (2002). Gwinner et al. (1998) identified three types of customer relational benefit including social benefits, confidence benefits, and special treatment benefits. These three core categories are viewed as the basis

Table 1 Definitions of categories and attributes for customer relational benefit

Categories	Definitions and attributes
Social benefits	Consumers perceive social benefits from forging a long-term relationship with a service provider, such as personal recognition with employees, customer familiarity, and the development of a friendship with the service provider (Berry 1995; Gremler and Gwinner 2000)
Confidence benefits	Confidence benefits will reduce anxiety levels associated with a service offering, increase perceived trust in the provider, diminish the perception of risk, and enhance knowledge of service expectations (Berry 1995; Hennig-Thurau et al. 2002)
Special treatment benefits	Consumers may attain special treatment benefits from prolonged relationships, such as economic and customization benefits that they do not receive from other service providers (Gwinner et al. 1998)

and constructed the attributes of customer relational benefit, which also helps understanding and evaluating the value of customer relationship. The brief explanations and definitions of three core categories for customer relational benefit are shown in Table 1 Moreover, this study modifies attributes by interviewing with senior managers of company in order to establish valid categories and attributes of customer relational benefit.

Step 2: Determining the importance degree of linguistic variables

Usually respondents respond questionnaires with imprecision and uncertainty. Thus, our questionnaire was designed by using fuzzy linguistic variables through the fuzzy set concept to avoid respondents' imprecision problem. Triangular fuzzy numbers can indicate the membership functions of the expression fuzzy linguistic variables. The triangular fuzzy numbers, $\tilde{1}$ to $\tilde{9}$, are utilized to improve the conventional nine-point scale and adopted to represent the respondents' voice in linguistic forms (including $\tilde{1}$ =Equally, $\tilde{3}$ =Moderately, $\tilde{5}$ =Strongly, $\tilde{7}$ =Very Strongly, and $\tilde{9}$ =Extremely) (Chan et al., 1999). The five triangular fuzzy numbers are defined with the corresponding membership functions as shown in Figure 2

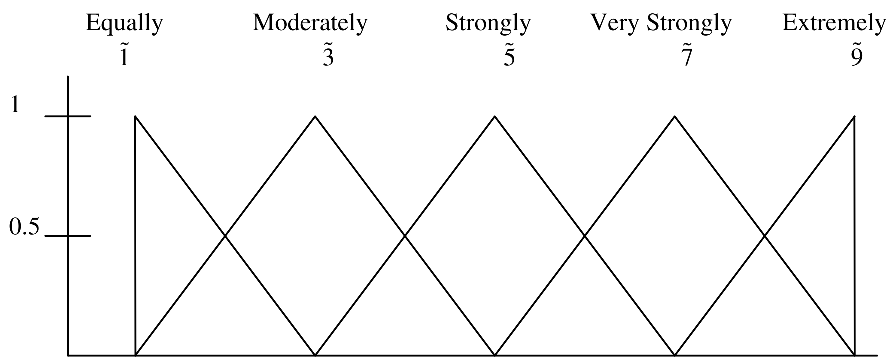


Fig. 2 The membership functions of triangular fuzzy numbers

Step 3: Developing fuzzy pairwise comparison matrices

In this step, all responses via the triangular fuzzy numbers will transform into fuzzy pairwise comparison matrices. If there are *n* categories of customer relational benefit (C_1, C_2, …, C_n), a fuzzy number can be defined as $\tilde{A} = (l, m, u)$. When one respondent evaluates *i* category is more important than *j* category for customer relational benefit, then the triangular fuzzy numbers $\tilde{a}_{ij} = \left(l_1, m_1, u_1\right)$ is displayed, and the reverse relationship is given by $\tilde{a}_{ji} = \left(1/u_1, 1/m_1, 1/l_1\right)$. Thus, this study utilizes triangular fuzzy number $\tilde{a}_{ij}$ to construct fuzzy comparison matrices for categories and attributes of customer relational benefit, as shown in Equation (3)

$$\tilde{A} = \begin{bmatrix} 1 & \tilde{a}_{12} & \tilde{a}_{13} & \cdots & \tilde{a}_{1(n-1)} & \tilde{a}_{1n} \\ \tilde{a}_{21} & 1 & \tilde{a}_{23} & \cdots & \tilde{a}_{2(n-1)} & \tilde{a}_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{a}_{(n-1)1} & \tilde{a}_{(n-1)2} & \tilde{a}_{(n-1)3} & \cdots & 1 & \tilde{a}_{(n-1)n} \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \tilde{a}_{n3} & \cdots & \tilde{a}_{n(n-1)} & 1 \end{bmatrix} \quad \text{where, } \tilde{a}_{ij} = \frac{1}{\tilde{a}_{ji}} \tag{1}$$

Step 4: Calculating relative importance weights of categories and attributes for customer relational benefit

This study adopts Chang's (1996) extent analysis approach and fuzzy comparative matrix to evaluate the weighting of customer relational benefit categories and attributes. First of all, let $X = \left\{x_1, x_2, \ldots, x_n\right\}$ be an object set, and $G = \left\{u_1, u_2, \ldots, u_m\right\}$ be a goal set. Then, each object is taken and extent analysis for each goal is performed respectively. Therefore, *m* extent analysis values for each object can be obtained, with the following signs:

$$\tilde{M}_{g_i}^1, \tilde{M}_{g_i}^2, \ldots, \tilde{M}_{g_i}^m \; ; \qquad \text{i=1, 2, …, n,}$$

where $\tilde{M}_{g_i}^j \; (j = 1, 2, ..., m)$ all are triangular fuzzy numbers. The steps of Chang's extent analysis method can be given as in the following:

(1) The value of fuzzy synthetic extent with respect to the *i* object is defined as

$$\tilde{M}_i = \sum_{j=1}^{m} \tilde{M}_{g_i}^j \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{M}_{g_i}^j \right]^{-1} \tag{2}$$

To obtain $\sum_{j=1}^{m} \tilde{M}_{g_i}^{j}$, the fuzzy addition operation of m extent analysis values for a particular matrix is performed, such that

$$\sum_{j=1}^{m} \tilde{M}_{g_i}^{j} = (\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j) \tag{3}$$

And to obtain $\left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{M}_{g_i}^{j} \right]^{-1}$, fuzzy addition operation of $\tilde{M}_{g_i}^{j} \, (j = 1, 2, ..., m)$ values is performed, such that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{M}_{g_i}^{j} = (\sum_{i=1}^{m} l_i, \sum_{i=1}^{m} m_i, \sum_{i=1}^{m} u_i) \tag{4}$$

and then the inverse of the vector above is computed, such that

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{M}_{g_i}^{j} \right]^{-1} = (\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i}) \tag{5}$$

(2) As $\tilde{M}_1 = (l_1, m_1, u_1)$ and $\tilde{M}_2 = (l_2, m_2, u_2)$ are two triangular fuzzy numbers, the degree of possibility of $\tilde{M}_2 = (l_2, m_2, u_2) \geq \tilde{M}_1 = (l_1, m_1, u_1)$ is defined as

$$V(\tilde{M}_2 \geq \tilde{M}_1) = \sup_{y \geq x} \left[\min(u_{\tilde{M}_1}(x), u_{\tilde{M}_2}(y)) \right] \tag{6}$$

and can be expressed as follows:

$$\begin{aligned} & V(\tilde{M}_2 \geq \tilde{M}_1) \\ & = hgt(\tilde{M}_1 \cap \tilde{M}_2) = u_{\tilde{M}_2}(d) \\ & = \begin{cases} 1, & if \quad m_2 \geq m_1, \\ 0, & if \quad l_1 \geq u_2, \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & otherwise, \end{cases} \end{aligned} \tag{7}$$

Figure 3 illustrates Equation (7) where d is the ordinate of the highest intersection point d between $u_{\tilde{M}_1}$ and $u_{\tilde{M}_2}$. To compare $\tilde{M}_2$ and $\tilde{M}_1$, both $V(\tilde{M}_1 \geq \tilde{M}_2)$ and $V(\tilde{M}_2 \geq \tilde{M}_1)$ values are needed preliminarily.

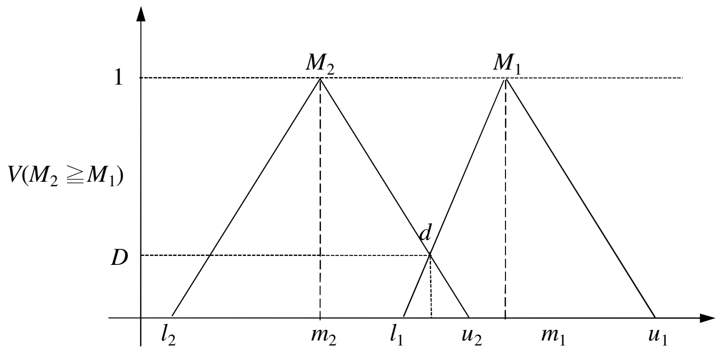


Fig. 3 The crossover of fuzzy numbers M_1 and M_2

(3) The degree of possibility for a convex fuzzy number to be greater than k convex fuzzy $\tilde{M}_i (i = 1, 2, ..., k)$ numbers can be defined by

$$\begin{aligned} & V(\tilde{M} \geq \tilde{M}_1, \tilde{M}_2, ..., \tilde{M}_k) \\ & = V\left[(\tilde{M} \geq \tilde{M}_1) \cap (\tilde{M} \geq \tilde{M}_2) \cap ... \cap (\tilde{M} \geq \tilde{M}_k)\right] \cdot \\ & = \min \ V(\tilde{M} \geq \tilde{M}_1), \qquad i = 1, 2, ..., k \end{aligned} \tag{8}$$

Assume that $d'(A_i) = \min \ V(\tilde{M}_i \geq \tilde{M}_k), k = 1, 2, ..., n; \ k \neq i$. Then the weight vector is given by

$$cW' = \left[d'(A_1), d'(A_2), ..., d'(A_n)\right]^T, \tag{9}$$

where $A_i (i = 1, 2, ..., n)$ are n elements.

(4) Ultimately, the normalized weight vectors of the kth level are derived after normalization as illustrated in Equation (10)

$$cW = \left[d(A_1), d(A_2), ..., d(A_n)\right]^T, \tag{10}$$

where cW is a non-fuzzy number and the relative importance weights of attributes for customer relational benefit.

According to the extent analysis method, the triangular fuzzy numbers illustrated in fuzzy pairwise comparison matrix can be converted into fuzzy synthetic values by using Equation (2), (3), (4), and (5) to calculate all responses for each category and attribute of customer relational benefit. The Equation (6), (7), and (8) are utilized to solve the problem of triangular fuzzy number overlapping (this problem will make no difference of relative importance weights for distinguish categories or attributes) which appeared in other Fuzzy AHP methods (Chang 1996; Kahraman et al. 2006). Finally, the normalized weight vectors for each category and attribute of customer relational benefit can be generated throughout Equation (9) and (10).

Step 5: Constructing the hierarchy of IMC strategy

In this step, we repeat similar process of ***Step 1*** (the difference is using strategies and activities of IMC instead of categories and attributes for customer relational benefit). We reviewed relative studies concerning IMC strategy to extract four types of IMC strategy including advertising strategy, sales promotion strategy, direct marketing strategy, and personal selling strategy (Rossiter and Bellman 1999; Reardon and McCorkle 2002; Honea and Dahl 2005; Kotler and Keller 2006; Belch and Belch 2007) to construct the hierarchy of IMC strategy. The brief explanations of four strategies are shown in Table 2. The comprehensive activities of four strategies are obtained by personal interviewing with senior marketing managers of company.

Table 2 Explanations of four IMC strategies

Strategies	Explanations
Advertising strategy	Advertising shows the sponsor's name, and allows for a non-personal presentation of ideas, goods, or services. Messages are usually conveyed through television, the Internet, magazines and other media to the target market (Rossiter and Bellman 1999).
Sales promotion strategy	Sales promotions utilize diverse short-term techniques to induce customer awareness, with the goal of interesting customers to purchase products or services (Honea and Dahl 2005).
Direct marketing strategy	Products/services are launched to the target market directly, through which there may be timely buying, selective contacts, savings of time, and an increase in convenience (Reardon and McCorkle 2002).
Personal selling strategy	Personal selling is where the salespeople communicate with the customers in the target market (Kotler and Keller 2006).

Step 6: Determine the relationship matrix between customer relational benefit and IMC strategy and completed an evaluation model

The relative importance weights of activities for IMC strategy can be obtained in this step. We create the fuzzy relationship matrix between the attributes of customer relational benefit and the managerial activities of IMC strategy evaluated by the senior managers through fuzzy linguistic variables. This fuzzy relationship matrix ($\tilde{R}_{ij}$) is constructed to relate the *i* attribute of customer relational benefit to the *j* activity of IMC strategy. The fuzzy relationship matrix is then defuzzified into a matrix (R_{ij}) with crisp values by using Equation (11) (Chen 1998).

$$M_{\_crisp} = \frac{(4m + l + u)}{6} \tag{11}$$

If there are *m* activities designated to meet *n* attributes, the relative importance weights of activities for IMC strategy are obtained through Equation (12)

$$u_j = \sum_{j=1}^{m} R_{ij} \cdot cW_i \tag{12}$$

where μ_j is the relative importance weights of the *j* activities for IMC strategy to develop customer relational benefit (j =1, 2, 3, …, n); cW_i denotes the relative importance weights of the *i* attribute for customer relational benefit; the relationship matrix (R_{ij}) is constructed to relate the *i* attribute of customer relational benefit to the *j* activity of IMC strategy. Then μ_j (j =1, 2, 3, …, n) were aggregated in order to identify the most important activities of IMC strategy. Ultimately, the evaluation model for selecting IMC strategy for CRM was completed.

4 Empirical Illustrations

The total revenue of the department store industry in southern Taiwan in 2008 was approximately US$700 million. The revenue of H Department Store was US$99.28 million (approximately 42.1% in total volume). The studied department store was the top department store in terms of its revenue. Moreover, H Department Store held a 40% market share in southern Taiwan while the other competitors had been gradually losing their customers. The extraordinaire performance indicates the outstanding customer relationship management and the excellent IMC strategies keeping existing customers that H Department Store executes. Thus, H Department Store is a good empirical example to demonstrate the evaluation model for our study.

4.1 The Hierarchy Construction of Customer Relational Benefit

Following ***Step 1*** of the modeling procedure, market segmentation is implemented and customer relational benefit attributes are acquired through the review of relevant studies and personal interviews with three senior marketing executives who work at H Department Store. In terms of market segmentations of H Department Store, three categories and nine attributes of customer relational benefit are settled upon according to the results of interviews with three floor supervisors who work at H Department Store over five years. The brief descriptions of categories and attributes of customer relational benefit are listed in Table 3. In addition, the main target customers of H Department Store are identified as 31- to 40-year-old females (70% of all customers) and the subordinate target customers are 21- to 30-year-old females (30% of all customers).

Table 3 Categories and attributes of customer relational benefit

Categories	Attributes	Descriptions
Social Benefits	Amicability	Service providers show every consideration to the loyal customers.
	Friendship	Service providers treat the loyal customers as friends when providing their service.
	Identification	Service providers could identify the loyal customers and know who they are.
Confidence Benefits	Anxiety Reduction	The loyal customers feel relieved and have decreased anxiety after receiving service.
	Trust	The loyal customers believe and trust service providers when they deliver their service.
Special Treatment Benefits	Price Discount	Service providers may give the loyal customers special discounts.
	Time Saving	Service providers might provide their service for the loyal customers more efficiently.
	Priority Service	The loyal customers will acquire first priority service.
	Customization	Service providers may provide personal or special service content to satisfy the loyal customers.

4.2 The Relative Importance Weights of Categories and Attributes for Customer Relational Benefit

According to ***Step 2***, ***3***, and ***4***, a matrix of aggregated customer relational benefit attributes and marketing segments is created (as listed in Table 5). Base on the senior marketing managers' suggestions, we interviewed 24 sales staffs with more than five years of work experience with luxury brands at H Department Store to receive frequent customers' voice. Questionnaire was designed as ***Step 2*** and distributed to the selected sales staffs in order to evaluate the relative weights of each category and attribute for customer relational benefit in different market segmentations. Regarding to how to obtain the value of relative weights, the calculating process was described as ***Step 3*** and ***4***.

Take the categories of customer relational benefit in the Age 21-30 market segmentation for example. According to ***Step 3***, we transformed 24 sales staffs' responses from fuzzy linguistic variables into triangular fuzzy numbers and aggregated those data into a fuzzy pairwise comparison matrix (as listed in Table 4) to evaluate the relative importance weights of categories for customer relational benefit in the Age 21-30 market segmentation.

Table 4 Fuzzy pairwise comparison matrix of the categories for customer relational benefit (Age 21-30 market segmentation)

Category of the customer relational benefit	Social Benefits	Confidence Benefits	Special Treatment Benefits
Social Benefits	(1, 1, 1)	(0.74, 1.10, 1.56)	(0.75, 1.09, 1.46)
Confidence Benefits	(0.64, 0.91, 1.35)	(1, 1, 1)	(0.68, 1.02, 1.40)
Special Treatment Benefits	(0.69, 0.92, 1.33)	(0.71, 0.98, 1.48)	(1, 1, 1)

Based on ***Step 4***, triangular fuzzy numbers, in the fuzzy pairwise comparison matrix are illustrated in Table 4, can be translated into fuzzy synthetic values by using Equation (2), (3), (4), and (5). The result of fuzzy synthetic values for each category of customer value is illustrated in Table 5.

Table 5 Fuzzy synthetic values of the categories for customer relational benefit (Age 21-30 market segmentation)

Category of the customer relational benefit	Fuzzy synthetic values
Social Benefits	(0.16, 0.28, 0.45)
Confidence Benefits	(0.16, 0.26, 0.43)
Special Treatment Benefits	(0.15, 0.23, 0.39)

According to Table 5, we found out that the membership functions of the fuzzy synthetic values for three categories are approximately equal (because three fuzzy synthetic values are very closer) and will consequently cause the problem of triangular fuzzy number overlapping. As a result, we utilized Equation (6), (7), and (8) to solve this problem. Meanwhile, the weight vector is generated by using Equation (9) and (10) to obtain the relative importance weights for each category of customer relational benefit illustrated in Table 6.

Table 6 The weight vector of initial and normalized relative importance weights for category of the customer relational benefit (Age 21-30 market segmentation)

Category of the customer relational benefit	Relative importance weights (Initial)	Relative importance weights (Normalized)
Social Benefits	1.000	0.281
Confidence Benefits	0.940	0.264
Special Treatment Benefits	0.834	0.234

The relative importance weights for each attribute of the customer relational benefit in Age 21-30 market segmentation can be conducted by using the same calculating process as well (as illustrated in the second column of Table 7). Moreover, we multiplied the relative importance weights for "category of the customer relational benefit" by the relative importance weights for "attribute of the customer relational benefit" to generate the incorporated relative importance weights for each attribute of the customer relational benefit in Age 21-30 market segmentation, illustrated in last column of Table 7, the same as numbers showed in the column 3 of Table 8. Regarding to the column 4 of Table 8, it demonstrates the relative importance weights for attributes of the customer relational benefit in the other market segmentation (Age 31-40) and the relative importance weights are resulted by duplicating the same computational process mentioned above.

Table 7 Relative importance weights of categories and attributes for the customer relational benefit (Age 21-30 market segmentation)

Category (*a*)		Attribute (*b*)		(*a*)×(*b*)
Social Benefits	(0.281)	Amicability	0.502	0.141
		Friendship	0.210	0.059
		Identification	0.288	0.081
Confidence Benefits	(0.26)	Anxiety reduction	0.319	0.084
		Trust	0.681	0.180
Special Treatment Benefits	(0.23)	Price discount	1.000	0.220
		Time saving	0.001	0.005
		Priority service	0.380	0.089
		Customization	0.351	0.082

In addition, the final relative importance weightings, as illustrated in the column 5 of Table 8, were gain as following principle:

Final relative importance weightings = Each attribute of customer relational benefit in Age 21-30 market segmentation × (Customers of Age 21-30/Total Customer) + Each attribute of customer relational benefit in Age 31-40 market segmentation × (Customers of Age 31-40/Total Customer).
(i.e., 0.117 = 0.141 × 30% + 0.107 × 70%).

Table 8 The relationship matrix between attributes of customer relational benefit and market segmentation

		Market segmentation		
1	**2**	**3**	**4**	**5**
Category of customer relational benefit	**Attribute of customer relational benefit**	**Age 21-30 (30%)**	**Age 31-40 (70%)**	**Final relative importance weightings**
Social benefits	Amicability	0.141	0.107	0.117
	Friendship	0.059	0.048	0.052
	Identification	0.081	0.128	0.114
Confidence benefits	Anxiety reduction	0.084	0.080	0.081
	Trust	0.180	0.189	0.186
Special treatment benefits	Price discount	0.220	0.074	0.118
	Time saving	0.005	0.100	0.070
	Priority service	0.089	0.119	0.110
	Customization	0.082	0.079	0.080

4.3 The Relationship Matrix between Customer Relational Benefit and IMC Strategy

Based on ***Step 5***, we reviewed relative studies and conducted personal interviews with three senior marketing executives and five senior marketing project managers who work at H Department Store over five years to acquire IMC strategies and activities. Nine activities of IMC strategy for H Department Store were obtained, including the followings:

(i) Advertising strategy—(1) advertising in shopping districts, including advertising of promotional programs, and increasing publicity for the shopping area; (2) news updates, including current news, Internet updates and D.M. issuing.

(ii) Sales promotion strategy—(3) promotional programs, including commercial advertisement allocation (television, newspaper, and magazine), sales popularization, the current season's product sales, last season's product promotion, etc.; (4) sales reports, including analysis of product sales, customer purchasing history, and outcomes of promotional programs, etc.

(iii) Direct marketing strategy—(5) VIP preference, including price givebacks, free gifts, free parking, exclusive price discounts, and so on; (6) VIP room, providing private space for VIP customers to rest; (7) establishing contact with customers, including sending birthday and greeting cards, sending promotional news via e-mail.

Table 9 The relationship matrix between attributes of customer relational benefit and activities of IMC strategy

		IMC strategy								
		Advertising strategy		**Sales promotion strategy**		**Direct marketing strategy**			**Personal selling strategy**	
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Categories of customer relational benefit	**Attributes of customer relational benefit**	Advertising in shopping district	News update	Promotional program	Sales report	VIP preference	Establishing contact with customers	VIP room	Service attitude	Response to customers
Social benefits	Amicability	4.43 / 0.52	5.29 / 0.62	4.57 / 0.53	4.29 / 0.50	4.57 / 0.53	6.43 / 0.75	6.14 / 0.72	7.00 / 0.82	6.43 / 0.75
	Friendship	2.86 / 0.15	4.29 / 0.22	4.29 / 0.22	4.57 / 0.24	4.00 / 0.21	6.43 / 0.22	6..43 / 0.33	6.71 / 0.35	5.29 / 0.27
	Identification	1.71 / 0.20	4.00 / 0.46	3.08 / 0.44	4.00 / 0.46	3.86 / 0.44	7.00 / 0.80	6.43 / 0.73	5.43 / 0.62	4.00 / 0.46
Confidence benefits	Anxiety reduction	2.71 / 0.22	2.71 / 0.22	4.29 / 0.35	4.29 / 0.35	4.00 / 0.33	6.14 / 0.50	4.57 / 0.37	6.71 / 0.55	6.43 / 0.52
	trust	2.71 / 0.51	4.00 / 0.75	4.86 / 0.90	4.57 / 0.85	4.29 / 0.80	5.86 / 1.09	4.86 / 0.90	7.00 / 1.30	6.43 / 1.20
Special treatment benefits	Price discount	1.57 / 0.19	4.71 / 0.56	5.29 / 0.62	4.00 / 0.47	6.14 / 0.73	4.43 / 0.52	1.71 / 0.20	3.71 / 0.44	2.57 / 0.30
	Time saving	1.14 / 0.08	4.57 / 0.32	4.29 / 0.30	4.00 / 0.28	4.00 / 0.28	5.57 / 0.39	2.71 / 0.19	4.57 / 0.32	3.57 / 0.25
	Priority service	1.29 / 0.14	4.29 / 0.47	4.57 / 0.50	5.29 / 0.58	4.86 / 0.53	6.43 / 0.71	6.14 / 0.67	6.71 / 0.74	5.14 / 0.56
	Customization	1.29 / 0.10	5.00 / 0.40	4.57 / 0.37	5.86 / 0.47	4.00 / 0.32	6.43 / 0.51	6.43 / 0.51	6.43 / 0.51	4.86 / 0.39

(iv) Personal selling strategy—(8) service attitudes, including improved attitudes of customer service personnel and counter service; (9) responses to customers, including an institutional alliance service center, a customer complaint handling system, and customer service personnel response.

Following ***Step 6***, the relationship matrix that aggregates the attributes of customer relational benefit and activities of IMC strategy is created, and the relative importance weights of this matrix for activities of IMC strategy are also evaluated (as illustrated in Table 9). The matrix of the relationship between customer relational benefit and IMC strategy was formulated by middle-managers with more than eight years of work experience at H Department Store. The values representing

Table 10 The evaluation model for selecting IMC strategy on CRM

		IMC strategy									Market segmentation		
		Advertising strategy		**Sales promotion strategy**		**Direct marketing strategy**			**Personal selling strategy**				
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Categories of customer relational benefit	**Attributes of customer relational benefit**	Advertising in shopping district	News update	Promotional program	Sales report	VIP preference	Establishing contact with customers	VIP room	Service attitude	Response to customers	**Age 21-30**	**Age 31-40**	**Final relative importance weighting**
Social benefits	**Amicability**	4.43 / 0.52	5.29 / 0.62	4.57 / 0.53	4.29 / 0.50	4.57 / 0.53	6.43 / 0.75	6.14 / 0.72	7.00 / 0.82	6.43 / 0.75	0.141	0.107	0.117
	Friendship	2.86 / 0.15	4.29 / 0.22	4.29 / 0.22	4.57 / 0.24	4.00 / 0.21	6.43 / 0.22	6..43 / 0.33	6.71 / 0.35	5.29 / 0.27	0.059	0.048	0.052
	Identification	1.71 / 0.20	4.00 / 0.46	3.08 / 0.44	4.00 / 0.46	3.86 / 0.44	7.00 / 0.80	6.43 / 0.73	5.43 / 0.62	4.00 / 0.46	0.081	0.128	0.114
Confidence benefits	**Anxiety reduction**	2.71 / 0.22	2.71 / 0.22	4.29 / 0.35	4.29 / 0.35	4.00 / 0.33	6.14 / 0.50	4.57 / 0.37	6.71 / 0.55	6.43 / 0.52	0.084	0.080	0.081
	Trust	2.71 / 0.51	4.00 / 0.75	4.86 / 0.90	4.57 / 0.85	4.29 / 0.80	5.86 / 1.09	4.86 / 0.90	7.00 / 1.30	6.43 / 1.20	0.180	0.189	0.186
Special treatment benefits	**Price discount**	1.57 / 0.19	4.71 / 0.56	5.29 / 0.62	4.00 / 0.47	6.14 / 0.73	4.43 / 0.52	1.71 / 0.20	3.71 / 0.44	2.57 / 0.30	0.220	0.074	0.118
	Time saving	1.14 / 0.08	4.57 / 0.32	4.29 / 0.30	4.00 / 0.28	4.00 / 0.28	5.57 / 0.39	2.71 / 0.19	4.57 / 0.32	3.57 / 0.25	0.005	0.100	0.070
	Priority service	1.29 / 0.14	4.29 / 0.47	4.57 / 0.50	5.29 / 0.58	4.86 / 0.53	6.43 / 0.71	6.14 / 0.67	6.71 / 0.74	5.14 / 0.56	0.089	0.119	0.110
	Customization	1.29 / 0.10	5.00 / 0.40	4.57 / 0.37	5.86 / 0.47	4.00 / 0.32	6.43 / 0.51	6.43 / 0.51	6.43 / 0.51	4.86 / 0.39	0.082	0.079	0.080
	Aggregated relative importance weightings of activities for IMC strategy	2.19	4.36	4.57	4.60	4.45	6.11	5.01	6.06	5.06			

relationship between attributes of customer relational benefit and activities of IMC strategy are calculated by using Equation (11), and showed in the left-top cells surrounded by a thick-lined rectangle in Table 9. Moreover, the overall relationship values listed in the bottom-right of the cells surrounded by a thick-lined rectangle in Table 9 are calculated by using Equation (12) (The values listed in the left-top cells are multiplied by the final relative importance weightings illustrated in column 5 of Table 8). The values of overall relationship also represented the relative importance weights of activities for IMC strategy.

4.4 The Completed Evaluation Model for Selecting IMC Strategy on CRM

The completed evaluation model is accomplished after aggregating the relative importance weights of activities for IMC strategy (The values are listed in the bottom-right of the cells surrounded by a thick-lined rectangle in Table 9) which are illustrated in the last row of Table 10. It can be seen from the matrix that all market segmentations of customers viewed "trust" as the most important attribute of customer relational benefit. Therefore, service employees of H Department Store should act professionally when they give service to customers and strengthen the attribute of "trust." Service staff at H Department Store should provide exclusive price discounts, free gifts, or free parking to achieve the purpose of the "price discount" attribute. In addition, H Department Store should concentrate on establishing several contacts to enhance the "identification" attribute. It could emphasize the establishment of several approaches to contact customers. For example, H Department Store could send greeting cards to specific customers and show them that service staff can recognize them.

Thus, H Department Store should strengthen its capabilities in order to create core strategies and activities for trust and amicability. Besides, the VIP preference of direct marketing strategy can improve the value of price discounts because of the strong relationship between strategy and the attributes of customer relational benefit.

5 Conclusion

Following the results of the analysis, H Department Store should pay special attention to the trust and price discount attributes. Hence, the management team determined the expected goal for each customer relational benefit attribute following the internal analysis to secure leading edge IMC strategy. The management team has put special emphasis on the attitude of service, establishing contact with and responding to customers, and the improvement of customer relational benefit in order to obtain a competitive advantage for IMC strategies and activities. H Department Store should allocate major resources to upgrading the values of friendship and identification, since customers are not satisfied with these two attributes. To develop the representative IMC strategy, resources should be assigned to the elements of customer relational benefit with the highest performance.

The "price discount" is an important attribute of customer relational benefit but it also needs to be improved. Thus, H Department Store should concentrate on upgrading the price discount capability to gain a competitive advantage. From the viewpoint of the managers, a strong relationship exists between price discount and sales promotion strategy, and direct marketing strategy including promotional programs and VIP preference. Hence, not only can improving the sales promotion and direct marketing strategy performance upgrade the price discount capability, but it can also obtain a superior customer relational benefit among customers.

The key initiatives for H Department store include (1) allocating major resources to upgrade the values of the attributes "trust" and "price discount"; (2) enhancing the "price discount," "identification," and "trust" attributes to boost customer satisfaction; (3) concentrating on upgrading sales promotion strategy and direct marketing strategy, including promotional programs and VIP preference, to gain a competitive advantage and retain customers; (4) mastering the personal selling strategy and direct marketing strategy, such as the attitude of service and establishing contact with customers to create a competitive advantage; and (5) upgrading the service attitude and responding to customers efficiently to improve the "trust" attribute.

6 Discussion and Implications

The evaluating model helps to identify the important attributes of customer relational benefit. The important attributes are then improved through activities of IMC strategy. These strategies and activities which were selected for improving the customer relational benefit can be validated using a set of matrices and working backwards from customer wants. The example presented in this paper has demonstrated the applicability of the model to a department's customer relationships, and the same model can be applied to other service organizations. Managers could anticipate the current and expected performance of customer relational benefit attributes and the effects of IMC strategy activities by employing the results of competitive evaluation. Thus, managers could apply the evaluating model described in this paper to re-examine their IMC strategies and ensure that these strategies satisfy their customers' needs in terms of relational benefits. Moreover, this evaluating model could assist a company in the comprehensive inspection of IMC strategies and the continual facilitation of strategic evolution.

There are several unique features about this proposed model. First, this model provides an effective and efficient tool to complement the IMC strategies. This proposed model does not eliminate subjectivity completely, but that is not an attainable or desirable end result. The advantage of this model is that it adds quantitative precision and fine-tuning to an otherwise qualitative decision-making process. Second, this proposed model not only incorporates marketing segmentation, attributes of customer relational benefit, and various techniques of IMC strategy, but also coordinates them through QFD matrices and suggests solutions through the result of interconnected matrices. These matrices, which can be formatted on a spreadsheet, allow the decision maker to examine the sensitivity of relationship marketing strategies with respect to changes in the market mix and its corresponding customer wants, as well as changes in the strengths and weaknesses of the competition. Finally, the example of the empirical case provided in this paper demonstrates the applicability and ease of use of the model with different service organizations.

References

Adiano, C.: Lawyers use QFD to gain a competitive edge. Quality Progress 31(5), 88–89 (1998)

Badri, M.A.: A combined AHP-GP model for quality control systems. International Journal of Production Economics 72(1), 27–40 (2001)

Berry, L.L.: Relationship marketing of services-growing interest, emerging perspectives. Academy of Marketing Science 23(4), 236–245 (1995)

Beskese, A., Kahraman, C., Irani, Z.: Quantification of flexibility in advanced manufacturing systems using fuzzy concepts. International Journal of Production Economics 89, 45–56 (2004)

Beynon, M.: DS/AHP method: A mathematical analysis, including an understanding of uncertainty. European Journal of Operational Research 140(1), 148–164 (2002)

Bitner, M.J.: Building service relationships: It's all about promises. Journal of the Academy of Marketing Science 23, 246–251 (1995)

Belch, G.E., Belch, M.A.: Adverting and Promotion: An Integrated Marketing Communications Perspective, 7th edn. McGraw-Hill/Irwin, New York (2007)

Boulding, W., Staelin, R., Ehret, M., Johnston, W.J.: A customer relationship management roadmap: what is known, potential pitfalls, and where to go. Journal of Marketing 69(4), 155–166 (2005)

Caliskan, N.: A decision support approach for the evaluation of transport investment alternatives. European Journal of Operation Research 175(3), 1696–1704 (2006)

Cathey, A., Schumanri, D.W.: Integrated marketing communications: Construct development and foundations for research. In: Proceedings of the 1996 Conference of the American Academy of Advertising (1996)

Chan, L.K., Wu, M.L.: Quality function deployment: A literature review. European Journal of Operation Research 143(3), 463–497 (2002)

Chan Felix, T.S., Chung, S.H., Wadhwa, S.: A hybrid genetic algorithm for production and distribution. Omega 33(4), 345–355 (2005)

Chang, D.Y.: Application of the extent analysis method on fuzzy AHP. European Journal of Operational Research 95, 649–655 (1996)

Chandran, B., Golden, B., Wasil, E.: Linear programming models for estimating weights in the analytic hierarchy process. Computer and Operations Research 32, 2235–2254 (2005)

Cohen, L.: Quality function deployment. How to make QFD work for you. Addison-Wesley Publishing Company, Reading (1995)

Colgate, M., Buchanan-Oliver, M., Elmsly, R.: Relationship benefits in an internet environment. Managing Service Quality 15(5), 426–436 (2005)

Crary, M., Nozick, L.K., Whitaker, L.R.: Sizing the US destroyer fleet. European Journal of Operational Research 136(3), 680–695 (2002)

Deng, H.: Multicriteria analysis with fuzzy pair-wise comparison. International Journal of Approximate Reasoning 21, 15–231 (1999)

Duncan, T.R.: IMC: Using Advertising and Promotion to Build Brands, International Edition. The McGraw-Hill, New York (2002)
Duncan, T.R.: IMC in industry: More talk than walk. Journal of Advertising 34(4), 5–6 (2005)
Ertay, T., Ruan, D., Tuzkaya, U.R.: Integrating data envelopment analysis and analytic hierarchy for the facility layout design in Manufacturing systems. Information Sciences 176(3), 237–262 (2006)
Griffin, A.: Evaluating QFD's use in US firms as a process for developing products. Journal of Product Innovation Management 9(2), 171–187 (1992)
Gremler, D.D., Gwinner, K.P.: Customer-employee rapport in service relationships. Journal of Service Research 3, 82–104 (2000)
Grégoire, Y., Fisher, R.: The Effects of Relationship Quality on Customer Retaliation. Marketing Letters 17(1), 31–46 (2006)
Gwinner, K.P., Grener, D.D., Bitner, M.J.: Relational benefit in services industries: The customer's perspective. Journal of the Academy of Marketing Science 26, 101–114 (1998)
Hauser, J.R., Don, C.: The house of quality. Harvard Business Review 66, 63–73 (1988)
Hennig-Thurau, T., Gwinner, K.P., Gremler, D.D.: Understanding relationship marketing outcomes: An integration of relational benefits and relationship quality. Journal of Service Research 4(3), 230–247 (2002)
Honea, H., Dahl, D.W.: The promotion affect sale: defining the affective dimensions of promotion. Journal of Business Research 58, 543–551 (2005)
Hsu, T.H., Lin, L.Z.: QFD with Fuzzy and Entropy Weight for Evaluating Retail Customer Values. Total Quality Management 17(7), 935–958 (2006a)
Hsu, T.H., Lin, L.Z.: Using fuzzy set theoretic techniques to analyze travel risk: An empirical study. Tourism Management 27(5), 968–981 (2006b)
Hsu, T.H., Tsai, T.N., Chiang, P.L.: Selection of the optimum promotion mix by integrating a fuzzy linguistic decision model with genetic algorithms. Information Sciences 179, 41–52 (2009)
Javalgi, R.G., Martin, C.L., Young, R.B.: Marketing research, market orientation and customer relationship management: A framework and implications for service providers. The Journal of Services Marketing 20, 12–23 (2006)
Jayachandran, S., Subhash, S., Peter, K., Pushkala, R.: The role of relational information processes and technology use in customer relationship management. Journal of Marketing 69(4), 177–192 (2005)
Kahraman, C., Cebeci, U., Ruan, D.: Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey. International Journal of Production Economics 87(2), 171–184 (2004)
Kahraman, C., Cebeci, U., Ulukan, Z.: Multi-criteria supplier selection using fuzzy AHP. Logistics Information Management 16(6), 382–394 (2003)
Kahraman, C., Tijen, E., Gülçin, B.: A fuzzy optimization model for QFD planning process using analytic network approach. European Journal of Operational Research 171(2), 390–411 (2006)
Kamakura, W., Mela, C., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., Naik, P., Neslin, S., Sun, B., Verhoef, P., Wedel, M., Wilcox, R.: Choice Models and Customer Relationship Management. Marketing Letters 16(3/4), 279–291 (2005)
Kerr, G., Schultz, D., Patti, C., Kim, I.: An inside-out approach to integrated marketing communication. International Journal of Advertising 27(4), 511–548 (2008)

Kinard, B.R., Capella, M.L.: Relationship marketing: the influence of consumer involvement on perceived service benefits. Journal of Service Marketing 20(6), 359–368 (2006)

Kotler, P., Armstrong, G.: Principles of Marketing, 9th edn. Prentice Hall, Englewood Cliffs (2005)

Kotler, P., Keller, K.L.: Marketing Management, 12th edn. Prentice Hall, NY (2006)

Krcmar, E., Stennes, B., van Kooten, G.C., Vertinsky, I.: Carbon sequestration and land management under uncertainty. European Journal of Operational Research 135(3), 616–629 (2001)

Kwong, C.K., Bai, H.: Determining the importance weights for the customer requirements in QFD using a fuzzy AHP with an extent analysis approach. Journal of Intelligent Manufacturing 35, 619–626 (2003)

Lanning, M.J.: Delivering Profitable Value. Perseus Publishing, Cambridge (1998)

Laroche, M., Pons, F., Zgolli, N., Cervellon, M.-C., Kim, C.: A mode of consumer response to two retail sales promotion techniques. Journal of Business Research 56, 513–522 (2003)

Lee, D.H., Park, C.W.: Conceptualization and measurement of multidimensionality of integrated marketing communications. Journal of Advertising Research 47(3), 222–236 (2007)

Lewis, M., Hartley, J.: Evolving forms of quality management in local government: Lessons from the best value pilot programme. Policy and Politics 29(4), 477–496 (2001)

Levitt, T.: The Marketing Mode: Pathways to Corporate Growth. McGraw-Hill, New York (1969)

Liljander, V., Roos, I.: Customer-relationship levels: from spurious to true relationships. Journal of Services Marketing 16(7), 593–614 (2002)

Lindberg-Repo, K., Grönroos, C.: Conceptualising communications strategy from a relational perspective. Industrial Marketing Management 33, 229–239 (2004)

Lin, L.Z., Hsu, T.H.: The qualitative and quantitative models for performance measurement systems: The agile service development. Quality & Quantity 42, 445–476 (2008)

Lovelock, C.: Competing on Service: Technology and Teamwork in Supplementary Services. Planning Review 23, 32–39 (1995)

Malladi, S., Min, K.J.: Decision support models for the selection of internet access technologies in rural communities. Telematics and Informatics 22(3), 201–219 (2005)

Marzo-Navarro, M., Pedraja-Iglesias, M., Rivera-Torres, M.P.: The benefits of relationship marketing for the consumer and for the fashion retailers. Journal of Fashion Marketing and Management 8(4), 425–436 (2004)

Martensen, A., Grønholdt, L., Bendtsen, L., Jensen, M.J.: Application of a model for the effectiveness of event marketing. Journal of Advertising Research 47(3), 283–301 (2007)

Mort, G.S., Drennan, J.: Mobile communications: A study of factors influencing consumer use of m-services. Journal of Advertising Research 47(3), 302–312 (2007)

Omkarprasad, S.V., Sushil, K.: Analytic hierarchy process: An overview of application. European Journal of Operational Research 169, 1–29 (2006)

Osarenkhoe, A., Bennani, A.-E.: An exploratory study of implementation of customer relationship management strategy. Business Process Management Journal 13(1), 139–164 (2007)

Nowak, G., Siraj, K.: Is Integrated marketing communications really affecting advertising and promotion? An exploratory study on national marketers promotional practices. In: Proceedings of the 1996 Conference of American Academy of Advertising (1996)

Palmatier, R.W., Scheer, L.K., Steenkamp, J.-B.E.M.: Customer loyalty to whom? managing the benefits and risks of salesperson-owned loyalty. Journal of Marketing Research 44(2), 185–199 (2007)

Partovi, F.Y.: An analytic model to quantify strategic service vision. International Journal of Service Industry Management 12(5), 476–499 (2001)

Payne, A., Frow, P.: A strategic framework for customer relationship management. Journal of Marketing 69(4), 167–176 (2005)

Pun, K.F., Chin, K.S., Lau, H.: A QFD/hoshin approach for service quality deployment: A case study. Managing Service Quality 10(3), 156–170 (2000)

Puntoni, S., Tavassoli, N.T.: Social context and advertising memory. Journal of Marketing Research 44(2), 284–296 (2007)

Reardon, J., McCorkle, D.E.: A consumer model for channel switching behavior. International Journal of Retail and Distribution Management 30(4), 179–185 (2002)

Reid, M.: Performance auditing of integrated marketing communication (IMC) actions and outcomes. Journal of Advertising 34(4), 41–54 (2005)

Reynolds, K.E., Beatty, S.E.: Customer benefits and company consequences of customer-salesperson relationships in retailing. Journal of Retailing 75(1), 11–32 (1999)

Reynolds, T.J.: Methodological and Strategy Development Implications of Decision Segmentation. Journal of Advertising Research 46(4), 445–461 (2006)

Rossiter, J.R., Bellman, S.: A proposed model for explaining and measuring web Ad effectiveness. Journal of Current Issues and Research in Advertising 21, 13–31 (1999)

Schultz, D.: IMC research must focus more on outcomes. Journal of Advertising 34(4), 5–9 (2005)

Sheu, J.B.: A hybrid fuzzy-based approach for identifying global logistics strategies. Transportation Research, Part E: Logistics and Transportation Review 40, 39–61 (2004)

Simon, J.B., Seigyoung, A., Karen, S.: Customer relationship dynamics: service quality and customer loyalty in the context of varying levels of customer expertise and switching costs. Academy of Marketing Science 33(2), 169–183 (2005)

Sissors, J.Z., Baron, R.: Advertising Media Planning. NTC Business Books, Lincolnwood (1996)

Srinivasan, R., Christine, M.: Strategic firm commitments and rewards for customer relationship management in online retailing. Journal of Marketing 69(4), 193–200 (2005)

Sun, B.: Technology innovation and implications for customer relationship management. Marketing Science 25(6), 594–598 (2006)

Tauder, A.R.: Getting Ready for the Next Generation of Marketing Communications. Journal of Advertising Research 45(1), 5–8 (2005)

Vorhies, D.W., Neil, A.M.: Benchmarking marketing capabilities for sustainable competitive advantage. Journal of Marketing 69(1), 80–94 (2005)

Wei, C.C., Chien, C.F., Wang, M.-J.J.: An AHP-based approach to ERP system selection. International Journal of Production Economics 96, 47–62 (2005)

Wu, W.Y., Shih, H.A., Chan, H.C.: A study of customer relationship management activities and marketing tactics for hypermarkets on membership behavior. The Business Review, Cambridge 10(1), 89–95 (2008)

Yang, Y.Q., Shou, Q.W., Mohammad, D., Sui, P.L.: A fuzzy quality function deployment system for buildable design decision-making. Automation in Construction 12, 381–393 (2003)

Direct Marketing Based on a Distributed Intelligent System

Virgilio López Morales* and Omar López Ortega

Universidad Autónoma del Estado de Hidalgo, CITIS,
Carr. Pachuca-Tulancingo Km. 4.5, 42184,
Pachuca, Hgo., México
Tel.: (+52)7717172197
e-mail: {virgilio,lopezo}@uaeh.edu.mx

Abstract. Within a more globalized and inter-connected world, it becomes necessary to optimize resources for locating final products to target market segments. Direct Marketing has benefited from computational methods to model consumer preferences, and many companies are beginning to explore this strategy to interact with customers. Nevertheless, it is still an open problem how to formulate, distribute and apply surveys to clients, and then gather their responses to determine tendencies in customers' preferences. In this paper we propose a distributed intelligent system as a technological innovation in this subject. Our main goal is to reach final consumers and correlate preferences by using an approach that combines Fuzzy-C Means and the Analytic Hierarchy Process. A Multi Agent System is used to support the definition of survey parameters, the survey itself and the intelligent processing of clients' judgements. Clusters are synthesized after processing customers preferences and they represent a useful tool to analyze their preferences towards products' features.

Keywords: Direct Marketing, Fuzzy Clustering, Analytic Hierarchy Process, Multi Agent System.

1 Introduction

Many companies obtain feedback through surveys that are sent to potential customers either by regular or electronic mail. The current globalized and interconnected economy makes it compulsory to expend less energy, time and resources to drive the right products to the final consumers. Even though such forms of contact have been useful so far, the type of economy that surges in an interconnected world demands the interaction of business systems analysts, database developers, statisticians, graphic designers and client service professionals [1].

* Corresponding author.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 255–271.
springerlink.com

More companies are exploring strategies such as Costumer Relationship Management or Direct Marketing for reducing costs and increase profitability by acquiring information directly from data sources. A recent survey indicates that the following issues are the top three executive concerns: Customer satisfaction, customer retention, and marketing return of investment. This is so because they are undoubtedly critical to current rapidly evolving marketing tactics: Web 2.0 (19.4 percent), Social Networking (12.2 percent), and Social Media (11.3 percent) (Cf. [2]). Given the fact that Information Technologies (IT) are playing a major role to interact with clients, customer-specific information can be collected and used for analyzing markets, and drive promotion campaigns based on such analysis [3].

Many techniques have been applied to select target markets in commercial applications, such as statistical regression [4], regression trees [5], neuronal computing, [6, 7], fuzzy clustering and the called Recency, Frequency, and Monetary Value (RMF) variable [8, 9, 10]. On the other hand, Web sites in combination with IT's have become an appealing and world-wide media to final users: When all pretense of limiting commercial use was removed in 1995 when the National Science Foundation ended its sponsorship of the Internet backbone, marketers employed this powerful medium, and Internet commerce was born [11]. The impact of electronic markets on a firm's product and marketing strategies have been examined empirically by [12] and [13]. The impact on price of reduced buyer search cost, allocation efficiency, and different incentives to invest in electronic markets are examined in [14]. In [15] it is analyzed the competition between conventional retailers and direct marketers. Even though such techniques have been valid, paradigms such as Multi-Agent Systems (MAS) and clustering provide useful techniques to improve business intelligence by facilitating management interaction with customers subjective judgements.

Therefore, we explore the combination of soft-computing algorithms to interact with clients. We propose the usage of MAS, the Analytic Hierarchy Process (AHP) [16] and the Fuzzy C-Means ([17]) to define survey's parameters, distribute such criteria to point sales, gather customers judgements, and obtain the pattern of clients' preferences.

More specifically, our system consists of the following modules. The module that is used to define survey's criteria resides at the management's site. It is also employed to publicize the survey to point-sales. Point-sales, which are located in different regions, possess an evaluation module that helps collecting customers' judgements on an evaluation sheet. Raw data is stored in an evaluation blackboard residing at the management side. A third module is in charge of processing the evaluations provided by customers. The processing of raw data is carried by combining Fuzzy C-Means and the AHP. Fuzzy C-Means contributes with a classification of similar families of customers, while the AHP offers the final ranking of products based of the clusters that are synthesized. Altogether, the distributed and intelligent system that we proposed is useful to elucidate the patterns associated with a given market segment.

This paper has the following structure. In Section 2, we delineate broadly how to integrate the Analytic Hierarchy Process and the Fuzzy C-Means to Direct Marketing. Section 3 formally describes the AHP, Fuzzy C-Means and the algorithm we developed to merge both techniques. Section 4 describes the Distributed Intelligent System structure and dynamics. Experimental results are depicted in Section 5. Finally, conclusions and future work are presented in Section 6.

2 Formation of Clusters to Boost Direct Marketing

As we stated previously, one major issue related to direct marketing is how to process a (normally large) number of clients' evaluations of products. The Analytical Hierarchy Process (AHP) ([16]) is employed for ranking a finite set of m alternatives, which are evaluated (subjectively) over a finite set of p evaluation criteria.

The AHP is suitable for processing surveys because, on the one hand, it allows management to define what set of products are to be evaluated along with the set of evaluation criteria. However, the AHP was originally devised for individual judgements. When it comes to be used as a tool for group decision making, it surges the question of how to process every individual evaluation. Our solution is explained next.

When the size of a market segment is established, customers are required to complete the evaluation sheet of the system we developed. Such evaluation complies to the structure of the AHP. That is to say, each client must evaluate a set of the company products (alternatives) by judging their relevant features (criteria). So far, so good. Nevertheless, management confronts a large number of raw data in order to elucidate how the company products are evaluated by the given market segment.

Let us suppose the market segment consists of z individuals. A matrix can be formed in order to compare criteria on a pairwise basis, as evaluated by each individual. This matrix is called Pairwise Comparison Matrix (**PCM**). Therefore, management will be forced to process z **PCM**'s. More specifically, all such matrices must be treated mathematically to obtain a value that truly reflects the likes and dislikes of the market segment.

The Fuzzy C-Means algorithm (FCM) is then applied to values of the **PCM** in order to define the largest cluster and its corresponding centroid. Thus, FCM yields a centroid for each entry of **PCM**, representing the most preferred value (tendency) of the group. Each global value is entered to the Global Pairwise Comparison Matrix $\mathbf{PCM^G}$. When matrix $\mathbf{PCM^G}$ is completed, the AHP is executed as if the group were a single evaluator.

Consequently, grouping individual judgements gives management a solid knowledge regarding how the target market segment perceives the company products.

3 Formal Presentation of Methods

3.1 The Analytical Hierarchy Process

It consists of three major stages. First, an evaluator judges the relative importance of evaluation criteria on a pair-wise basis. This leads to a Pairwise Comparison Matrix (**PCM**), possessing the following structure:

$$\mathbf{PCM} = \begin{vmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{vmatrix}, \tag{1}$$

where c_{ij} is a numeric value that shows the relative importance of criterion c_i to criterion c_j. This first stage completes with the calculation of the eigenvector of the **PCM**.

$$\mathbf{eigenCriteria} = \begin{vmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{vmatrix}, \tag{2}$$

Eigenvector **eigenCriteria** defines the actual priority obtained by each criterion. On a second stage, the evaluator decides to what extent one alternative over another complies with a given criteria.

$$\mathbf{PCM}^{\mathbf{criterion}}_{\mathbf{alternative}} = \begin{vmatrix} 1 & a_{12} & \dots & a_{1m} \\ a_{21} & 1 & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & 1 \end{vmatrix}, \tag{3}$$

where a_{ij} is a numeric evaluation that reflects to what extent alternative a_i complies with criterion c_k when compared to alternative a_j. The eigenvector of matrix 3 is computed.

$$\mathbf{eigenAC_k} = \begin{vmatrix} eac_{1k} \\ eac_{2k} \\ \vdots \\ eac_{mk} \end{vmatrix}, \tag{4}$$

In $\mathbf{eigenAC_k}$, eac_{jk} represents how alternative j ranks when it is evaluated against criterion k. The second step is repeated as many times as criteria exist, terminating when all the resultant eigenvectors are arranged orderly in matrix **EIGENAC**.

The third and final step of the AHP consists of multiplying matrix **EIGENAC** times eigenvector **eigenCriteria** calculated in step one.

$$\mathbf{EIGENAC} \cdot \mathbf{eigenCriteria} \tag{5}$$

The result is vector **W**:

$$\mathbf{W} = \begin{vmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{vmatrix}, \tag{6}$$

where w_l represents the final and definite ranking obtained by each alternative. The alternative with the highest score gets the highest rank.

3.2 Fuzzy C Means Clustering Algorithm

Data clustering is concerned with the partitioning of a data set into several groups such that the similarity within a group is larger than among groups. This implies that the data set to be partitioned has to have an inherent grouping to some extent; otherwise if the data is uniformly distributed, trying to find clusters will fail, or will lead to artificially introduced partitions. Another problem that may arise is the overlapping of data groups. Overlapping groupings sometimes reduce the efficiency of the clustering method, and this reduction is proportional to the amount of overlap between groupings.

The approach of the clustering technique here presented is to find cluster centers that will represent each cluster. A cluster center is a way to tell where the heart of each cluster is located, so when presented with an input vector, the system can tell to which cluster such vector belongs by measuring a similarity metric between the input vector and all the cluster centers, and determining which cluster is the nearest or most similar one.

In the following, the well-known Fuzzy C- Means Clustering algorithm is shown ([17]). Fuzzy C-means clustering (FCM), relies on the basic idea of Hard C-means clustering (HCM) [17]. Bezdek proposed this algorithm in 1973 [18], with the difference that in FCM each data point belongs to a cluster to a degree of membership grade, while in HCM every data point either belongs to a certain cluster or not.

So FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1.

Let us define a set of n vectors, x_i, $i = 1, \ldots, n$ are to be partitioned into c fuzzy groups G_i, $i = 1, \ldots, c$, and find a cluster center on each group such that a cost function of dissimilarity measure is minimized. Imposing normalization stipulates that the summation of degrees of belongingness for a data set always be equal to unity:

$$\sum_{i=1}^{c} \mu_{ij} = 1, \quad \forall\ j = 1, \ldots, n. \tag{7}$$

The cost function (or objective function) measures a fuzzy distance between a vector x_k in group j and the corresponding cluster center c_i , can be defined by:

$$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m d_{ij}^2, \tag{8}$$

where μ_{ij} is between 0 and 1, c_i es the cluster center of fuzzy group i, $d_{ij} = ||c_i - x_j||$ is the Euclidean distance between ith clusters center and jth data point; and $m > 1$, is called a weighted exponent, which is judiciously chosen. Observe matrix U being defined by an $c \times n$ membership matrix, where the element $\mu_{ij} \in [0, 1]$ is defined by a membership function for the jth data point x_j belonging to group i, as:

$$\mu_{ij} = \begin{cases} 1\ if\ ||x_j - c_i||^2 \leq ||x_j - c_k||^2, & \text{for each } k \neq i, \\ 0 \ , \ \text{otherwise} \end{cases} \tag{9}$$

The necessary conditions for Eq. (8) to reach a minimum can be found by forming a new objective function *barJ* as follows:

$$\begin{aligned} \bar{J}(U, c_1, c_2, \ldots, c_c, \quad & \lambda_1, \ldots, \lambda_n) = \\ & = J(U, c_1, c_2, \ldots, c_c) + \textstyle\sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} \mu_{ij} - 1) \\ & = \textstyle\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m d_{ij}^2 + \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} \mu_{ij} - 1), \end{aligned} \tag{10}$$

where λ_j, $j = 1$ to n, are the Lagrange multipliers for the n constraints in Eq. (7). By differentiating $\bar{J}(U, c_1, c_2, \ldots, c_c, \lambda_1, \ldots, \lambda_n)$ with respect to all its input arguments, the necessary conditions for Eq. (8) to reach its minimum are

$$c_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_{ij}}{\sum_{j=1}^{n} \mu_{ij}^m}, \tag{11}$$

and

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{dkj})^{\frac{2}{m-1}}}, \tag{12}$$

In the following, the clustering algorithm is stated.

Algorithm 1 (Fuzzy C Means). Given the data set Z, choose the number of cluster $1 < c < N$, the weighting exponent $m > 1$, a constant for a cost function minimum $\varepsilon > 0$, and a constant Th which is a termination tolerance threshold. Initialize the partition matrix U randomly, such that $\mu_{ij}(0) \in [0, 1]$

Step 1. Compute clusters prototypes: Calculate c fuzzy cluster centers c_i, $i = 1 \ldots, c$ using Eq. (11).

Step 2. Compute the cost function According to Eq. (8). Stop if either it below the tolerance ε or its improvement over previous iteration is below the threshold Th.

Step 3. Compute a new U using Eq. 12. Go to **Step 2.**

End of the FC-Means algorithm

3.3 The Hybrid Approach to Process Customers Evaluations

We describe the usage of Fuzzy C-Means and the AHP to process customers' judgements. The combined usage of Fuzzy C-Means and the AHP to Direct Marketing is explained next.

Let $\xi = \{e_1, e_2. \cdots, e_n\}$ be the set of clients' evaluations, each of whom must compare the relative importance of a finite set of criteria $C = \{c_1, c_2, \cdots, c_p\}$ on which products are judged. This results in:

$$\mathbf{PCM^k} = \begin{vmatrix} 1 & a^k_{12} & \ldots & a^k_{1p} \\ a^k_{21} & 1 & \ldots & a^k_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a^k_{p1} & a^k_{p2} & \ldots & 1 \end{vmatrix}, \tag{13}$$

where $k = 1, 2, \cdots, n$ is the k_{th} client's evaluation; a^k_{ij} is the relative importance of criterion i over criterion j as determined by client's evaluation e_k.

When all the n Pairwise Comparison Matrices are formed, it remains to construct matrix $\mathbf{PCM^G}$ that reflects the pattern associated with the totality of the clients' evaluations.

The algorithm to construct the Global Pairwise Comparison Matrix is as follows.

1. The cardinality p of set C is computed.
2. A matrix $\mathbf{PCM^G}$ of dimensions $p \times p$ is formed.
3. The diagonal of matrix $\mathbf{PCM^G}$ is filled with 1.
4. Vector α_{ij} is formed with entries $a^k_{ij}, k = 1, 2, \cdots, n$.
5. $a^G_{ij} = FuzzyCMeans(\alpha_{ij})$
6. Method *countIncidences* is called for determining the quantity of evaluators inside each cluster. Cluster with the highest number of incidences is selected. Cluster centroid is obtained.
7. Repeat steps 4, 5, 6 $\forall(i, j) = 1, 2, \cdots, p; \forall(PCM^k), k = 1, 2, \cdots, n$

Thus,

$$\mathbf{PCM^G} = \begin{vmatrix} 1 & a_{12}^G & \dots & a_{1p}^G \\ a_{21}^G & 1 & \dots & a_{2p}^G \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1}^G & a_{p2}^G & \dots & 1 \end{vmatrix}. \tag{14}$$

Equation (14) is the resultant Global Pairwise Comparison Matrix that serves as basis to execute the AHP once all the customers's evaluations are processed.

Next, we illustrate how a Multi-Agent System fully automates the processing of data. Specifically, the entire set of activities, from data gathering, processing and final calculation is performed by the distributed and intelligent multi-agent system.

4 The Multi-Agent System

This section depicts the Multi-Agent System structure and dynamics. The MAS is fixed by the following agents, whose structure is shown in Fig. (1) by means of a deployment diagram:

- A *coordinator agent*,
- A set of *evaluator agents*,
- A *clustering agent*,
- An *AHP agent*.

These agents altogether posses the following dynamics:

1. The *coordinator agent* acquires problem variables i.e. the set of criteria associated to the survey, the set of products to be evaluated, as well as the number of clients that will perform the evaluation. It leaves a message on the *Evaluation Blackboard* to inform each of the *evaluator agents* about the newly input survey.
2. Each of the *evaluator agents* assists in the evaluation of criteria and products, as each client provides his/her judgement.
3. The *coordinator agent* corroborates that every *evaluator agent* has completed its task, by querying the *Evaluation Blackboard*.
4. The *coordinator agent* informs *clustering agent* upon verification of data completeness. Then, *clustering agent* processes clients's evaluation with Fuzzy C-Means to build clusters.
5. The *clustering agent* informs the *coordinator agent* upon completion of its assignment.
6. The *coordinator agent* request the *AHP agent* to compute the final prioritization of products by running the AHP. Then, it informs when the task is achieved.

The previous list of activities is formally represented in the communication diagram of Figure (2). Those two types of diagrams are part of UML 2.0 [19].

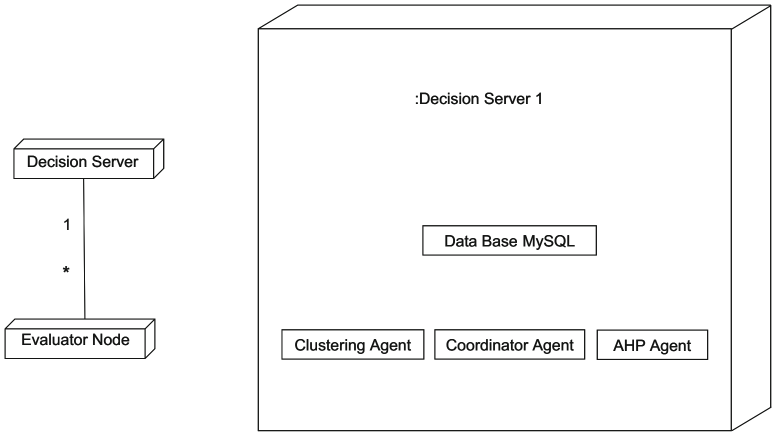


Fig. 1 Structure of the Multi-Agent System

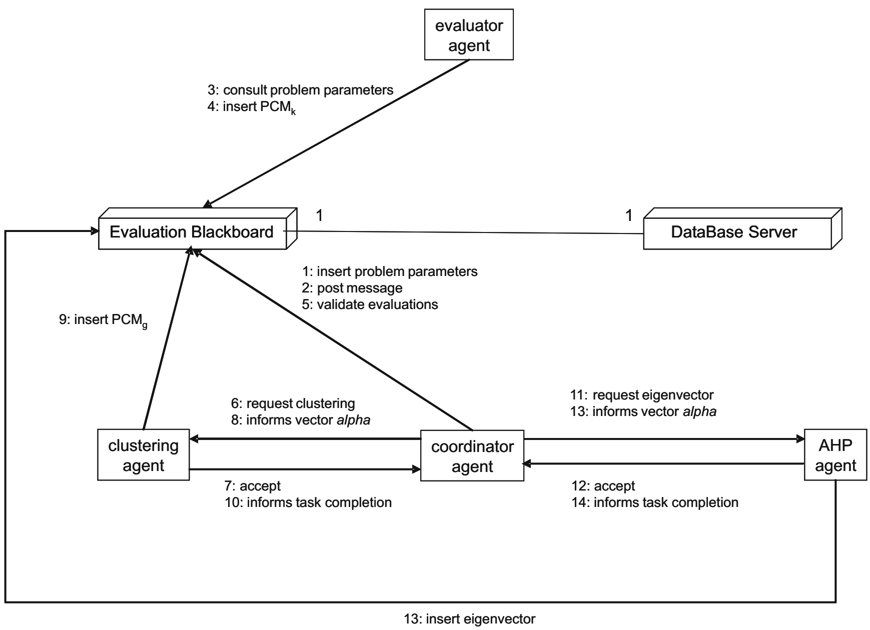


Fig. 2 Communication diagram of the Multi-Agent System

The implementation of the MAS is done on the JADE platform [20]. JADE is a useful tool because it allows to promote intelligent behavior to a given agent, while providing a rich set of communication capabilities based on FIPA-ACL. Both, the Fuzzy C-Means clustering technique and the AHP were developed on Java so *clustering agent* and *AHP agent*, respectively, call the coding transparently. The MAS is a distributed architecture because each agent resides in its own processing unit, and communication is done over the TCP/IP protocol, for which JADE possesses powerful libraries.

As it can be seen in Fig. (1), the *coordinator agent* communicates directly with both, the *clustering agent* and the *AHP agent*. It is not so regarding the *evaluator agents*. In this latter case, communication is done by posting messages on the *Evaluation Blackboard*. This *Evaluation Blackboard* is represented in Fig. (2) as an *artifact*. Such blackboard is actually a database implemented on MySQL, whose structure is shown in Fig. (3).

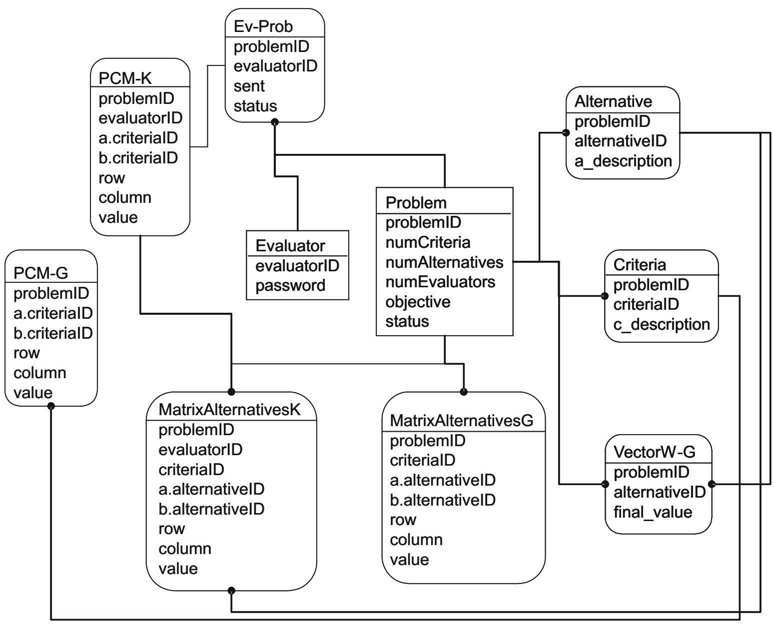


Fig. 3 IDEF1x model of the Evaluation Blackboard

Being the MAS a distributed architecture, it results a very useful tool for modern organizations because management and point sales are geographically separate entities. However, they must share the same information in order to achieve direct marketing. At this regard, management defines the set of criteria to evaluate products, what products must be evaluated, and the size of the population that will provide judgements. This is done at one physical location. The *coordinator agent* assists management directly.

On the other hand, actual salesmen or women are in touch with clients, yet they must adhere to the criteria fixed by management. The *evaluator agent* is running inside the computer used by the sales force, and gathers the criteria that was decided by management. There is one *evaluator agent* assisting every salesman or woman regardless their actual location. This is helpful to interview the clients they talk to. In this way, the clients opinions are fed to the central repository in real time.

When the totality of opinions are input, the *coordinator agent* orders the *clutering agent* and the *AHP agent* to process clients' data so management can visualize the manner in which a given market segment judges the company's products.

Such tasks are exemplified in section 5.

5 Experimental Results

In this section we present a case-study to validate the combined Fuzzy C-Means - AHP -MAS approach to direct marketing. The case study refers at determining what car model out of a list is best judged by a number of potential clients belonging to a specific market segment. To show the validity of the approach, we only provide data given by ten different clients, whom were asked to judge five different cars models on five different criteria. Management and salesmen or women were asked to employ the MAS. We present, step by step, the usage of the MAS and the final results.

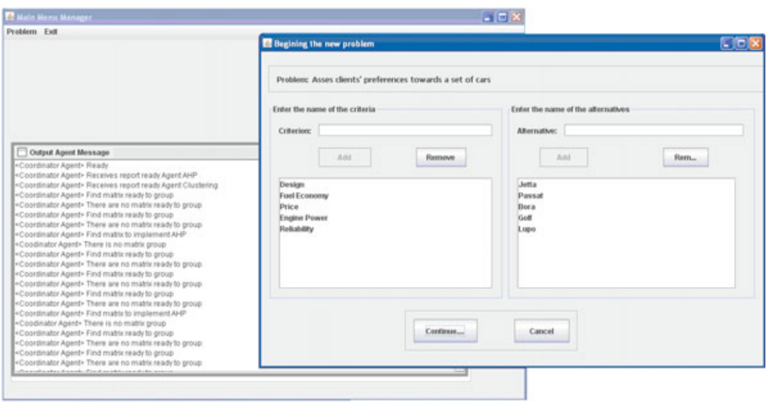


Fig. 4 Coordinator Agent. Entering survey parameters

Let $\xi = \{e_1, e_2. \cdots, e_{10}\}$ be the set of clients, and $C = \{c_1, c_2, c_3, c_4, c_5\}$ the set of criteria where: c_1 = Design, c_2 = Fuel Economy, c_3 = Price, c_4 = Engine Power, and c_5 = Reliability. Five different alternatives are evaluated, which are labeled A_1 = Jetta, A_2 = Passat, A_3 = Bora, A_4 = Golf, and A_5 = Lupo.

Management, comfortably sitting in their headquarters, introduce the survey parameters in a Graphical User Interface associated to the *coordinator agent*. Firstly, they establish the ID associated with the problem, along with the number of criteria, alternatives and population size (total number of evaluators). Afterwards, they introduce the objective of the problem, description of criteria, and the products to be evaluated (Fig. 4). These parameters are stored in Table *Problem* of the *Evaluation Blackboard* previously described. Accordingly, Fig. (5) displays the final definition of the survey parameters.

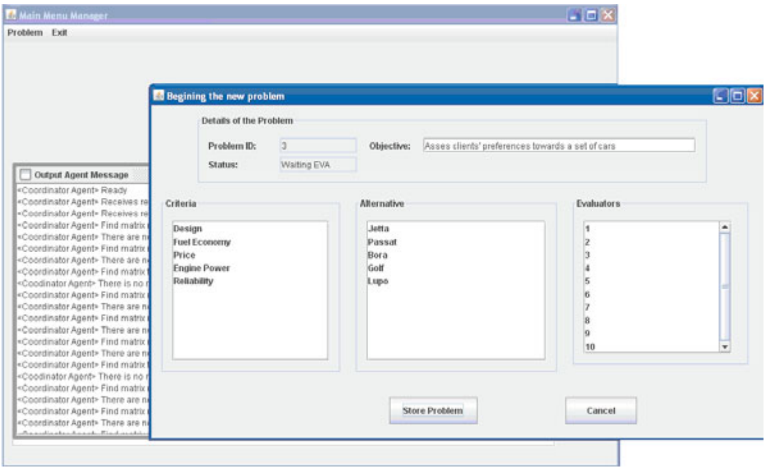


Fig. 5 Coordinator Agent. Summary of survey parameters.

Once the problem parameters are introduced, the *coordinator agent* posts a message on the *Evaluation Blackboard*, which will be read by each of the *evaluator agents* on their own network location. Thus, each *evaluator agent* constantly verifies whether a new problem has been introduced.

When a new survey is encountered (Fig. 6), its parameters are displayed so that the evaluator proceeds to determine the absolute importance of every criterion (Fig. 7).

Here we would like to elaborate on this way of evaluation. According to empirical usage of the system, human evaluators complaint about the time consuming process and the inability to keep track of their own judgements when they were requested to pair-wise compare both, criteria and alternatives. They also expressed that the numbers they were facing lacked meaning at some point. Instead, all of them agreed that it is more intuitive to make an absolute judgement on a 1-10 scale, and automate the pairwise comparisons as part of the system. The construction of the pair-wise comparison matrix for criteria is transparent to the evaluator. It also guarantees consistency of the PCM. Consequently, this process yields a PCM matrix for each evaluator, which is stored in the table *PCM-K* of the *Evaluation Blackboard*.

Upon completion of the entire set of evaluations, the *coordinator agent* informs the *clustering agent* that it must initiate the calculation of the clusters (Fig. 8).

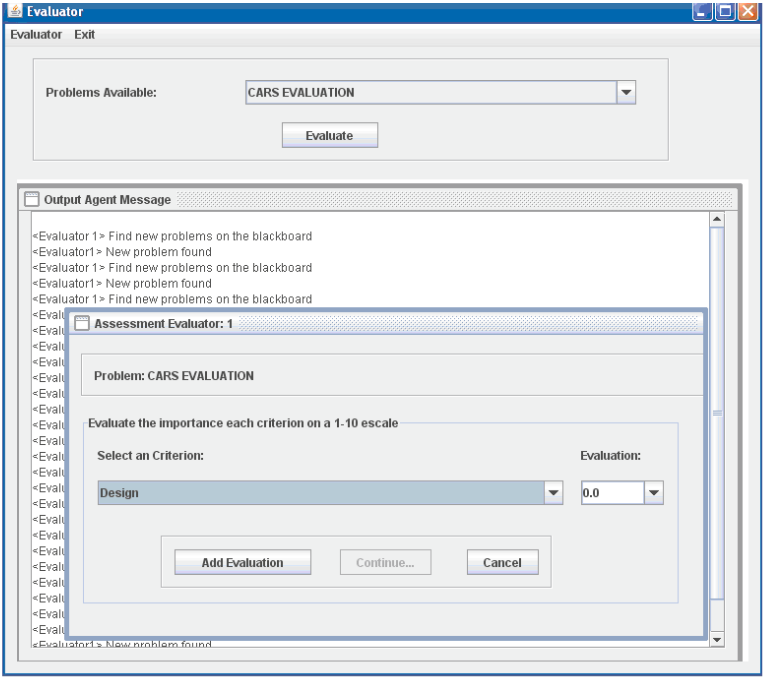


Fig. 6 Evaluator Agent. Finding a new survey at Sales Point.

Then, *clustering agent* acknowledges receipt and proceeds to build clusters, and then stores the Global PCM in table *PCM-G* of the *Evaluation Blackboard*. A summary of the final results for this particular case are displayed in Fig. (9), while the details can be analyzed as presented in Fig. (10).

5.1 Clients' Evaluation

The actual judgements given by the clients are depicted in the following table. First, they were asked to evaluate on a scale from 0 to 10, how important is Design (c_1), Fuel economy (c_2), Price (c_3), Engine power (c_4), and Reliability (c_5) at the moment of selecting a car.

c_i	e_k									
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
c_1	8	9	9	9	10	7	7	6	10	7
c_2	9	8	9	7	6	8	9	9	7	8
c_3	10	10	8	9	7	10	6	10	6	8
c_4	10	9	8	7	8	6	10	6	10	6
c_5	8	8	9	8	7	6	8	10	8	7

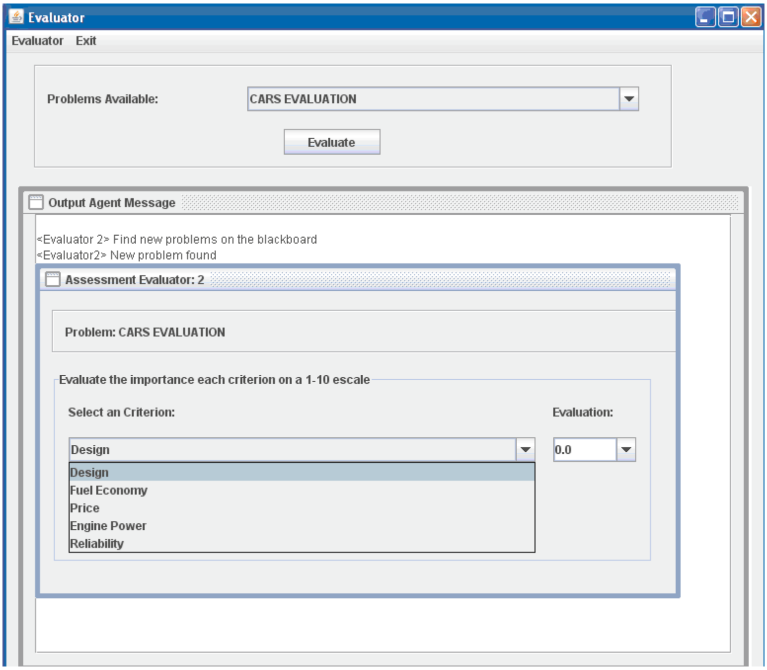


Fig. 7 Evaluator Agent. Criterion evaluation.

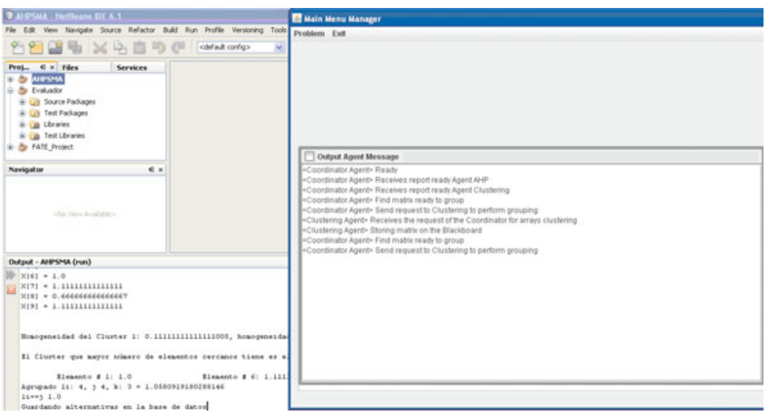


Fig. 8 Coordinator Agent informs Clustering Agent

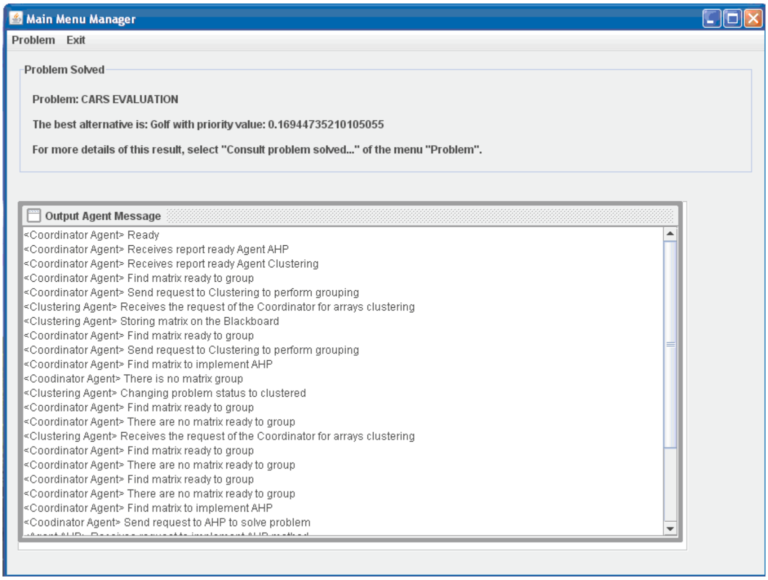


Fig. 9 Summary of final results

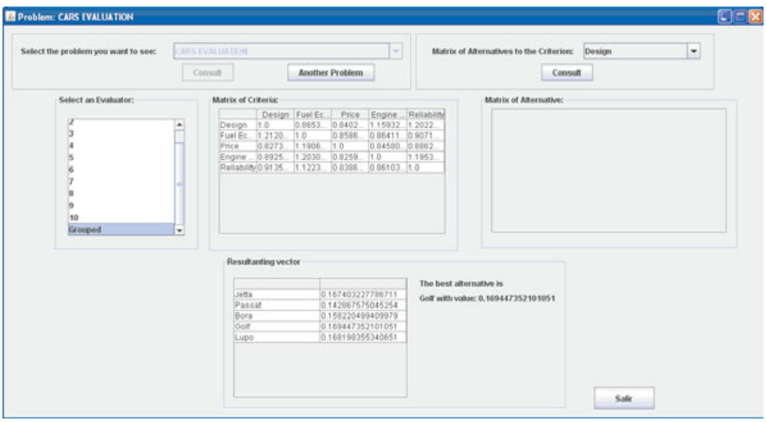


Fig. 10 Details of final results

Once every client has established how important every criteria he/she considers to be in for purchasing a car, clients are asked to evaluate to what extend they think alternative cars comply to the evaluation criteria. In the following table we present only one example of how one client ranked the five different car models on each criteria.

e_1	a_i				
	Jetta	Passat	Bora	Golf	Lupo
c_1	8	10	10	8	6
c_2	9	4	5	8	10
c_3	7	4	6	7	10
c_4	8	10	9	8	5
c_5	9	10	10	8	8

According to the previous table, client number one considers that *Jetta* evaluates with an 8 for its design, a 9 for its fuel economy, 7 for the price, 8 for the engine power, and a 9 for the reliability. There is one instance of the previous table for every one of the clients that participate in the survey. The totality of the evaluations are stored in the the *Evaluation Blackboard* (Fig. 3).

Once the target population evaluated (subjectively) the range of products, then the *coordinator agent*, running on the management node, validates that all the evaluations are complete. Shortly after, it requests that *clustering agent* and *AHP agent* achieve their own tasks by processing the raw data.

Knowledge obtained by management is a final ranking, which determines what product appeals the most to the target market segment. In this case, A_4 = Golf best balances the five features evaluated, as evidenced by ranking $R = \{A_1 : 0.1674, A_2 : 0.1428, A_3 : 0.1582, A_4 : 0.1684, A_5 : 0.1681\}$.

6 Concluding Remarks

We have presented an intelligent and distributed Multi-Agent System that incorporates the Analytical Hierarchy Process and the Fuzzy C-Means algorithm to enhance direct marketing. Particularly, the system is aimed at facilitating surveys and processing the large amounts of raw data that is generated.

The results provided with the case-study are very promising, because it has been shown that management can establish direct contact with a large group of customers. Every individual, in turn, is left free to evaluate the company products according to his or her personal criteria.

This is very valuable per se. Yet, the system also proved capable of processing the totality of the evaluations. With this, the perceptions of a market segment are deeply scrutinized by forming clusters. In this sense, the market segment is treated as a single unit because the perceptions of the majority are discovered.

It is intended to improve the MAS we have presented here by including different soft-computing techniques, such as neural networks and Case-Based reasoning. These techniques will provide more facilities so that management can compare and analyze the market behavior.

Acknowledgments. The authors would like to thank the anonymous reviewers because of their useful comments that improved this paper.

The work of Virgilio López Morales is partially supported by the PAI2006 - UAEH - 55A, SEP - SESIC - PROMEP - UAEH - PTC - 1004 and ANUIES - SEP - CONACyT/ECOS - NORD, M02:M03, research grants.

References

1. Hanson, C.: What are the best methods for increasing diversity at a digital marketing company? DMNews 1(1) (January 26, 2009)
2. Tsai, J.: Marketing trends for 2009. Costumer Relationship Management Magazine 1(1) (January 2009)
3. Pijls, W., Potharst, R.: Number 2000-40-LIS in ERS. In: Classification and target group selection based upon frequent patterns. ERIM Report Series Research in Management, The Netherlands, 1–16 (2000)
4. Bult, J., Wansbeek, T.: Optimal selection for direct mail. Journal Marketing Science 14, 378–394 (1995)
5. Haughton, D., Oulabi, S.: Direct marketing modeling with Cart and CHAID. Journal of Direct Marketing 7, 16–26 (1993)
6. Zahavi, J., Levin, N.: Issues and problems in applying neuronal computing to target marketing. Journal of Direct Marketing 9(3), 33–45 (1995)
7. Zahavi, J., Levin, N.: Aplying neuronal computing to target marketing. Journal of Direct Marketing 11(1), 5–22 (1997)
8. Kaymak, U.: Fuzzy Target Selection Using RFM Variables. In: Proc. Of Joint 9th IFSA World Congress and 20th NAFIS Int. Conference, Vancouver, Canada, pp. 1038–1043 (2001)
9. Setnes, M., Kaymak, U.: Fuzzy Modeling of Client Preference from Large Data Sets: An Application to Target Selection in Direct Marketing. IEEE Transactions on Fuzzy Systems 9(1) (2001)
10. Sousa-João, M., Kaymak, U., Madeira, S.: A Comparative Study of Fuzzy Target Selection Methods in Direct Marketing. In: Proceedings, IEEE World Congress on Computational Intelligence, pp. 1251–1256 (2002)
11. Site, C.: Cyber Atlas On line, http://www.cyberatlas.com/
12. Rayport, J., Sviokla, J.: Managing in the Marketspace. Harvard Business Review 1(1) (1994)
13. Bessen, J.: Riding the Marketing Information Wave. Harvard Business Review 1(1) (1993)
14. Bakos, Y.: Reducing Buyer Search Costs: Implications for Electronic Marketplaces. Management Science 43(12) (1997)
15. Balas, S.: Direct Marketer vs. Retailer: A Strategic Analysis of Competition and Market StructureMail vs. Mall. Marketing Science 17(3) (1998)
16. Saaty, T.L.: A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15(3), 234–281 (1977)
17. Roger Jang, C.T.S., Mizutani, E.: Neuro-fuzzy and Soft Computing. Prentice Hall, New York (1997)
18. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York (1981)
19. Bauer, B., Odell, J.: UML 2.0 and agents: how to build agent-based systems with the new UML standard. Engineering Applications of Artificial Intelligence 18, 141–157 (2005)
20. Fabio Bellifemine, G.C., Greenwood, D.: Developing Multi-Agent Systems with JADE. Addison Wesley, London (2007)

Direct Marketing Modeling Using Evolutionary Bayesian Network Learning Algorithm

Man Leung Wong

Department of Computing and Decision Sciences,
Lingnan University, Tuen Mun, Hong Kong
e-mail: mlwong@ln.edu.hk

Abstract. Direct marketing modeling identifies effective models for improving managerial decision making in marketing. This paper proposes a novel system for discovering models represented as Bayesian networks from incomplete databases in the presence of missing values. It combines an evolutionary algorithm with the traditional *Expectation-Maximization(EM)* algorithm to find better network structures in each iteration round. A data completing method is also presented for the convenience of learning and evaluating the candidate networks. The new system can overcome the problem of getting stuck in sub-optimal solutions which occurs in most existing learning algorithms and the efficiency problem in some existing evolutionary algorithms. We apply it to a real-world direct marketing modeling problem, and compare the performance of the discovered Bayesian networks with other models obtained by other methods. In the comparison, the Bayesian networks learned by our system outperform other models.

Keywords: Direct Marketing Modeling, Data Mining, Bayesian Networks, Evolutionary Algorithms.

1 Introduction

The objective of the direct marketing modeling problem is to predict and rank potential buyers from the buying records of previous customers. The customer list will be ranked according to each customer's likelihood of purchase. The decision makers can then select the portion of customer list to roll out. An advertising campaign including mailing of catalogs or brochure is targeted on the most promising prospects. Hence, if the prediction is accurate, it can help to enhance the response rate of the advertising campaign and increase the return of investment.

In real-life applications, the databases containing the buying records of customers may contain missing values. Irrelevant records or trivial items with

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 273–294.
springerlink.com

missing values can be simply discarded from the raw databases in the data preprocessing procedure. However, in most cases, the variables are related to each other and the deletion of incomplete records may lose important information. This will affect performance dramatically especially if we want to discover some knowledge "nuggets" from the databases and they happen to be contained in the incomplete records. Usually, people may alternatively replace the missing values with certain values, such as the mean or mode of the observed values of the same variable. Nevertheless, it may change the distribution of the original database.

Bayesian networks are popular within the community of artificial intelligence and data mining due to their ability to support probabilistic reasoning from data with uncertainty. They can represent the co-related relationships among random variables and the conditional probabilities of each variable from a given data set. With a network structure at hand, people can conduct probabilistic inference to predict the outcome of some variables based on the values of other observed ones. Hence, Bayesian networks are widely used in many areas, such as diagnostic and classification systems [1, 2, 3], information retrieval [4], troubleshooting, and so on. They are also suitable for reasoning with incomplete information.

Many methods have been suggested to learn Bayesian network structures from complete databases without missing values, which can be classified into two main categories [5]: the dependency analysis method [6] and the score-and-search approach [7, 8, 9]. For the former approach, the results of dependency tests are employed to construct a Bayesian network conforming to the findings. For the latter one, a scoring metric is adopted to evaluate candidate network structures while a search strategy is used to find a network structure with the best score. Decomposable scoring metrics, such as MDL and BIC, are usually used to deal with the problem of time consuming score evaluation. When the network structure changes, we only need to re-evaluate the score of the corresponding nodes related to the changed edges, rather than the scores of the whole nodes. And stochastic search methods which employ evolutionary algorithms have been used in the latter approach for complete data, such as Genetic Algorithms [10, 11], Evolutionary Programming [12], and hybrid evolutionary algorithms [13].

Nevertheless, learning Bayesian networks from incomplete data is a difficult problem in real-world applications. The parameter values and the scores of networks cannot be computed directly on the records having missing values. Moreover, the scoring metric cannot be decomposed directly. Thus, a local change in the network structure will lead to the re-evaluation of the score of the whole network, which is time-consuming considering the number of all possible networks and the complexity of the network structures. Furthermore, the patterns of the missing values also affect the dealing methods. Missing values can appear in different situations: *Missing Completely At Random*, or *Not Ignorable* [14]. In the first situation, whether an observation is missing or not is independent of the actual states of the variables. So the incomplete databases may be representative samples of the complete databases. However, in the second situation, the observations are missing to some specific states for some variables. Different approaches should be adopted for different situations, which again complicates the problem.

Many researchers have been working on parameter learning and structure learning from incomplete data. For the former, several algorithms can be used to estimate or optimize the parameter values of the known Bayesian network structures, such as Gibbs sampling, EM [8], and *Bound-and-Collapse* (BC) method [15, 16]. For structure learning from incomplete data, the main issues are how to define a suitable scoring metric and how to search for Bayesian network structures efficiently and effectively. Concerning the score evaluation for structure learning, some researchers proposed to calculate the expected values of the statistics to approximate the score of candidate networks. Friedman proposed a Bayesian *Structural Expectation-Maximization* (SEM) algorithm which alternates between the parameter optimization process and the model search process [17, 18]. The score of a Bayesian network is maximized by means of the maximization of the expected score. Peña *et al.* used the BC+EM method instead of the EM method in their *BS-BC+EM* algorithm for clustering [19, 20]. However, the search strategies adopted in most existing SEM algorithms may not be effective and may make the algorithms find sub-optimal solutions. Myers *et al.* employed a genetic algorithm to learn Bayesian networks from incomplete databases [21]. Both network structures and the missing values are encoded and evolved. The incomplete databases are completed by specific genetic operators during evolution. Nevertheless, it has the efficiency and convergence problems because of the enlarged search space and the strong randomness of the genetic operators for completing the missing values.

In this study, we propose a new learning system that uses EM to handle incomplete databases with missing values and uses a hybrid evolutionary algorithm to search for good candidate Bayesian networks. The two procedures are iterated so that we can continue finding a better model while optimizing the parameters for a good model to complete the database with more accurate information. In order to reduce the time for statistics computation, the database is preprocessed into two parts : records with and without missing values. Instead of using the expected values of statistics as in most existing SEM algorithms, our system applies a data completing procedure to complete the database and thus decomposable scoring metrics can be used to evaluate the networks. The MDL scoring metric is employed in the search process to evaluate the fitness of the candidate networks.

We apply our system to a direct marketing modeling problem, which requires to rank the previous customers according to their probability of potential purchasing. The results show that the performance of the evolved Bayesian networks obtained by our system is better than the models learned by several other learning algorithms.

The rest of this paper is organized as follows. In Section 2, we will present the backgrounds of direct marketing modeling, Bayesian networks, the missing value problem, and some Bayesian network learning algorithms. In Section 3, our new learning system for incomplete databases, *EBN*, will be described in details. In Section 4, we use our system to discover Bayesian networks from a real-life direct marketing database. We will conclude the paper in the last section.

2 Background

2.1 Direct Marketing Modeling

Direct marketing concerns communication with prospects, so as to elicit response from them. In contrast to the mass marketing approach, direct marketing is targeted at a group of individuals that are potential buyers and are likely to respond. In retrospect, direct marketing emerged because of the prevalence of mail ordering in the nineteenth century [22]. As technology advances, marketing is no longer restricted to mailing but includes a variety of media. Nevertheless, the most important issue in the business remains to be the maximization of the profitability, or ROI, of a marketing campaign.

In a typical scenario, we often have a huge list of customers. The list could be records of existing customers or data bought from *list brokers*. But among the huge list, there are usually few real buyers which amount to only a few percents [23]. Since the budget of a campaign is limited, it is important to focus the effort on the most promising prospects so that the response rate could be improved.

Before computers became widely used, direct marketers often used simple heuristics to enhance the response rate. One straightforward approach is to use common sense to make the decision. In particular, we could match prospects by examining the demographics of the customers in the list. For example, in the life insurance industry, it is natural to target the advertising at those who are rich and aging. Another common approach to enhance the response rate is to conduct list testing by evaluating the response of samplings from the list. If a certain group of customers gives a high response rate, the actual campaign may be targeted at the customers similar to this group. A more systematic approach, which was developed in 1920s but is still being used today, is to differentiate potential buyers from non-buyers using the recency-frequency-monetary model (RFM) [22]. In essence, the profitability of a customer is estimated by three factors including the recency of buying, the frequency of buying, and the amount of money spent. Hence, only individuals that are profitable will be the targets of the campaign.

With the advancement of computing and database technology, people seek for computational approaches to assist in decision making. From the data set that contains demographic details of customers, the objective is to develop a *response model* and use the model to predict promising prospects. In certain sense, response models are similar to classifiers in the classification problem. However, unlike the classifier which makes a dichotomous decision (i.e. active or inactive respondents), the response model needs to score each customer in the data set with the likelihood of purchase. The customers are then ranked according to the score. A ranked list is desirable because it allows decision makers to select the portion of customer list to roll out [24]. For instance, out of the 200,000 customers on the list, we might wish to send out catalogs or brochures to the most promising 30% of customers so that the advertising campaign is cost-effective (the 30% of the best customers to be mailed is referred to as the *depth-of-file*) [25]. Hence, one way to evaluate the response model is to look at its performance at different depth-of-file. In the literature, there are various approaches proposed for building the response model. Here, we give a brief review in the following paragraphs.

Earlier attempts often adopted a statistical analysis approach. Back in 1967, a company already used multiple regression analysis to build the response model. In 1968, the Automatic Interaction Detection (AID) system was developed which essentially uses tree analysis to divide consumers into different segments [22]. Later, the system was modified and became the Chi-Squared Automatic Interaction Detector (CHAID). One statistical analysis technique, which is still widely used today, is logistic regression. Essentially, the logistic regression model assumes that the *logit* (i.e. the logarithm of the *odd ratios*) of the dependent variable (active or inactive respondents) is a linear function of the independent variables (i.e. the attributes). Because the approach is popular, newly proposed models are often compared with the logistic regression model as the baseline comparison [25, 26, 27].

Zahavi and Levin [27] examined the possibility of learning a back-propagation neural network as the response model. However, due to a number of practical issues and that the empirical result did not improve over a logistic regression model, it seems that the neural network approach does not bring much benefit.

Because there are striking similarities between classification and the direct marketing problem, it is straightforward to apply classification algorithms to tackle the problem. As an example, Ling and Li [28] used a combination of two well-known classifiers, the naïve Bayesian classifier and C4.5, to construct the response model. Because scoring is necessary, they modified the C4.5 classifier so that a prediction (i.e. active and inactive respondents) comes with a *certainty factor*. To combine the two classifiers, they applied ada-boosting [29] to both classifiers in learning. When they evaluated their response model across three different real-life data sets, the result showed that their approach are effective for solving the problem.

Bhattacharyya formulated the direct marketing problem as a multi-objective optimization problem [25, 26]. He noted that the use of a single evaluation criterion, which is to measure the model's accuracy, is often inadequate [26]. For practical concern, he suggested that the evaluation criterion needs to include the performance of the model at a given depth-of-file. In an early attempt, he proposed to use GAs to learn the weights of a linear response model while the evaluation function is a weighted average of the two evaluation criteria. When comparing the learnt model with the logit model on a real-life data set, the new approach indicates a superior performance [25]. Recently, he attempted to use genetic programming to learn a tree-structured symbolic rule form as the response model [26]. Instead of using a weighted average criterion function, his new approach searches for *Pareto-optimal* solutions. From the analysis, he found that the GP approach outperforms the GA approach and is effective at obtaining solutions with different levels of trade-offs [26].

2.2 Bayesian Networks

A Bayesian network, G, has a directed acyclic graph (DAG) structure. Each node in the graph corresponds to a discrete random variable in the domain. An edge, $Y \rightarrow X$, on the graph, describes a parent and child relation in which Y is the parent and X is the child. All parents of X constitute the parent set of X which is denoted by Π_X. In addition to the graph, each node has a conditional probability table (CPT) specifying the probability of each possible state of the node given each possible combination of states of its parents. If a node contains no parent, the table gives the marginal probabilities of the node.

Let U be the set of variables in the domain, $U = \{X_1,\ldots,X_n\}$. Following Pearl's notation [30], a conditional independence (CI) relation is denoted by $I(X, Z, Y)$ where X, Y, and Z are disjoint subsets of variables in U. Such notation says that X and Y are conditionally independent given the *conditioning set*, Z. Formally, a CI relation is defined with:

$$P(x \mid y, z) = P(x \mid z) \quad \text{whenever} \quad P(y, z) > 0 \tag{1}$$

where x, y, and z are any value assignments to the set of variables X, Y, and Z respectively. A CI relation is characterized by its *order*, which is the number of variables in the conditioning set Z. By definition, the joint probability distribution of U can be expressed as:

$$P(X_1, \ldots, X_n) = \prod_i P(X_i|\Pi_{X_i}) \tag{2}$$

For simplicity, we use $X_i = k$ to specify that the i-th node takes the k-th possible state in its value domain, $\Pi_{X_i} = j$ to represent Π_{X_i} being instantiated to the j-th combinational state, and N_{ijk} to represent the counts of $X_i = k$ and $\Pi_{X_i} = j$ appearing simultaneously in the data. The conditional probability $p(X_i = k|\Pi_{X_i} = j)$, also denoted as *parameter* θ_{ijk}, can be calculated from complete data by:

$$\theta_{ijk} = \frac{N_{ijk}}{\sum_k N_{ijk}} \tag{3}$$

As mentioned before, there are two main categories of Bayesian network learning algorithms. The dependency analysis approach constructs a network by testing the validity of any independence assertion $I(X, Z, Y)$. If the assertion is supported by the data, edges cannot exist between X and Y on the graph [5, 6]. The validity of $I(X, Z, Y)$ is tested by performing a CI-test, in which statistical hypothesis testing procedure could be used. Suppose that the likelihood-ratio χ^2 test is used and the χ^2 statistics is calculated by:

$$G^2 = -2 \sum_{x,y,z} P(x, y, z) * \log \frac{P(x, y, z)}{P(y, z)P(x|z)} \tag{4}$$

Checking the computed G^2 against the χ^2 distribution, we can obtain the p-value [13]. If the p-value is less than a predefined *cutoff value* α, the assertion $I(X, Z, Y)$ is not valid; otherwise, it is valid and edges cannot exist between X and Y.

The score-and-search approach uses a scoring metric to evaluate candidate networks [7]. Take the decomposable MDL scoring metric for example [9], the MDL score of network G with every node N_i in the domain U can be described as:

$$MDL(G) = \sum_{N_i \in U} MDL(N_i, \Pi_{N_i}) \tag{5}$$

The MDL score of a network is smaller than that of another network if the former network is better. With the scoring metric, the learning problem becomes a search problem. It should be noted that since the metric is node-decomposable, it is only necessary to re-calculate the MDL scores of the modified nodes when the network structure is changed. However, the metric cannot be used directly if the databases have missing values.

2.3 The Missing Value Problem

In real-world applications, the databases may contain incomplete records which have missing values. People may simply discard incomplete records, but relevant information may be deleted. Alternatively, they can complete the missing values with the information of the databases such as the mean values of other observed values of the variables. However, the distribution of the data may be changed. Advanced approaches including maximum likelihood estimation [14], Bayesian multiple imputation [31], machine learning [32], Bayesian networks [33, 34], k-nearest neighbour, regression [35, 36], and singular value decomposition [37] have been applied to complete the missing values in databases and microarray gene expression data sets.

One advantage of Bayesian networks is that they support probabilistic reasoning from data with uncertainty. However, for learning Bayesian networks from incomplete databases, the parameter values and the scores of networks cannot be computed directly on the records having missing values. Moreover, the scoring metric cannot be decomposed directly. Thus, a local change in the network structure will lead to the re-evaluation of the score of the whole network.

For parameter learning, existing methods either use different inference algorithms to get the expected values of statistics or complete the missing values. Two commonly adopted methods are Gibbs sampling and EM [8]. Gibbs sampling tries to complete the database by inferring from the available information and then learns from the completed database [38]. On the other hand, EM calculates the expected values of the statistics via inference and then updates the parameter values using the previously calculated expected values [39, 40]. It will converge to a local maximum of the parameter values under certain conditions. Furthermore, EM usually converges faster than Gibbs sampling. Both Gibbs sampling and EM assume that the missing values appear randomly or follow a certain distribution. In order to encode prior knowledge of the pattern of missing data, Ramoni and Sebastinani proposed a new deterministic *Bound-and-Collapse* (BC) method that does not need to guess the pattern of missing data [15, 16, 41]. It firstly bounds the possible estimates consistent with the probability interval by computing the maximum and minimum estimates that would have been inferred from all possible completions of the database. Then the interval is collapsed to a unique value via a convex combination of the extreme estimates using information on the assumed pattern of missing data.

For structure learning from incomplete databases, the score-and-search approach can still be employed. The main issues are how to define a suitable scoring metric and how to search for Bayesian networks efficiently and effectively. Many variants of *Structural Expectation Maximization* (SEM) were proposed for this kind of learning in the past few years [17, 18].

2.4 Basic SEM Algorithm

The basic SEM algorithm cam learn Bayesian networks in the presence of missing values and hidden variables [17]. It alternates between two steps: an optimization for the Bayesian network parameters conducted by the EM algorithm, and a search for a better Bayesian network structure using a hill climbing strategy. The two steps iterate until the whole algorithm is stopped. The score of a Bayesian network is approximated by the expected value of statistics. Friedman extended his SEM to directly optimize the true Bayesian score of a network in [18]. The framework of the basic SEM algorithm can be described as follows:

1. let M_1 be the initial Bayesian network structure.
2. for t=1,2,...

- Execute EM to approximate the maximum-likelihood parameters Θ_t for M_t.
- Perform a greedy hill-climbing search over Bayesian network structures, evaluating each structure using approximated score $Score(M)$.
- let M_{t+1} be the Bayesian network structure with the best score.
- If $Score(M_t)$ =$Score(M_{t+1})$ then return M_t and Θ_t.

2.5 HEA

HEA is proposed by Wong and Leung for learning Bayesian networks from complete databases [13]. It employs the results of lower order CI-tests to refine the search space and adopts a hybrid evolutionary algorithm to search for good network structures. Each individual in the population represents a candidate network which is encoded by a connection matrix. Besides, each individual has a cutoff value α which is also subject to be evolved. At the beginning, for every pair of nodes (X,Y), the highest p-value returned by the lower order CI-tests is stored in a matrix P_v. If the p-value is greater than or equal to α, the conditional independence assertion *I(X,Z,Y)* is assumed to be valid, which implies that the nodes X and Y cannot have a direct edge between them. By changing the α values dynamically, the search space of each individual can be modified and each individual conducts its search in a different search space. Four mutation operators are used in HEA. They add, delete, move, or reverse edges in the network structures either through a stochastic method or based on some knowledge. A novel merge operator is suggested to reuse previous search results. The MDL scoring metric is used for evaluating candidate networks. The cycle prevention method is adopted to prevent cycle formation in the network structures.

The experimental results demonstrate that HEA has better performance on some benchmark data sets and real-world data sets than other state-of-the-art algorithms [13].

3 Learning Bayesian Networks from Incomplete Databases

3.1 The EBN Algorithm

Although HEA outperforms some existing approaches, it cannot deal with incomplete databases. Thus, we propose a novel evolutionary algorithm, called EBN (*Evolutionary Bayesian Network learning method*), that utilizes the efficient and effective global search ability of HEA and applies EM to handle missing values. Some strategies are also introduced to speed up EBN and to improve its performance. EBN is described in Fig. 1.

In EBN, there are two special kinds of generations. *SEM generation* refers to one generation in the SEM framework (step 9 of Fig. 1) while *HEA generation* refers to the iteration in HEA search process (step 9(g) of Fig. 1).

Firstly, the database is separated and stored into two parts in the data preprocess phase. The set of records having missing values is marked as H and the set of records without missing values is marked as O. Order-0 and order-1 CI tests are then conducted on O and the results are stored in the matrix P_v for refining the search space of each individual in the following procedures.

At the beginning of the SEM phase, for each individual, we check a randomly generated α value with the stored values in the matrix P_v to refine its search space. It should be noted that the search space will not be refined if O is not available. A DAG structure is then randomly constructed from the refined search space for this individual. Thus, the initial population is generated (step 7 of Fig. 1). Through some specific strategies, an initial network structure is generated for the current best network which is denoted as G_{best}. EBN will then be executed for a number of SEM generations until the stopping criteria are satisfied, that is, the maximum number of SEM generations is reached or the log-likelihood of G_{best} does not change for a specified number of SEM generations (step 9 of Fig. 1). The log-likelihood of G_{best} in the t-th SEM generation can be computed by:

$$ll(G_{best}(t)) = \sum_{i,j,k} [E(N_{ijk})log(\theta_{ijk})] \quad (6)$$

Within each SEM generation, EM will be conducted first to find the best values for the parameters of G_{best} (step 9(a) of Fig. 1). The missing values in H will be filled according to G_{best} and its parameters (step 9(c) of Fig. 1). Combining the newly completed result of H with O, we get a new complete data set O'. Then, the HEA search process will be executed on O' for a certain number of HEA generations to find a better network to replace G_{best}. The MDL scoring metric is again employed in the search process to evaluate the fitness of the candidate networks. The whole process will iterate until it stops. Some techniques are depicted in following subsections.

3.2 The EM Procedure in EBN

EM is employed here for parameter estimation of the input Bayesian network. The procedure is described in Fig. 2.

In order to facilitate the converge of the EM procedure, we choose G_{best} as the input network. In step 1 of Fig. 2, the initial parameter values of G_{best} are computed on data $O*$. For the first execution of EM in the first SEM generation, O is used as $O*$ (step 9(a) of Fig. 1). In the other SEM generations, $O*$ is the completed data O' from the previous SEM generation (step 9(a) of Fig. 1).

Data Preprocess

1. Store incomplete records together, mark the whole set as H.
2. Store other records together, mark the whole set as O.

CI test Phase

3. If O is available
   a. Perform order-0 and order-1 CI tests on O.
   b. Store the highest p-value in the matrix P_v.

   else store negative values in the matrix P_v.

SEM phase

4. Set t, the generation count, to 0.
5. Set t_{SEM}, the SEM generation count, to 0.
6. Set t_{uc}, the count of generations with unchanged log-likelihood, to 0.
7. For each individual G_i in the population $Pop(t)$
   - Initialize the α value randomly, where $0 \leq \alpha \leq 1$.
   - Refine the search space by checking the α value against the stored P_v value.
   - Inside the reduced search space, create a DAG randomly.
8. Generate the initial network structure for G_{best}.
9. While t_{SEM} is less than the maximum number of SEM generations or t_{uc} is less than MAX_{uc},
   a. If $t_{SEM} = 0$, execute **EM**(G_{best}, O, H);
      else execute **EM**(G_{best}, O', H).
   b. If the log-likelihood of G_{best} does not change, increment t_{uc} by 1;
      else set t_{uc} to 0.
   c. Complete missing data in H using G_{best} and its parameters, and get updated complete data O'.
   d. Execute order-0 and order-1 CI-tests on O', and store the highest p-value in P_v.
   e. For each individual G_i in the population $Pop(t)$
      - Refine the search space by checking the α value against the P_v value.
      - Evaluate G_i using the MDL metric on O'.
   f. Set t_{HEA}, the HEA generation count in each SEM generation, to 0.
   g. While t_{HEA} is less than the maximum number of HEA generations in each SEM generation ,
      - execute **HEA search phase**.
      - increment t_{HEA} and t by 1, respectively.
   h. Pick up the individual that has the lowest MDL score on O' to replace G_{best}.
   i. increment t_{SEM} and t by 1, respectively.
10. Return the individual that has the lowest MDL score in any HEA generation of the last SEM generation as the output of the algorithm.

Fig. 1 EBN Algorithm

EM contains two steps: the E-step and the M-step. In the E-step, the expected values of statistics of unobserved data (often called *sufficient statistics*) are estimated using probabilistic inference based on the input G_{best} and its parameter assignments. For each node X_i and record l^*, we can calculate the expected value of N_{ijk} using the following equation:

$$E(N_{ijk}) = \sum_{l^*} E(N_{ijk}|l^*) \tag{7}$$

where

$$E(N_{ijk}|l^*) = p(X_i = k, \Pi_{X_i} = j|l^*) \tag{8}$$

Let l represents the set of all other observed nodes in l^*. When both X_i and Π_{X_i} are observed in l^*, the expected value can be counted directly which is either 0 or 1. Otherwise, $p(X_i = k, \Pi_{X_i} = j|l^*) = p(X_i = k, \Pi_{X_i} = j|l)$, and it can be calculated using any Bayesian inference algorithm [42]. In our experiments, the junction tree inference algorithm is adopted [43]. Since the database is preprocessed, we just need to run the E-step on H.

Then, in the M-step, the parameters θ_{ijk} are updated by

$$\theta_{ijk} = \frac{E'(N_{ijk})}{\sum_k E'(N_{ijk})} \tag{9}$$

where $E'(N_{ijk})$ is the sum of the sufficient statistics calculated on H in the E-step and the statistics calculated on O which are evaluated and stored at the beginning.

The two steps will iterate until the EM procedure stops. EM will terminate when either the value of the log-likelihood does not change in two successive iterations, or the maximum number of iterations is reached.

Procedure EM(G_{best}, $O*$, H)

1. If data $O*$ is not empty, calculate the parameter values of G_{best} on $O*$; else the parameter values of G_{best} are generated randomly.
2. Set t, the EM iteration count, to 0.
3. While not converged,
   - For every node N_i,
     - calculate the expected statistics on H;
     - update θ_{ijk} using $E'(N_{ijk})$.
   - Calculate the log-likelihood of G_{best}.
   - Increment t by 1.
4. Output G_{best} and its parameters.

Fig. 2 Pseudo-code for the EM procedure

3.3 The Initial Network Structure for G_{best}

After executing the HEA search procedure, G_{best} is updated by the best candidate network having the lowest MDL score in the population (step 9(h) of Fig. 1) and then the newly found G_{best} is used in the EM procedure for the next SEM generation (step 9(a) of Fig. 1). However, we have to generate an initial network structure G_{best} for the first execution of the EM procedure for the first SEM generation. The quality of this network structure is crucial, because EBN is conducted on the database whose missing values are filled by performing inference using G_{best} and its parameters. In other words, the inference procedure may take a long time if G_{best} is not good enough.

In EBN, the initial network structure is obtained from a modified database. Considering the missing values in the original database as an additional state for each variable, we can get a new complete database. Then a network structure can be learned from the new complete database by HEA. The initial network structure G_{best} induced by HEA will not be put into the initial population after the execution of the EM procedure for the first SEM generation, so that the diversity of the population will not be destroyed at an early stage. The advantage of this method is that we can still find a good network structure even when data O is not available.

3.4 Data Completing Procedure

One of the main problems in learning Bayesian networks from incomplete databases is that the node-decomposable scoring metric cannot be used directly. In order to utilize HEA in EBN, we complete the missing data after each execution of the EM procedure so that the candidate networks can be evaluated efficiently on a complete database.

When more than one node are unobserved in a record, we fill the missing data according to the topological order of the current best network G_{best}. For example, if node X_i and X_j are both unobserved in record l^* and $X_i \rightarrow X_j$ exists in G_{best}, we first fill the value of X_i and put it back into the junction tree, and then find a value for X_j.

A Bayesian inference algorithm is again employed to obtain the probabilities of all possible states of an unobserved node under the current observed data. We can simply pick up the state having the highest probability. Alternatively, we can select a state via a roulette wheel selection method. Suppose the value of node X_i is unobserved in current record l^*, and X_i has k possible states in its value domain. We use $\{p_1, p_2,...,p_k\}$ to represent the set of the probability of each of its state appearing under current observed data in l^* respectively. In this approach, a random decimal r between 0 and 1 is generated, and then the m-th state will be chosen if

$$m = 1, r \leq p_1. \tag{10}$$

or

$$1 < m \leq k, \sum_{i=1}^{m-1} p_i < r \leq \sum_{i=1}^{m} p_i. \quad (11)$$

In EBN, we adopt the second completing approach so that the states with lower probabilities may also be selected.

3.5 HEA Search Procedure

With a complete data O', HEA can be utilized to learn good Bayesian networks. The lower order CI-test will be conducted again on O' and the highest p-values are stored in the matrix P_v, just as mentioned in subsection 2.5. Hence, each individual will refine its search space according to the new results on the new data O'. All the candidate networks are evaluated on O' using the MDL scoring metric. In each HEA generation, the mutation operators and the merge operator will be applied on each individual to generate a new offspring. The cycle prevention method is adopted to prevent cycle formation in the network structures. Individuals are then selected through a number of pairwise competitions over all the DAGs in the old population and the offspring to form a new population for the next HEA generation. This process continues until the maximum number of HEA generations is reached. Finally, the best network with the lowest MDL score on O' will be returned by the HEA search procedure.

4 Application in Direct Marketing Modeling

In this section, we apply EBN in a real-world direct marketing modeling problem. We compare the performance of the Bayesian networks evolved by EBN (EBN models) with those obtained by LibB[1] and Bayesware Discoverer[2] from incomplete real-world data sets, as well as the performance of *Bayesian neural network* (BNN) [44], *logistic regression* (LR), *naïve Bayesian classifier* (NB) [45], and *tree-augmented naïve Bayesian network classifier* (TAN) [45].

We also present the performance of the Bayesian networks evolved by HEA using two missing values handling methods. They transform an incomplete data set into a completed one and employ HEA as search method for learning Bayesian networks from the new data set.

In the first method, denoted as HEA1, we simply replace missing values for each variable with the mode of the observed data of the variable (the state that has the largest number of observations). In the second method, denoted as HEA2, we consider the missing values as a new additional state for each variable, thus a new completed data set is generated.

[1] LibB is available at http://compbio.cs.huji.ac.il/LibB/.
[2] A trial version of Bayesware Discoverer is available at http://www.bayesware.com/.

4.1 Methodology

The response models are evaluated on a real-life direct marketing data set. It contains records of customers of a specialty catalog company, which mails catalogs to good customers on a regular basis. In this data set, there are 5,740 active respondents and 14,260 non-respondents. The response rate is 28.7%. Each customer is described by 361 attributes. We applied the forward selection criteria of logistic regression [46] to select nine relevant attributes out of the 361 attributes.

Missing values are then introduced randomly into the data set. The percentages of the missing values in our experiments are 1%, 5%, and 10%, respectively. In our experiments, EBN, LibB and Bayesware Discoverer are executed directly on the data sets with missing values. For BNN, LR, NB, TAN and HEA1, we replace the missing values with the mean value or the mode for each variable. For HEA2, the missing values are treated as an additional new state for each variable.

For EBN, the maximum number of iterations in EM is 10, the maximum number of HEA generations in each SEM generation is 100, the maximum number of SEM generations is 50, the population size is 50, tournament size is 7, and MAX_{uc} is set to 10. For both HEA1 and HEA2, the maximum number of generations is set to 5000, the population size is 50, and the tournament size is 7.

To compare the performance of different response models, we use decile analysis which estimates the enhancement of the response rate for ranking at different depth-of-file. Essentially, the descending sorted ranking list is equally divided into 10 deciles. Customers in the first decile are the top ranked customers that are most likely to give response. Correspondingly, customers in the last decile are least likely to buy the specified products. A *gains table* will be constructed to describe the performance of the response model. In a gains table, we tabulate various statistics at each decile, including [47]:

- **Predicted Probability of Active:** It is the average of the predicted probabilities of active respondents in the decile by the response model.
- **Percentage of Active:** It is the percentage of active respondents in the decile.
- **Cumulative Percentage of Active:** It is the cumulative percentage of active respondents from decile 0 to this decile.
- **Actives:** It is the number of active respondents in this decile.
- **Percentage of Total Actives:** It is the ratio of the number of active respondents in this decile to the number of all active respondents in the data set.
- **Cumulative Actives:** It is the number of active respondents from decile 0 to this decile.
- **Cumulative Percentage of Total Actives:** It is the ratio of the number of cumulative active respondents (from decile 0 to this decile) to the total number of active respondents in the data set.

- **Lift:** It is calculated by dividing the percentage of active respondents by the response rate of the file. Intuitively, it estimates the enhancement by the response model in discriminating active respondents over a random approach for the current decile.
- **Cumulative Lift:** It is calculated by dividing the cumulative percentage of active respondents by the response rate. This measure evaluates how good the response model is for a given depth-of-file over a random approach. It provides an important estimate of the performance of the model.

4.2 Cross-Validation Results

In order to compare the robustness of the response models, we adopt a 10-fold cross-validation approach for performance estimation. A data set is randomly partitioned into 10 mutually exclusive and exhaustive folds. Each time, a different fold is chosen as the test set and other nine folds are combined together as the training set. Response models are learned from the training set and evaluated on the corresponding test set.

In Table 1, the average of the statistics of the EBN models for the 10 test sets of the data set with 1% missing values at each decile are tabulated. Numbers in the parentheses are the standard deviations. The EBN models have the cumulative lifts of 320.62 and 232.24 in the first two deciles respectively, suggesting that by mailing to the top two deciles alone, the Bayesian networks generate over twice as many respondents as a random mailing without a model. For this data set, the average learning time of EBN is 49.1 seconds on a notebook computer with an Intel$^{(R)}$ Core$^{(TM)}$ 2 Duo 1.8GHz processor and 3 GB of main memory running Windows XP operating system.

For the sake of comparison, the average of the cumulative lifts of the models learned by different methods from data sets with different percentages of missing values are summarized in Tables 2, 3, and 4, respectively. Numbers in the parentheses are the standard deviations. For each data set, the highest cumulative lift in each decile is highlighted in bold. The superscript $^+$ represents that the cumulative lift of the EBN models from the corresponding data set is significant higher at 0.05 level than that of the models obtained by the corresponding methods. The superscript $^-$ represents that the cumulative lift of the EBN models is significant lower at 0.05 level than that of the corresponding models.

In Table 2, the average and the standard deviations of the cumulative lifts of the models learned by different methods for the data set with 1% missing values are shown. In the first two deciles, the networks learned by LibB have cumulative lifts of 211.19 and 185.59, respectively; and 213.04 and 189.43 respectively for Bayeseware Discoverer models. It can be observed that EBN models get the highest cumulative lifts in the first three deciles, and the cumulative lifts of the EBN models in the first two deciles are significantly higher at 0.05 level than those of the other eight models.

In Table 3, the average and the standard deviations of the cumulative lifts for different models learned from the data set with 5% missing values are shown. In the first two deciles, the EBN models have the highest cumulative lifts of 320.27 and 224.07 respectively, and they are significantly higher than those of the other eight methods at 0.05 level. The average learning time of EBN is 200.5 seconds for this data set.

In Table 4, the average and the standard deviations of the cumulative lifts for different models discovered from the data set with 10% missing values are shown. Again, it demonstrates that the discovered EBN models have the highest cumulative lifts in the first two deciles, which are 320.18 and 212.88 respectively. The cumulative lifts of EBN models in the first two deciles are significantly higher at 0.05 level than those of the other eight methods. For this data set, the average learning time of EBN is 559.2 seconds.

To summarize, the networks generated by EBN always have the highest cumulative lifts in the first two deciles. Moreover, the cumulative lifts of the EBN models are significantly higher at 0.05 level than those of the other models in the first two deciles. Thus, we can conclude that EBN is very effective in learning Bayesian networks from data sets with different missing value percentages.

Since an advertising campaign often involves huge investment, a Bayesian network which can categorize more prospects into the target list is valuable as it will enhance the response rate. From the experimental results, it seems that EBN are more effective than the other methods.

Table 1 Gains Table of the EBN models for the 10 test sets of the data set with 1% missing values

Decile	Prob. of Active	% of Active	Cum. % of Active	Actives	% of Total Actives	Cum. Actives	Cum. % of Total Actives	Lift	Cum. Lift
0	44.61% (1.66%)	91.96% (6.41%)	91.96% (6.41%)	183.00 (12.75)	31.90% (2.35%)	183.00 (12.75)	31.90% (2.35%)	320.62 (23.64)	320.62 (23.64)
1	43.23% (0.82%)	41.37% (8.45%)	66.67% (5.55%)	82.33 (16.81)	14.31% (2.76%)	265.33 (22.10)	46.22% (3.54%)	143.86 (27.78)	232.24 (17.78)
2	42.92% (1.95%)	2.09% (7.63%)	45.14% (2.65%)	4.17 (15.19)	0.72% (2.62%)	269.50 (15.83)	46.94% (2.24%)	7.26 (26.30)	157.25 (7.50)
3	31.20% (1.72%)	30.30% (3.20%)	41.43% (1.91%)	60.30 (6.37)	10.51% (1.10%)	329.80 (15.22)	57.45% (1.94%)	105.60 (11.03)	144.33 (4.88)
4	24.61% (0.33%)	27.92% (3.55%)	38.73% (1.44%)	55.57 (7.07)	9.69% (1.30%)	385.37 (14.33)	67.14% (1.85%)	97.40 (13.03)	134.95 (3.71)
5	23.17% (0.37%)	58.26% (12.40%)	41.98% (1.99%)	115.93 (24.67)	20.20% (4.22%)	501.30 (23.70)	87.33% (3.47%)	202.99 (42.43)	146.29 (5.81)
6	22.69% (0.24%)	1.99% (6.05%)	36.27% (1.25%)	3.97 (12.05)	0.69% (2.08%)	505.27 (17.47)	88.02% (1.92%)	6.90 (20.90)	126.38 (2.76)
7	22.45% (0.55%)	4.30% (8.76%)	32.28% (1.19%)	8.57 (17.43)	1.48% (3.00%)	513.83 (18.98)	89.50% (1.89%)	14.91 (30.14)	112.44 (2.37)
8	17.12% (0.61%)	24.29% (4.65%)	31.39% (0.83%)	48.33 (9.25)	8.43% (1.66%)	562.17 (14.90)	97.94% (0.63%)	84.74 (16.65)	109.36 (0.70)
9	14.96% (0.87%)	5.66% (1.71%)	28.70% (0.71%)	11.83 (3.58)	2.06% (0.63%)	574.00 (14.17)	100.00% (0.00%)	19.75 (6.01)	100.00 (0.00)

Table 2 Cumulative lifts of the networks learned by different methods for the real-world data sets with **1%** missing values. The statistics are obtained from the 10 test sets.

Decile	EBN	LibB	Bayesware Discoverer	BNN	LR	NB	TAN	HEA1	HEA2
0	**320.62**	211.19$^+$	213.04$^+$	200.11$^+$	188.30$^+$	198.50$^+$	195.80$^+$	195.60$^+$	195.10$^+$
	(23.64)	(28.00)	(41.61)	(11.00)	(12.23)	(9.99)	(6.41)	(9.03)	(11.17)
1	**232.24**	185.59$^+$	189.43$^+$	171.01$^+$	168.80$^+$	169.70$^+$	168.30$^+$	169.80$^+$	170.10$^+$
	(17.78)	(17.44)	(14.53)	(9.76)	(9.73)	(7.15)	(7.35)	(8.65)	(6.79)
2	**157.25**	156.79	155.99	156.56	152.30$^+$	154.30	150.90$^+$	154.00	155.00
	(7.50)	(7.08)	(7.46)	(5.74)	(6.72)	(4.45)	(4.89)	(5.54)	(5.83)
3	144.33	**146.54**	146.07	144.26	141.40$^+$	139.40$^+$	139.70$^+$	142.60	144.30
	(4.88)	(5.56)	(7.90)	(4.67)	(3.13)	(2.55)	(2.75)	(5.23)	(3.80)
4	134.95	136.43	**140.78**	135.60	132.80$^+$	131.20$^+$	132.50	132.70$^+$	134.30
	(3.71)	(6.92)	(12.08)	(1.98)	(1.23)	(1.75)	(4.17)	(3.09)	(3.02)
5	**146.29**	134.65$^+$	136.09$^+$	127.33$^+$	125.80$^+$	124.70$^+$	124.10$^+$	126.40$^+$	128.30$^+$
	(5.81)	(10.05)	(4.35)	(2.15)	(2.86)	(2.79)	(2.69)	(2.88)	(2.45)
6	**126.38**	119.16$^+$	119.63$^+$	120.20$^+$	118.30$^+$	116.70$^+$	118.70$^+$	120.80$^+$	118.80$^+$
	(2.76)	(4.11)	(1.82)	(2.02)	(2.26)	(1.64)	(1.70)	(3.01)	(1.48)
7	112.44	113.69	112.53	**113.80**$^-$	112.50	111.90	113.40$^-$	113.10	112.90$^-$
	(2.37)	(3.87)	(1.84)	(1.61)	(1.35)	(1.45)	(1.17)	(1.52)	(0.57)
8	**109.36**	108.58	107.64$^+$	107.71$^+$	106.60$^+$	106.20$^+$	106.20$^+$	106.20$^+$	106.10$^+$
	(0.70)	(2.03)	(1.86)	(0.98)	(1.07)	(0.92)	(1.03)	(1.14)	(0.88)
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 3 Cumulative lifts of the networks learned by different methods for the real-world data sets with **5%** missing values. The statistics are obtained from the 10 test sets.

Decile	EBN	LibB	Bayesware Discoverer	BNN	LR	NB	TAN	HEA1	HEA2
0	**320.27**	217.63$^+$	246.59$^+$	199.37$^+$	188.50$^+$	195.40$^+$	197.80$^+$	193.30$^+$	192.40$^+$
	(22.43)	(47.64)	(31.34)	(10.33)	(11.45)	(10.27)	(9.84)	(5.79)	(12.97)
1	**224.07**	186.30$^+$	165.69$^+$	171.09$^+$	167.80$^+$	170.30$^+$	169.60$^+$	167.90$^+$	169.90$^+$
	(16.29)	(21.35)	(19.94)	(9.50)	(9.20)	(6.33)	(7.38)	(6.82)	(7.58)
2	153.53	155.28	152.60	**155.97**	151.40	152.60	151.50	153.30	153.80
	(6.98)	(6.96)	(7.80)	(5.60)	(4.77)	(4.14)	(5.23)	(4.47)	(5.85)
3	143.41	**145.15**	143.24	143.21	140.40$^+$	139.50$^+$	139.90$^+$	143.60	142.90
	(5.83)	(8.33)	(6.71)	(3.67)	(2.67)	(2.72)	(2.85)	(3.89)	(4.51)
4	135.63	136.75	**144.16**$^-$	134.18	132.40$^+$	130.50$^+$	131.30$^+$	133.00$^+$	133.10$^+$
	(3.67)	(6.21)	(5.18)	(2.61)	(1.58)	(1.27)	(3.27)	(2.54)	(3.38)
5	**145.72**	133.47$^+$	124.27$^+$	126.88$^+$	125.60$^+$	125.00$^+$	123.60$^+$	126.30$^+$	128.10$^+$
	(5.51)	(10.49)	(3.38)	(2.49)	(2.67)	(2.62)	(1.65)	(2.95)	(3.21)
6	**126.11**	118.90$^+$	118.10$^+$	120.07$^+$	118.40$^+$	117.00$^+$	118.10$^+$	119.30$^+$	118.90$^+$
	(2.86)	(4.94)	(1.85)	(2.29)	(2.41)	(1.70)	(1.66)	(2.06)	(1.79)
7	111.74	113.57$^-$	113.09$^-$	**113.73**$^-$	112.40	111.50	112.50	112.70	113.20$^-$
	(2.05)	(3.69)	(2.18)	(1.48)	(1.17)	(1.35)	(1.27)	(1.42)	(0.79)
8	**109.20**	108.08$^+$	106.80$^+$	107.64$^+$	106.60$^+$	106.00$^+$	106.10$^+$	106.20$^+$	105.90$^+$
	(1.11)	(1.89)	(1.56)	(0.87)	(0.97)	(1.15)	(1.10)	(1.23)	(0.88)
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 4 Cumulative lifts of the networks learned by different methods for the real-world data sets with **10%** missing values. The statistics are obtained from the 10 test sets.

Decile	EBN	LibB	Bayesware Discoverer	BNN	LR	NB	TAN	HEA1	HEA2
0	**320.18** (24.36)	239.06^+ (64.44)	196.86^+ (18.50)	195.71^+ (13.60)	185.10^+ (12.56)	190.40^+ (13.55)	194.90^+ (11.43)	194.10^+ (9.87)	195.80^+ (9.27)
1	**212.88** (15.96)	188.42^+ (21.09)	171.22^+ (9.13)	169.89^+ (9.75)	164.90^+ (10.46)	167.70^+ (6.29)	167.20^+ (8.83)	167.00^+ (6.36)	168.50^+ (7.63)
2	152.76 (5.65)	153.36 (6.38)	152.20 (6.40)	**154.32** (6.76)	149.30 (8.11)	151.30 (3.95)	151.30 (5.38)	152.40 (6.06)	153.00 (5.50)
3	141.78 (4.40)	**142.46** (9.31)	139.63 (4.50)	142.28 (4.66)	138.90^+ (3.57)	138.40^+ (2.91)	139.40 (3.63)	139.90 (3.96)	141.00 (4.29)
4	**136.15** (5.39)	134.86 (5.83)	131.55^+ (4.84)	133.14^+ (3.55)	130.70^+ (2.31)	128.60^+ (1.78)	129.80^+ (4.16)	132.30^+ (2.67)	132.40^+ (3.86)
5	**143.02** (6.50)	134.62^+ (10.86)	124.17^+ (5.17)	125.38^+ (1.82)	123.60^+ (2.01)	123.50^+ (1.72)	123.20^+ (1.99)	124.50^+ (2.37)	125.50^+ (2.46)
6	**125.51** (3.20)	119.65^+ (5.40)	117.23^+ (2.73)	119.27^+ (2.25)	117.70^+ (2.67)	116.10^+ (2.33)	117.30^+ (1.42)	118.30^+ (2.26)	118.40^+ (1.84)
7	111.58 (2.08)	**112.61** (4.21)	112.36 (1.85)	113.25^- (1.28)	111.90 (1.85)	111.20 (1.81)	112.50^- (1.27)	112.30 (1.25)	113.10^- (0.88)
8	**109.35** (0.91)	108.97 (1.81)	105.51^+ (1.22)	107.09^+ (0.67)	106.40^+ (0.84)	105.60^+ (0.97)	106.30^+ (0.82)	106.20^+ (0.92)	106.30^+ (0.82)
9	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)

5 Conclusion

In this paper, we have described a new evolutionary algorithm called EBN that applies EM, a strategy for generating an initial network structure, and a data completing procedure to learn Bayesian networks from incomplete databases. To explore its interesting applications for real-life data mining problems, we have applied EBN to a real-world data set of direct marketing and compared the performance of the networks obtained by EBN with the models generated by other methods. The experimental results demonstrate that EBN outperforms other methods in the presence of missing values.

The main advantage of EBN lies in the integration of EM and a hybrid evolutionary algorithm. While using EM and Bayesian inference to complete the missing values of a variable, the relationships of this variables with other variables are also considered instead of examining only the observed values of the variable. Thus better missing value imputations can be obtained. At the same time, the hybrid evolutionary algorithm facilitates the discovery of much better Bayesian network structures effectively and efficiently.

In this work, the missing values in the data sets are introduced randomly. In the future, studies will be conducted to facilitate EBN for incomplete data sets with other patterns of missing values.

Acknowledgments. This work is supported by the Lingnan University Direct Grant DR04B8.

References

1. Jensen, F.V.: An Introduction to Bayesian Network. University of College London Press, London (1996)
2. Andreassen, S., Woldbye, M., Falck, B., Andersen, S.: MUNIN: A causal probabilistic network for interpretation of electromyographic findings. In: Proceedings of the Tenth International Joint Conference on Artificial Intelligence, pp. 366–372 (1987)
3. Cheeseman, P., Kelly, J., Self, M., Stutz, J., Taylor, W., Freeman, D.: Autoclass: A Bayesian classification system. In: Proceedings of the Fifth International Workshop on Machine Learning, pp. 54–64 (1988)
4. Heckerman, D., Horvitz, E.: Inferring informational goals from free-text queries: A Bayesian approach. In: Cooper, G.F., Moral, S. (eds.) Proceedings of the Fourteenth Conference of Uncertainty in Artificial Intelligence, pp. 230–237. Morgan Kaufmann, Wisconsin (July 1998)
5. Cheng, J., Greiner, R., Kelly, J., Bell, D., Liu, W.: Learning Bayesian networks from data: An information-theory based approach. Artificial Intelligence 137, 43–90 (2002)
6. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search, 2nd edn. MIT Press, MA (2000)
7. Cooper, G., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. Machine Learning 9(4), 309–347 (1992)
8. Heckerman, D.: A tutorial on learning Bayesian networks. Microsoft Research Adv. Technol. Div., Redmond, WA, Tech. Rep. MSR-TR-95-06 (1995)
9. Lam, W., Bacchus, F.: Learning Bayesian belief networks-an approach based on the MDL principle. Computer Intelligence 10(4), 269–293 (1994)
10. Larrañaga, P., Poza, M., Yurramendi, Y., Murga, R., Kuijpers, C.: Structural learning of Bayesian network by genetic algorithms: A performance analysis of control parameters. IEEE Trans. Pattern Anal. Machine Intell. 18, 912–926 (1996)
11. Larrañaga, P., Kuijpers, C., Mura, R., Yurramendi, Y.: Learning Bayesian network structures by searching for the best ordering with genetic algorithms. IEEE Transactions on System, Man and Cybernetics 26(4), 487–493 (1996)
12. Wong, M.L., Lam, W., Leung, K.S.: Using evolutionary programming and minimum description length principle for data mining of Bayesian networks. IEEE Trans. Pattern Anal. Machine Intell. 21, 174–178 (1999)
13. Wong, M.L., Leung, K.S.: An efficient data mining method for learning Bayesian networks using an evolutionary algorithm-based hybrid approach. IEEE Trans. Evolutionary Computation 8(4), 378–404 (2004)
14. Schafer, J.L., Graham, J.W.: Missing data: Our view of the state of art. Psychological Methods 7(2), 147–177 (2002)
15. Ramoni, M., Sebastinani, P.: Efficient parameter learning in Bayesian networks from incomplete databases, Tech. Rep. KMI-TR-41 (1997)
16. Ramoni, M., Sebastinani, P.: The use of exogenous knowledge to learn Bayesian networks from incomplete databases, Tech. Rep. KMI-TR-44 (1997)

17. Friedman, N.: Learning belief networks in the presence of missing values and hidden variables. Machine Learning (1997)
18. Friedman, N.: The Bayesian Structural EM algorithm. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (1998)
19. Peña, J.M., Lozano, J.A., Larrañaga, P.: An improved Bayesian Structural EM algorithm for learning Bayesian networks for clustering. Pattern Recognition Letters, 779–786 (2000)
20. Peña, J.M., Lozano, J.A., Larrañaga, P.: Learning recursive Bayesian multinets for data clustering by means of constructive induction. Machine Learning, 63–89 (2002)
21. Myers, J.W., Laskey, K.B., DeJong, K.A.: Learning Bayesian networks from incomplete data using evolutionary algorithms. In: Proceedings of the First Annual Conference on Genetic and Evolutionary Computation Conference 1999, pp. 458–465. Morgan Kauffman, Orlando (July 1999)
22. Petrison, L.A., Blattberg, R.C., Wang, P.: Database marketing: Past present, and future. Journal of Direct Marketing 11(4), 109–125 (1997)
23. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zansi, A.: Discovering Data Mining: From Concept to Implementation. Prentice-Hall Inc., Englewood Cliffs (1997)
24. Zahavi, J., Levin, N.: Issues and problems in applying neural computing to target marketing. Journal of Direct Marketing 11(4), 63–75 (1997)
25. Bhattacharyya, S.: Direct marketing response models using genetic algorithms. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 144–148 (1998)
26. Bhattacharyya, S.: Evolutionary algorithms in data mining: Multi-objective performance modeling for direct marketing. In: Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining, pp. 465–473 (August 2000)
27. Zahavi, J., Levin, N.: Applying neural computing to target marketing. Journal of Direct Marketing 11(4), 76–93 (1997)
28. Ling, C.X., Li, C.: Data mining for direct marketing: Problems and solutions. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 73–79 (1998)
29. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on Machine Learning, pp. 148–156 (1996)
30. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988)
31. Rubin, D.B.: Multiple Imputation for nonresponse in Surveys. Wiley, Chichester (1987)
32. Lakshminarayan, K., Harp, S.A., Samad, T.: Imputation of missing data using machine learning techniques. In: Proceeding of the second International Conference on Knowledge Discovery and data Mining, pp. 140–146 (1996)
33. Zio, M.D., Scanu, M., Coppola, L., Luzi, O., Ponti, A.: Bayesian networks for imputation. Journal of the Royal Statistical Society (A) 167(2), 309–322 (2004)
34. Hruschka, E.R.J., Ebecken, N.F.F.: Missing values prediction with K2. Intelligent Data Analysis 6(6), 557–566 (2002)
35. Kim, H., Golub, G.H., Park, H.: Imputation of missing values in DNA microarray gene expression data. In: Proceedings of IEEE Computational Systems Bioinformatics Conference, pp. 572–573. IEEE Press, Los Alamitos (2004)

36. Cai, Z., Heydari, M., Lin, G.: Microarray missing value imputation by iterated local least squares. In: Proceedings of the Fourth Asia-Pacific Bioinformatics Conference, pp. 159–168 (2006)
37. Liu, L., Hawkins, D.M., Ghosh, S., Young, S.: Robust singular value decomposition analysis of microarray data. Proceedings of the National Academy of Sciences of the United States of America 100(23), 13167–13172 (2003)
38. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions and the bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 6, 721–742 (1984)
39. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society(B) 39(1), 1–38 (1977)
40. Lauritzen, S.L.: The EM algorithm for graphical association models with missing data. Computational Statistics and Data Analysis, 191–201 (1995)
41. Ramoni, M., Sebastiani, P.: Robust learning with missing data. Machine Learning 45, 147–170 (2001)
42. Lauritzen, S.L., Spiegelhalter, D.J.: Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society 50(2), 157–224 (1988)
43. Huang, C., Darwinche, A.: Inference in belief networks: a procedural guide. International Journal of Approximate Reasoning 15(3), 225–263 (1996)
44. Neal, R.M.: Bayesian Learning for Neural Networks. Springer, Heidelberg (1996)
45. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning 29, 131–163 (1997)
46. Agresti, A.: Categorical Data Analysis. Wiley, New York (2002)
47. Rud, O.P.: Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management. Wiley, New York (2001)

Designing Optimal Products: Algorithms and Systems

Stelios Tsafarakis and Nikolaos Matsatsinis

Technical University of Crete, Department of Production and
Management Engineering,Decision Support Systems Laboratory,
Chania, Greece
e-mail: tsafarakis@isc.tuc.gr
nikos@ergasya.tuc.gr

Abstract. The high cost of a product failure makes it imperative for a company to assess the market penetration of a new product at its early design. In this context, the Optimal Product Line Design problem was formulated thirty five years ago, and remains a significant research topic in the area of quantitative marketing until today. In this chapter we provide a brief description of the problem, which belongs to the class of NP-hard problems, and review the optimization algorithms that have been applied to it. The performance of the algorithms is evaluated, and the best two approaches are applied to simulated data, as well as a real world scenario. Emphasis is placed on Genetic Algorithms, since the results of the study indicate them as the approach that better fits to the specific marketing problem. Finally, the relevant marketing systems that deal with the problem are presented, and their pros and cons are discussed.

1 Introduction

Nowadays the economic environment where companies operate has become more competitive than ever. The globalization of the markets, the shorter product life cycles, and the rapid technology development, put high pressure on firms' profitability. In order to survive under such circumstances a company must continuously launch new products or redesign its current ones. However, such procedures entail risk. A new product is costly and difficult to change, hence if it ends up as a commercial failure, the firm's viability may be put in danger. Bad design constitutes one of the most frequent reasons for the failure of a new product (Kotler and Armstrong, 2008). In order to avoid such situations, managers try to design optimal products and assess their market penetration before their entrance to the production stage. This constitutes a wide area of research in quantitative marketing for over thirty years, known as the *Optimal Product (Line) Design Problem*. Here, each product consists of a number of attributes that take specific levels, and the consumer preferences regarding the various attribute levels are considered known. Taking as input these consumer preferences, optimization algorithms are used in order for optimal product profiles to be designed. Different objectives can be employed, such as the maximization of the products' market share or

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 295–336.
springerlink.com

the maximization of the company's profit. In real world applications, as the number of attributes and levels increases, the number of candidate product profiles can grow uncontrollable large, making the managerial task for selecting the appropriate combination of attribute levels practically infeasible. Actually the optimal product line design problem has been proved to be NP-hard, which means that the complete solution space cannot be enumerated in polynomial time. In this context, a number of different heuristic approaches have been applied to solve the problem from 1974 until today. Some of the algorithms have been incorporated to intelligent marketing systems, which assist a manager in such problems of high complexity. Several papers have been published that present a specific algorithm and compare its performance with one or more other approaches. However, the comparison concerns only the approximation of the optimal solution, whereas marketing practitioners who work on real problems are interested in a number of other more qualitative issues. No study has been published that reviews the advantages and disadvantages of the algorithms that have been applied to the problem, while little work has been done regarding the evaluation of the relevant marketing systems.

In this chapter, we aim at filling this gap by presenting an integrated work, which provides a detailed description of the product line design problem, reviews all the algorithms that have been applied to the problem along with the related marketing systems that incorporate them, and tries to draw valuable insights for the manager as well as the researcher. Emphasis is placed on the application of Genetic Algorithms to the problem, since it is the mostly used method in the literature, and constitutes the most advanced algorithm that has been incorporated into a marketing system. This work will help the marketing manager understand the problem, select the method that mostly fits to his requirements, and decide on whether he will use one of the existing marketing systems or he will need to develop a new system from the scratch, which will better satisfy his company's requirements. A real world application is presented extensively, in order to support the manager in such a decision. Furthermore, this chapter provides a guide to the interested researchers, describing the optimization algorithms applied to the problem, comparing their performance, and pointing out the potential areas for future work. The rest of the chapter is organized into five sections as follows. Section 2 provides a brief description of the problem and its main properties. In Section 3, the different formulations of the problem are described. The optimization algorithms that have been applied to the problem are presented extensively in Section 4, and their pros and cons are evaluated. In Section 5 we compare the performance of Genetic Algorithms and Simulated Annealing, using a real data set as well as a Monte Carlo simulation. The relevant marketing systems and programs that have been developed are presented in Section 6, and some conclusions are drawn. Finally, Section 7 provides an overview of the main conclusions of the study, and suggests areas of future research.

2 The Optimal Product (Line) Design Problem

The goal of the optimal product (line) design problem is the design of one (or more products), the introduction of which to the market will maximize a firm's objective (usually market share). This requires the proper modeling of customer preferences concerning the various product features. In particular, each product is represented as a bundle of attributes (features) which can take specific levels. A personal computer for example, consists of the attributes monitor, processor, hard disk, memory etc., the levels of which are illustrated in Table 1. Every individual has its own preferences; for example a civil engineer will probably choose a large monitor, whereas a mathematician may select a fast processor. Customer preferences are represented as values (called *part-worths*) for each attribute level. An example is given in Table 1.

Table 1 Part-worths for each attribute level of a personal computer

Attributes	Levels	Partworths	
		Customer1	*Customer2*
Monitor	17''	0.8	0.1
	19''	0.2	0.3
	20''	0.3	0.4
	24''	0.5	0.9
Processor	Single-core 3,8 GHz	0.1	0.2
	Core-2 2,6 GHz	0.3	0.3
	Core-4 2Ghz	0.9	0.5
Hard disk	200 GB	0.4	0.2
	500 GB	0.6	0.3
	750 GB	0.7	0.5
	1 T	0.4	0.8
Memory	2 GB	0.2	0.1
	4 GB	0.4	0.3
	6 GB	0.9	0.4
Mouse	Cable	0.3	0.1
	Wireless	0.4	0.9
Camera	Embedded	0.3	0.8
	No camera	0.2	0.2

Before making a choice among competing products, a consumer is assumed to implicitly assign a utility value to each, by evaluating all its attributes in a simultaneous compensative manner. On the basis of the above representation scheme, the utility value of a product is the sum of the part-worths of the corresponding attribute levels. The higher the product's utility, the higher the probability to be chosen. Suppose that the two customers whose part-worths are presented in Table 1, have to select between PC1 (17'', core-4 2GHz, HD 750 GB, 6 GB RAM, cable mouse, no camera) and PC2 (24'', Single-core 3,8 GHz, HD 200 GB, 6 GB RAM, wireless mouse, embedded camera). Customer 1 will probably choose PC1 (utility=3.8) over PC2 (utility=2.5), whereas Customer 2 will probably choose PC2 (utility=3.4) over PC1 (utility=1.8). The utilities are converted to choice probabilities for each product through *choice models*, and are then aggregated across the whole customer base to provide hypothetical market shares. If we know the part-worths for a population of consumers, we can simulate the introduction of different product configurations (combinations of attribute levels) to the market and estimate conditional market shares. With the use of optimization algorithms we can find the product(s) that maximizes a firm's market share, given the customer preferences and the configuration of the competitive products in the market. An example could be a new car manufacturer who is interested in introducing 3 new car models in different categories (Sport, SUV, Station Wagon) that will provide him with the highest possible volume sales. The customer preferences are usually revealed through market surveys, the results of which are entered into preference disaggregation methods like Conjoint Analysis, which estimate part-worths for each individual. In the Optimal Product (line) Design problem the part-worths for each customer, as well as the competitive product profiles are considered known, and the aim is to find the product(s) configuration that maximizes a specific firm's criterion. Next, we describe the different properties of the optimal product (line) design problem.

2.1 Choice Rule

The *choice rule* (or *choice model*) is the underlying process by which a customer integrates information to choose a product from a set of competing products. Different choice rules have been developed with varying assumptions and purposes and they differ in the underlying logic structure that drives them (Manrai, 1995). The choice rule models the consumer's purchasing pattern by relating preference to choice. It is a mathematical model which converts the product utilities that an individual assigns to the set of alternatives under consideration, to choice probabilities for each alternative. Choice rules can be either *deterministic* or *probabilistic*. The *first choice* or *maximum utility* is a deterministic rule, which assumes that the individual will always purchase the product with the highest utility. In this case the highest utility alternative receives probability of choice equal to 1, while the rest of the alternatives get a zero probability. Probabilistic rules on the other hand, assume that all alternatives receive a probability of choice in proportion to their utility value. Popular probabilistic choice models are:

- the Bradley-Terry-Luce (1952; Luce, 1959), $P_{ij} = U_{ij} \Big/ \sum_{j=1}^{n} U_{ij}$,
- and the MultiNomial Logit (McFadden, 1974), $P_{ij} = e^{U_{ij}} \Big/ \sum_{j=1}^{n} e^{U_{ij}}$,

where P_{ij} is the probability that consumer i selects product j, U_{ij} is the utility he assigns to product j, and n is the number of competing products. The first approaches applied to the problem employed the maximum utility rule, which is still widely used in product design applications due to its simple form. Its main limitation is that it tends to exaggerate the market share of popular alternatives while underestimating the unpopular ones. Probabilistic models have not received much attention in the specific problem, as they increase the algorithm's complexity (the problem becomes non-linear). The kind of choice model used affects the problem formulation, as we will see in a later section.

2.2 Optimization Criteria

The first criterion introduced was the maximization of a company's market share, also known as *share of choices* (Shocker and Srinivasan, 1974), which remains the most frequently used objective until today. Later, two more criteria were presented, the *buyer's welfare* (Green and Krieger, 1988) and the *seller's welfare* (Green *et al.*, 1981). In the former, no competition is assumed, and the aim is the maximization of the sum of the utilities that products under design offer to all customers. This is the least frequently used criterion, which mainly concerns product and services offered by public organizations. In the seller's welfare, the goal is the maximization of a firm's profit. This is the most complicated criterion since it requires the incorporation of the marginal return that the firm obtains from each attribute level into the objective function.

2.3 Number of Products to be Designed

The *optimal product design* problem (one product to be designed) was first formulated by Zufryden (1977). Eight years later Green & Krieger (1985) introduced the *optimal product line design problem* (two or more products to be designed), which is the main focus of the specific research area today.

2.4 Procedure Steps

The optimal product line design problem can be formulated either as a one-step or a two-step approach. In the latter, a reference set of candidate alternatives is first defined, and the items that optimize a certain criterion are selected next with the use of a specific algorithm (Green & Krieger, 1985). The problem here is to decide on the size of the reference set of products, and the way that it will be constructed in order to include all potential good solutions. Nowadays, the increase in computers' speed, as well as the development of advanced optimization algorithms, has enabled the design of the items that comprise the line directly from part-worth data in a one-step approach (Green & Krieger, 1988).

2.5 Optimization Algorithm

In real world applications, as the number of attributes and levels increases, the number of different product profiles raises exponentially, making the selection of the appropriate combination of attribute levels a very complex managerial task. For example in a product category with 7 attributes each taking 6 different levels, the number of possible product profiles is 279,936, while for designing a line of 3 products the number of candidate solutions is over a trillion. The exponential increase in the number of candidate solutions with the increase in the number of attributes and levels is illustrated in Table 2 (Alexouda, 2004), where K is the number of attributes, and J is the number of levels.

Kohli and Krishnamurti (1989) proved that the share of choices for the single product design problem is NP-hard, which means that the complete enumeration of the solution space is practically infeasible in tractable time. Kohli and Sukumar (1990) proved the same for the buyer's welfare and the seller's welfare, also for the single product design. In this context many different heuristic approaches have been applied to the problem from 1974 until today, the most significant of which are illustrated in Table 3.

Table 2 The number of possible solutions (products and product lines) of different problem sizes (source: Alexouda, 2004)

Products in line	K	J	Number of possible products	Number of possible product lines
2	3	4	64	2016
2	4	3	81	3240
2	4	4	256	32,640
2	5	3	243	29,403
3	3	4	64	41,664
3	4	3	81	85,320
3	4	4	256	2,763,520
3	5	3	243	2,362,041
2	5	4	1024	523,776
2	5	5	3125	4,881,250
2	5	6	7776	30,229,200
2	6	4	4096	8,386,560
2	6	5	15,625	122,062,500
2	6	6	46,656	1,088,367,840
2	7	4	16,384	134,209,536
2	7	5	78,125	3,051,718,750
2	7	6	279,936	39,181,942,080
2	8	4	65,536	2,147,450,880
2	8	5	390,625	76,293,750,000
2	8	6	1,679,616	1,410,554,113,920
3	5	4	1024	178,433,024
3	5	5	3125	5,081,381,250
3	5	6	7776	78,333,933,600
3	6	4	4096	11,444,858,880
3	6	5	15,625	635,660,812,500
3	6	6	46,656	16,925,571,069,120
3	7	4	16,384	732,873,539,584
3	7	5	78,125	79,469,807,968,750
3	7	6	279,936	3,656,119,258,074,240
3	8	4	65,536	46,910,348,656,640
3	8	5	390,625	9,934,031,168,750,000
3	8	6	1,679,616	789,728,812,499,209,000

Table 3 Approaches applied to the optimal product (line) design problem

Paper	**Choice rule**	**Objective**	**Steps**	**Algorithm**	**Prod-ucts**	**System**
Shocker & Srinivasan 1974	Deterministic	Share, Profit	One	Gradient search, Grid search	Single	
Zufryden 1977	Deterministic	Share	One	Mathematical programming	Single	ZIPMAP
Green, Carroll & Goldberg 1981	Deterministic, Probabilistic	Share, Profit	One	Response Surface methods	Single	QUALIN
Green & Krieger 1985	Deterministic	Share	Two	Greedy Heuristic, Interchange Heurist	Line	DESOP, LINEOP
Kohli & Krishnamusti 1987	Deterministic	Share	One	Dynamic Programming	Single	
Green & Krieger 1988	Probabilistic	Share, Profit, Buyers welfare	One	Divide & Conquer	Line	SIMOPT
McBride & Zufryden 1988	Deterministic	Share	Two	Mathematical programming	Line	
Sudharshan, May & Gruca 1988	Deterministic, Probabilistic	Share	One	Non linear programming	Line	DIFFSTRAT
Green, Krieger & Zelnio 1989	Probabilistic	Share	Two	Coordinate Ascent	Line	PROSIT
Kohli & Sukumar 1990	Deterministic	Share, Profit, Buyers welfare	One	Dynamic Programming	Line	
Dobson & Kalish 1993	Deterministic	Share, Profit, Buyers welfare	Two	Greedy Heuristic	Line	
Nair, Thakur & Wen 1995	Deterministic	Share, Profit	One	Beam Search	Line	

Table 3 *(Cont.)*

Balakrishnan & Jacob 1996	Deterministic	Share, Buyers welfare	One	Genetic algorithm	Single	
Chen & Hausman 2000	Probabilistic	Profit	Two	Non linear programming	Line	
Alexouda & Paparrizos 2001	Deterministic	Profit	One	Genetic algorithm	Line	MDSS
Shi, Olafsson & Chen 2001	Deterministic	Share	One	Nested Partitions	Line	
Krieger & Green 2002	Probabilistic	Share	One	Greedy Heuristic	Single	
Steiner & Hruschka 2003	Probabilistic	Profit	One	Genetic algorithm	Line	
Alexouda 2004	Deterministic	Share	One	Genetic algorithm	Line	MDSS
Balakrishnan, Gupta & Jacob 2004	Deterministic	Share	One	Genetic algorithm	Line	
Camm, Cochran, Curry & Kannan 2006	Deterministic	Share	One	Branch and Bound with Lagrangian relaxation	Single	
Belloni, Freund, Selove & Simester 2008	Deterministic	Profit	One	Branch and Bound with Lagrangian relaxation		

3 Problem Formulation

The formulation of the problem depends on the employed choice rule and the selected optimization criterion.

3.1 Deterministic Choice Rules

The most common approach found in the literature is the share of choices problem for the optimal product line design using the first choice rule.

3.1.1 Share of Choices

Here, each individual is assumed to have an existing favorite product called *status quo*. The problem can be formulated as a 0-1 integer program, with the use of the following parameters (Kohli & Sukumar, 1990):

$\Omega = \{1,2,\ldots, K\}$ is the set of K attributes that comprise the product.
$\Phi_\kappa = \{1,2,\ldots, J_\kappa\}$ is the set of J_κ levels of attribute k.
$\Psi = \{1,2,\ldots M\}$ is the set of products to be designed.
$\theta = \{1,2,\ldots I\}$ is the set of I customers.
w_{ijk}= is the part-worth that customer $i \in \theta$ assigns to level $j \in \Phi_\kappa$ of attribute $k \in \Omega$.

j^*_{ki} = is the level of attribute $k \in \Omega$ of customer's $i \in \theta$ status quo product.
$c_{ijk} = w_{ijk} - w_{ij*k}$ is the relative difference in the part-worth that customer $i \in \theta$ assigns between level j and level $j*$ of attribute $k \in \Omega$.

Since the firm may already offer a number of products, we index as $\theta' \subset \theta$ the set of customers whose current status quo product is offered by a competitor. In this way the company aims at gaining the maximum possible number of clients from its competitors, without cannibalizing its existing product line. Three decision variables are also used:

$$x_{jkm} = \begin{cases} 1, \text{if the level of product's m attribute k is j,} \\ 0, \text{otherwise} \end{cases}$$

$$x_{im} = \begin{cases} 1, \text{if product's m utility for customer i is less than his status quo,} \\ 0, \text{otherwise} \end{cases}$$

$$x_i = \begin{cases} 1, \text{if customer i does not choose to switch from his status quo,} \\ 0, \text{otherwise} \end{cases}$$

In this context the share of choices problem in the product line design using a deterministic rule is formulated as follows:

$$\min \sum_{i\in\theta'} x_i \tag{1}$$

subject to

$$\sum_{j\in\Phi_\kappa} x_{jkm} = 1, \quad k\in \Omega,\ m\in \Psi, \tag{2}$$

$$\sum_{k\in\Omega}\sum_{j\in\Phi_\kappa} c_{ijk}x_{jkm} + y_{im} > 0, \quad i\in \theta',\ m\in \Psi, \tag{3}$$

$$x_i - \sum_{m\in\Psi} x_{im} \geq 1 - M, \quad \forall\, i\in \theta' \tag{4}$$

$$x_{jkm},\ x_{im},\ x_i = 0, 1 \text{ integer},\ i\in \theta',\ j\in \Phi_\kappa,\ k\in \Omega,\ m\in \Psi. \tag{5}$$

Constraint (2) requires each product in the line to be assigned exactly one level of each attribute. Constraint (3) restricts x_{im} to be 1 only if customer i prefers his status quo to product m. Constraint (4) forces x_i to be 1 only if x_{im} = 1 for all $m\in \Psi$, that is if customer i prefers his status quo to all products in the line. Constraint (5) represents the binary restrictions regarding the problem's decision variables. The objective function (1) minimizes the number of instances for which x_i = 1, and hence minimizes the number of customers who decide to be loyal to their status quo products (which is equivalent to maximizing the number of customers who switch from their status quo to a product from the company's line).

3.1.2 Buyer's Welfare

In this case no status quo product is assumed for the customer (buyer), who will select the item from the offered line that maximizes his utility. The following decision variable is used:

$$x_{ijkm} = \begin{cases} 1, \text{if level j of attribute k appears in product m, and buyer i,} \\ 0, \text{otherwise} \end{cases}$$

The problem can be formulated as a 0-1 integer program as follows (Kohli & Sukumar, 1990):

$$\max \sum_{i\in\theta}\sum_{m\in\Psi}\sum_{k\in\Omega}\sum_{j\in\Phi_\kappa} w_{ijk}x_{ijkm} \tag{6}$$

subject to

$$\sum_{j\in\Phi_\kappa}\sum_{m\in\Psi} x_{ijkm} = 1, \quad i\in \theta,\ k\in \Omega, \tag{7}$$

$$\sum_{j\in\Phi_\kappa} x_{ijkm} - \sum_{j\in\Phi_{\kappa'}} x_{ijk'm} = 0, \quad k'>k,\quad k,k'\in \Omega,\ i\in \theta,\ m\in \Psi, \tag{8}$$

$$x_{ijkm} + x_{i'j'km} \leq 1, \quad i'>i, j'>j, \ i,i' \in \theta, j,j' \in \Phi_\kappa, k \in \Omega, \ m \in \Psi, \tag{9}$$

$$x_{ijkm} = 0, 1 \text{ integer}, \ i \in \theta, j \in \Phi_\kappa, \ k \in \Omega, \ m \in \Psi \tag{10}$$

Constraint (7) requires that, across products, only one level of an attribute be associated with a specific buyer. Constraint (8) requires that, across attributes, the level assigned to buyer $i \in \theta$ must correspond to the same product. Constraint (9) requires that for all buyers assigned to a specific product, the same level of an attribute must be specified. Together, these three constraints require that each buyer be assigned one of the products in the line. The objective function (6) selects the products (attribute levels combination) to maximize the total utility across buyers.

3.1.3 Seller's Welfare

Kohli & Sukumar (1990) provide a detailed description of the seller's welfare problem, where the firm wants to maximize the marginal utility obtained by the introduction of a line of M new products. The seller may already offer some products in the market, and competition is represented through the existence of a current status quo product for each customer. If customer $i \in \theta$ selects a product in which level $j \in \Phi_\kappa$ of attribute $k \in \Omega$ appears, the seller is assumed to obtain a part-worth value u_{ijk}. The seller's marginal return obtained from level $j \in \Phi_\kappa$ of attribute $k \in \Omega$ is:

- $d_{ijk} = u_{ijk} - u_{ij*k}$, if customer $i \in \theta$ switches from a product offered by the seller
- $d_{ijk} = u_{ijk}$ if customer $i \in \theta$ switches from a product offered by a competitor

The problem can be formulated as a 0-1 integer program as follows:

$$\max \sum_{i\in\theta}\sum_{m\in\Psi}\sum_{k\in\Omega}\sum_{j\in\Phi_\kappa} d_{ijk}\, x_{ijkm}\, y_i \tag{11}$$

subject to

$$\sum_{m\in\Psi}\sum_{k\in\Omega}\sum_{j\in\Phi_\kappa} w_{ijk}\,(x_{ijkm} - x_{i'jkm}) \geq 0, \quad i \neq i', \ i \in \theta, \tag{12}$$

$$y_i \sum_{m\in\Psi}\sum_{k\in\Omega}\sum_{j\in\Phi_\kappa} w_{ijk}\, x_{ijkm} \geq y_i u_i^*, \quad i \in \theta \tag{13}$$

$y_i = 0, 1$ integer, and (7), (8), (9), (10).

Constraints (7)-(10) require, as in the buyer's welfare problem, that a specific product is assigned to each customer, and that each product in the line be assigned exactly one level of each attribute. Constraint (12) requires that each customer is assigned to the product that maximizes his utility. Constraint (13) requires that the seller obtains a return from customer i only if the utility of the new item assigned to the customer is higher than the utility of the u_i^* of his status quo product. The objective function (11) selects the products to maximize the seller's total return from the products in the line.

3.2 Probabilistic Choice Rules

When probabilistic choice rules are used, the market is assumed to consist of N competitive products with known configurations, including the M candidate items for the firm's line:

$\Xi = \{1,2,\ldots N\}$ is the set of products that comprise the market.

3.2.1 Share of Choices

As before $\Psi \subset \Xi$ is the set of products to be designed. Customers do not have a status quo product, and do not deterministically choose the highest utility product. Instead, we assume that each of the N alternatives has a certain probability to be selected, which is calculated with the use of a probabilistic choice model. Using BTL for example, the probability that customer i will choose product m is estimated as follows:

$$P_{im} = U_{im} \Big/ \sum_{n \in \Xi} U_{in}, \quad i \in \theta,\ m \in \Psi,\ n \in \Xi, \tag{14}$$

where U_{im} the utility that customer i assigns to product m (sum of its part-worths):

$$u_{im} = \sum_{k \in \Omega} \sum_{j \in \Phi_\kappa} w_{ijk} x_{jkm}, \quad i \in \theta,\ j \in \Phi_\kappa,\ k \in \Omega,\ m \in \Psi.$$

In this context the problem is formulated as the following non-linear program:

$$\max \sum_{m \in \Psi} \sum_{i \in \theta} P_{im} \tag{15}$$

subject to

$x_{jkm} = 0$, 1 integer, and (2).

The objective function (15) maximizes the market share of the m products (probability to be purchased) of the company's line.

3.2.2 Seller's Welfare

Green and Krieger (1992) presented the seller's welfare problem, in an application of the SIMOPT program to pharmaceutical products. In order for the company's profit to be maximized, variable (depending on attribute levels) and fixed costs for each product must be included in the objective function. The variable cost per unit for a product m is given by the following linear additive function:

$$c_m^{(\mathrm{var})} = \sum_{k \in \Omega} \sum_{j \in \Phi_\kappa} c_{jk}^{(\mathrm{var})} x_{jkm}, \quad j \in \Phi_\kappa,\ k \in \Omega,\ m \in \Psi,$$

where

$c_{jk}^{(\mathrm{var})}$ the variable cost of attribute's k level j for the seller.

A similar function is used for the fixed cost of product m:

$$c_m^{(fix)} = \sum_{k \in \Omega} \sum_{j \in \Phi_K} c_{jk}^{(fix)} x_{jkm} \quad , \; j \in \Phi_K,\ k \in \Omega,\ m \in \Psi.$$

If p_m denotes item's m price, the problem is formulated as the following non-linear program:

$$\max \sum_{m \in \Psi} \left[(p_m - c_{jk}^{(\mathrm{var})}) \sum_{i \in \theta} P_{im} I - c_m^{(fix)} \right] \tag{16}$$

subject to

$x_{jkm} = 0, 1$ integer, and (2).

The objective function (16) maximizes the total seller's profit obtained from the introduction of a line of M products.

4 Optimization Algorithms Applied to the Problem

In this section we review and evaluate the most important algorithms that have been applied to the optimal product line design problem.

4.1 Greedy Heuristic

Introduced by Green & Krieger (1985), this heuristic proceeds in two steps. At the first step a "good" set of reference products is created. The second step begins by choosing the best alternative among the candidate products. Then, the second alternative is selected from the reference set, which optimizes the objective function provided that the first product is already included in the market. The procedure iterates by adding one product at a time until the desired number of products in the line has been reached. In another paper, Green and Krieger (1987) describe the "best in heuristic" for developing the set of reference products. Initially the product profile that maximizes the utility u_{1max} of customer 1 is found through complete enumeration of the attribute levels. If customer's 2 utility for customer's 1 best product is within a user specified fraction ε of u_{2max}, then customer's 2 best product is not added to the set; otherwise it is. As the method proceeds through the group of customers, all of the products currently on the set are tested to see if any are within ε of u_{kmax} for customer k, and the previous rule is applied. The process is usually repeated through randomized ordering of the customers, and different values of ε, depending on the desired size of the set. Local optimality is not guaranteed, as it depends on the first product added to the line.

4.2 Interchange Heuristic

In the same paper, Green and Krieger (1985) introduced another method where initially, a product line is randomly selected and its value is estimated. Next, each alternative from the reference set is checked to see whether there exists a product in the line, the replacement of which by the specific alternative will improve the line's value. If this condition holds, the alternative is added, and the product that is removed is the one that results in the maximum improvement of the line's value. The process is repeated until no further improvement is possible. The authors recommend the use of the solution provided by the Greedy Heuristic, as the initial product line. The Interchange Heuristic guarantees local optimality, where the local neighborhood includes all solutions that differ from the existing by one product.

4.3 Divide and Conquer

In this approach, developed by Green and Krieger (1988), the set of attributes *K* that comprise the product line is divided into two equal subsets *K1* and *K2*. First, the levels of attributes belonging to *K1* that are good approximations of the optimal solution are estimated. The authors suggest averaging the part-worths within each level of each attribute, and selecting for each attribute the level with the highest average. In each iteration, the values of the attributes belonging to the one subset are held fixed, while the values of the other subset are optimized through an exhaustive search. If the search space is too large for completely enumerating half of the attributes, the set of attributes can be divided into more subgroups, at the risk of finding a worst solution. Local optimality is guaranteed, where the local neighborhood depends on the number of subsets.

4.4 Coordinate Ascent

Green *et al.* (1989), propose a heuristic that can be considered as a Coordinate Ascent implementation. A product line is initially formed at random and evaluated. The algorithm then iterates through each product attribute in a random order, and assesses each possible level. The altering of an attribute's level is acceptable if it improves the solution's quality. Only a single attribute change is assessed at a time (one opt version), and the algorithm terminates when no further improvement is possible. Local optimality is guaranteed, with the local neighborhood including all solutions that differ from the existing one by a single attribute.

4.5 Dynamic Programming

Kohli and Krishnamusti (1987), and Kohli and Sukumar (1990) use a dynamic programming heuristic for solving the optimal product and product line design problems respectively. Here, the product (line) is built one attribute at a time. Initially, for each level of attribute B, the best level of attribute A is identified, forming in this way a number of partial product profiles, equal to attribute's B number of levels. Next, for each level of attribute C, the best partial profile (consisting of attributes A and B) that was built in the previous step is identified. The method proceeds until all product(s) attributes have been considered. Finally, the product (line) that optimizes the desired criterion is selected among the full profiles constructed. The quality of the final solution is highly dependant to the order in which the attributes are considered, thus multiple runs of the heuristic using different attribute orderings are recommended. No local optimality is guaranteed.

4.6 Beam Search

Nair *et al.* (1995) solved the product line design problem using Beam Search. BS is a breadth-first process with no backtracking, where at any level of the search only the b (Bean Width) most promising nodes are further explored in the search tree. The method begins with K relative part-worth matrices $C(k)$ (with elements $c_{ij} = w_{ij} - w_{ij*}$), and initializes work matrices $A_1(\bullet)$ based on C. At each stage l (layer), matrices $E_l(\bullet)$ of combined levels are formed, by combining two matrices $A_l(\bullet)$ at a time in the given order. Then, the b most promising combinations of levels are selected to form columns in new matrices $A_{l+1}(\bullet)$ in the next layer, where it remains approximately half of the number of matrices in the previous layer. In this way, unpromising attribute levels are iteratively pruned, until a single work matrix remains. This final matrix consists of b columns, each containing a full product profile. These are the candidate alternatives for the first product in the line. For the first of the b alternatives, the data set is reduced by removing the customers who prefer this product over their status quo. The previous process is repeated for finding one second-product in the line, and iterated until M products are build that form a complete product line. The same procedure is repeated, until b complete product lines are designed, from which the one that gives the best value in the objective function is selected. The final solution depends on the way of pairing the different attribute combinations at each layer. The authors suggest a best-worst pairing, which gives better results than the random one. No local optimality is guaranteed.

4.7 Nested Partitions

In the Nested Partitions implementation (Shi *et al.*, 2001), a region is defined by a partial product line profile, for example all products that contain a specific attribute level. In each iteration a subset of the feasible region is considered the most promising, which is further partitioned into a fixed number of subregions, by determining the level of one more attribute, and aggregating what remains of the feasible region into one surrounding region. In each iteration therefore, the feasible region is covered by disjoint subregions. The surrounding region and each of the subregions are sampled using a random sampling scheme, through which random levels are assigned to the remaining attributes. The randomly selected product profiles are evaluated, in order for an index to be estimated that determines which region becomes the most promising in the next iteration. This region is then nested within the last one. If the surrounding region is found to be more promising than any of the regions under consideration, the method backtracks to a larger region using a fixed backtracking rule. NP combines global search through partitioning and sampling, and local search through calculation of the promising index. The method can incorporate other heuristics to improve its performance. The authors tried a Greedy Heuristic, as well as a Dynamic Programming into the sampling step, and a Genetic Algorithm into the selection of the promising region. The results of their study indicate that the incorporation of each of the three heuristics is beneficial, with GA giving the best performance.

4.8 Genetic Algorithms

Genetic Algorithms are optimization techniques that were first introduced by Holland (1975). They are based on the principle of "natural selection" proposed by Darwin a long time ago, and constitute a special case of Evolutionary Programming algorithms. In accordance with Biology science, GAs represent each solution as a chromosome that consists of genes (variables), which can take a number of values called *alleles*. A typical GA works as illustrated in Figure 1. Initially a set of chromosomes (population) is generated. If prior knowledge about the problem exists, we use it to create possible "good" chromosomes; else the initial population is generated at random. Next, the problem's objective function is applied to every chromosome of the population, in order for its fitness (performance) to be evaluated. The chromosomes that will be *reproduced* to the next generation are then selected according to their fitness score, that is, the higher the chromosome's fitness the higher the probability that it will be copied to the subsequent generation. Reproduction ensures that the chromosomes with the best performance will survive to the future generations, a process called "survival of the fittest", so that high quality solutions will not be lost or altered.

A mating procedure follows, where two parents are chosen to produce two offspring with a probability p_c, through the application of a *crossover* operator. The logic behind crossover is that a chromosome may contain some "good" features (genes) that are highly valued. If two chromosomes (parents) exchange their good features then there is a great possibility that they will produce chromosomes (offspring) that will combine their good features, thus creating higher performance solutions. The expectation is that from generation to generation, crossover will produce new higher quality chromosomes. Subsequently, each of the newly formed chromosomes is selected with a probability p_m to be *mutated*. Here one of its genes is chosen randomly and its value is altered to a new one randomly generated. Mutation produces new chromosomes that would never be created through crossover. In this way, entirely new solutions are produced in each generation, enabling the algorithm to search new paths and escape from possible local minima. Whereas reproduction reduces the diversity of the population, mutation maintains a certain degree of heterogeneity of solutions, which is necessary to avoid premature convergence of the evolutionary process. However, mutation rates must be kept low, in order to prevent the disturbance of the search process that would lead to some kind of random search (Steiner and Hruschka, 2003). Finally, if the convergence criterion has been met, the algorithm stops and the best solution so far is returned; else it continues from the population's evaluation step.

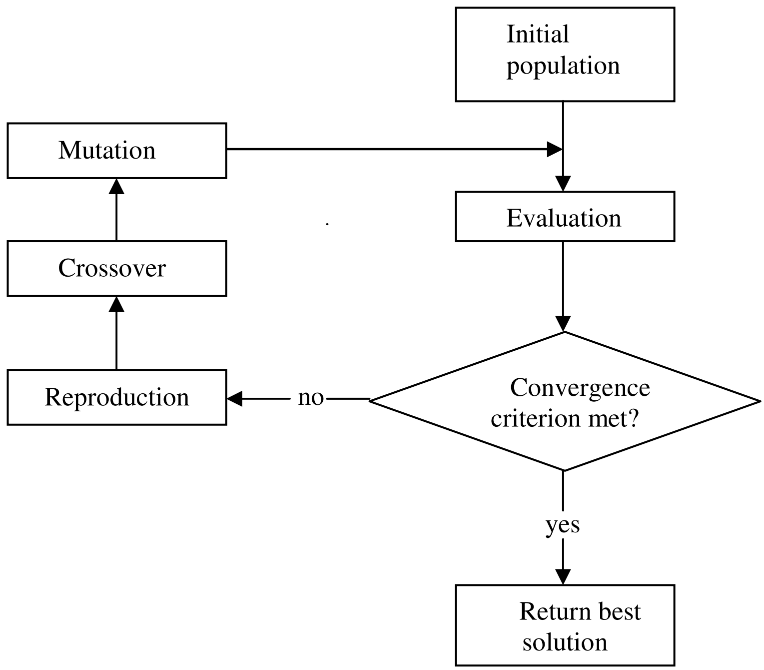


Fig. 1 Genetic Algorithm flowchart

4.8.1 Type of Problems Solved

GAs were first applied to the optimal product design problem by Balakrishnan and Jacob (1996), who dealt with the share of choices and the buyer's welfare problem, by employing the first choice rule. The authors provide a number of advantages that leaded them to use this approach. The search is implemented from a set of points (equal to the size of the population) rather than a single point, increasing in this way the method's exploratory capability. GAs do not require additional knowledge, such as the differentiability of the function; instead they use the objective function directly. GAs do not work with the parameters themselves but with a direct encoding of them, which make them especially suited for discontinuous, high-dimensional, and multimodal problem domains, like the optimal product design. Later, Alexouda and Paparrizos (2001) applied GAs to the seller's welfare problem for the optimal product line design, while Alexouda (2004), as well as Balakrishnan *et al.* (2004) dealt with the share of choices problem. All three approaches employed the first choice rule. The only approach that uses probabilistic choice rules is that of Steiner and Hruschka (2003), who dealt with the seller's welfare problem.

4.8.2 Problem Representation

Except for Balakrishnan *et al.* (2004), all other approaches adopted a binary representation scheme. In Balakrishnan and Jacob (1996), each product is represented by a chromosome, which is divided into K substrings that correspond to the product's attributes. Each substring consists of J_K (the number of attribute's k levels) genes that take values (alleles) 0 or 1. Hence the length of a chromosome is P=$\sum_{k\in\Omega} j_k$. A value of 1 denotes the presence of the specific level in the corresponding attribute, and a value of 0 its absence. This representation has the restriction that exactly one gene must take the value of 1 in each substring. Lets for instance assume that a personal computer consists of the attributes processor (Single-core 3,8 GHz, Core-2 2,6 GHz, Core-4 2Ghz), monitor (17'', 19'', 20'', 24''), and hard disk (200 GB, 500 GB, 750 GB). Then a Core-2 2,6 GHz with 20'' monitor and 750 GB hard disk will be represented by the chromosome C={010 0010 001}. In Alexouda and Paparrizos (2001), Steiner and Hruschka (2003), and Alexouda (2004), a chromosome corresponds to a line of products. Each chromosome is composed of $M*K$ substrings that represent the product's attributes, each consisting of J_K genes that take values 0 or 1. As before, a value of 1 denotes the presence of the specific level in the corresponding attribute, and a value of 0 its absence. The restriction that exactly one gene must take the value of 1 in each substring also holds here. The length of each chromosome is $P=M\sum_{k\in\Omega} j_k$. Referring to the personal computer example, the chromosome D={010 0010 001|100 0001 010} represents a line of two products; a Core-2 2,6 GHz with 20'' monitor and 750 GB hard disk, and a single-core 3,8 GHz with 24'' monitor and 500 GB hard disk. Balakrishnan *et al.* (2004) use an integer representation, where a chromosome corresponds to a line of products, a gene to an attribute, and the gene's values to attribute levels. Hence, each chromosome is of length $M*K$, and is divided into M substrings, each representing a product in the line. Within each substring, gene k can take J_K different values. The line of the two products described by chromosome D above, is represented in this case by chromosome E={233|142}. Here, the authors raise an issue concerning the principal of *minimal redundancy*, according to which each member of the space being searched should be represented by one chromosome only (Radcliffe, 1991). The integer representation scheme does not adhere to this principle, since the same line of products can be represented by M! different chromosomes. The previous PC product line, for instance, can also be represented by the chromosome E'={142|233} (the two products exchange their positions). This could cause inefficiencies in the search process, as the crossover between two identical products (E and E') may result in two completely different sets of offspring. On the other hand, it may prove to be an advantage, as more members of the search space will probably be explored. In order to alleviate this concern, they adopt an alternative representation scheme where the substrings (products) in a chromosome are arranged in lexicographic order. That is, product 111 is before 112 which is before 121 etc. In this encoding, called *sorted* representation, the chromosome E would not exist. They tested both the sorted and the unsorted representations.

4.8.3 Genetic Algorithm's Parameters

Balakrishnan and Jacob (1996) represented the problem with the use of matrices. The GA population (number of chromosomes) has a size of N, and is stored in the matrix POP_{N*P}. Customers' preferences (part-worths for each attribute level) are maintained in the matrix $BETA_{I*P}$. The utilities that each of the I customers assigns to each of the N products (represented by chromosomes) are estimated in each generation, and stored in the matrix PRODUTIL= $BETA*POP^T$. For the share of choices problem the utility of each customer's status quo product is maintained in the matrix STATQUO. The chromosome n is evaluated through the comparison of the n-th column in PRODUTIL with the corresponding in STATQUO. The fitness of the chromosome is the number of times that PRODUTIL(i,n)>STATQUO(i,n), i=1…I, that is the number of customers that prefer the new product to their status quo. For the buyer's welfare problem the fitness of the chromosome n is the sum of elements of the n-th column in PRODUTIL, that is the aggregate utility value for the whole set of customers.

4.8.3.1 Initialization of the population

All five approaches initialize the GA population in a totally random manner. Furthermore, Alexouda and Paparrizos (2001), Alexouda (2004), and Balakrishnan *et al.* (2004), also assess the performance of a hybrid strategy in respect to the initialization of the population. Before running the GA, a Beam Search heuristic is applied and the best solution found is seeded into the genetic algorithm's initial population, while the remaining N-1 chromosomes are randomly generated. The population size is set to 100 (Balakrishnan and Jacob, 1996), 150 (Alexouda and Paparrizos, 2001; Steiner and Hruschka, 2003), 180 (Alexouda, 2004), or 400 (Balakrishnan *et al.*, 2004).

4.8.3.2 Reproduction

Except for Steiner and Hruschka (2003), all other approaches adopt an elitist strategy for the process of reproduction, where the F fittest chromosomes are copied intact into the next generation. Such an approach ensures that the best chromosomes will survive to the subsequent generations. The value of F ranges from $4N/10$ (Alexouda and Paparrizos, 2001; Alexouda, 2004), to $N/2$ (Balakrishnan and Jacob, 1996; Balakrishnan *et al.*, 2004). Steiner and Hruschka (2003) employ a binary tournament selection procedure, where $N/2$ pairs of chromosomes are randomly selected with replacement, and from each pair only the chromosome with the higher fitness value survives to the succeeding generation. This is a semi-random process, which ensures that the chromosomes with higher fitness values have more probabilities to survive.

4.8.3.3 Crossover

In the approaches that adopt a binary representation scheme, the unit of interest in the crossover procedure is the substring, in order for feasible solutions to be produced. In Steiner and Hruschka (2003) for example, who use one-point crossover with probability p_c=0.9 and random selection of the cross site, the crossover of the two parents

A = {010 0010 001|100 0001 010} and
B = {100 0100 010|010 0010 100},

after the second substring will generate the two offspring

A' = {010 0010 010|010 0010 100} and
B' = {100 0100 001|100 0001 010}.

Except for the above approach, the other ones employ a uniform crossover with the probability p_c taking the values 0.4 (Alexouda and Paparrizos, 2001), 0.45 (Alexouda, 2004) and 0.5 (Balakrishnan and Jacob, 1996). In the approach that employs an integer representation scheme, the unit of interest in crossover is the gene. If for instance, the two parents

S={122|323} and
T={141|421},

exchange their second and sixth genes, this will generate the offspring

S'={142|321} and
T'={121|423}.

When the sorted representation is used, the offspring are sorted in lexicographic order after the crossover operation. According to Radcliffe (1991), a *forma* specifies at certain chromosome's positions (called defining positions) particular values that all its instances must contain. That is, if a chromosome η is an instance of a forma β, then η and β both contain the same values at the specified positions. Chromosomes S and T, for example, both belong to the forma:

$$\beta = 1** *2*,$$

where the * denotes a "don't care" value. The principle of *respect* defines that the crossover of two chromosomes that belong to same forma must produce offspring also belonging to the same forma. Whereas in the unsorted representation the crossover is "respectful", the property does not hold in the sorted representation, due to the ordering of the attributes after the crossover.

4.8.3.4 Mutation

Except for the one with the integer representation scheme, in all other approaches the mutation operator is applied at the substring level. Chromosomes are randomly selected (without replacement) with a probability p_m (mutation rate). An attribute (substring) of the selected chromosome is randomly picked and its level is altered. If, for instance, chromosome A is chosen to be mutated at the second substring, a potential mutated chromosome will be A''={010 1000 001|100 0001 010}. In Balakrishnan *et al.* (2004), the mutation takes place at the gene level, while two different mutation operators are used. Except for the standard mutation operator, a hybridized one is employed, which uses as a *mutator* chromosome the best solution found by the Beam Search heuristic. Whenever a chromosome is selected for mutation, a gene is randomly selected and its value is either randomly changed using the standard mutation operator, or altered to the value contained in the specific attribute of the mutator chromosome. In this way the good attribute values of the BS best solution will be copied to the GA population. On the other hand, this may result in premature convergence to the alleles of the mutator string. In order to avoid this, the two mutator operators have equal probability to be applied. The mutation rate takes a wide range of values: 0.05 (Steiner and Hruschka, 2003), 0.1 (Alexouda, 2004), 0.2 (Alexouda and Paparrizos, 2001), 0.3 (Balakrishnan and Jacob, 1996), or 0.4 (Balakrishnan *et al.*, 2004).

4.8.3.5 Stopping criterion

From the reproduced chromosomes, plus the offspring plus the mutated chromosomes, only the *N* fittest are maintained to the next generation, and the algorithm iterates until a stopping criterion is met. Balakrishnan and Jacob (1996), Steiner and Hruschka, (2003), and Balakrishnan *et al.* (2004) employ a moving average rule, where the algorithm terminates when the percentage change in the average fitness of the best three chromosomes over the five previous generations is less than 0.2% (convergence rate). In the other two approaches the procedure terminates when the best solution does not improve in the last 10 (Alexouda and Paparrizos, 2001), or 20 (Alexouda, 2004) generations.

4.8.4 Performance Evaluation

4.8.4.1 Genetic Algorithm vs. Dynamic Programming

Balakrishnan and Jacob (1996) compared the results of their approach and the Dynamic Programming approach (Kohli and Krishnamusti, 1987) with the complete enumeration solutions in 192 data sets, in both the share of choices and buyer's welfare problems. A full factorial experimental design was generated using the factors and levels presented in Table 4.

Table 4 Factors and levels used in the experiment

<table>
<tr><th>Factor</th><th colspan="12">Levels</th></tr>
<tr><td>Number of attributes</td><td colspan="4">4</td><td colspan="4">6</td><td colspan="4">8</td></tr>
<tr><td>Number of attribute levels</td><td colspan="3">2</td><td colspan="3">3</td><td colspan="3">4</td><td colspan="3">5</td></tr>
<tr><td>Number of customers</td><td colspan="3">100</td><td colspan="3">200</td><td colspan="3">300</td><td colspan="3">400</td></tr>
</table>

The part-worths were randomly generated following a normal distribution, and normalized within each customer to sum to 1. Random was also the generation of each customer's status quo product. Four replications were performed in each case resulting in a total of 192 data sets. In the share of choices problem, the average best solution provided by GA was 99.13% of the optimal product profile found by complete enumeration, while the same value for the DP was 96.67%. GA also achieved a tighter standard deviation (0.016) than that of DP (0.031). In the buyer's welfare problem the respective values were 99.92% for the GA with 0.0028 std, and 98.76% for the DP with 0.0165 std. The number of times that the optimal solution was found (hit rate) was 123 for the GA and 51 for the DP in the share of choices, and 175 for the GA and 82 for the DP in the buyer's welfare. The performance of GA was also compared with that of DP in two larger problems of sizes 326,592 and 870,912, where an exhaustive search was infeasible in tractable time. The data sets consisted of 200 customers, and 9 attributes that take (9,8,7,6,2,2,3,3,3) or (9,8,8,7,6,2,2,3,3) levels, while ten replications for each data set were performed. GA showed a better, worse, and equal performance compared to DP in 11, 3, and 6 data sets for the share of choices, and in 8, 3, and 9 data sets respectively for the buyer's welfare.

4.8.4.2 Genetic Algorithm vs. Greedy Heuristic

Steiner and Hruschka (2003) compared the results of their approach and the Greedy Heuristic approach (Green and Krieger 1985) with the complete enumeration solutions, in the seller's welfare problem. A factorial experimental design was generated using the factors and levels presented in Table 5.

Table 5 Factors and levels used in the experiment

Factor	**Levels**		
Number of attributes	3	4	5
Number of attribute levels	2	3	4
Number of products in the line	2	3	4
Number of competing firms	1	2	3

From the 81 different cases a subset of 69 was considered. Four replications were performed under each case, resulting in a total of 276 problems solved, where customer part-worths, attribute level costs, and competitive products configuration were randomly generated. The value of the solution found by GA was never less than 96.66% of the optimal (minimum performance ratio), while the corresponding value for the GH was 87.22%. The optimal solution was found in 234 cases by the GA, and in 202 cases by the GH, which corresponds to a hit ratio of 84.78% and 73.19% respectively. The solution found by GA was strictly better than that found by GH in 66 cases, and strictly worse in only 25.

4.8.4.3 Genetic Algorithm vs. Beam Search

Alexouda and Paparrizos (2001), Alexouda (2004), and Balakrishnan *et al.* (2004) compared the performance of GA with that of BS, which was considered the state of the art approach of the time. The first two approaches make a comparison of the two methods with a full search method in the seller's welfare and share of choices problems respectively. Eight small problems were solved using different values for the number of products in the line (2, 3), number of attributes (3, 4, 5, 6, 7, 8), and number of levels (3, 4, 5, 6). Ten replications were performed in each case, while the number of customers was kept constant to 100. The results are shown in Table 6.

Table 6 Results of the comparison of the two methods

	Seller's welfare	**Share of choices**
GA found optimal	73.75%	77.50%
BS found optimal	41.25%	45%
GA outperforms BS	53.75%	33.75%
BS outperforms GA	12.50%	12.50%
GA/optimal	0.9958	0.9951
BS/optimal	0.9806	0.9882

Furthermore, they compared the performance of a GA with completely random initialization (GA1), a GA where the initial population is seeded with the best BS solution (GA2), and a BS heuristic, in problems with larger sizes where complete enumeration is unfeasible. The number of customers was set to either 100 or 150 (Table 7).

Table 7 Results of the comparison of the three methods

	Seller's welfare		**Share of choices**	
	I=100	*I=150*	*I=100*	*I=150*
GA1 outperforms BS	93.88%	93.33%	47.92%	53.33%
BS outperforms GA1	6.11%	5.83%	33.33%	31.25%
GA2 outperforms BS	86.66%	80.83%	40%	43.33%
GA1 outperforms GA2	-	-	31.67%	35%
GA2 outperforms GA1	-	-	45.83	43.33%
GA1/ BS	1.0962	1.0794	-	-
GA2/ BS	1.0853	1.0702	-	-

Balakrishnan *et al.* (2004) defined eight different types of GA and hybrid GA procedures (Table 8).

Table 8 Genetic Algorithm techniques defined

Type	**Repre-sentation**	**Integration with BS**	
		Hybrid Mutation	*Seed with BS*
GASM	Unsorted	No	No
GASSM	Sorted	No	No
GAHM	Unsorted	Yes	No
GASHM	Sorted	Yes	No
GASMBS	Unsorted	No	Yes
GASSMBS	Sorted	No	Yes
GAHMBS	Unsorted	Yes	Yes
GASHMBS	Sorted	Yes	Yes

A 2x2 full factorial experimental design was employed using the factors number of products in the line (4 or 7), and number of attributes (7 or 9), with respective attribute levels (6 3 7 4 5 3 3) and (7 3 5 5 6 3 3 7 5), while the number of customers was 200. Two replications were performed in each case. The values of GA parameters are illustrated in Table 9.

Table 9 Values of the Genetic Algorithm parameters

Parameter	**Value**
Mutation rate	0.04
Population size	400
Number of attributes to crossover (N=4, K=7)	10
Number of attributes to crossover (N=4, K=9)	17
Number of attributes to crossover (N=7, K=7)	12
Number of attributes to crossover (N=7, K=9)	21
Number of generations	500

After experimentation it was found that a mutation rate less than 0.04 resulted in a premature convergence to suboptimal solutions, while higher values did not offered a substantial improvement. In addition, higher number of attributes to crossover was more beneficial in problems with smaller number of products in the line, as compared to problems with larger product lines. The results are presented in Table 10.

Table 10 Results of the comparison of the 10 methods

Method	**Best solution found (percentage of cases)**	**Average approximation of best solution**
GASM	12.5%	94.44%
GASSM	12.5%	94.21%
GAHM	12.5%	94.16%
GASHM	12.5%	94.15%
GASMBS	25%	94%
GASSMBS	0	93.35%
GAHMBS	0	92.82%
GASHMBS	0	92.32%
BS	0	89.53%
CPLEX	50%	82.68%

Another full factorial design (2x2x2) was employed, in order to asses the impact of the number of products in line (4 or 7), the number of attributes (7 or 9), and the presence or absence of attribute importance, to the following variables of interest:

- The best GA solution.
- The ratio of the best GA solution to the best BS solution.
- The number of unique chromosomes in the final population:
    - With the best fitness.
    - With fitness within the 5% of the best solution.
    - With fitness between the 5% and 10% of the best solution.
- The worst chromosome in the final population.
- The average fitness in the final population.
- The standard deviation of chromosomes' fitness in the final population.
- The number of generation at which the best solution was found.

Two product lines are considered different when at least one product exists in the one but not in the other, while two products are considered to be different if they differ in the level of at least one attribute. Ten replications were performed in each case resulting in a total of 80 data sets. The eight GA instances, as well as the BS heuristic, were run 10 times for each data set, hence 6400 different GA runs were performed. The results showed that GA techniques performed better or equally well as compared to BS in the 6140 cases (95.93%), performed strictly better in 5300 (82.81%), and underperformed in 260 (4.07%). The best GA solution reached a maximum difference of 12.75% with that of the BS, and was on average 2.6% better. The maximum difference reached when the BS solution was better was 6.1%. The hybridized GA methods always produced solutions at least as good as the BS solution, and in the 80.2 % of cases produced strictly better solutions. An interesting finding is that GA techniques which employ the unsorted representation, the standard mutation, and do not seed initial population with the best BS solution, showed the best average performance. A possible reason is the fact that the sorted representation scheme does not adhere to the principle of respect regarding the crossover operation. In addition, the incorporation of the best BS solution into the initial GA population, as well as the hybrid mutation operator probably make the algorithm converge to an area of solutions around the seeded BS solution, which in some cases may be suboptimal. Some loss in diversity of the final population may also be exhibited, as the integrated techniques displayed the worst results in respect to the number of unique chromosomes in the final population. Furthermore, integrated techniques suffer from premature convergence, as they tend to produce the best solution earlier, and result in the lowest standard deviation of chromosomes' fitness in the final population. Particularly, GA techniques without any hybridization (GASM, GASSM) provided final solutions at least as good as that of the hybridization techniques in 52.37% of cases on average, and strictly better on 35.12%. This indicates that the integration with the BS heuristic does not improve the quality of the solution. The number of products and number of attributes significantly affect ($p<0.0001$) the best GA solution, the ratio of the best GA solution to the best BS solution, all three measures of unique chromosomes in the final population, the standard deviation of chromosomes' fitness in the final population, and the number of generation at which the best solution was found; all in the positive direction. Finally, the presence of attribute importance has a statistically significant impact on the best GA solution, and the ratio of the best GA solution to the best BS solution.

4.8.5 Sensitivity Analysis

Balakrishnan and Jacob (1996) conducted a sensitivity analysis of the GA performance to changes in the values of its parameters, employing both the share of choices and the buyer's welfare criterion. A full factorial experimental design was generated using the factors and levels presented in Table 11.

Table 11 Factors and levels included in the experiment

Factor	Levels				
Mutation rate	0	0.01	0.1	0.25	0.3
Attributes participating in the crossover	0	K/4	K/2	3K/4	
Population size	50	100	200		
Degree of improvement in stopping rule	2%	0.2%			

The product category was assumed to consist of 8 attributes, each taking 5 levels, while the number of customers was set to 400. For each of the two problems a total of 120 GA runs were performed. In the share of choices, the average best solution provided by GA was 96.8% of the optimal product profile found by complete enumeration, and was found after 7.35 iterations (generations) on average. Hence GA reaches a near optimal solution by evaluating only the one fourth of the percent of the total number of possible solutions, which for the specific problem is 390625. Analyses of variance were performed to assess the impact of the four parameters to the quality of the solution. A main effects model had an R^2 of 0.504 and was statistically significant ($p<0.05$). Larger population sizes result in higher fitness of the best chromosome. As the number of attributes participating in the crossover increase, the quality of the solution also increases. As it was expected the tightening of the convergence parameter from 2% to 0.2% improves the fitness of the best solution. Whereas mutation rate had no significant main effect ($p=0.175$), the best algorithm's performance was achieved at the highest mutation rate. Similar results concerning the parameters' impact in the solution's quality were exhibited in the buyer's welfare problem, where a main effects model had an R^2 of 0.724 and was statistically significant ($p<0.05$). The average best solution provided by GA was 97.9% of the optimal product profile found by complete enumeration, and was found after 8.48 iterations on average. Steiner and Hruschka (2002) in another paper studied the sensitivity of the approximation of the optimal solutions w.r.t. varying parameter values for different problem sizes. A 12x5x3 factorial experiment was designed with 12 values of population size in the range [30, 250] at increments of 20 chromosomes, 5 different crossover probabilities (0.6, 0.7, 0.8, 0.9, 1), and 3 values of mutation rate (0, 0.01, 0.05). The size of the search space varied from 12650 to 10586800 feasible product lines, depending on the number of products in the line (2, 3, 4), number of attributes (2, 3, 4), and number of levels (4, 5). The recommended GA parameter values depending on the problem size after more than 1500 test runs are illustrated in Table 12.

Table 12 Recommended GA parameter values

Problem size	***12650***	***79800***	***161700***	***3921225***	***10586800***
Population size	130	150	230	250	250
Approximation of optimal	99.5%	99%	98.3%	99.2%	97.5%
Crossover probability	1	0.9	1	1	1
Approximation of optimal	99%	98.4%	97.6%	98.6%	96.8%
Mutation rate	0.05	0.05	0.05	0.01	0.01
Approximation of optimal	98.9%	98.8%	97.7%	98.5%	96.8%

4.9 Lagrangian Relaxation with Branch and Bound

Camm *et al.* (2006) introduced a computationally efficient algorithm that guarantees global optimality in the share of choices problem for designing a single product. They developed an exact method that uses *Lagrangian Relaxation* with *Branch and Bound* for finding provable optimal solutions to large scale problems, using a deterministic choice rule. Branch and Bound (Land and Doig, 1960) constitutes an optimization algorithm mainly used in discrete and combinatorial problems, which attempts to discard large subsets of the entire set of feasible solutions without enumeration, by proving that the global optimal solution cannot be contained in them. This procedure requires the estimation of lower and upper bounds of the objective being optimized, so that the search is limited to promising regions only. When the lower bound exceeds the upper bound in a certain branch, then this branch is excluded from further search. In order to calculate upper bounds the authors use Lagrangian Relaxation, a method that "relaxes" hard problem constraints in order to create another problem that is less complex than the initial. The constraints are moved into the objective function and a penalty is added to the fitness of the solution if they are violated. The upper bounds provide an indication of the quality of any feasible solution compared to the (unknown) optimal. The lower bounds are created using heuristics that generate feasible solutions. The proposed method is initialized with the use of a greedy algorithm that finds a feasible solution. Next, a lagrangian dual problem is defined, by relaxing constraint (3), and the subgradient optimization procedure of Downs and Camm (1996) is used for the estimation of the values of the associated lagrangian multipliers. They use this lagrangian problem as a quick attempt to improve on the initial greedy solution. The search tree is initialized with the use of the best solution between the greedy and lagrangian generated one, and a depth first strategy is employed. The algorithm branches on constraints (2) in ascending order with respect to their cardinality (number of levels within attribute). In this way, each level of the search tree corresponds to an attribute. The authors use several logic rules to develop and prune the search tree, in order to significantly decrease the number of variables on which they branch, thereby reducing the time required to solve problems to verifiable optimality. The algorithm found and verified the global optimum solution to 1 real and 32 simulated problems with as many as 32 attributes and 112 levels. The required time ranged from 1.4 seconds to 40 minutes, depending on the problem complexity.

Belloni *et al.* (2008) proposed a Lagrangian Relaxation with Branch and Bound method for identifying global optimal solutions in the seller's welfare problem for designing a line of products, using a deterministic choice rule. As the authors mention, the lagrangian relaxation itself is not a practical algorithm, and most managers would consider it too complicated and computationally intensive for implementation and practical use. However, they use it to compute guaranteed optimal solutions, which are then used to benchmark the solutions generated by other heuristic algorithms. Heuristics are used to generate a feasible solution that has a fitness value (profit) of *f*. If it is shown that any feasible solution which includes a certain product generates a fitness value of less than *f*, then all solutions that contain the particular product are excluded from further search. Lagrangian relaxation is employed for the estimation of an upper bound on the fitness score that can be generated by a given set of solutions. The constraint relaxed is that each consumer can purchase exactly one product. Hence, for any solution in which the consumer

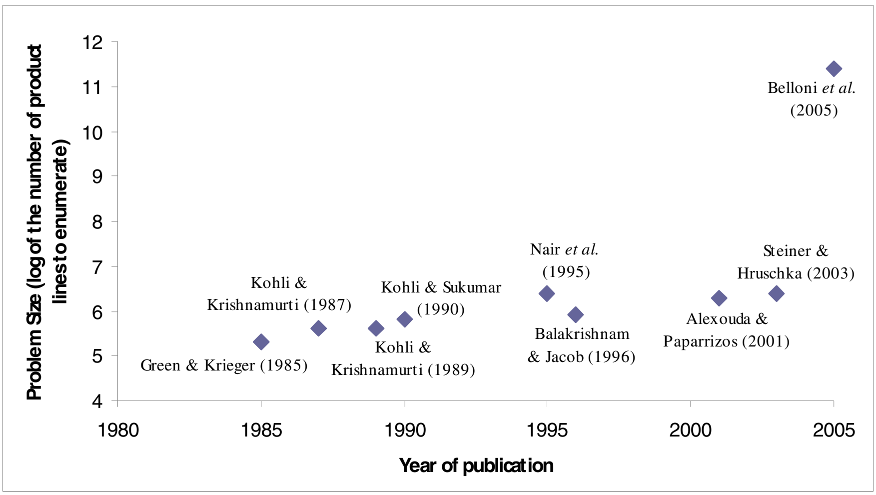


Fig. 2 Size of problems solved (source: Belloni *et al.*, 2005)

selects more than one product, a penalty is subtracted from the fitness of that solution. Similarly, when a consumer chooses less than one product, a reward is added to the solution's fitness. The method seeks for the tightest possible upper bounds by varying the penalties which are applied to the objective function when a solution does not satisfy the relaxed constraints. Finding tight upper bounds helps ruling out portions of the feasible set as fast as possible. The algorithm was applied to 12 simulated problems, as well as 2 versions of a real world problem. The full problem had almost $5*10^{15}$ feasible solutions and the truncated problem had over

147 billion feasible solutions. With a computer that evaluates 30,000 solutions per second, it would take 57 days to completely enumerate the truncated problem, and over 5,000 years to exhaustively search the full problem. The method solved in about 24 hours the truncated and in approximately one week the full problem.

4.10 Comparison of the Algorithms

Belloni *et al.* (2005) measured the complexity of the problems evaluated in previous studies from 1985 to 2005 using the log of the number of feasible product lines (Figure 2). They considered a problem as solved if there is a guarantee that the solution is globally optimal.

Table 13. Comparison of methods on the actual data set (source: Belloni *et al.*, 2008)

Method	**Average performance (%)**	**Best performance as % of the optimal**	**CPU time**	**Subjective difficulty**
Lagrangian relaxation with branch and bound	100	-	1 week	Very high
Coordinate ascent	98.0	98.6	5.4 sec	Low
Genetic algorithm	99.0	100	16.5 sec	Medium
Simulated annealing	100	100	128.7 sec	Medium
Divide and conquer	99.6	100	12.5 sec	Low
Greedy heuristic	98.4	98.4	3.5 sec	Low
Product swapping	99.9	99.9	14.1 sec	Low
Dynamic programming	94.4	97.4	5.5 sec	High
Bean search	93.9	98.6	1.9 sec	High
Nested partitions	96.7	98.4	8.4 sec	High

Belloni *et al.* (2008) compared the performance of 9 different algorithms both in actual and simulated data sets. The real problem had over $4.9*10^{15}$ feasible solutions, and the lagrangian relaxation with branch and bound took over a week to find the global optimum. Except for algorithms' performance, they report a subjective assessment of relative difficulty, where "medium" or "high" level of difficulty denotes methods that require some problem-specific fine tuning of parameter values. Table 13 illustrates the results for ten trials of each method.

As the authors comment, among the more practical methods, the genetic algorithm, simulated annealing, divide and conquer, and product swapping perform best, reaching solutions that are on average within 1% of the optimum. The methods' performance was also evaluated using 12 simulated data sets. Table 14 presents the results for 10 problem instances for each data set.

Table 14 Comparison of methods on the simulated data sets (source: Belloni *et al.*, 2008)

Method	**Average performance (%)**	**Finds optimal solution (%)**	**Finds solution >95% of optimal (%)**	**Average CPU time (sec)**
Lagrangian relaxation with branch and bound	100	100	100	659.4
Coordinate ascent	96.0	15.8	65.8	0.6
Genetic algorithm	99.9	81.7	100	11.8
Simulated annealing	100	100	100	131.8
Divide and conquer	98.7	45.8	97.5	0.7
Greedy heuristic	97.5	23.3	82.5	0.2
Product swapping	98.5	39.2	95.8	0.8
Dynamic programming	96.3	10.0	70.8	0.9
Bean search	99.1	46.7	99.2	0.4
Nested partitions	93.9	4.2	44.2	2.2

In the simulated data sets the genetic algorithm and the simulated annealing manage to accomplish at least as good performance as on the actual data set, whereas the divide and conquer, and the product swapping perform slightly worse. The simulated data sets enable us to make more general conclusions about the algorithms performance than the single real data set. Hence, we can recommend the genetic algorithm and the simulating annealing as the best methods to be applied to the optimal product line design problem, since they provide excellent performance as well as the highest stability among all data sets. Simulated annealing always reaches the global optimum (but cannot guarantee it) with a small cost in time (more than two minutes), while genetic algorithm finds or comes very close to the global optimum, requiring much less time (11-16 sec).

5 A Comparison of Genetic Algorithm to Simulated Annealing

In this section we will apply the two methods that were evaluated as best in section 1.4.10 (genetic algorithm and simulated annealing) to a real data set, as well as to a number of simulated data sets. Since the problem is easily formulated with the use of matrices (see section 1.4.8.3) we implemented the two algorithms using MATLAB. The algorithms were run in a desktop computer with a core 2 duo processor at 2,4 Ghz and 4GB memory.

5.1 Genetic Algorithm Implementation

An integer representation scheme is adopted for the GA implementation as in Balakrishnan *et al.* (2004). We chose this kind of representation (instead of a binary), since all genetic operators (crossover, mutation) must only be applied to entire attributes (genes), and not to a portion of binary bits that compose an attribute, in order for feasible solutions to be produced. As we saw earlier, hybridization strategies for the initialization of the population do not improve the quality of the solution so we initialize the population totally at random. In accordance with the findings of the sensitivity analysis performed by Steiner and Hruschka (2002), we set the population size to 250, and the crossover probability to 1. Uniform crossover was adopted, where half of the attributes exchange values between the two parents. The mutation rate was set to 0.04 instead of 0.01, as the length of the chromosome is shorter due to the different representation scheme. As a reproduction process we employed both an elitist strategy and a roulette wheel selection procedure. The latter constitutes a semi-random process where the chromosomes are selected to survive into the succeeding generation with a probability proportional to their fitness value. In particular, we divide the roulette into 250 (the size of the population) sections. Each section captures a percentage of the roulette's space equal to the ratio of each chromosome's fitness to the total fitness of the population (sum of all chromosomes' fitness values). The roulette wheels 250 times and the chromosome that corresponds to the area the roulette stops each time is chosen for reproduction. While the roulette wheel approach provided comparable final solutions with the elitist strategy, it resulted in faster convergence, hence it is the one that is adopted. As convergence criterion we employed the moving average rule, where the GA stops iterating when the increase of the mean fitness value of the best 3 chromosomes is less than 0.2%, in comparison to last 5 generations.

5.2 Simulated Annealing Implementation

Simulated annealing is similar to coordinate ascent. A product line is initially formed at random and evaluated. The algorithm then tests random product attribute changes, and not only accepts changes that improve the objective function (market share), but may also accept changes that decrease the objective. This helps the algorithm to escape from local minima. Following the method's formulation by Kirkpatrick *et al.* (1983), if a change of an attribute level improves market share then it is accepted, else the probability P_a that the change is accepted is given by the following equation:

$$P_a = \exp\left(\frac{MA - MA'}{T}\right),$$

where MA the market share after the change, MA' the market share before the change, and T is a control parameter (called the "temperature") that is decreased gradually during the process. As the value of T is decreased, the probability of changes that reduce market share to be accepted is decreased as well. We divided the problem into 25 time stages, and tested 7,500 feature changes in stage, which results in a total of 187,500 tested attribute changes of Like in Belloni *et al.* (2008), we initially set T=1443 and calculated each subsequent value by multiplying the existing value by 0.8. In this way, we begin with a large value of T so that almost all attribute changes are accepted, and end with a very small value of T so that very few negative changes are accepted (this procedure is called the "cooling schedule"). The best solution found after the 187,500 attribute changes is returned as the final solution of the algorithm.

5.3 Monte Carlo Simulation

In order to compare the performance of the two algorithms in artificial data sets we will employ a Monte Carlo simulation. A fractional factorial experiment was designed with four factors varying at two levels each (Table 15).

Table 15 Factors and levels used in the experiment

Factor		Levels	
A	Number of attributes	5	9
B	Number of attribute levels	4	8
C	Number of products in the line	4	7
D	Number of customers	100	500

Each data set consists of simulated part-worths, as well as a status quo product for each customer. The hypothetical company plans to introduce 4/7 different products, each consisting of 5/9 attributes which can take 4/8 different levels. The individual-level part-worths for each attribute level, are randomly drawn from a uniform distribution in the range [0,1]. The part-worths are standardized within each customer, with the lowest level for each attribute set to zero, and the sum of the highest levels of each attribute set to one. A deterministic choice rule is employed for the share of choices problem. Eight profiles were generated and 10 replications were performed to each resulting in a total of 80 data sets. Table 16 presents the results, where fitness values are represented as a percentage of the best value obtained from the two algorithms.

Table 16 Genetic Algorithm vs. Simulated Annealing in artificial data sets

Method	Average Performance (%)	Average Time (sec)
Genetic Algorithm	98.75	10.4
Simulated Annealing	100	85.3

The two algorithms achieve equal performance in all 80 data sets except for 1, where SA outperforms GA. SA needs about 8 times more computational time compared to GA. Furthermore, since GA is a population based algorithm, we examine 4 extra variables of interest: The *average fitness of the chromosomes in the final population*, the *fitness of the worst chromosome in the final population*, the *number of unique chromosomes in the final population* and the *standard deviation of their fitness*.

Table 17 Mean values of the variables of interest (only for GA)

Variable	Value
Number of unique solutions in the final population	92/250
Worst chromosome's fitness in the final population	90.64%
Average fitness of the chromosomes in the final population	94.95%
Standard deviation of chromosomes fitness in final population	0.0018

As we can observe (Table 17), about 37% of the proposed solutions in the final population are unique, with a high average fitness (94.95%). Even the worst solution has a good average fitness (90.64%), while the standard deviation in the final population is 0.0018. The results have important implications for marketing managers, since while it is important for them to be provided with the highest performance product line, it is just as critical to present them with a wide range of feasible and high quality product lines. Given such a set of choices, the manager can subsequently utilize any number of subjective criteria to assess these different solutions and select the one that perceives as best (Balakrishnan *et al.*, 2004).

5.4 A Real World Case

A large Greek firm from the food and drinks sector is planning to enter the milk market. In order to decide the product line that will initially introduce, the firm conducted a market survey concerning customer preferences, in a number of geographical areas around the country. The survey lasted from April to June 2008, and took place in different super markets. A total of 334 consumers belonging to the target group (frequent milk buyers) were interviewed providing general information, as well as specifying their status quo product. Furthermore, each consumer completed a conjoint exercise in order for his part-worth utilities to be revealed. The exercise included the ranking from the most to the least preferred of 18 cards, containing profiles of hypothetical milk products. Each profile was represented through four attributes, the levels of which are presented in Table 18.

Table 18 Attributes and levels included in the study

Attribute	**Levels**			
Quantity (l)	0.5	1	1.5	2
Milk type	Fresh	High pasteurized	Goat	
Fat	1,5%	3,5%		
Package	Paper	Plastic		

Focus group interviews were organized for the set of attributes and their corresponding levels to be determined. The final judgment was made by the firm's marketing managers. Conjoint analysis, provided by the SPSS 16 software package, was run on each respondent's data, in order for individual part-worths for every level of each attribute to be estimated. The part-worths were then normalized within each respondent, by setting the lowest level of each attribute to zero, and rescaling the sum of the best attribute levels to unity. The firm wants to introduce a product line that will maximize its market share in the short run. Hence it needs to decide the length as well as the composition of the line. As with the synthetic data sets, we will employ a deterministic choice rule for the share of choices problem.

Since the company has not decided the number of products that will be introduced in the line, we run the algorithm multiple times, with M taking values 1, 2… 10. Five replications for each of the two algorithms were performed in each case and the best solution across them was retained. A complete enumeration was also performed in each case, since the size of the problem is relatively small. The results are shown in Table 19, where fitness scores are represented as a percentage of the optimal solution.

Table 19 Genetic Algorithm vs. Simulated Annealing in the real data set

Method	Average Performance (%)	Average Time (sec)
Genetic Algorithm	100	8.1
Simulated Annealing	100	79.5

The two algorithms manage two find the (unique) global optimum in all cases, but GA converges much faster. The product lines and the corresponding market shares that derived from the real world application are illustrated in Table 20. The increase in market share for a line of more than five products was too small so those cases were omitted.

Table 20 Results of the application

Products in line	Market share	Products configuration			
		Quant	*Milk type*	*Fat*	*Pack-age*
1	7.8%	1	Fresh	1,5%	Paper
2	12.7%	1	Fresh	1,5%	Paper
		1	High past.	1,5%	Paper
3	16.1%	1	Fresh	1,5%	Paper
		1	High past.	3,5%	Paper
		0.5	Goat	1.5%	Paper
4	18.3%	1	Fresh	1,5%	Paper
		1	High past.	3,5%	Paper
		0.5	Goat	1.5%	Paper
		2	Fresh	3.5	Plastic
5	19.4%	1	Fresh	1,5%	Paper
		1	High past.	3,5%	Paper
		0.5	Goat	1.5%	Paper
		2	Fresh	3.5	Plastic
		1.5	High past.	3,5%	Plastic

The firm's executives will decide the exact length of the line, taking into account some extra parameters such as production costs. An interesting finding is that in each case the products formed in the previous case are always included, and only one new item is formed. This is the result of the employment of a deterministic choice rule to the share of choices problem. This limitation can be overcome with the use of a probabilistic choice rule, which however requires different data from the market survey. Instead of a status quo product for each respondent, a set of common competitive products must be defined for the entire group of respondents. Since the product category under investigation is quite broad, the task of defining the exact number and configuration of the products that will directly compete with the ones in the line is very tricky.

6 Programs and Systems

In this section we present the programs and systems that deal with the optimal product (line) design problem. All systems have been developed using one or more of the algorithms discussed in the previous section.

6.1 DESOP-LINEOP

DESOP and LINEOP are the two modules that comprise the program developed by Green & Krieger (1985), which was the first that dealt with the optimal product line design problem. The choice rule is deterministic, the objective is the maximization of market share, and the approach proceeds in two steps. In the first step, a reference set of promising products is constructed through the use of DESOP. The input is a matrix containing the part-worths of the I customers for each level of each attribute as well as a matrix containing the configuration of each customer's status quo product. The program accepts up to 400 customers and 20 attributes, each taking up to 9 levels, while the total number of levels must not exceed 80. The customers whose status quo product has higher utility than the best possible product profile are removed. The user is provided with summary descriptive data regarding the frequency with which each attribute displays the highest part-worth, and he is able to remove a subset of levels or fix an attribute at a certain level. Using the best in heuristic, the program generates the reference set of products, as well as an IxM matrix with the utilities each customer assigns to each of the candidate products. This utilities matrix along with the status quo matrix are entered at the second step to the LINEOP, which selects the products from the reference set that will comprise the product line. The program accepts up to 64 candidate products and produces a line of a maximum length of 30, using either the Greedy or the Interchange heurist.

6.2 SIMOPT

SIMOPT (Green & Krieger, 1988) solves all three problems, directly from part-worth data in a one step approach, using the Divide and Conquer heuristic. The user can specify the subset compositions of the heuristic, which, according to authors, should be formed so as to minimize the correlation of part-worths across subsets of attributes. Attributes that are more closely related to each other should be assigned to the same subset. Except for the customer part-worths matrix, the set of competitive product profiles is also required, as the system uses probabilistic choice rules. Furthermore, the user may optionally provide importance weights for each customer (reflecting the frequency and/or the amount of purchase), background attributes or demographic weights for use in market segment selection and market share forecasting. When the Seller's welfare is selected, costs/return data measured at the individual-attribute level are required. The system provides the user with the capability to perform a sensitivity analysis, in order to observe how market shares (or return) change for all competitors as one varies the levels within each attribute in turn. Since in practice a manager will not probably be interested just in maximizing market share or return, but needs to have a picture of the trade off between them, SIMOPT also supports a Pareto frontier analysis. The user is provided with all the undominated profiles with respect to return and share, and can simulate giving up an amount of the one objective for an increase in the other.

6.3 GENESYS

Balakrishnan and Jacob (1995) developed the GENEtic algorithms based decision support SYStem, which uses the triangulation methodology to increase the confidence in the quality of the obtained solution for the single product design problem. According to it the solution obtained with a certain method is considered "good", if it is in the ball park of the solution obtained through a maximally different heuristic. Using complete enumeration for small problems, Genetic Algorithm, and Dynamic Programming, GENESYS enables the user to avoid the solutions that are caught in local optima. The system consists of a menu driven user interface, where the user can select a single heuristic or the triangulation approach, as well as whether the share of choices or buyer's welfare problem will be solved. Customer part-worths and status quo products are stored in a database, and the three solution methods are stored in a model base. The DP implementation is as in Kohli & Krishnamusti (1987), and the GA as in Balakrishnan and Jacob (1996).

6.4 MDSS

Alexouda (2004) developed a Marketing Decision Support System for solving all the three problems in the optimal product line design, using a deterministic choice rule. The system employs a one-step approach through a GA implementation (Alexouda and Paparizzos, 2001). Borland C+ Builder 3 has been used for the construction of the system, which consists of a database where the seller's return data, as well as the customer part-worths and status quo products are stored, a model base that contains the GA implementation for each of the three problems as well as a complete enumeration method for small problems, and a graphical user interface. Emphasis has been placed on the friendliness of the user interface, which is menu-driven with common easy-of-use features like grid formats, navigators for grids, and pop-up menus. Tools that provide an easy to understand visible way to present options to the user are available, as well as shortcuts that perform actions quickly. Except for the attribute optimization, the system also offers a market simulation module that provides the user with the capability to perform what-if analysis, and assess the likely degree of success of different product line configurations to the market.

6.5 Advanced Simulation Module

ASM is a commercial system that was launched by Sawtooth Software in January 2003. All three problems of the optimal product line design are supported, as well as a market simulation module. The user can select between a deterministic and a probabilistic choice rule, as well as among five different optimization methods: Complete Enumeration, Grid Search, Gradient Search, Stochastic Search, and Genetic Algorithms. Grid Search is similar to the Coordinate Ascent approach by Green *et al.* (1989). In the Gradient Search, a combination of attributes to be altered simultaneously is found, through a Steepest Ascent method that locates the top of a peak in a response surface. An initial solution is generated randomly or specified by the user. Each attribute is changed (one at a time) and the resulting gain or loss in the objective is measured. Then, the direction for changing all attributes simultaneously that results in the largest improvement per unit change is decided. This is the direction of locally Steepest Ascent for the response surface, called Gradient. A line search is conducted next, beginning from the existing solution and moving in the direction specified by the gradient. The first move is very small, and each subsequent move is twice as far from the starting point. The results from the final three points are used to fit a quadratic curve to the response surface, and the point that maximizes the quadratic function is located. The response surface is evaluated at that point, and the solution is retained if it is better than the previous best. When no improvement is achieved from one iteration to the next, the algorithm terminates. In Stochastic Search one attribute is randomly altered at a time and if it results in an objective's improvement the change is acceptable. The process iterates for a prespecified number of times. The authors recommend using either Grid or Stochastic Search from different starting points. If the same solution is always obtained then this is probably the global optimum. Otherwise the search domain should be reduced using the experience obtained, in order to conduct a complete enumeration. When continuous attributes exist (e.g. price), the Gradient Search is the most appropriate. Genetic Algorithms should be used when conditions limit the capabilities of the other methods, for instance when the response surface is very irregular with multiple peaks.

6.6 Discussion

Intelligent marketing systems that deal with the optimal product line design problem have evolved considerably among the past 25 years. Among the five systems presented, four are purely academic and only one is a commercial product. A lot of work has been done since the launch of the first program (Green and Krieger, 1985), which could only solve problems of limited size (not more than 400 customers and 80 attribute levels in total). However, among the algorithms that achieved the highest performance in the problem, as reported in section 4.10, only GAs have been incorporated into marketing systems. Whereas GAs have been used to systems that solve both the single (GENESYS) and the product line design problem (MDSS and ASM), all systems provide the decision maker only with a single best solution. As mentioned in a previous section, the manager that will make the final decision is usually interested in having a range of good solutions, so that he can select the one which satisfies a number of subjective criteria. The existing systems should incorporate multi-objective optimization modules, in order to give the manager the ability to design an optimal product line using some other secondary criteria except for the main one. Furthermore, new systems that will incorporate methods such as the simulated annealing, or the newly introduced lagrangian relaxation with branch and bound, should be developed.

7 Conclusions

The optimal product line design problem has been studied for over thirty years, and several approaches have been applied to solve it. The selection of the right approach has several important practical implications for marketing managers, since an inappropriate approach may produce a bad product line, or a product line of the wrong length. A bad designed product line may decrease the expected profits, and a product line of the wrong length may reduce the expected market share by cannibalizing the firm's existing products. The manager should carefully compare the different approaches and systems and choose the one that better fits his requirements. This constitutes a quite complex task, especially for marketing managers who usually do not have special knowledge concerning optimization algorithms and soft computing. This chapter provided a useful companion that will guide the decision maker through this tricky process.

When beginning the formulation of the problem, the first critical property that the decision maker must select is the choice rule. Most approaches have employed deterministic choice rules, in order to reduce the problem's complexity. However deterministic choice rules suffer from serious limitations. Hence, more emphasis should be placed on probabilistic choice rules, as they provide a better representation of the actual consumer choice process. The large increase in computers' speed, as well as the advances in optimization algorithms, can now compensate the extra complexity that probabilistic rules add to the problem. As far as optimization algorithms are concerned, survey results have shown that methods that work with full product profiles (Genetic Algorithms, Simulated Annealing) perform better than methods that work with partial product profiles (Dynamic Programming, Beam Search). This holds because the latter methods investigate in each iteration, only the most promising solutions and disregard the others, thus it is possible that they disregard (near) optimal solutions in a very early stage. Until now, global optimality has been guaranteed in tractable time, only for the single product design problem, with the use of lagrangian relaxation with branch and bound. Such a method is extremely difficult to implement, and is mainly of academic interest. Furthermore, it has not been incorporated yet into a marketing system. Among the methods that have been applied to the optimal product line design problem, GAs and SA have shown the best performance. GAs have an extra benefit compared to SA, as they work with a set of candidate solutions rather than a single one. In this way they provide the decision maker with a wide range of different product lines, which constitutes an important issue in real world marketing problems. The results of our study indicated that the average performance score of the final set of solutions provided by GA is very high. This provides the manager with the capability to select among a set of high quality product lines the one that best satisfies his personal objectives such as production costs, strategic fit, and technological considerations. GAs constitute also the most advanced optimization method that have been incorporated into a marketing system that deals with the problem. However the GA-based marketing systems have been implemented in such a way that provide the decision maker with only a single best solution, thus they fail to capitalize on the method's main advantage. There is still a lot of work to be done in the area of marketing systems that deal with the optimal product line design problem. Actually, except for one all systems are academic ones, mainly used for illustrative purposes.

As for future research we would suggest that new methods that work with a set of candidate solutions, inspired from natural intelligence (Ant Colony Optimization, Particle Swarm Optimization) can be applied to the problem. Studies have shown that these methods achieve better performance than GAs in a number of discrete NP-hard optimization problems similar to the product line design, such as the Travelling Salesman problem, the Flowshop Scheduling problem, the Task Assignment problem, or the Single Machine Total Weighted Tardiness problem. Furthermore, there is still work to be done in the application of lagrangian relaxation with branch and bound in the product line design, in order for proved global optimal solutions to be provided in tractable time. Regarding marketing systems that deal with the problem, new algorithms should be incorporated such as SA or lagrangian relaxation. In addition, the systems should provide the marketing manager with multiple good solutions, and multiobjective optimization modules should be added, so that secondary criteria could be incorporated.

References

Alexouda, G., Paparrizos, K.: A Genetic Algorithm approach to the product line design problem using the Seller's Return criterion: an exhaustive comparative computational study. European Journal of Operational Research 134(1), 165–178 (2001)

Alexouda, G.: An evolutionary algorithm approach to the Share of Choices problem in the product line design. Computers and operational research 31, 2215–2229 (2004)

Alexouda, G.: A user-friendly marketing decision support system for the product line design problem using evolutionary algorithms. Decision support systems 38, 495–509 (2005)

Balakrishnan, P., Gupta, R., Jacob, V.: Development of hybrid genetic algorithms for product line designs. IEEE Transactions on Systems, Man, and Cybernetics 34(1), 468–483 (2004)

Balakrishnan, P., Jacob, V.: Genetic algorithms for product design. Management Science 42(8), 1105–1117 (1996)

Belloni, A., Freund, R., Selove, M., Simester, D.: Optimizing Product Line Designs: Efficient Methods and Comparisons. Working Paper, MIT Sloan School of Management (2005)

Belloni, A., Freund, R., Selove, M., Simester, D.: Optimizing Product Line Designs: Efficient Methods and Comparisons. Management Science 54(9), 1544–1552 (2008)

Bradley, R.A., Terry, M.E.: Rank analysis of incomplete block designs I: The method of paired comparisons. Biometrika 39, 324–345 (1952)

Camm, J.D., Cochran, J.J., Curry, D.J., Kannan, S.: Conjoint Optimization: An Exact Branch-and-Bound Algorithm for the Share-of-Choice Problem. Management Science 52(3), 435–447 (2006)

Chen, K.D., Hausman, W.H.: Technical Note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. Management Science 46(2), 327–332 (2000)

Dobson, G., Kalish, S.: Heuristics for pricing and positioning a product line using conjoint and cost data. Management Science 39(2), 160–175 (1993)

Downs, B.T., Camm, J.D.: An exact algorithm for the maximal covering problem. Naval Research Logistics 43, 435–461 (1996)

Green, P.E., Carroll, J.D., Goldberg, S.M.: A general approach to product design optimization via conjoint analysis. Journal of Marketing 45, 17–37 (1981)

Green, P.E., Krieger, A.M.: Models and heuristics for product line selection. Marketing Science 4(1), 1–19 (1985)

Green, P.E., Krieger, A.M.: A consumer-based approach to designing product line extensions. Journal of product innovation management 4, 21–32 (1987)

Green, P.E., Krieger, A.M.: An application of a product positioning model to pharmaceutical products. Marketing Science 11, 117–132 (1992)

Green, P.E., Krieger, A.M., Zelnio, R.N.: A componential segmentation model with optimal design features. Decision Sciences 20(2), 221–238 (1989)

Holland, J.H.: Adaptation in Natural and Artificial systems. The University of Michigan Press, Ann Arbor (1975)

Kohli, R., Krishnamusti, R.: A heuristic approach to product design. Management Science 33(12), 1523–1533 (1987)

Kohli, R., Krishnamusti, R.: Optimal product design using conjoint analysis: Computational complexity and algorithms. European Journal of Operational Research 40(2), 186–195 (1989)

Kohli, R., Sukumar, R.: Heuristics for product line design using conjoint analysis. Management Science 36(12), 1464–1478 (1990)

Kotler, P., Armstrong, G.: Principles of Marketing, 12th edn. Prentice Hall, New Jersey (2008)

Krieger, A.M., Green, P.E.: A decision support model for selecting product/service benefit positionings. European Journal of Operational Research 142, 187–202 (2002)

Land, A.H., Doig, A.G.: An automatic method for solving discrete programming problems. Econometrica 28, 497–520 (1960)

Luce, R.D.: Individual choice behavior: a theoretical analysis. Wiley, New York (1959)

Manrai, A.K.: Mathematical models of brand choice behaviour. European Journal of Operational Research 82, 1–17 (1995)

McBride, R.D., Zufryden, F.: An integer programming approach to the optimal line selection problem. Marketing Science 7, 126–140 (1988)

McFadden, D.: Conditional Logit Analysis of Qualitative Choice Behaviour. In: Zaremka, P. (ed.) Frontiers in Econometrics, Academic Press, N.Y. (1973)

Nair, S.K., Thakur, L.S., Wen, K.: Near optimal solutions for product line design and selection: Beam Search heuristics. Management Science 41(5), 767–785 (1995)

Radcliffe, N.J.: Forma analysis and random respectful recombination. In: Proceedings of the 4th International Conference on Genetic Algorithms (1991)

Sawtooth Software, Advanced Simulation Module (ASM) for product optimization. Sawtooth Software technical paper series (2003)

Shi, L., Olafsson, S., Chen, Q.: An optimization framework for product design. Management Science 47(12), 1681–1692 (2001)

Shocker, A.D., Srinivasan, V.: A consumer-based methodology for the identification of new product ideas. Management Science 20, 927–937 (1974)

Steiner, W., Hruschka, H.: A probabilistic one-step approach to the optimal product line design problem using conjoint and cost data. Review of Marketing Science working papers 1(4), 1–36 (2002)

Steiner, W., Hruschka, H.: Generic Algorithms for product design: how well do they really work? International Journal of Market Research 45(2), 229–240 (2003)

Zufryden, F.: A conjoint measurement-based approach for optimal new product design and market segmentation. In: Shocker, A.D. (ed.) Analytical approaches to product and marketing planning, Marketing Science Institute, Cambridge (1977)

Zufryden, F.: ZIPMAP: a zero-one integer programming model for market segmentation and product positioning. Journal of the Operational Research Society 30, 63–70 (1979)

PRODLINE: Architecture of an Artificial Intelligence Based Marketing Decision Support System for PRODuct LINE Designs

P.V. (Sundar) Balakrishnan[1,*], Varghese S. Jacob[2], and Hao Xia[3]

[1] Business Program
University of Washington, MS 358500
Bothell, WA 98011
e-mail: sundar@u.washington.edu
[2] School of Management
University of Texas at Dallas
P.O. Box 830688, SM 40
Richardson, TX 75083-0688
[3] School of Management
University of Texas at Dallas
P.O. Box 830688, SM 33
Richardson, TX 75083-0688

Abstract. Product line design is one of the most important decisions for an organization in today's hypercompetitive world. Product line designs are NP-hard, which implies that it requires an unacceptable amount of time to obtain the guaranteed optimal solution to a problem of reasonable scale. Machine learning techniques such as genetic algorithms can provide very "good" solutions to these problems. In this chapter, we describe the architecture and user interface of a multi-feature decision support system, PRODLINE, which allows the decision maker to address the decision problem of product line designs. A key feature of the system is its ability to provide users with solutions using different solution techniques as well as the ability to change easily the algorithm parameters to assess if improvements in the solution are possible. A final novel and major advantage of the PRODLINE system is that it permits the user to consider strategic competitive responses to the optimal product line design problem.

1 Introduction

1.1 The Product Line Problem

Product line design is generally acknowledged to be one of the most important decisions for a firm. Such decisions are becoming more critical with the rapid pace of technological changes, and the swift evolution of consumer tastes. The

* Corresponding author.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 337–363.
springerlink.com

increasingly fierce crucible of global competition requires firms to work hard to develop a line of products that are perfectly aligned to customer preferences. To compete for consumers in this environment requires an explicit acknowledgement that customers' preferences and dictates are the overriding imperative. Any attempt at making products that are the equivalent of "lukewarm tea" is prone to disaster. On the other hand, catering to the whims of each segment by providing "hot coffee" or "cold tea" can be extremely lucrative. Consequently, new products development is a potentially high-yielding investment but correspondingly fraught with very high risk and cost. The consequences of failure can be catastrophic to the firm. The need for the consumer's voice as input into the process of designing products that will have market acceptance becomes ever more critical in hypercompetitive environments (Balakrishnan, Gupta and Jacob 2006, Shocker and Srinivasan 1974).

Increasingly, the diversity of consumers' preferences makes the design of product line, instead of merely a single product, both necessary and complicated. Therefore, product line design is a process that requires significant attention from a firm before the actual production and marketing can take place. Besides, the highly dynamic nature of the modern market requires firms' rapid response to possible opportunities or challenges, thus product line designs are also expected to be solved within an acceptable period, leaving enough time for managers to make further decisions based on its output. To help mangers address critical problems in real time, more powerful desktop based tools are required that that help to analyze and tackle such a complex decision problem. To this end, Decision Support Systems (DSSs) offer such a tool that can be easy to employ and interact with for managers. In this chapter, we describe our DSS implementation to support an Artificial Intelligence based modeling and solution approach using consumer part-worths data for the product line design problems.

First, before we describe the architecture and the user interaction with our decision support system, it is necessary to specify the product line design problem more clearly. Kotler (1997) defined product line as "*a group of products that are closely related because they perform a similar function, are sold to the same customer groups, are marketed through the same channels, or fall within given ranges*". Then, by product line design, we mean the set of interrelated decisions regarding the selection of specific profiles of a certain number of substitute products that form a product line for possible introduction (or modification) into the marketplace. More specifically, using the generally accepted setup for product design problem that a product category has a specific number of attributes and each of these attributes has a certain number of levels, a product's profile is defined as a 0-1 matrix denoting which level (one and only one level) of each attribute is chosen for the product.

The following simple example is employed to illustrate the basic concepts of product line design. Consider a Television manufacturer who prepares to market a product line consisting of three models of TV sets. Assume that, in this product category, there are four relevant attributes that are critical to the customers in the marketplace. In the design and production of a specific TV model, the attributes of Screen size and Sound quality can each be set to one of three possible levels. The attributes of Parental Controls and Program guides, on the other hand, are features that can be either be incorporated into the design of a specific model or not. The specific attributes and attribute levels that can be employed in the design of the different TV models are displayed below (see Table 1).

Table 1 Attributes and Attribute Levels for the TV Product Category Example

Attributes	Level 1	Level 2	Level 3
Screen	36” Plasma	32” LCD	30” CRT
Sound	Dolby Sound	Stereo Sound	Surround Sound
Parental Controls	Present	Not Present	-
Program guide	Present	Not Present	-

In this trivial problem, the universe of potential models consists of only 36 different products that can be uniquely designed based on these four attributes each with 3, 3, 2 and 2 levels. The single product design problem would require a manufacturer to choose one from among these 36 different products, i.e., the order of magnitude of this small problem is approximately of size 1 (i.e., log of 36) . The product design to be selected would be based on the particular managerial criteria of interest such as maximizing profit or market share (to be discussed later). On the other hand, if a manufacturer wanted to introduce simultaneously three different models, the product line design problem becomes a little more difficult as there are now 7140 distinct combinations of three models to evaluate and select from (i.e., size is of an order of magnitude of 4 approximately). Clearly, such combinatorial problems can explode in size as the length of the product line, the number of attributes and attribute levels increase. Consequently, in this complex problem domain soft computing methodologies that can provide good answers but not necessarily the optimal solution may be of great interest to practitioners and managers.

Employing a zero-one coding scheme the product line can be represented as a string of zero-one values indicating the presence or absence of a specific level of a particular attribute. Continuing with the example, a specific product line of three TV models that a manufacturer might consider introducing into the marketplace is represented as a string of digits as shown below.

100,100,10,10 | 010,010,10,10| 001, 001, 01, 01

In this coding scheme, the first set in the product line features a 36”plasma screen, with Dolby sound, parental controls and program guide. The second model in the product line features a 32” LCD screen, with stereo sound, parental controls and program guide; while the third product has a 30”CRT screen, with surround sound, but has no parental controls or program guide.

The above representation is only one of many alternate coding schemes that could be employed. This coding is scheme employed in this paper, both because it is popular and because it works well within the conjoint analysis framework.

Conjoint analysis (and hybrid conjoint analysis) is a widely accepted technique to estimate the idiosyncratic preferences, i.e., an individual consumer's part-worth values for each attribute level (Zufryden 1977, Green et al. 1981, Green and Srinivasan 1990). The estimation of part-worth values is an increasingly important topic in the marketing literature, but for our exposition, we treat these estimated values as known and errorless. Using these idiosyncratic preference data, we can compute each consumer's utility of any product profile. However, knowing the specific utility of a particular product profile by itself is generally not an end in itself. Such computed product utilities are then employed in a market simulator to predict which of a set of competing products an individual would prefer. This enables the analyst to obtain an estimate of market share of any given product profile in a given competitive environment defined by a set of competing products. Alternatively, it is sometimes known that each consumer has a current favorite brand (i.e., their status-quo product), and that he or she will switch to a new product that provides him/her with the highest utility if and only if the utility of this new product is greater than that of status-quo product.

There are three basic managerial criteria that are typically employed in the product line design problem domain. The first is the buyer's welfare criterion, i.e. to maximize the sum of all buyers' utility, which is generally more useful for non-profit organizations. The second objective is the share-of-choices criterion, which maximizes the total number of consumers who switch to the new product line. The third objective is the seller's return criterion, which, requires additional information on the sellers' marginal utility for each consumer and attribute level. Of the three, the share-of-choices problem is arguably the most important and most widely studied in the literature in this area. In this paper, for the purposes of expositional clarity, we focus our attention only on the share-of-choices problem. We begin with a brief review of existing approaches to this problem in the next section.

1.2 Existing Approaches to Product Line Design Problem

As stated in the last section, a product's profile is defined as a 0-1 matrix, thus the product line design problem can be formulated as a 0-1 integer programming problem (Kohli and Sukumar, 1990). Product design and product line design have long been proved to be NP-hard (Kohli and Krishnamurti, 1989), which implies that it requires unacceptable time to get the guaranteed optimal solution to a problem of reasonable scale.

Early literature (Kohli and Sukumar, 1990) has tried the brute force approach of complete enumeration. This is an option only when the number of feasible product lines is very small. Commercial integer programs such as CPLEX were also tried to solve this problem but have proved to be impractical (Balakrishnan et al. 2004). Methods that combine Lagrangian relaxation with branch-and-bound are more efficient and have been used to solve the single product design problem (Camm et al., 2006), and the profit-maximization product line design problems (Belloni et al., 2008). Unfortunately, these are still too computationally complicated; for example, in the latter one it took one week to find the optimal solution to a small size problem. Therefore, the mainstream approach is to develop heuristic algorithms to get near-optimal solutions. An increasingly popular alternative approach to mathematical programming procedures are Artificial Intelligence (AI) based techniques (Smith, Palaniswami and Krishnamoorthy 1996, Goldberg 1989). Machine learning techniques such as genetic algorithms (GA) or neural networks provide very "good" solutions to these problems. There has also been increasing interest in developing hybrid combinations of various techniques (Coit and Smith 1996, Balakrishnan, Gupta and Jacob 2006).

Some researchers have developed a category of heuristic algorithms that searches in the product space, i.e. choose complete products for product line, which include the Greedy heuristic (Green and Krieger, 1985), the Divide-and-conquer heuristic (Green and Krieger, 1993), etc. However, with a large number of attributes and levels, the enumeration of feasible products is computationally infeasible. Hence, researchers have chosen to develop algorithms that work with partial product profiles, or attributes levels. The most important algorithms in the build up approach are Heuristic Dynamic-Programming (DP) of Kohli and Sukumar (1990) and Beam Search (BS) proposed by Nair et al. (1995).

Genetic Algorithms approach was first applied successfully to the domain of single product design problem by Balakrishnan and Jacob (1992, 1995, 1996). GA has subsequently been extended to address the more complex product line designs problem (Alexouda and Paparrizos 2001, Balakrishnan et al. 2004, etc.). Compared to deterministic algorithms such as heuristic dynamic programming (DP) and beam search (BS), Artificial Intelligence approaches have several advantageous features. GA for instance, is more flexible and can be customized in several ways; users can even change its stopping condition to get a desirable balance between processing time and solutions' quality. One drawback is that GA's processing time is typically longer than that of DP or BS. However, such CPU times still tend to be within an acceptable range for a wide range of problem sizes. On the positive side, the quality of solutions provided by GA is usually superior to that obtained by BS and DP, a result that has now been demonstrated by a number of studies in this literature employing conjoint data (Balakrishnan et al. 2004, Alexouda 2004, Belloni et al. 2008, etc.). The build-up nature of DP and BS leads to possibility that they can get rid of optimal solutions at early stages of the search and result in an arbitrarily bad solution (Nair et al. 1995). This is less of a problem for the GA approach and consequently it always has a positive probability to find the optimal solution in every generation. Consequently, a relevant issue herein is that, unlike in the case of GA, for heuristics such as DP the order in which the attributes are arranged and considered in the analyses tends to impact significantly the quality of the obtained solution (Balakrishnan and Jacob 1996). In addition, instead of just one solution, the GA based approach can provide a family of solutions that can be provided to managers for further examination, an aspect that is an important practical consideration in the development of a decision support system (DSS). In next two sections, we will introduce the architecture and the user interface of such a DSS, the analytics of which were developed by Balakrishnan, Gupta and Jacob (2004, 2006) and has recently been released for pedagogical purposes (Balakrishnan, 2009).

2 Architecture of PRODLINE

A decision support system/software with a graphical user interface, PRODLINE, is developed to tackle the product line design problems using Heuristic Dynamic Programming and Genetic Algorithms. The PRODLINE system's basic architecture is presented in Fig. 1.

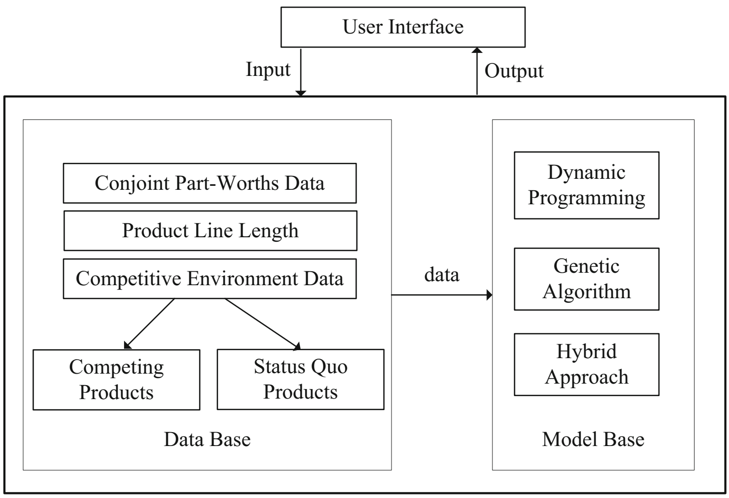


Fig. 1 The Architecture of PRODLINE

The specific aspects of some of the more critical modules that make up the heart of the PRODLINE system are discussed below.

2.1 Database

The database provides PRODLINE with the necessary inputs that consist of three pieces of information. These include the descriptions of the product attributes and attribute levels; consumer utilities; and the competitive product environment. Users need to provide this data by loading two data files: conjoint part-worths data file and the competitive environment data file.

2.1.1 The Conjoint Part-Worths Data

This data file contains two parts, basic information and utility matrix. The first part describes basic information about the product line problem to be solved, namely: the number of consumers, the number of attributes and the number of levels for each attribute, as well as the names of these feature levels. The utility matrix contains the consumer part-worth utilities, i.e. each consumer's utility for each level of each attribute. Every row of this matrix denotes one consumer's part-worth utilities of the attributes levels in sequence, all normalized to sum up to 1. This data is typically captured from a conjoint analysis or hybrid conjoint analysis marketing research study.

2.1.2 Product Line Length

A critical decision that is on many occasions made prior to the market simulation stage by the management team is the length of the product line. The key decision makers may decide *ex ante* based on market conditions and/or the organizational resources available as to the number of items in the product line that they might be willing to support. The resource constraint, in addition to budgetary aspects, may include the human capital available in terms of brand management talent. One extremely interesting and useful aspect of the DSS is that through the process of interaction with the system, the managers are now able to reevaluate *ex post* their original choice of the product line length. The prior decision of the management with respect to the length of the product line can be re-examined on the basis of the resulting market shares. The management team may decide, for instance, that the incremental market share obtained by, say, the fourth item in the line is not sufficiently large to pull its weight, given the infrastructure costs associated with establishing and maintaining the fourth brand.

2.1.3 Competitive Environment Data File

Any new products that are designed will be introduced into a competitive environment except in the case where the goal for, say, the non-profit sector may be to maximize consumer welfare. Therefore, there has to be a mechanism for representing the competitive environment as well as the specific products that are extant in the marketplace that the proposed new products would compete against. There are two different ways provided to the user to describe the competitive environment in the system. One approach allows the user to specify all the existing substitute products in the marketplace. Each consumer will then choose a product that provides him or her the highest utility from across all the competing product as well as from the newly proposed items in the product line. The second approach is to pre-specify each consumer's status quo product, which could be operationalized as his or her current favorite product. The behavioral "first choice" rule at the market simulation stage that could then be invoked is that the consumer will switch to an item in the new product line if the utility of any of the products in the line surpasses that of the status quo product.

For either the competing product or the status-quo situation, the environment data file contains the same basic product category information as the conjoint part-worths data file. In addition, this file includes information on the competitive products. The product matrix is a 0-1 integer file that indicates which of each attributes' level appears in each of the available products. Obviously, for each attribute of each product, one and only one level can appear and this takes on the unitary value. The absence of other attribute levels is indicated by zero. The competitive product landscape information is typically obtained from the results of a market research survey. The product landscape could also be set to represent managerial judgment about future competitor product availability. In essence, we have a flexible mechanism to simulate alternative environments to which the species (product line) must adapt in order to survive.

2.2 Model Base

The version of PRODLINE presented here is developed to solve the product line design problem based on the share-of-choices criterion (the function of maximizing buyers' welfare is similar in user interface approach and is not described in this paper). The PRODLINE system provides for the two alternative heuristic approaches of DP and GA. The system also allows for the third possibility of employing a hybrid approach that seeds the results from DP with the GA approach. As will be shown in detail in a later section, the users are first asked to specify the length of product line; then they are required to select one of the two main heuristic approaches before the system can process the data. On completion of user specification of the heuristic parameters, DSS will provide its solution to the product line design problem. That is, PRODLINE will provide the user with the profiles of each product, as well as the resulting computation of the predicted market share for the new product line as well as the share for each item in the product line.

2.2.1 Dynamic Programming

The Heuristic Dynamic Programming approach works on partial product profiles in product line design problem (Kohli and Krishnamurthi 1987). Hence, there are different ways to apply this approach according to the sequencing rules used to decide the order of attributes selected for branching. More specifically, as Fig.2 shows, rules that are provided for consideration within our DSS include: 1. Original attribute sequence; 2. Descending attribute levels (attributes with the most number of levels considered first); 3. Ascending attribute levels (converse of descending rule); 4. Random generated sequence; and 5. User specified order.

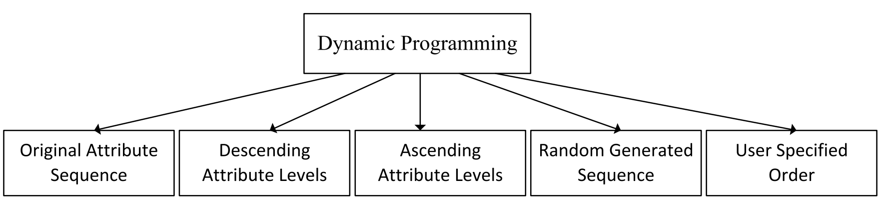


Fig. 2 Attribute Sequencing Options for Dynamic Programming

2.2.2 Genetic Algorithms

In our GA approach, a string is used to represent a potential product line solution. The chromosome represents a particular product in this line. Each gene (or sub-string) represents an attribute in a particular product. An allele represents the absence or presence of a specific level of an attribute. From the example in section 1.1, we have the following mapping:

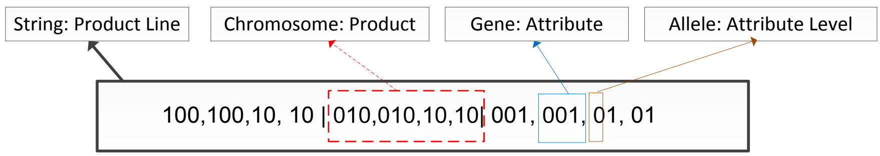


Fig. 3 Coding Scheme Product Lines Genetic Algorithms

Note that with an alternative coding implementation the strings can be represented in a non-binary integer format. Employing such a format, the second product in the string would now be represented as 2 2 1 1, indicating the presence of the second level in attributes one and two and the first level in attributes three and four. This coding scheme makes for ease of data storage as we have a non-sparse matrix. Such an integer coding approach was adopted in the published research of Balakrishnan et al (2004, 2006).

The fitness of each string is then evaluated by the specific criterion, such as market share, that a candidate product line can get in a predefined competitive environment. The market share metric is elegant, as it is both managerially important and simple to compute. The detailed analytical math programming formulation is provided in Balakrishnan, Jacob and Gupta (2006). This metric merely requires keeping count of the number of people in the sample of consumers in our database who switch to the proposed new offerings in the product line from the set of all other products in the market place offered by the competitors. In case of the status quo approach, we count the number of customers who switch from their current favorite brand to one our proposed items in new set of offerings.

The basic architecture of our artificial intelligence based decision support system is as depicted next in Fig.4.

The genetic algorithms approach implemented here can be customized in many different ways. PRODLINE allows users the flexibility to make multiple modifications as needed based on the specifics of the problems they are tackling. Users have the freedom to modify the various GA parameters or they can employ the default options. In the architecture depicted in fig 2.4 the dashed-line boxes in the inputs module provide specific choices that can be changed by the users. The details of the choices that can be employed are discussed briefly next.

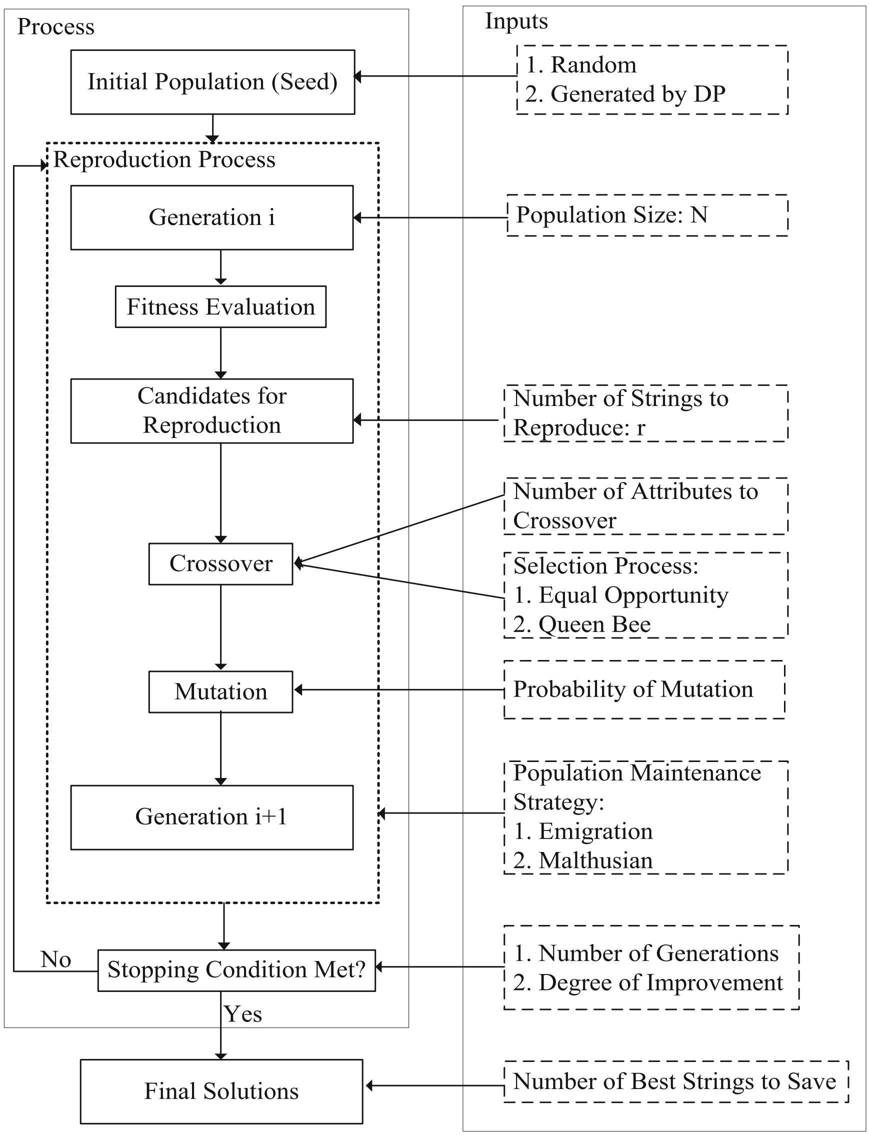


Fig. 4 Architecture of the Genetic Algorithms implementation

2.2.3 GA-DP Hybrid Approach

The performance of GA depends to a considerable degree on the quality of its initial population of product line solutions (i.e. Seeds). As shown in Fig. 4, PRODLINE allows users considerable flexibility with respect to the set of initial strings that are candidate solutions to the product line design problem. The initial population of potential solution strings can be specified in one of two ways. The more common approach would be to have a randomly generated set of strings (potential solutions) from within the DSS model to ensure a highly diverse initial population. Alternatively, the candidate solutions in the initial population can be completely pre-specified in an external file based on managerial instincts and/or the result of solutions obtained from alternative heuristics.

The advantage of employing such a flexible input approach is that it helps to provide the user with an alternative heuristic. Specifically, a GA-DP hybrid approach can be employed from within the software that uses the result of DP heuristic as the seed for the GA model. To employ the hybrid algorithm approach, users will only need to solve the product line problem by first using the heuristic DP approach and then save the resulting output as one of input strings for use with the GA approach. This allows the user to assess for themselves the value of employing hybridized algorithms -- whether it is for managerial assessment of solution quality or it is as a pedagogical device and instruction.

2.2.4 Parameter Specifications

The dotted box in Fig.4 that depicts the PRODLINE architecture encompasses a number of decisions with respect to the various GA parameters. The specific choices that are available which can be selected by the user for each parameter are described briefly below. For greater details about each of the parameters and about a larger range of potential choices that can be made available to the decision makers, the reader is referred to the papers by Balakrishnan, Gupta and Jacob (2004, 2006).

Population Size: One of the GA parameter inputs that could be specified relates to the size of population, i.e., the number of strings that are stored and evaluated in each generation. The number of candidate solution strings that is to be evaluated at each generation can be flexibly changed as a function of the size of the problem being addressed by the managers. This allows for a wider search for larger sized problems.

Selection methods: PRODLINE allows the user a choice of two different types of selection methods. One option is termed “Equal Opportunity”. Here the parent strings that are identified for mating are chosen with equal probability from the entire population of parent strings. This method is helpful in keeping the population of candidate strings heterogeneous over time, which in turn can increase the probability of finding solutions with higher quality. On the other hand, we allow the user the choice of an alternative selection method that we call the “Queen bee”. In this option, the model always chooses the best string from the current generation as one of the two parents for mating. The other parent, however, is selected randomly from the set of parent strings. The concern from prior research has been that the use of a Queen Bee like strategy could lead to the possibility that the population of solutions will over successive generations become increasingly homogeneous and thus get trapped in a local maxima.

Crossover: The number of items in the product line combined with the information on the number of attributes that make up the product category have been previously input into the system. These two pieces of information together determine the maximum number of attributes that can participate in the exchange of genetic material between two parent strings. Typically, based on prior work, the DSS suggests to the user that one-third of the total number of genes that characterize the solution be employed in the crossover operation. The decision support system provides both these values here to the user. However, it is still left to the discretion of the user to input any value for the number of genes, “r”, which will be employed in the crossover operation. This user specified number of attributes is then randomly selected and the genetic information is exchanged between the two parent strings to produce two new offspring strings. In figure 2.5 we illustrate the crossover operation on two parent strings A and B. Assuming that the user specified an “r” of three, then three attributes are randomly selected and the genes at those positions are swapped between the two parent strings (A and B). This crossover operation leads to the birth of two offspring (A’ and B’) strings whose fitness can then be evaluated. Specifically, the italicized genes from parent string A and the genes in bold from parent string B which are swapped are now found in offspring B’ and offspring A’ respectively.

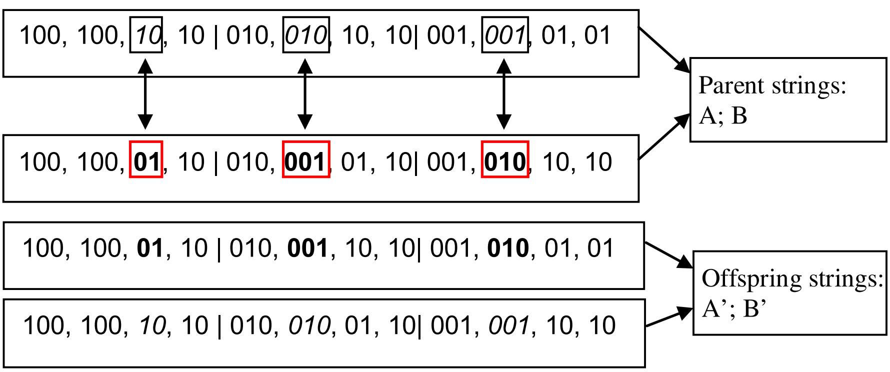


Fig. 5 Example of the attributes crossover operation

Mutation: The probability of mutation of the offspring can be specified by the user. We implement the mutation in such a manner that the resulting candidate solution strings continue to be feasible (as long as all initial generation strings are feasible). The strings to be mutated are randomly selected without replacement, from the population of candidate solutions, with some user specified probability. In this chosen string, a single attribute is randomly picked and the level of that attribute is changed to a randomly chosen level that is within the range of feasible levels. This ensures that if we begin with a set of feasible candidate solutions, the mutated strings are also feasible.

Population Maintenance Strategy: In this decision support system, we provide for two alternative approaches for con trolling the size of the population in each generation. We term these choices the “Emigration” and “Malthusian” for strategies for their descriptive nomenclature as it pertains to population maintenance. These two choices permit the user to specify the degree of relative harshness of the environmental condition, i.e., with respect to how individual product lines in one generation survive to the next generation. The selection of the maintenance strategy impacts how the candidate solutions are culled and maintained over the course of all generations till the simulation concludes. In the emigration strategy, the best strings are selected for reproduction and their offspring form the members of new generation with population size of N. On the other side, in the Malthusian strategy the offspring of reproduction are added back into the previous generation and then the best N of this larger population is selected to form the new generation. The choice of the Malthusian strategy results in a higher likelihood of culling the weaker strings. However, this could result in a particular relatively high fitness string to propagate its genes through the population. This may lead to decreasing the diversity of the population leaving fewer different choices at the end.

3 PRODLINE: User Interaction

In this section, we will describe the PRODLINE interface and show a sample interaction with an example of the Decision Support System in operation. The detailed procedures for PRODLINE to apply the three approaches to product line design problem using conjoint data are either illustrated or discussed. The data employed here for purposes of illustration is based on a case study provided by Balakrishnan and Roos (2008) that describes attributes and attribute levels for televisions as well as the consumer preferences data.

3.1 Inputs

The heart of the system is the GA module whose architecture was depicted in Fig.4. PRODLINE allows the user to specify the system inputs in an intuitive and simple manner. Once the inputs are specified, the system will execute the selected heuristics and provide the outputs. When the DSS is initiated it opens with a screen as shown in Fig.6

To begin interacting with the DSS, the user needs only to click on the “OK to Start” button shown on the opening screen. This then brings up the screen shown in Fig.7 which allows the user to specify the input files in which the data needed for the analysis is available.

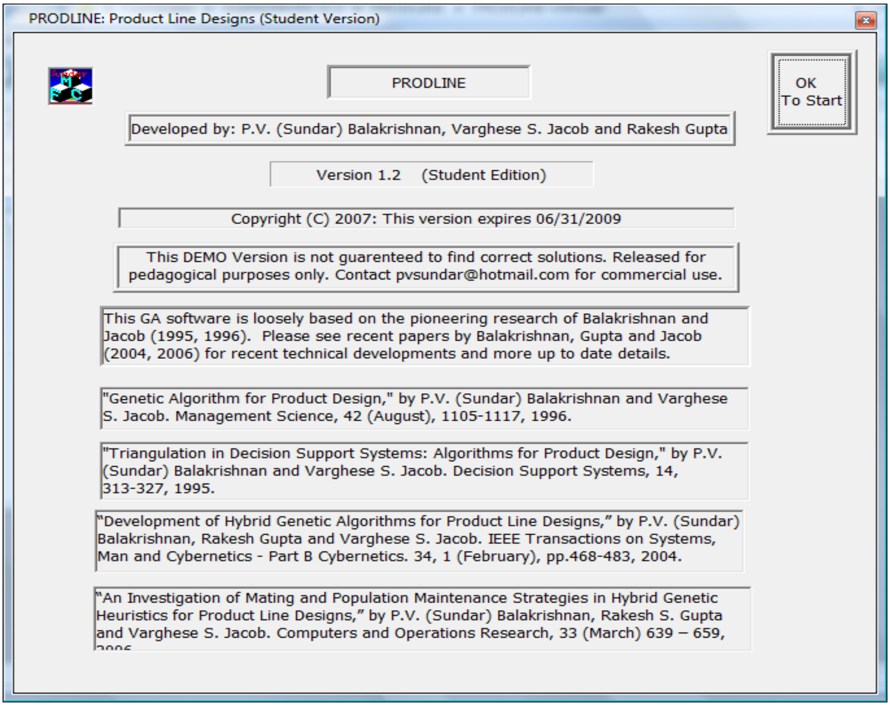


Fig. 6 Opening Screen of PRODLINE

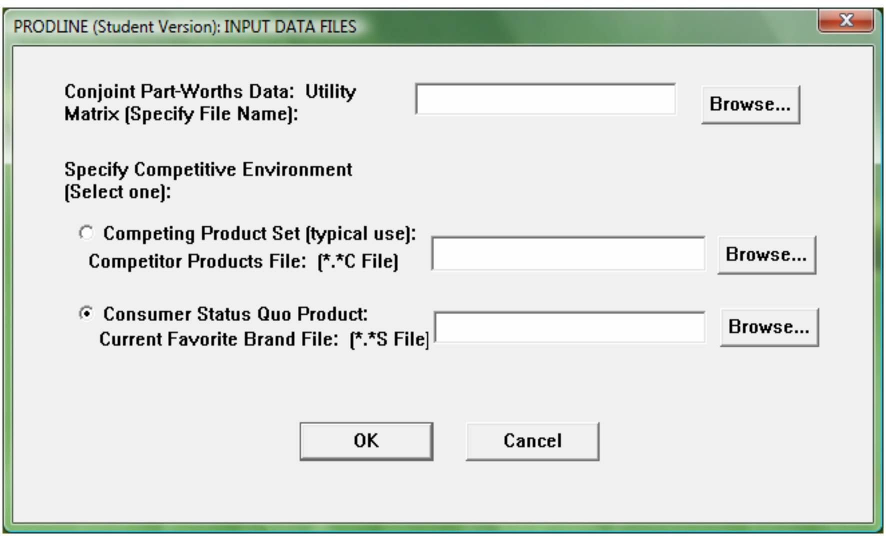


Fig. 7 Input Screen

The format of the consumer utility file that contains the conjoint part-worths data is described in Appendix A. Note that the user has the option of describing a product set that characterizes the competitive landscape or specifying each consumer's status quo product. The status quo, i.e., the current favorite brand, is the benchmark for comparison as to whether a newly designed product will result in the customer switching to the new product or staying with product they currently use. In Appendix B the format of the competitive environment file in which the data on the competing products is input is provided. In Appendix C a snippet of the utility data file containing the part-worths for 200 consumers employed in this paper is provided as an illustration. In Appendix D a sample competitive environment file specifying the six attributes and the five different competing products employed in this case study is presented.

Once the names of the consumer utility (see Appendix C for an example) and competitive environment (see Appendix D) input data files and their location are specified, the user is asked to specify the parameters needed to execute the model to solve the problem. Fig.8 shows the main screen. The first step in this modeling process is setting up the problem size and the location where the output file should be saved. This achieved by choosing the drop down box **Parameters ➔ Simulation** from the screen shown in Fig.8, which will result in the Screen shown in Fig.9.

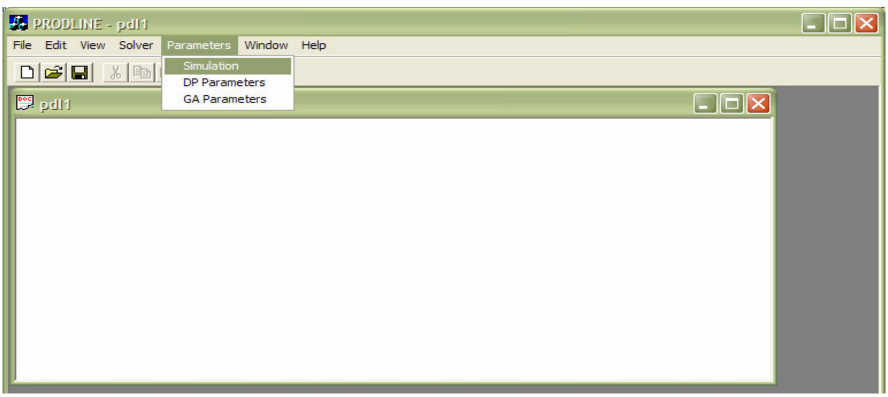


Fig. 8 Model Parameters and Solver Selection Screen

The user can now specify, as shown in Fig.9, the location of the output of the simulation, as well as the specific problem objective employed, whether it is maximizing the share of choices or buyers' welfare.

One of the other key decisions to be made by the managers prior to their interaction with the system relates to the length of the product line, i.e., how many products the user would like to consider in the product line. This information is input at this stage into the system.

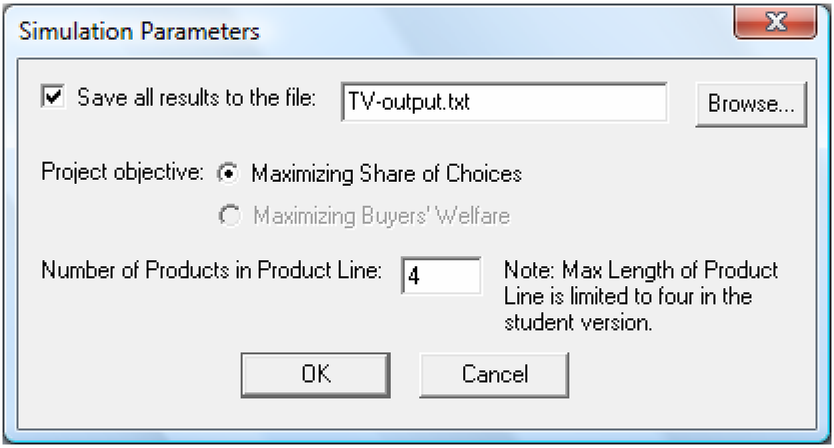


Fig. 9 Simulation Objectives Specification Screen

The next step in the process is to decide whether to use DP or GA to solve the problem. If DP is being used then one would choose **Parameters ➔ DP Parameters** from the options shown in Fig.8 This results in the screen shown in Fig.10.

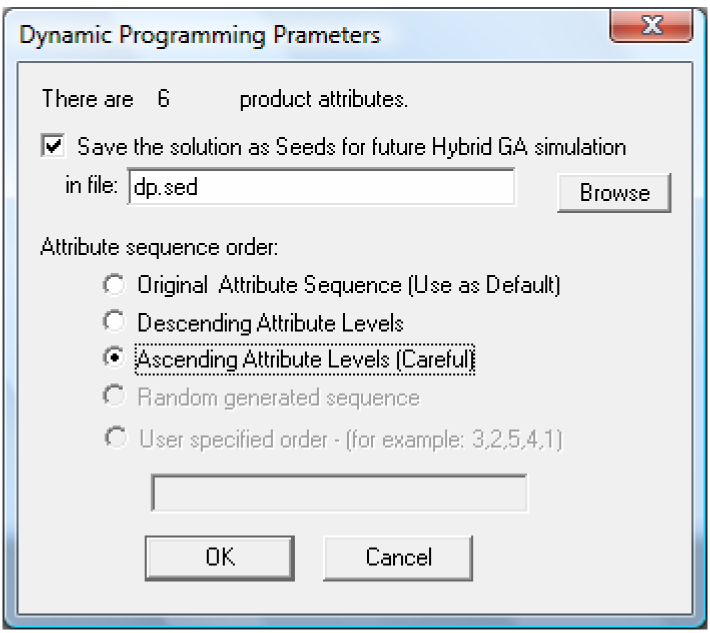


Fig. 10 DP Parameter Selection Screen

The input choice in Fig 10 allows one to decide whether to save the results of the DP solution for further analysis. More importantly, the resulting product line solution obtained by the heuristic DP approach can then be used as an input string by the GA approach. This allows the user to employ a hybrid technique wherein the DP results are used to see the initial GA population. In addition, as seen in Fig 3.5 the user is permitted to select the appropriate attribute sequencing approach, as specified earlier in Fig. 2 The choice of the attribute-sequencing rule invoked can be fairly critical in the quality of the resulting solutions (Balakrishnan and Jacob 1996). More detailed research is needed to determine the appropriate combination of the alternative sequencing rules and the problem structure that consistently results in the highest quality of solutions. Exhaustive Monte Carlo simulation studies might help to provide some generalized guidance and the availability of a DSS such as this can ease considerably the analyst's burden.

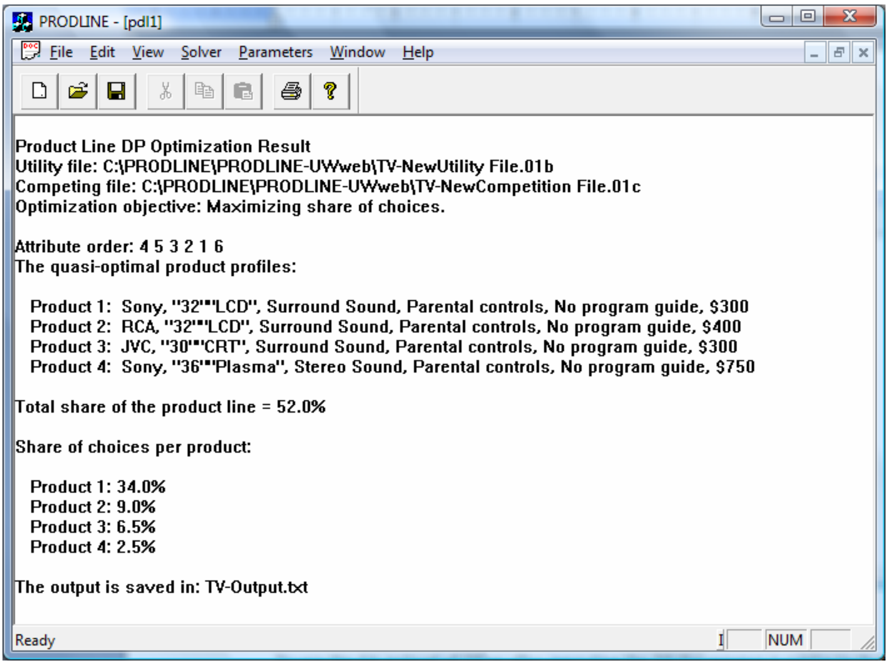


Fig. 11 DP Results Screen

The next step is to choose the solver method as DP to execute the DP model from within the DSS model base. This is achieved by choosing the **Solver ➔ DP** option from the drop down screen shown in Fig.8 This will execute the DP model and the resulting output with the product line solution as well as the market share of each product will be displayed on the screen as seen in Fig 11 These results can also be saved to the output file specified on the screen in Fig.9 This particular output screen (see Fig 11) shows that the DP heuristic using the Ascending attribute order rule to design a four item product line results in a 52% market share. The market shares of each individual item in this new product line range from 2.5% to 34%. For this specific problem, simply by changing the attribute sequence rule from Ascending to the Descending attribute rule and rerunning the DP model results in a dramatic performance improvement in terms of the resulting market shares. It must be recalled that these market share predictions are based on the environment that is specified with respect to the competing products that are currently available in the market place as well as the idiosyncratic preferences of the sample of consumers from the target market who were surveyed. The resulting product line results from multiple runs of DP employing the alternative sequencing rules can be saved and employed as input into the initial GA population.

To use the GA model instead of DP or after executing the DP, the process is similar to that described earlier for the DP. The first step would be to choose the parameters for the GA by choosing **Parameters ➔ GA Parameters** from the model parameter selection screen shown in Fig.8 The user is given the option of loading the GA parameters from a file or specifying the parameters interactively. This screen is shown in Fig.12.

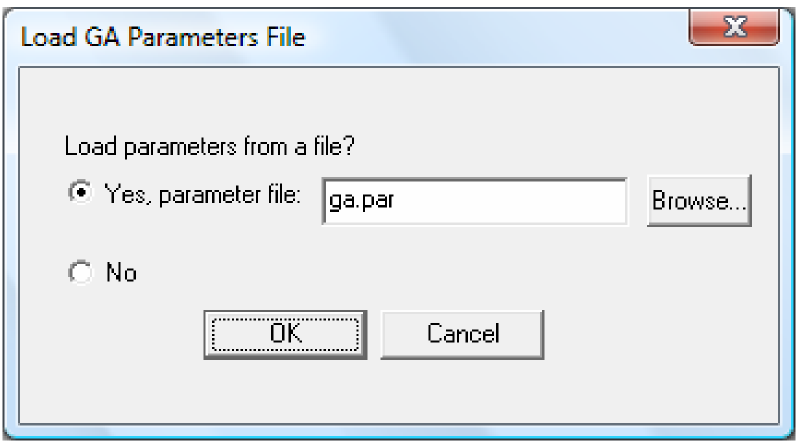


Fig. 12 Option of Loading GA Parameters from a File Screen

Whether or not a parameter file is specified, the screen depicted in Fig.13 will open up. The dialog box will have the parameter values populated from a specified file if a parameter input file is chosen. On the other hand, if a file is not specified, default values of the parameter will be presented to the user. Some of the default parameter values that are presented for consideration to the user for any modification are based on the size of the problem and other values that were input in earlier screens. The user has the choice of accepting the default parameter values or input other values based on their local knowledge. Even if the information is populated from a file, the user has the option to change the values on the screen and save it to a new parameter file. This feature allows the user to experiment with different parameter values for the GA to test if there is an improvement in the results with changes in the GA parameters.

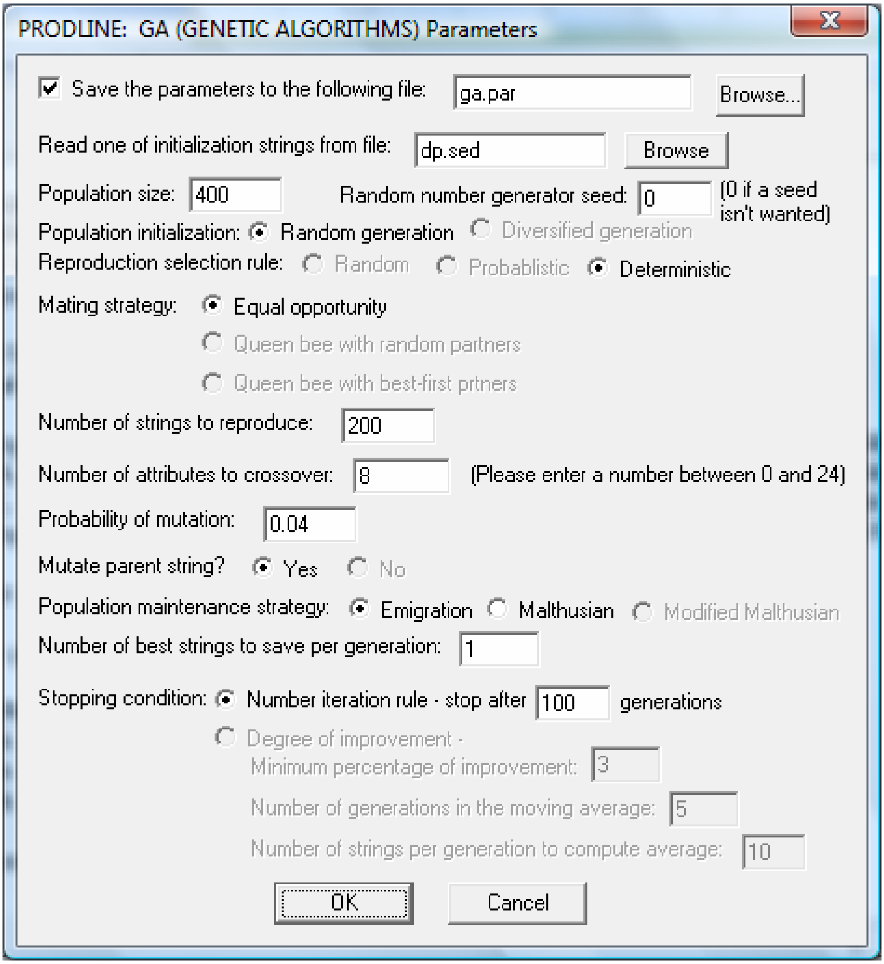


Fig. 13 GA Parameter Specification Screen

To run the GA one would need to choose the option **Solver → GA** from the screen shown in Fig.8. Once the GA is executed the results are presented to the user as shown in Fig.14 The result will provide information on the specific features of each item in the newly proposed product line, the share of choices per product, the total share of the product line. Along with this, information relating to the certain specifics of the GA, namely the generation in which the best solution string (i.e., the product line) first appeared is also presented. The user should then see an output screen that looks something as follows in Fig.14.

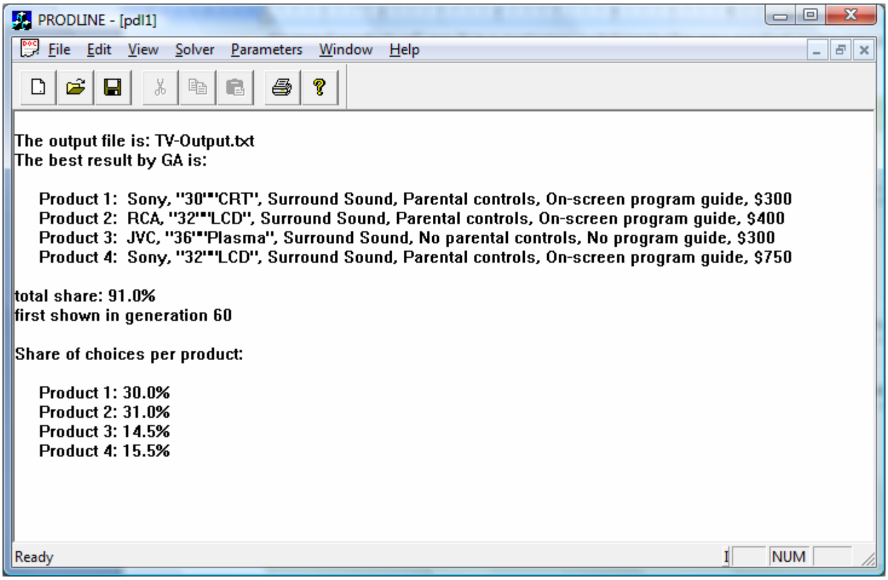


Fig. 14 Genetic Algorithms Results Screen

More detailed process results for each generation are stored in a process output file. A sample snippet of that process tracing output is provided in Appendix E. The information in this file contains the value of the best string, mean of the population of string, their standard deviation, maximum and minimum values obtained at the end of each generation of the GA. In addition, the system also provides the managers a number of good solution options that are within a certain percentage of the best value obtained to date. The DSS allows the user to save the results of as many of the best solutions as needed for further processing and consideration.

4 Discussion and Future Directions

In this chapter, we have provided the architecture of a Decision Support System that will allow the user to tackle easily the product line design problem using multiple approaches. The PRODLINE system allows the user to vary with relative ease the various DP or GA parameters to determine if the product line solutions can be improved. One key feature of the DSS is that it also allows the output of one technique DP become an input to the solution by the GA, this guarantees that the worst solution that the system will provide through this hybrid technique is no worse than that provided by the heuristic DP alone.

The sample interaction of PRODLINE detailed here shows that the results are promising. The architecture is flexible to accommodate multiple competitive scenarios. The product line results obtained by employing the genetic algorithm heuristic detailed in this paper for a specific case data are relatively good and are superior to the results obtained by dynamic programming. The DP heuristic seems to quite sensitive to the ordering of the attributes while the GA is insensitive such trivial changes. The DP heuristic results using the Ascending order attribute sequence in particular seems to be significantly worse than when employing the descending order sequence. This seems to be consistent with the simulation results obtained by Balakrishnan, Jacob and Gupta (2006). In that paper which employed an integer coding scheme but similar architecture the GA model was able to handle problem sizes of 36 orders of magnitude without much difficulty or significant increase in computational time. This suggests that these soft computing approaches as discussed here have significant potential to address really complex and large problems.

The suggestion by prior scholars that triangulation of multiple approaches be employed to tackle difficult problems is taken to heart in the design of this system (Campbell and Fiske 1959, Denzin 1970, Balakrishnan and Jacob 1994). The PRODLINE system described here clearly allows the analysts to employ such a concept. The advantage of employing maximally different multiple approaches helps to overcome problems that result from overt dependence on any single method. In particular, as has been shown repeatedly in many "real life" situations problems that are characterized by large size product line design problems, it is simply not feasible to obtain any solution through exhaustive searches (Belloni et al. 2008) in any reasonable amount of time of even. Consequently, it is critical to engender greater managerial confidence before the use of DSS for real problems becomes widespread. One approach is that it becomes imperative to show that solutions obtained by a specific heuristic is not very different from solutions obtained employing considerably different methods.

The non-obvious advantage of the PRODLINE system is that it permits the user to consider strategic responses to optimal product line design problem. Till now, such a managerial problem was merely a theoretical question that could not be answered in real time. The provision of a system such as this allows for a "What-If" game theoretic scenario analysis. The management team could first deploy this system to design their best product line to the current competitive landscape. They could then consider competitor response explicitly by inputting their new product line into the competing product set data. Having thus described the new environment with their new products, the users can determine the best product line response from the competitor. Now having determined the best competing response, the managers can then input these product reactions into the competing product set data matrix. The system can then be re-invoked to determine the managers' best response to the competitors' response. Note that this game can even be played over a number of rounds to see if there is a stable equilibrium. This could be a valuable addition as an analysis tool in today's hypercompetitive environment.

Acknowledgments. The software PRODLINE was developed by Balakrishnan, Jacob and Gupta. The assistance of Jason Roos in preparing a case for pedagogical purposes for demonstrating an academic version of the software is gratefully acknowledged. The PRODLINE (2009) pedagogical version software can be obtained by registering at the web site of the first author, P.V. (Sundar) Balakrishnan.

References

[1] Alexouda, G., Paparrizos, K.: A genetic algorithm approach to the buyers' welfare problem of product line design: a comparative computational study. Yugoslav J. Operat. Res. 9, 223–233 (1999)
[2] Alexouda, G., Paparrizos, K.: A genetic algorithm approach to the product line design problem using the seller's return criterion: an extensive comparative computational study. Eur. J. Operat. Res. 134, 167–180 (2001)
[3] Alexouda, G.: An evolutionary algorithm approach to the share of choices problem in the product line design. Comput. & Operat. Res. 31, 2215–2229 (2004)
[4] Balakrishnan, P.V., Jacob, V.S.: A Genetic Algorithm for Product Design. In: paper presented at the INFORMS Marketing Science Conference, London (1992)
[5] Balakrishnan, P.V., Jacob, V.S.: Triangulation in decision support systems: algorithms for product design. Decis. Support Syst. 14, 313–327 (1995)
[6] Balakrishnan, P.V., Jacob, V.S.: Genetic algorithms for product design. Manag. Sci. 42, 1105–1117 (1996)
[7] (Sundar) Balakrishnan, P.V., Jacob, V.S.: Development of hybrid genetic algorithms for product line designs. IEEE Trans. Systems, Man, Cybernetics 34, 468–483 (2004)
[8] (Sundar) Balakrishnan, P.V., Gupta, R., Jacob, V.S.: An investigation of mating and population maintenance strategies in hybrid genetic heuristics for product line designs. Comput. & Operat. Res. 33, 639–659 (2006)
[9] (Sundar) Balakrishnan, P.V., Roos, J.M.T.: Case: Televisions 4'Us Optimal Product Line Designs. Unpublished Case, `http://faculty.washington.edu/sundar/PRODLINE-RELEASE` (2008)
[10] (Sundar) Balakrishnan, P.V.: PRODLINE: Pedagogical version software. Download, from (2009), `https://catalysttools.washington.edu/webq/survey/sundar/45656`
[11] Belloni, A., Freund, R., Selove, M., Simester, D.: Optimizing Product Line Designs: Efficient Methods and Comparisons. Manag. Sci. 54, 1544–1552 (2008)
[12] Camm, J.D., Cochran, J.J., Curry, D.J., Kannan, S.: Conjoint optimization: An exact branch-and-bound algorithm for the share-of-choice problem. Manag. Sci. 52, 435–447 (2006)
[13] Campbell, D.T., Fiske, D.W.: Convergent and Discriminant Validity by Multitrait-Multimethod Matrix. Psychological Bulletin 56, 81–105 (1959)
[14] Coit, D.W., Smith, A.: Solving the redundancy allocation problem using a combined neural network/genetic algorithm approach. Comput. & Operat. Res. 23, 515–526 (1996)
[15] Denzin, N.: The Research Act. Aldine, Chicago (1970)
[16] Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading (1989)

[17] Green, P.E., Carrol, J.D., Goldberg, S.M.: A general approach to product design optimization via conjoint analysis. J. Marketing 45, 38–48 (1981)
[18] Green, P.E., Kreiger, A.M.: Recent contributions to optimal product positioning and buyer segmentation. Eur. J. Operat. Res. 41, 127–141 (1989)
[19] Green, P.E., Srinivasan, V.: Conjoint analysis in consumer research: New developments and directions. J. Marketing 54, 3–19 (1990)
[20] Kotler, P.: Marketing management: analysis, planning, implementation and control, 9th edn. Prentice-Hall International, New Jersey (1997)
[21] Kohli, R.: Rajeev Ramesh Krishnamurti A Heuristic Approach to Product Design. Manag. Sci. 33, 1523–1533 (1987)
[22] Kohli, R., Sukumar, R.: Heuristics for product line design using conjoint analysis. Manag. Sci. 36, 311–322 (1990)
[23] Nair, S.K., Thakur, L.S., Wen, K.W.: Near optimal solutions for product line design and selection: beam search heuristics. Manag. Sci. 41, 767–785 (1995)
[24] Shocker, A.D., Srinivasan, V.: A Consumer-Based Methodology for the Identification of New Product Ideas. Manag. Sci. 20, 921–937 (1974)
[25] Smith, K., Palaniswami, M., Krishnamoorthy, M.: A hybrid neural approach to combinatorial optimization. Comput. & Operat. Res. 23, 597–610 (1996)
[26] Zufryden, F.: A conjoint-measurement-based approach to optimal new product design and market segmentation. In: Shocker, A.D. (ed.) Analytical Approaches to Product and Market Planning, Cambridge, MA (1977)

Appendix A

CONSUMER PARTS-WORTHS: DATA FILE FORMAT

- The first line indicates:
  the number of consumers in the data set (200);
  the number of attributes (6);
  and the number of levels for each of these attributes.
- The next set of lines indicates the idiosyncratic preferences (scaled to sum to 1.0) for each individual for each attribute level.
- The last set of lines indicates the **names** of the attribute levels.

Appendix B

COMPETITIVE PRODUCTS MATRIX: FILE FORMAT

- The first line indicates the number of competing products in the market place (5); the number of attributes (6); and the number of levels for each of these attributes.
- The next set of lines indicates the specific levels of each attribute in each of the competing models.
- The last set of lines indicates the names of the attribute levels.

Appendix C

CONSUMER PARTS-WORTHS: Sample Data

200,6,3,3,3,2,2,4,,,,,,,,
0.176,0.057,0.097,0.240,0.006,0.000,0.000,0.070,0.077,0.192,0.000,0.009,0.00
0,0.027,0.000,0.024,0.025
0.200,0.122,0.072,0.000,0.013,0.033,0.000,0.042,0.056,0.213,0.000,0.139,0.00
0,0.039,0.068,0.003,0.000
0.227,0.032,0.000,0.000,0.020,0.037,0.000,0.118,0.129,0.227,0.000,0.057,0.00
0,0.071,0.051,0.030,0.000
<SNIP>
0.311,0.104,0.000,0.000,0.022,0.043,0.000,0.050,0.039,0.123,0.000,0.001,0.00
0,0.127,0.115,0.065,0.000
JVC,
RCA,
Sony,
"30""CRT",
"36""Plasma",
"32""LCD",
Dolby Sound,
Stereo Sound
Surround Sound
No parental controls
Parental controls
No program guide
On-screen program guide
$300
$400
$500
$750

Appendix D

COMPETITIVE PRODUCTS MATRIX: SAMPLE DATA

5,6,3,3,3,2,2,4,
1,0,0,1,0,0,1,0,0,1,0,1,0,1,0,0,0
1,0,0,0,1,0,0,1,0,0,1,0,1,0,1,0,0
0,1,0,0,0,1,0,0,1,1,0,0,1,0,0,1,0
0,1,0,1,0,0,1,0,0,1,0,1,0,1,0,0,0
0,0,1,0,1,0,0,0,1,0,1,0,1,0,0,0,1
JVC,
RCA,
Sony,
"30""CRT",
"36""Plasma",
"32""LCD",
Dolby Sound,
Stereo Sound
Surround Sound
No parental controls
Parental controls
No program guide
On-screen program guide
$300
$400
$500
$750

Appendix E

Detailed Results from Process Output File:

```
Generation no: 60
        Number of unique strings: 380
        Current best value: 0.910000
        Number of unique strings of the best value: 1
        Number of unique strings within 5% of the best value: 56
        Number of unique strings between 5% and 10% of the best value:201
        The best evaluation in this generation: 0.910000
        The worst evaluation in this generation: 0.570000
        Average evaluation for the whole population: 0.823675
        Standard deviation for the whole population: 0.051186
        Average evaluation for the unique strings: 0.821408
        Standard deviation for the unique strings: 0.051461
        No.1 best string:

                Product 1: 3 1 3 2 2 1          0.300000
                Product 2: 2 3 3 2 2 2          0.310000
                Product 3: 1 2 3 1 1 1          0.145000
                Product 4: 3 3 3 2 2 4          0.155000
         total share: 0.910000
```

A Dempster-Shafer Theory Based Exposition of Probabilistic Reasoning in Consumer Choice

Malcolm J. Beynon[1], Luiz Moutinho[2], and Cleopatra Veloutsou[2]

[1] Cardiff Business School, Cardiff University, CF10 3EU, Wales, UK
e-mail: BeynonMJ@Cardiff.ac.uk
[2] School of Business and Management, University of Glasgow, Scotland UK

Abstract. This chapter considers a probabilistic reasoning based investigation of an information system concerned with consumer choice. The DS/AHP technique for multi-criteria decision making is employed in this consumer analysis, and with its development formed from the Dempster-Shafer theory of evidence and the well known Analytical Hierarchy Process, it is closely associated with the notion of soft computing (in particular probabilistic reasoning). Emphasis in the chapter is on the elucidation of a marketing information system (expert system), which includes results on; the levels of judgements made by consumers, the combination of their preference judgements and results in the formulation and size of consideration sets of cars (within the considered car choice problem). Tutorial, graphical and tableau results are presented to enable the reader, unfamiliar with this form of soft computing, the clearest opportunity to follow its novel form of analysis and information content.

1 Introduction

Probabilistic reasoning forms a general description to one of the methodologies closely associated with soft computing (Roesmer, 2000). Bonissone (1998, p. 6) offers a succinct description of probabilistic reasoning;

> "Probabilistic reasoning's main characteristic is its ability to update previous outcome estimates by conditioning them with newly available evidence."

The basis of probabilistic reasoning lies in the two approaches, Bayesian theory (Bayes, 1963) and the Dempster-Shafer theory of evidence (Dempster, 1968; Shafer, 1976). In this chapter, Dempster-Shafer theory forms the mathematical rudiments about which a marketing intelligent system is described. The particular marketing areas considered here, using Dempster-Shafer theory, are; aiding the understanding of consumer behaviour and brand (product) positioning through the

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 365–387.
springerlink.com

exposition of consideration sets. The association of these areas with marketing information systems is outlined in Talvinen (1995), in the case of product positioning, in marketing based expert systems.

Within a marketing context, the understanding of consumer behaviour acknowledges the needs of consumers to make judgments during their process of purchasing specific brands (products) from amongst a range of brands. From the marketing management perspective, there is an incentive to understand the positioning of their brand(s) with respect to their competitors, so to achieve the optimum number of sales etc. A concomitant intelligent system should be able to examine the levels of judgements made by consumers, where consumers have been allowed to control their level of judgement making instead of having to obey defined overriding external remits, an acknowledgement of the well-known bounded rationality problem (Simon, 1955; Miller, 1956; Hogarth, 1980).

Throughout this chapter, the marketing problem considered is with regard to the preference judgements made by potential consumers towards a number of cars, based on certain criteria describing the cars. Understanding how consumers' preference specific groups of brands of cars are critical for companies, especially when the available number of brands competing with one another is large. The stated car choice problem considered is an often investigated problem and closely inset in the general study of consumer brand choice (Punj and Brookes, 2001). This problem brings with it the notion of emotional decision making (Luce *et al.*, 1999), where familiarity with the problem and the social stereotypes are prevalent. Also the implication of brand cues and product positioning, with the advertisement on the different makes of cars influential in the judgements made by a consumer (Hastak and Mitra, 1996; Shapiro *et al.*, 1997; Wedel and Pieters, 2000).

Consumers use various criteria to analyse their options when they are making a purchase decision, such as price and comfort in the case of the car choice problem. In this chapter, a nascent method of multi-criteria decision-making is exposited, namely DS/AHP, introduced in Beynon *et al.* (2000) and Beynon (2002a), with a model structure similar to the well known Analytic Hierarchy Process - AHP (Saaty, 1977), but whose analytical foundation based on the Dempster-Shafer theory of evidence. In summary, like the AHP, it enables consumers to deconstruct the considered problem hierarchically, with the judgements made over the different criteria and between the considered decision alternatives (DAs). However, in contrast to the AHP, the judgement making is controlled by the consumer (Beynon *et al.*, 2000), a consequence of the role played by the Dempster-Shafer theory in the acceptance of ignorance and non-specificity in the judgement making process (see later). The role of Dempster-Shafer theory in a development of AHP is illustrative of the notion of soft computing, whereby a hybrid technique is constructed.

The subsequent marketing intelligent system described, at the consumer level, aims to exposit the levels of preference judgements made by consumer(s), as well as findings on the best DA or groups of DAs amongst those considered (the term best meaning most preferred across a single criterion or number of criteria). In terms of the best DA(s), the extant marketing literature on this issue is the notion of consideration sets, an active and ongoing area of consumer research (Roberts and Nedungadi, 1995; Roberts and Lattin, 1997). Whereby, groups of DAs form the knowledge structure of consumers and are identified for overall preference over all the DAs considered.

Using DS/AHP, two directions of understanding consideration sets are exposited in the underlying marketing intelligent system. Firstly, those sets of DAs which are memory-based, and subsequently brought to a consumer choice problem by the consumer (Desai and Hoyer, 2000). Secondly, the results from a DS/AHP analysis are in the form of levels of preference on different sized groups of DAs (future consideration and choice sets etc.). Within this chapter, with the car choice problem, examples of consideration sets, as a fundamental aspect to the DS/AHP methodology are exposited.

The intention of the chapter is to elucidate how the soft computing orientated methodology probabilisitic reasoning, more specifically Dempster-Shafer theory, when employed in a development (DS/AHP) of an existing technique (AHP in this case), affords results making up an information system associated with the fundamental marketing problem of consumer behaviour (choice). The tutorial calculations included in the chapter are intended to offer the reader the opportunity to understand how Dempster-Shafer theory, in particular, facilitates the notions of ignorance and non-specificity in quantifying the undertaken consumer judgement making etc. (Beynon, 2005), as well as results in the form of consideration sets of different sizes.

2 Background

The background section in this chapter outlines the fundamentals of the DS/AHP method of multi-criteria decision making. Within this exposition, the rudiments of the general methodology, the Dempster-Shafer theory of evidence, are first exposited, followed by DS/AHP, and where appropriate inference on the marketing implications of the processes encompassed with DS/AHP are also included.

2.1 Dempster-Shafer Theory

Central to the DS/AHP method of multi-criteria decision making utilised here, is the Dempster-Shafer theory of evidence. The origins of Dempster-Shafer theory came from the seminal work of Dempster (1968) and Shafer (1976), and often considered as a generalisation of Bayesian theory that can robustly deal with incomplete and imprecise data (Shafer, 1990).[1] Dempster-Shafer theory offers a number of advantages, with respect to multi-criteria decision making, including the opportunity to assign measures of exact belief to *focal elements* (groups of DAs), and allows for the attachment of belief to the *frame of discernment* (all DAs). Bloch (1996) presents a description of the basic principles of Dempster-Shafer theory, including its main advantages (see also Bryson and Mobolurin, 1999).

[1] From its origins, Dempster-Shafer theory has been developed to the point of there are a number of interpretations, including reliance on probabilistic quantification or belief functions (Shafer, 1990; Smets, 1994).

The rudiments of Dempster-Shafer theory are next briefly described. Let $\Theta = \{h_1, h_2, \ldots, h_n\}$ be a finite set of n hypotheses (frame of discernment). A *basic probability assignment* or *mass value* is defined by a function m: $2^\Theta \rightarrow [0, 1]$ such that $m(\emptyset) = 0$, ($\emptyset$ - empty set) and $\sum_{x\in 2^\Theta} m(x) = 1$ (the notation 2^Θ relates to the power set of Θ). Any subset x of the frame of discernment Θ, for which the mass value $m(x)$ is non-zero is called a *focal element*, with the mass value representing the exact belief in the proposition depicted by x. A collection of mass values is denoted a *body of evidence* (BOE), defined $m(\cdot)$, with $m(\Theta)$ considered the amount of ignorance (also called uncertainty), since it represents the level of exact belief that cannot be discerned to any proper subsets of Θ (Bloch, 1996).

Further functions have been constructed that aim to extract additional information regarding the evidence contained in a BOE, see Klir and Wierman (1998). One measure employed with DS/AHP, here, is a non-specificity measure, denoted $N(\cdot)$, within Dempster-Shafer theory, which was introduced by Dubois and Prade (1985), defined as $N(m(\cdot)) = \sum_{x_1\in 2^\Theta} m(x_1)\log_2 |x_1|$ (a form of entropy measure describing information content). The $N(\cdot)$ is considered the weighted average of the focal elements, with $m(\cdot)$ the degree of evidence focusing on x_1, while $\log_2|x_1|$ indicates the lack of specificity of this evidential claim. The general range of this measure (given in Klir and Wierman, 1998) is $[0, \log_2|\Theta|]$, where $|\Theta|$ is the number of DAs in the frame of discernment (Θ). In general, measurements such as non-specificity are viewed as a higher uncertainty type, encapsulated by the term *ambiguity*, Klir and Wierman (1998) state;

> "..the latter (ambiguity) is associated with any situation in which it remains unclear which of several alternatives should be accepted as the genuine one."

Within a marketing information system, a measure such as $N(\cdot)$ is able to quantify the level of judgements made by a consumer, or group of consumers (see later).

To facilitate more pertinent evidence from a BOE on specific focal elements, further measures of total belief can be found surrounding a BOE. A *belief* measure is a function *Bel*: $2^\Theta \rightarrow [0, 1]$, and is drawn from the sum of exact beliefs (mass values) associated with focal elements that are subsets of the focal element x_1 in question, defined by $Bel(x_1) = \sum_{x_2\subseteq x_1} m(x_2)$ for $x_1 \subseteq \Theta$. It represents the confidence that a proposition y lies in x_1 or any subset of x_1. Moreover, $m(x_1)$ measures the assignment of belief exactly to x_1, with $Bel(x_1)$ measuring the total assignment of belief to x_1 (Ducey, 2001). A *plausibility* measure is a function *Pls*: $2^\Theta \rightarrow [0, 1]$, defined by $Pls(x_1) = \sum_{x_2\cap x_1=\emptyset} m(x_2)$ for $x_1 \subseteq \Theta$. Clearly $Pls(x_1)$ represents the extent to which we fail to disbelieve x_1, the total assignment which does not exclude x_1. In a marketing context, these two functions have connection with the consumer choice process as suggested in Park *et al.* (2000), whose title included the phrase "choosing what I want versus rejecting what I do not want" (see also Chakrovarti and Janiszewski, 2003).

The Dempster-Shafer theory provides a method to combine different sources of evidence (BOEs), using Dempster's rule of combination. The combination rule, within Dempster-Shafer theory, is a form of updating information, a fundamental aspect of probabilistic reasoning. This rule assumes that the sources of evidence are independent, then the function $[m_1 \oplus m_2]: 2^\Theta \rightarrow [0, 1]$, combining the evidence in the BOEs $m_1(\cdot)$ and $m_2(\cdot)$ (updating one with the other), defined by;

$$[m_1 \oplus m_2](y) = \begin{cases} 0 & y = \varnothing \\ \dfrac{\sum_{x_1 \cap x_2 = y} m_1(x_1) m_2(x_2)}{1 - \sum_{x_1 \cap x_2 = \varnothing} m_1(x_1) m_2(x_2)} & y \neq \varnothing \end{cases}$$

is a mass value, where x_1 and x_2 are focal elements. An important feature in the denominator part of $[m_1 \oplus m_2]$, is $\sum_{x_1 \cap x_2 = \varnothing} m_1(x_1) m_2(x_2)$, often denoted by k, considered representative of conflict between the independent sources of evidence ($m_1(\cdot)$ and $m_2(\cdot)$). The larger the value of k the more conflict in the evidence, and less sense there is in their combination (Murphy, 2000). In the limit $k = 1$ (complete conflict), it indicates no focal elements intersect between sources of evidence, and the combination function is undefined (Bloch, 1996).

To clarify the understanding surrounding the technical details associated with Dempster-Shafer theory (the expressions presented here), a small example is next considered (a version taken from Beynon *et al.*, 2000). Suppose that the choice of motorcycle by a consumer has been narrowed down to three motorcycles, D, R and S (only labels considered here). Hence the frame of discernment is represented by $\Theta = \{D, R, S\}$. Let us assume that a consumer is going to base their final decision on their preference information on two criterion, I1 (fuel efficiency) and I2 (mechanical reliability).

The information (evidence) from these two criteria, using Dempster-Shafer theory, are formulated in terms of two BOEs, defined $m_{\mathrm{I1}}(\cdot)$ and $m_{\mathrm{I2}}(\cdot)$, with the respective example focal elements and mass values reported in Table 1.

Table 1 Allocation of mass values to focal elements in BOEs, $m_{\mathrm{I1}}(\cdot)$ and $m_{\mathrm{I2}}(\cdot)$

	∅	{D}	{R}	{S}	{D, R}	{D, S}	{R, S}	{D, R, S}
$m_{\mathrm{I1}}(\cdot)$			0.1	0.2			0.4	0.3
$m_{\mathrm{I2}}(\cdot)$		0.1	0.2		0.2		0.3	0.2

From Table 1, by definition, neither $m_{\mathrm{I1}}(\cdot)$ nor $m_{\mathrm{I2}}(\cdot)$ can place any probability mass value to the proposition ∅ (the empty set). The BOE $m_{\mathrm{I1}}(\cdot)$, evidence from I1, distributes its mass predominantly amongst the focal elements {S}, {R, S} and {D, R, S}. The BOE $m_{\mathrm{I2}}(\cdot)$, evidence from I2, distributes its mass predominantly amongst {R}, {D, R}, {R, S} and {D, R, S}. That is, the positive values in Table 1, are mass values representing the exact belief in the preferment of the associated focal element (subset of {D, R, S}).

If these BOEs had come from judgements made by a consumer over the different criteria, I1 and I2, then the levels of the non-specificity ($N(\cdot)$) in these judgements can be evaluated. For the BOE $m_{I1}(\cdot)$, the value of the expression $N(\cdot)$ is found to be;

$$\begin{aligned} N(m_{I1}(\cdot)) &= \sum_{x_1 \in 2^\Theta} m_{I1}(x_1)\log_2 |x_1| \\ &= m_{I1}(\{R\})\log_2|\{R\}| + m_{I1}(\{S\})\log_2|\{S\}| + m_{I1}(\{R, S\})\log_2|\{R, S\}| \\ &\quad + m_{I1}(\{D, R, S\})\log_2|\{D, R, S\}| \\ &= 0.1 \times \log_2 1 + 0.2 \times \log_2 1 + 0.4 \times \log_2 2 + 0.3 \times \log_2 3 \\ &= 0.875. \end{aligned}$$

In contrast $N(m_{I2}(\cdot)) = 0.817$, showing the evidence associated with I1 is less specific than that associated with I2.

Either of the BOEs, $m_{I1}(\cdot)$ or $m_{I2}(\cdot)$, could be separately used to elucidate preferences on the motorcycles D, R and S. However, an updated view of the available information to the consumer is possible, if the information in I1 and I2 could be combined, here using Dempster's rule of combination next described. Let $m_I(\cdot)$ represent the BOE established from the combined evidence of $m_{I1}(\cdot)$ and $m_{I2}(\cdot)$, assuming that $m_{I1}(\cdot)$ and $m_{I2}(\cdot)$ represent items of evidence which are independent of one another. The BOE $m_I(\cdot)$ is given by Dempster's rule of combination; $m_I(\cdot) = [m_{I1} \oplus m_{I2}](\cdot)$, an intermediate stage of this combination process is presented in Table 2.

Table 2 Intermediate stage of combination of BOEs, $m_{I1}(\cdot)$ and $m_{I2}(\cdot)$ (in each cell - focal element followed by associated mass value)

$m_{I2}(\cdot) \setminus m_{I1}(\cdot)$	{R}, 0.1	{S}, 0.2	{R, S}, 0.4	{D, R, S}, 0.3
{D}, 0.1	∅, 0.01	∅, 0.02	∅, 0.04	{D}, 0.03
{R}, 0.2	{R}, 0.02	∅, 0.04	{R}, 0.08	{R}, 0.06
{D, R}, 0.2	{R}, 0.02	∅, 0.04	{R}, 0.08	{D, R}, 0.06
{R, S}, 0.3	{R}, 0.03	{S}, 0.06	{R, S}, 0.12	{R, S}, 0.09
{D, R, S}, 0.2	{R}, 0.02	{S}, 0.04	{R, S}, 0.08	{D, R, S}, 0.06

Table 2 shows an intermediate stage of the combination of the BOEs, $m_{I1}(\cdot)$ and $m_{I2}(\cdot)$, namely the intersection and multiplication of the respective focal elements and mass values in the BOEs. To illustrate the combination process, for the individual mass values $m_{I1}(\{R, S\}) = 0.4$ and $m_{I2}(\{D, R\}) = 0.2$ from the I1 and I2 criterion BOEs, respectively, their combination results in a focal element $\{R, S\} \cap \{D, R\} = \{R\}$ with a value $0.4 \times 0.3 = 0.12$.

Amongst the findings, a number of the focal elements found are empty (∅), it follows, the level of conflict $\sum_{x_1 \cap x_2 = \varnothing} m_{I1}(x_1)m_{I2}(x_2) = 0.15$ (part of the denominator of the combination rule - see previously), then the resultant BOE, defined $m_I(\cdot)$, can be taken from the summing of the values associated with the same focal elements in Table 2, and then divided by 1 – 0.15 = 0.85. The subsequent, newly formed BOE $m_I(\cdot)$, is;

$$m_I(\{D\}) = 0.03/0.85 = 0.035,\ m_I(\{R\}) = 0.364,\ m_I(\{S\}) = 0.118,$$
$$m_I(\{D, R\}) = 0.071,\ m_I(\{R, S\}) = 0.341 \text{ and } m_I(\{D, R, S\}) = 0.071.$$

The established BOE $m_I(\cdot)$ is made up of six focal elements (and mass values), with predominance of exact belief (mass) assigned to the focal elements {R} and {R, S}. In the case of ignorance {D, R, S}, $m_I(\{D, R, S\}) = 0.071$ (= $m_I(\Theta)$), its value is lower than the respective $m_{I1}(\Theta)$ and $m_{I2}(\Theta)$, a consequence of the combination of the evidence associated with I1 and I2. Indeed, this combination rule is illustrative of the updating of evidences mentioned at the start of the chapter, a main characteristic of probabilistic reasoning.

As it stands, the evidence in the BOE $m_I(\cdot)$ cannot be directly examined to indicate overall preference associated with single or groups of motorcycles (D, R, and S). To gauge the specific belief in chosen focal elements (subsets of D, R and S), the belief and plausibility measures defined previously can be used. For the case of the focal element {R, S};

$$\begin{aligned}
Bel(\{R, S\}) &= \textstyle\sum_{x_2 \subseteq \{\mathbf{R},\mathbf{S}\}} m_I(x_2),\\
&= m_I(\{R\}) + m_I(\{S\}) + m_I(\{R, S\}),\\
&= 0.364 + 0.118 + 0.341 = 0.823,\\
Pls(\{R, S\}) &= \textstyle\sum_{x_2 \cap \{\mathbf{R},\mathbf{S}\} = \varnothing} m_I(x_2),\\
&= m_I(\{R\}) + m_I(\{S\}) + m_I(\{D, R\}) + m_I(\{R, S\}) + m_I(\{D, R, S\}),\\
&= 0.364 + 0.118 + 0.071 + 0.341 + 0.071 = 0.965.
\end{aligned}$$

From the results concerning the focal element {R, S}, it can be seen that $Bel(\{R, S\})$ is less than or equal to $Pls(\{R, S\})$, as would be expected (in general). Through the comparison of $Bel(\cdot)$ and $Pls(\cdot)$ values over different focal elements, the most preferred motorcycle (or motorcycles) could be identified.

2.2 Formulisation of DS/AHP and Consumer Choice

This sub-section briefly outlines the rudiments of the DS/AHP technique for multi-criteria decision making, using the Dempster-Shafer theory of evidence described previously. Where appropriate, its description within a marketing context (with the actual car choice problem later described).

The introduction of DS/AHP, in Beynon et al. (2000), Beynon (2002a) and Beynon (2006), was to investigate (model) subjective preference judgements made by decision makers (DMs) on groups of DAs, with respect to all the DAs under consideration. Moreover, it was introduced with a view to offering an aid to multi-criteria decision making, which, when large numbers of DAs are considered, in particular, does not require the relatively large amount of judgements to be made as would be necessary if employing the more well known Analytic Hierarchy Process (Saaty, 1977), see Beynon et al. (2000) for further comparative discussion. This was due to its operations made in the presence of ignorance, through the employment of Dempster-Shafer theory. In a marketing context, most DMs screen DAs on more than one criterion, mostly on well known characteristics of the brands rather than novel characteristics (Gilbride and Allenby, 2004).

With DS/AHP, for a single DM, there are two sets of judgements to be made, firstly knowledge judgements on the importance of each considered criterion, then preference judgements on identified groups of DAs over the different criteria. From these judgements, concomitant criterion BOEs are then constructed on the individual criteria, followed by their combination, using Dempster's rule of combination, creating a BOE within which the evidence exists to identify a best DA (or DAs). The term best here, succinctly refers to the most preferred DA(s) identified using the Bel(·) and Pls(·) measures, as illustrated previously.

To quantify the importance of the individual criterion in this chapter, within a considered car choice problem, each DM (participants in later presented study) was asked to allocate a weight of between 0 and 100 towards each criterion based on their perceived importance, which are then normalised, so they sum to unity (see Beynon, 2002a). For each criterion, the importance value is terms the criterion priority value. If a participant decided to assign a criterion priority value of 0 to any criterion then he/she was not required to make judgements for that criterion on the DAs considered.

To discern the preferences of DAs on an individual criterion, a number of groups of DAs are identified by a DM (consumer) and assigned a (positive) preference scale value (see Figure 1 later). In this chapter a seven-unit scale is used (integer values 2, 3, …, 8), to allow a DM to discern levels of preference on the groups of DAs identified (ranging from "moderately preferred" to "extremely preferred"), is for each group against all the DAs considered (frame of discernment). This positively skewed measurement scaling procedure was tailored to the prerequisites of DS/AHP, and is in line with the well-known work of Miller (1956), and Beynon (2002a; 2002b). Indeed, Beynon (2002b) derived expressions for the scale values to employ, which are dependent on the level of ignorance associated with a decision problem. Table 3 reports a presentation of the relative meaning of the verbal statements to the associated numerical values (with certain verbal statements not given).

Table 3 Connection between numerical values and verbal statements

Numerical value	2 · · 5 · · 8
Verbal statement	Moderately preferred · · Strongly preferred · · Extremely preferred

In Table 3, the numeric values from two to eight indicate from the associated verbal statements an increase in the level of preference to an identified group of DAs. For example, in the case of a group of DAs s (focal elements) being assigned the numerical scale value 5, over an individual criterion, this would indicate the group of DAs has been identified as strongly preferred when compared to the whole set of DAs considered (frame of discernment Θ). This approach to preference judgement making, to a frame of reference is not uncommon, see Lootsma (1993).

Given these two sets of judgements, Beynon (2002a) showed that for a single criterion, if a list of d focal elements (groups of DAs) s_1, s_2, …, s_d are identified and assigned the scale values a_1, a_2, …, a_d, respectively, defining $m(\cdot)$ as the relevant mass values making up the *criterion* BOE for the specific criterion, then;

$$m(s_i) = \frac{a_i p}{\sum_{j=1}^{d} a_j p + \sqrt{d}},\ i = 1, 2, \ldots, d \text{ and } m(\Theta) = \frac{\sqrt{d}}{\sum_{j=1}^{d} a_j p + \sqrt{d}},$$

where Θ is the frame of discernment and p is the associated criterion priority value. A mass value $m(s_i)$ is considered the level of exact belief in the associated focal element (s_i) being most preferred on that criterion. The measure $m(\Theta)$ is defined the level of local ignorance here, since it is the value assigned to Θ, based on the judgements towards a single criterion only.

These exact belief values are found without direct comparison between identified groups of DAs. This relates to the incompleteness in judgements, which is acknowledged and incumbent in the concomitant ignorance. The utilisation of Dempster-Shafer theory, in DS/AHP, brings an allowance of ignorance throughout the judgement making process, which may encapsulate the notions of incompleteness, imprecision and uncertainty (see Smets, 1991). An example of the incompleteness is in preference judgements not having to be made on individual DAs, this could be due to forestalling or doubt by the consumer (Lipshitz and Strauss, 1997). The implication here, in a marketing context, is that a consumer may not exactly know what the reasons are for their possible non-specificity in the judgements they make.

Following the construction of the individual criterion BOEs, they can be combined, using Dempster's combination rule, presuming the judgements across the different criteria are independent, to formulate a final BOE. The DM, can identify the best DA (or DAs), from this final BOE, using the belief and plausibility measures previously defined. Within consumer behaviour, the belief and plausibility measures are related to additive and subtractive choice framing (Shafir, 1993). Further, these measures aid in the identification of the awareness, consideration and choice sets for the DM and an associated decision-making group (see later).

Throughout the rest of the chapter the details from a DS/AHP will be considered in terms of a marketing intelligent system, which offers information on the level of judgements made by a DM, as well as information of the best DA or group of DAs identified over evidence from different criteria.

3 DS/AHP Analysis of Car Choice Problem

The research experiment chosen to elucidate the use of DS/AHP, to formulate a marketing intelligent system, was based on the conduct of a group discussion with eleven consumers - three couples and five single individuals. The focus of the experiment was on their preferences of a number of different makes of cars.

The ages of the members of the decision-making group ranged from 25 to 60. The majority of the participants had a university degree and the group as a whole had a good level of education. The moderation of the group discussion was performed by the researchers in order to ensure a required level of investigator triangulation. A number of projective techniques were used in particular through the utilisation of pictorial information and visual aids pertaining to the subject under investigation: the formation of consumer consideration sets with regard to choice criteria utilised in car purchase decision-making.

The pairing of the stimuli focused on three analytical dyads shown to the participants in three folders (one for each dyad), containing extensive information and pictures about each car model under study, as suggested by Raffone and Wolters (2001). Each car model was labelled with a letter and the comparative dyads were designed in terms of level of consumer familiarity (Aurier *et al.*, 2000), product diversity as well as price quality tiers. Therefore, the first dyad to be analysed included SMART and IGNIS (a new model just launched) whereas the second grouping dealt with ALFA 156 and VOLVO S60, and finally, the last pairing contained a "sports car" cluster - TOYOTA MR2 and BMW 3. The setting of these research stimuli was also designed to manipulate brand name valence as well as testing the subjects' processing task. Furthermore, the selected research design rested upon the notion that buyers have category specifics based on "mentally defined" price-quality tiers (following the experiment of Verwey (2003)). The five criteria selected for analyses of consumers' decision-making, with regard to car purchase were, comfort, economy, performance, price and safety.

Time was spent introducing DS/AHP to the participants (consumers), including the types of judgement making required (at criteria and DA levels), thorugh a non-related example. After having analysed all the provided information for the dyads for a considerable period of time, a very short research instrument was applied in order to gauge and quantify their perceptions of choice criteria leading to the potential formation of consideration sets. The rest of this section exposits the DS/AHP analysis on the judgements made by the 11 consumers, including an elucidation of the judgements made by a single consumer and the construction of the subsequent results describing the choice process in identifying the best car (or cars). As stated previously, the elucidation of the judgements made, and subsequent results, are all part of the concomitant information system here.

Each consumer was allowed to control the level of judgement making, to what they felt confident to undertake (see Chase and Simon, 1973), as outlined in the description of the DS/AHP given previously. The participants were informed that the levels of preference for each criterion analysed should be considered in relation to all the available cars in the experiment. No cars were allowed to appear in more than one group identified over a single criterion, and not all cars needed to have preference judgements made on them. Within the DS/AHP analysis of the car choice problem, the six cars SMART, IGNIS, ALFA 156, VOLVO S60, TOYOTA MR2 and BMW 3 considered are labelled A, B, C, D, E and F respectively, collectively defined the frame of discernment Θ (= $\{A, B, C, D, E, F\}$). The judgements made by one individual consumer (labelled DM1) are reported in Figure 1.

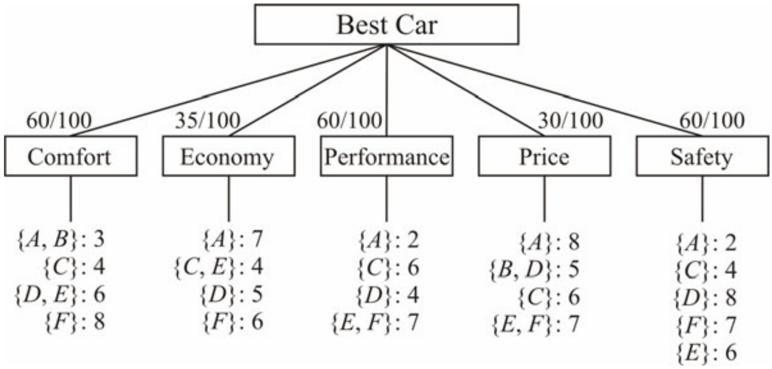


Fig. 1 Hierarchy of judgements made by DM1, when using DS/AHP, on car choice problem

In Figure 1, a hierarchical structure (as in AHP) is used to present the judgements made by DM1. Moving down from the focus 'best car' to the identified groups of DAs over each criterion, there are two different sets of judgements made by DM1 (a single consumer). Firstly, there is the set of criterion priority values, which indicate the levels of importance or perceived knowledge a consumer has towards the criteria (each assigned a value which collectively sum to 100 – then divided by a 100).

Normalising the weights shown in Figure 1, the criterion priority values (from DM1) for the criteria; comfort ($p_{1,\mathrm{C}}$), economy ($p_{1,\mathrm{E}}$), performance ($p_{1,\mathrm{PE}}$), price ($p_{1,\mathrm{PR}}$) and safety ($p_{1,\mathrm{S}}$) are, 0.2449, 0.143, 0.245, 0.122, and 0.245, respectively. With the criterion priority value assigned for each criterion, it is required for DM1 to make preference judgements towards groups of cars on those criteria with positive criterion priority values.

With respect to the car choice problem, defining $m_{1,\mathrm{C}}(\cdot)$ as the criterion BOE for the judgements made by DM1 on the comfort criterion, termed *comfort criterion* BOE, from Figure 1, $s_1 = \{A, B\}$, $s_2 = \{C\}$, $s_3 = \{D, E\}$ and $s_4 = \{F\}$ with $a_1 = 3$, $a_2 = 4$, $a_3 = 6$ and $a_4 = 8$, respectively. For a general criterion priority value $p_{1,\mathrm{C}}$, then;

$$m_{1,\mathrm{C}}(\{A,B\}) = \frac{3p_{1,\mathrm{C}}}{21p_{1,\mathrm{C}} + \sqrt{4}},\ m_{1,\mathrm{C}}(\{C\}) = \frac{4p_{1,\mathrm{C}}}{21p_{1,\mathrm{C}} + \sqrt{4}},$$

$$m_{1,\mathrm{C}}(\{D,E\}) = \frac{6p_{1,\mathrm{C}}}{21p_{1,\mathrm{C}} + \sqrt{4}},\ m_{1,\mathrm{C}}(\{F\}) = \frac{8p_{1,\mathrm{C}}}{21p_{1,\mathrm{C}} + \sqrt{4}} \text{ and}$$

$$m_{1,\mathrm{C}}(\Theta) = \frac{\sqrt{4}}{21p_{1,\mathrm{C}} + \sqrt{4}}.$$

These mass values are dependent only on the criterion priority value $p_{1,\mathrm{C}}$, for the comfort criterion $p_{1,\mathrm{C}} = 0.245$, hence $m_{1,\mathrm{C}}(\{A, B\}) = 0.103$, $m_{1,\mathrm{C}}(\{C\}) = 0.137$,

$m_{1,\mathrm{C}}(\{D, E\}) = 0.206$, $m_{1,\mathrm{C}}(\{F\}) = 0.274$ and $m_{1,\mathrm{C}}(\Theta) = 0.280$. Using the more general values of $m_{1,\mathrm{C}}(\cdot)$, Figure 2*a* illustrates the effect of the criterion priority value $p_{1,\mathrm{C}}$ on the comfort criterion BOE (also shown in Figure 2*b* is the respective graph for the price criterion BOE, defined $m_{1,\mathrm{PR}}(\cdot)$, with associated criterion priority value $p_{1,\mathrm{PR}}$).

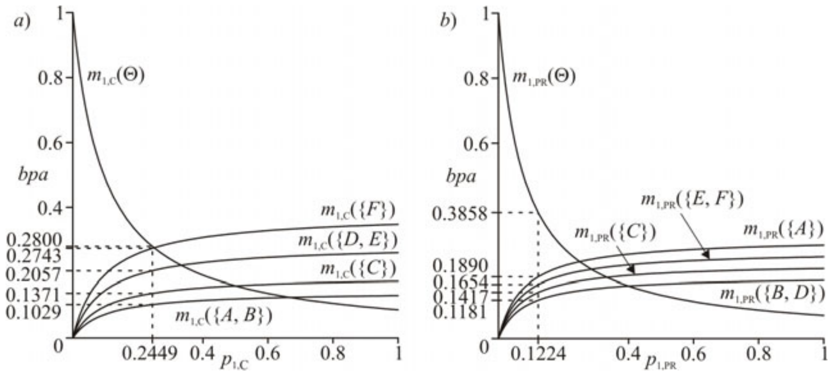


Fig. 2 Criterion BOE $m_{1,\mathrm{C}}(\cdot)$ and $m_{1,\mathrm{PR}}(\cdot)$ values, as $p_{1,\mathrm{C}}$ and $p_{1,\mathrm{PR}}$ go from 0 to 1

In Figure 2*a*, as $p_{1,\mathrm{C}}$ tends to 0 (little importance/priority) more exact belief value would be assigned to the associated local ignorance $m_{1,\mathrm{C}}(\Theta)$, and less to the identified groups of cars. The reciprocal is true, as $p_{1,\mathrm{C}}$ tends to 1, when there is perceived importance on the comfort criterion, so the level of local ignorance decreases. The values of $m_{1,\mathrm{C}}(\cdot)$ for when $p_{1,\mathrm{C}} = 0.245$ are also confirmed in Figure 2*a*. In Figure 2*b*, a similar set of graphs are constructed for the mass values making up the price criterion BOE (with general criterion priority value $p_{1,\mathrm{PR}}$). The graphs representing the $m_{1,\mathrm{PR}}(\cdot)$ values for the identified groups of cars, in Figure 2*b*, are closer together than in Figure 2*a*.

Inspection of the judgements made by DM1 in Figure 1 elucidates the range of scale values used on the comfort criterion is larger than those scale values used on the price criterion. For the price criterion with $p_{1,\mathrm{PR}} = 0.122$ then $m_{1,\mathrm{PR}}(\{A\}) = 0.189$, $m_{1,\mathrm{PR}}(\{E, F\}) = 0.165$, $m_{1,\mathrm{PR}}(\{C\}) = 0.142$, $m_{1,\mathrm{PR}}(\{B, D\}) = 0.118$ and $m_{1,\mathrm{PR}}(\Theta) = 0.386$, as shown in Figure 2*b*.

Criterion BOE can be found for the other three criteria: economy - $m_{1,\mathrm{E}}(\cdot)$, performance - $m_{1,\mathrm{PE}}(\cdot)$ and safety - $m_{1,\mathrm{S}}(\cdot)$, based on the judgements made by DM1 shown in Figure 1 (using their respective criterion priority value: $p_{1,\mathrm{E}} = 0.143$, $p_{1,\mathrm{PE}} = 0.245$ and $p_{1,\mathrm{S}} = 0.245$):

Economy: $m_{1,\mathrm{E}}(\{A\}) = 0.194$, $m_{1,\mathrm{E}}(\{C, E\}) = 0.111$, $m_{1,\mathrm{E}}(\{D\}) = 0.139$, $m_{1,\mathrm{E}}(\{F\}) = 0.167$ and $m_{1,\mathrm{E}}(\Theta) = 0.389$.

Performance: $m_{1,\mathrm{PE}}(\{A\}) = 0.074$, $m_{1,\mathrm{PE}}(\{C\}) = 0.221$, $m_{1,\mathrm{PE}}(\{D\}) = 0.147$, $m_{1,\mathrm{PE}}(\{E, F\}) = 0.258$ and $m_{1,\mathrm{PE}}(\Theta) = 0.300$.

Safety: $m_{1,\mathrm{S}}(\{A\}) = 0.055$, $m_{1,\mathrm{S}}(\{C\}) = 0.111$, $m_{1,\mathrm{S}}(\{D\}) = 0.221$, $m_{1,\mathrm{S}}(\{E\}) = 0.166$, $m_{1,\mathrm{S}}(\{F\}) = 0.194$ and $m_{1,\mathrm{S}}(\Theta) = 0.253$.

To offer information on the homogeneity and intensity of the consumer's choice process, the conflict levels between the judgements made by DM1 over the different criteria can be calculated, see Table 4. With respect to DS/AHP, the level of conflict (k) relates to how different the judgements made are over the different criteria (see the previous description of Dempster-Shafer theory).

Table 4 Conflict values between criterion BOEs for DM1

Criteria	Economy	Performance	Price	Safety
Comfort	0.308	0.312	0.288	0.384
Economy	-	0.297	0.261	0.352
Performance	-	-	0.324	0.369
Price	-	-	-	0.347

In Table 4, the higher the conflict value (within the domain [0, 1]), the more conflict there exists between the judgements made over the two criteria. The most conflict evident is between the comfort and safety criteria (with k = 0.384), indicating the most difference in the judgements made is over these criteria. Since the conflict levels are relatively low between criteria (nearer 0 than 1), it strengthens the validity of the results found from the intended combination of the five criterion BOEs (see later). This is the first set of results associated with a marketing intelligent system, from which a DM (or external analyst), would view and make conclusions (such as what is the inference to be taken from the differing levels of conflict between criteria).

A further measure defined, namely non-specificity, here relates to the level of grouping apparent in the groups of cars identified for preference by DM1, over the different criteria. With six cars considered, the domain on the level of non-specificity is [0, 2.585]. In Table 5, the levels of non-specificity on the judgements made by DM1, over the five criterion BOEs, are reported.

Table 5 Levels of non-specificity on judgements made by DM1 on criterion BOEs

Evidence	Comfort	Economy	Performance	Price	Safety
Non-specificity	1.032	1.116	1.035	1.281	0.653

From Table 5, the largest and least levels of non-specificity amongst the criterion BOE are associated with the price (1.281) and safety (0.653) criteria respectively. To illustrate the calculation of these non-specificity values ($N(\cdot)$), for the price criterion:

$$
\begin{aligned}
N(m_{1,\mathbf{PR}}(.)) &= \textstyle\sum_{x_1 \in 2^{\Theta}} m_{1,\mathbf{PR}}(x_1)\log_2 |\, x_1 \,|, \\
&= m_{1,\mathbf{PR}}(\{A\})\log_2 |\{A\}| + m_{1,\mathbf{PR}}(\{E,F\})\log_2 |\{E,F\}| \\
&\quad + m_{1,\mathbf{PR}}(\{C\})\log_2 |\{C\}| + m_{1,\mathbf{PR}}(\{B,D\})\log_2 |\{B,D\}| \\
&\qquad + m_{1,\mathbf{PR}}(\Theta)\log_2 |\Theta|, \\
&= 0.189\log_2 1 + 0.165\log_2 2 + 0.142\log_2 1 + 0.118\log_2 2 \\
&\quad + 0.386\log_2 6, \\
&= 1.281.
\end{aligned}
$$

A comparison between the judgements made on the price and safety (and other) criteria (given in Figure 1), shows the price criterion includes two groups of cars identified with two cars in each, whereas only singleton groups of cars are identified with the safety criterion. This follows the premise that information chunk boundaries have psychological reality (Gobet and Simon, 1998). One further important factor is the value of the associated criterion priority value, since with a low criterion priority value more mass value is assigned to Θ, hence a higher non-specificity value.

The goal for DM1 is to consolidate their evidence on the best car to choose, based on all the criteria considered. Using DS/AHP, this necessitates the combining of the associated criterion BOE using Dempster's combination rule presented in the description of Dempster-Shafer theory given previously (the $[m_1 \oplus m_2](\cdot)$ expression). In Table 6, the intermediate values from the combination of the two criterion BOEs, comfort $m_{1,\mathrm{C}}(\cdot)$ and price $m_{1,\mathrm{PR}}(\cdot)$, are reported. That is, using Dempster's combination rule, the combination is made up of the intersection and multiplication of focal elements and mass values, respectively, from the two different criterion BOEs considered.

Table 6 Intermediate values from combination of comfort and price BOEs, for DM1

$m_{1,C}(\cdot) \backslash m_{1,PR}(\cdot)$	{A}, 0.189	{E, F}, 0.165	{C}, 0.142	{B, D}, 0.118	Θ, 0.386
{A, B}, 0.103	{A}, 0.019	∅, 0.017	∅, 0.015	{B}, 0.012	{A, B}, 0.040
{C}, 0.137	∅, 0.026	∅, 0.023	{C}, 0.019	∅, 0.016	{C}, 0.053
{D, E}, 0.206	∅, 0.039	{E}, 0.034	∅, 0.029	{D}, 0.024	{D, E}, 0.079
{F}, 0.274	∅, 0.052	{F}, 0.045	∅, 0.039	∅, 0.032	{F}, 0.106
Θ, 0.280	{A}, 0.053	{E, F}, 0.046	{C}, 0.040	{B, D}, 0.033	Θ, 0.108

To illustrate the combination process, for the individual mass values $m_{1,\mathrm{C}}(\{A, B\}) = 0.103$ and $m_{1,\mathrm{PR}}(\{A\}) = 0.189$ from the comfort and price criterion BOEs, respectively, their combination results in a focal element $\{A, B\} \cap \{A\} = \{A\}$ with a value $0.103 \times 0.189 = 0.019$. The ∅ term present in Table 6 is the empty set and the sum of these values (in italics) represents the level of associated conflict (k in the description of Dempster-Shafer theory), in the combination of these two criterion BOE, in this case $k = 0.288$ (see Table 4). The final mass value constructed for a particular focal element is illustrated for the $\{A\}$ focal element, which is given by;

$$[m_{1,\mathbf{C}} \oplus m_{1,\mathbf{P}R}](\{A\}) = \frac{m_{1,\mathbf{C}}(\{A,B\})m_{1,\mathbf{PR}}(\{A\}) + m_{1,\mathbf{C}}(\Theta)m_{1,\mathbf{PR}}(\{A\})}{1-0.288},$$

$$= \frac{0.019+0.053}{0.713} = 0.102.$$

To re-iterate, Dempster's rule of combination is used to aggregate the evidence from a consumer's judgements on the five different criteria considered. Similar mass values can be found for the other focal elements (shown in Table 6), to form a temporary BOE. The combination rule can be used iteratively to combine all the criterion BOEs to the successively created temporary BOE. Defining $m_{1,\mathrm{CAR}}(\cdot)$ as the post combination *consumer* BOE from all the criterion BOEs for DM1, its associated focal elements (groups of cars) and mass values are reported in Table 7.

Table 7 Individual groups of cars (focal elements) and mass values in the $m_{1,\mathrm{CAR}}(\cdot)$ BOE

{A}, 0.090	{D}, 0.181	{A, B}, 0.010	{D, E}, 0.020
{B}, 0.003	{E}, 0.169	{B, D}, 0.008	{E, F}, 0.046
{C}, 0.145	{F}, 0.292	{C, E}, 0.008	Θ, 0.028

In Table 7, 12 groups of cars (focal elements including Θ) and mass values, making up the consumer BOE for DM1 are shown. To illustrate, the focal element $m_{1,\mathrm{CAR}}(\{B, D\}) = 0.008$, implies the exact belief in the group of cars $\{B, D\}$ including the best car from the combined evidence is 0.008. Furthermore, the level of ignorance $m_{1,\mathrm{CAR}}(\Theta) = 0.028$, from the combination of all the judgements of DM1 towards their choice of best car.

The non-specificity of the consumer BOE $m_{1,\mathrm{CAR}}(\cdot)$, $N(m_{1,\mathrm{CAR}}(\cdot)) = 0.160$, is lower than the non-specificity levels for the individual criterion BOE. This is a direct consequence of the utilisation of Dempster's combination rule, which apportions mass values to smaller groups of cars through the intersection of groups of cars from the different criterion BOE (see Table 6).

To consider total beliefs to groups of cars, the belief (*Bel*) and plausibility (*Pls*) functions are utilised on $m_{1,\mathrm{CAR}}(\cdot)$ associated with DM1. Rather than present the belief and plausibility values for each possible subgroup of cars considered (62 in number), a specific reduced number are described. Moreover, Table 8 reports those groups of cars that have the largest belief and plausibility values from all those groups of cars of the same size.

Table 8 Groups of DAs with largest belief and plausibility values from the $m_{1,\text{CAR}}(\cdot)$ BOE (from the combination of the criterion BOEs associated with the judgements from DM1)

Size of car group	Belief	Plausibility
1	{F}, 0.292	{F}, 0.366
2	{E, F}, 0.507	{D, F}, 0.575
3	{D, E, F}, 0.708	{D, E, F}, 0.752
4	{C, D, E, F}, 0.861	{C, D, E, F}, 0.897
5	{A, C, D, E, F}, 0.951	{A, C, D, E, F}, 0.997

To illustrate the results in Table 8, considering all groups of cars made up of three cars, those with the largest belief and plausibility values are $\{D, E, F\}$ in both cases, with $Bel(\{D, E, F\}) = 0.708$ and with $Pls(\{D, E, F\}) = 0.752$. These values are calculated from the information reported in Table 7, and are constructed as shown below;

$$\begin{aligned} Bel(\{D,E,F\}) &= \textstyle\sum_{x_2 \subseteq \{D,E,F\}} m_{1,\mathbf{CAR}}(x_2), \\ &= m_{1,\mathbf{CAR}}(\{D\}) + m_{1,\mathbf{CAR}}(\{E\}) + m_{1,\mathbf{CAR}}(\{F\}) \\ &\quad + m_{1,\mathbf{CAR}}(\{D,E\}) + m_{1,\mathbf{CAR}}(\{E,F\}), \\ &= 0.181 + 0.169 + 0.292 + 0.020 + 0.046, \\ &= 0.708, \end{aligned}$$

and

$$\begin{aligned} Pls(\{D,E,F\}) &= \textstyle\sum_{\{D,E,F\} \cap x_2 \neq \varnothing} m_{1,\mathbf{CAR}}(x_2), \\ &= m_{1,\mathbf{CAR}}(\{D\}) + m_{1,\mathbf{CAR}}(\{E\}) + m_{1,\mathbf{CAR}}(\{F\}) + m_{1,\mathbf{CAR}}(\{B,D\}) \\ &\quad + m_{1,\mathbf{CAR}}(\{C,E\}) + m_{1,\mathbf{CAR}}(\{D,E\}) + m_{1,\mathbf{CAR}}(\{E,F\}) \\ &\quad + m_{1,\mathbf{CAR}}(\Theta), \\ &= 0.181 + 0.169 + 0.292 + 0.008 + 0.008 + 0.020 + 0.046 + 0.028, \\ &= 0.752. \end{aligned}$$

The results in Table 8 highlight the use of DS/AHP to identify a reduced number of cars to possibly further consider, put simply they are potential consideration sets (of different sizes). For the car choice problem here, if considering finding only the single best car, the measures of belief and plausibility both indicate the car F (BMW 3) is best, based on all the judgements from DM1. This discussion and results in Table 8 illustrate the possible role of DS/AHP as a method to identify choice sets from consideration and/or awareness sets (see later for further discussion). Indeed, it is a further consideration in terms of a marketing intelligent system, with results able to be viewed by the DM (or external analyst) with respect to the best car (or group of cars).

To set against the analysis on the judgements made by DM1, so far described, a further series of results are briefly reported based on the judgements of a second consumer (labelled DM2), see Figure 3.

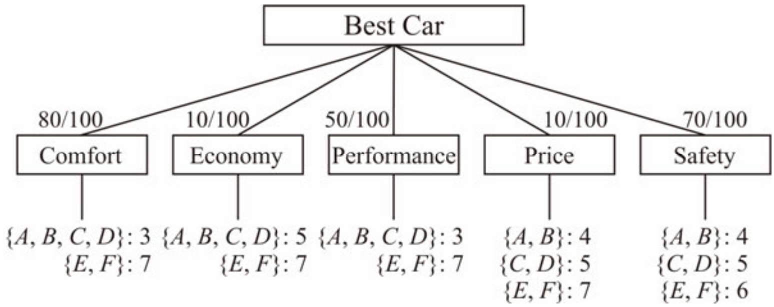


Fig. 3 Hierarchy of judgements made by DM2, when using DS/AHP, on car choice problem

DM2s' judgements are considerably less specific than those with DM1 (see Figures 1 and 3), with larger sized groups of cars identified by DM2 over the five criteria. Incidentally the judgements made by DM2 are consistent with the dyad grouping of the cars presented to the consumers. With the cars grouped by, {*A*, *B*}, {*C*, *D*} and {*E*, *F*}, suggesting a level of brand name valence by this consumer. Their judgements exhibiting influence by the price quality tiers of the three dyad groups of cars. Reinforcing the notion of flat chunks organisation and its relation to retrieval structures (Gobet, 2001). As with DM1, the criterion BOE graphs for the comfort and price criteria for DM2 are reported in Figure 4.

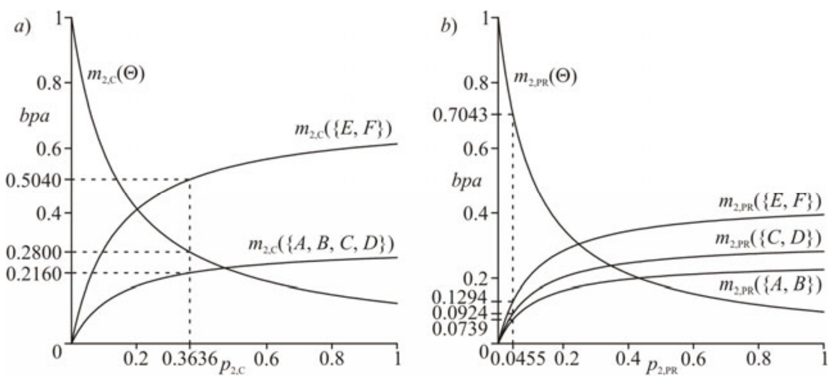


Fig. 4 Criterion BOE $m_{2,C}(\cdot)$ and $m_{2,PR}(\cdot)$ mass values, as $p_{2,C}$ and $p_{2,PR}$ go from 0 to 1

Comparing the results reported in Figures 2 and 4, the separation between the $m_{2,C}(\{E, F\})$ and $m_{2,C}(\{A, B, C, D\})$ lines in Figure 4*a* is a consequence of the large difference between the scale values 3 and 7 assigned to the two groups of cars $\{A, B, C, D\}$ and $\{E, F\}$, respectively. The non-specificity levels on the criterion BOEs for DM2 are reported in Table 9 and exhibit consistently higher values than those associated with DM1 (see Table 5). This is a consequence of the larger sized groups of cars identified across all the criteria by DM2. A further consequence of the less specific judgements made is the non-specificity of consumer BOE for DM2 (1.137) is considerably larger than that for DM1 (0.160).

Table 9 Levels of non-specificity on judgements made by DM2

Evidence	Comfort	Economy	Performance	Price	Safety
Non-specificity	1.660	2.260	1.793	2.116	1.422

While the views of the individual consumers are of interest, the combination of the evidence from the 11 consumers would offer information on the overall levels of belief towards the identification of the best car(s) from the six cars considered over the five criteria. That is, the combination of all the consumer BOEs from the 11 consumers enables a novel approach to the evaluation of results from group decision-making with DS/AHP. This is undertaken by the utilisation of Dempster's combination rule (as described previously). For brevity we do not present the final group BOE from all consumers (consumer BOEs), instead the identified best groups of cars of different sizes, based on the belief and plausibility measures, are reported in Table 10.

Table 10 Groups of cars with largest belief and plausibility values from final group BOE

Size of car group	Belief	Plausibility
1	{D}, 0.665	{D}, 0.655
2	{D, F}, 0.975	{D, F}, 0.975
3	{C, D, F}, 0.991	{C, D, F}, 0.991
4	{C, D, E, F}, 1.000	{C, D, E, F}, 1.000
5	{B, C, D, E, F}, 1.000	{B, C, D, E, F}, 1.000

From Table 10, irrespective of whether belief or plausibility measures are considered the same group of cars is identified for each specific size of group. The best single car is identified as the car *D* (VOLVO S60), if a choice set of say three cars was considered then the group of cars $\{C, D, F\}$ should be chosen (as the respective consideration sets). The results in Table 10 exhibit the possible consideration or choice sets that the consumers could further consider (see later).

At each stage of the DS/AHP analysis certain BOE are constructed and can be combined in a number of different ways to allow further understanding of the prevalent judgements made (offering alternative insights making up the marketing information system). For example, each consumer BOE was found from the combination of criterion BOEs, and the group BOE found from the combination of the consumer BOEs. To gauge a measure on the judgements made specifically over the different criteria, the criterion BOEs associated with a single criterion from the eleven consumers can be combined. The result is five BOEs (one for each criterion), Table 11 reports their concomitant levels of non-specificity.

Table 11 Levels of non-specificity for the different criteria (from group evidence)

Evidence	Comfort	Economy	Performance	Price	Safety
Non-specificity	0.130	0.322	0.088	0.176	0.051

An inspection of the results in Table 11 shows the criterion with overall least and largest levels of non-specificity in the judgements made are safety (0.051) and economy (0.322), respectively. This result is interesting in that overall safety was judged on most discernibly in terms of both the grouping of cars under this criterion and the level of criterion priority value each consumer assigned to it, whereas the economy criterion was most non-specific. This could be a direct consequence of the information made available to the consumers not including all that was necessary for them to make more specific judgements. The combined judgements of the eleven consumers over the different criterion are next exposited in Table 12. That is, for each criterion using the defined combined BOE the different sized groups of cars with highest belief and plausibility values (not given) are shown.

Table 12 Groups of cars with largest belief and plausibility values from different criteria

Belief	Comfort	Economy	Performance	Price	Safety
1	{D}	{D}	{F}	{A}	{D}
2	{D, F}	{C, D}	{C, F}	{A, B}	{D, F}
3	{D, E, F}	{C, D, E}	{C, E, F}	{A, B, C }	{D, E, F}
4	{C, D, E, F}	{B, C, D, E}	{C, D, E, F}	{A, B, C, D}	{C, D, E, F}
5	{B, C, D, E, F}	{B, C, D, E, F}	{A, C, D, E, F}	{A, B, C, D, E}	{A, C, D, E, F}
Plausibility	Comfort	Economy	Performance	Price	Safety
1	{D}	{D}	{F}	{A}	{D}
2	{D, F}	{B, D}	{C, F}	{A, C}	{D, F}
3	{D, E, F}	{B, D, E }	{C, E, F}	{A, B, C }	{C, D, F}
4	{C, D, E, F}	{B, D, E, F}	{C, D, E, F}	{A, B, C, D}	{C, D, E, F}
5	{B, C, D, E, F}	{B, C, D, E, F}	{A, C, D, E, F}	{A, B, C, D, E}	{A, C, D, E, F}

From Table 12, in terms of a single best car to identify, three of the five criteria (comfort, economy and safety) all suggest the car *D* as best choice of car. With cars *F* and *A* identified as best from the criteria performance and price respectively (based on belief or plausibility values). The results from the price criterion are interesting and also in some way different to those from the other criteria. That is, (considering only the belief value) the best two cars to consider under the price criterion are *A* and *B* - the cheapest two of the six cars considered. Also (for the price criterion) the best four cars to further consider are *A*, *B*, *C* and *D*, the cheapest four of the six cars. The reader is reminded the six cars considered were presented to the consumers in the dyad groups {*A*, *B*}, {*C*, *D*} and {*E*, *F*} based primarily on their prices.

The results presented here show the individual consumers generally followed this dyadic grouping. This highlights the effect of brand cues, which in this case were in the form of the folders containing extensive information and pictures about each car. Indeed with the price clearly included in the cue information, the results on the price criterion indicate the consumers have exhibited "mentally defined" price-quality tiers during their judgement making. This finding is supported by the research study conducted by Mehta *et al.* (2003).

4 Future Trends

The central element in the DS/AHP analysis, using the Dempster-Shafer theory of evidence of evidence, is the body of evidence (BOE), with certain BOE constructed at different stages in the analysis, with a number of different sets of results able to be found. The descriptive measures, conflict and non-specificity, allow a novel insight into the judgement making by the individual consumers and well as the combined judgements of the consumer group. Further analysis could include the investigation of the levels of conflict between the individual members of the group and looking into the possible identification of subgroups of a group with the most similar series of judgements.

With a large emphasis given on the elucidation of consideration sets of DAs, it would be interesting to investigate, from a more marketing direction, the pertinence of the identified groups of DAs. Moreover, the idea of consideration sets is exhibited in the judgement making opportunities of the consumer and in the interpretation of the final results, based on the levels of belief and plausibility in the best car existing in a group of cars, which could be compared with the algorithm developed by Gensch and Soofi (1995) to estimate the inclusion of DAs in a consideration set (average of the selected alternative probabilities was proposed as a statistic by which the predictive quality of various consideration set can be compared).

5 Conclusions

This chapter has utilised a nascent approach to multi-criteria decision-making, namely DS/AHP in the area of consumer choice. With the fundamentals of DS/AHP based on the Dempster-Shafer theory of evidence, the analysis is undertaken in the presence of ignorance and non- specificity. Indeed, Dempster-Shafer theory is a core technique associated with probabilistic reasoning, itself one of the methodologies making up soft computing.

The chapter has attempted to convey a realistic approach for the individual consumer to undertake the required judgement making process. Importantly, the DS/AHP method allows the consumer to control the intensity of the judgement making they perform. The results (intermediate and final) elucidate a plethora of information for the consumer choice problem to be gauged on.

Allowance exists, using DS/AHP, for each consumer to assign levels of positive preference to groups of cars. The results also included information on groups of cars, hence the notion of consideration sets is firmly implanted in the fundamentals of DS/AHP.

References

Aurier, P., Jean, S., Zaichkowsky, J.L.: Consideration Set Size and Familiarity with Usage Context. Advances in Consumer Research 27, 307–313 (2000)

Bayes, T.: An essay toward solving a problem in the doctrine of chances. Phil. Trans. Roy. Soc. (London) 53, 370–418 (1963)

Beynon, M.: DS/AHP method: A mathematical analysis, including an understanding of uncertainty. European Journal of Operational Research 140(1), 149–165 (2002a)

Beynon, M.J.: An Investigation of the Role of Scale Values in the DS/AHP Method of Multi-Criteria Decision Making. Journal of Multi-Criteria Decision Analysis 11(6), 327–343 (2002b)

Beynon, M.J., Curry, B., Morgan, P.H.: The Dempster-Shafer Theory of Evidence: An Alternative Approach to Multicriteria Decision Modelling. OMEGA 28(1), 37–50 (2000)

Beynon, M.J.: Understanding Local Ignorance and Non-specificity in the DS/AHP Method of Multi-criteria Decision Making. European Journal of Operational Research 163, 403–417 (2005)

Beynon, M.J.: The Role of the DS/AHP in Identifying Inter-Group Alliances and Majority Rule within Group Decision Making. Group Decision and Negotiation 15(1), 21–42 (2006)

Bloch, B.: Some aspects of Dempster-Shafer evidence theory for classification of multimodality images taking partial volume effect into account. Pattern Recognition Letters 17, 905–919 (1996)

Bonissone, P.P.: Soft Computing: The Convergence of Emerging Reasoning Technologies. Soft Computing 1, 6–18 (1998)

Bryson, N., Mobolurin, A.: A process for generating quantitative belief functions. European Journal of Operational Research 115(3), 624–633 (1999)

Chakravarti, A., Janiszewski, C.: The Influence of Macro-Level Motives on Consideration Set Composition in Novel Purchase Situations. Journal of Consumer Research 30, 244–258 (2003)

Chase, W.G., Simon, H.K.: Perception in Chess. Cognitive Psychology 4, 55–81 (1973)

Dempster, A.P.: A generalization of Bayesian inference (with discussion). J. Roy. Stat. Soc., Series B 30(2), 205–247 (1968)

Desai, K.K., Hoyer, W.D.: Descriptive Characteristics of Memory-Based Consideration Sets: Influence of Usage Occasion Frequency and sage Location Familiarity. Journal of Consumer Research 27(3), 309–323 (2000)

Dubois, D., Prade, H.: A note on measures of specificity for fuzzy sets. International Journal of General Systems 10(4), 279–283 (1985)

Ducey, M.J.: Representing uncertainty in silvicultural decisions: an application of the Dempster-Shafer theory of evidence. Forest Ecology and Management 150, 199–211 (2001)

Gensch, D.H., Soofi, E.S.: Information-Theoretic Estimation of Consideration Sets. International Journal of Research in Marketing 12, 25–38 (1995)

Gilbride, T., Allenby, G.: A Choice Model with Conjuctive, Discjunctive, and Compensatory Screening Rules. Marketing Science 23(3), 391–406 (2004)

Gobet, F.: Chunk Hierarchies and Retrieval Structures: Comments on Saariluoma and Laine. Scandinavian Journal of Psychology 42, 149–155 (2001)

Gobet, F., Simon, H.A.: Expert Chess Memory: Revisiting the Chunking Hypothesis. Memory 6(3), 225–255 (1998)

Hastak, M., Mitra, A.: Facilitating and Inhibiting Effects of Brand Cues on Recall, Consideration Set and Choice. Journal of Business Research 37(2), 121–127 (1996)

Hogarth, R.M.: Judgement and Choice, 2nd edn. Wiley, New York (1980)

Kastiel, D.L.: Computerized consultants. Business Marketing 72(3), 52–74 (1987)

Klir, G.J., Wierman, M.J.: Uncertainty-Based Information: Elements of Generalized Information Theory. Physica-Verlag, Heidelberg (1998)

Lipshitz, R., Strauss, O.: Coping with Uncertainty: A Naturalistic Decision-Making Analysis. Organisational Behaviour and Human Decision Processes 69(2), 149–163 (1997)

Lootsma, F.A.: Scale Sensitivity in the Multiplicative AHP and SMART. Journal of Multi-Criteria Decision Analysis 2, 87–110 (1993)

Luce, M.F., Payne, J.W., Bettman, J.R.: Emotional Trade-off Difficulty and Choice. Journal of Marketing Research 36(2), 143–159 (1999)

Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. The Psychological Review 63, 81–97 (1956)

Murphy, C.K.: Combining belief functions when evidence conflicts. Decision Support Systems 29, 1–9 (2000)

Park, C.W., Jun, S.Y., MacInnis, D.J.: Choosing What I Want Versus Rejecting What I Do Not Want: An Application of Decision Framing to Product Option Choice Decisions. Journal of Marketing Research XXXVII, 187–202 (2000)

Punj, G., Brookes, R.: Decision constraints and consideration-set formation in consumer durables. Psychology and Marketing 18(8), 843–863 (2001)

Raffone, A., Wolters, G.: A Cortical Mechanism for Binding in Visual Working Memory. Journal of Cognitive NeuroScience 13(6), 766–785 (2001)

Roberts, J.H., Lattin, J.M.: Consideration: Review of Research and Prospects for Future Insights. Journal of Marketing Research XXXIV, 406–410 (1997)

Roberts, J., Nedungadi, P.: Studying consideration in the consumer decision process: Progress and challenges. International Journal of Research in Marketing 12, 3–7 (1995)

Roesmer, C.: Nonstandard Analysis and Dempster-Shafer Theory. International Journal of Intelligent Systems 15, 117–127 (2000)

Saaty, T.L.: A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology 15, 59–62 (1977)

Shafer, G.: A Mathematical theory of Evidence. Princeton University Press, Princeton (1976)

Shafer, G.: Perspectives in the theory of belief functions. International Journal of Approximate Reasoning 4, 323–362 (1990)

Shafir, E.: Choosing versus rejecting: Why Some Options are Better and Worse than others. Memory and Cognition 21(4), 546–556 (1993)

Shapiro, S., MacInnis, D.J., Heckler, S.E.: The effects of incidental ad exposure on the formation of consideration sets. Journal of Consumer Research 24(1), 94–104 (1997)

Simon, H.A.: A Behavioral Model of Rational Choice. Quarterly Journal of Economics 69(1), 99–118 (1955)

Smets, P.: Varieties of ignorance and the Need for Well-founded Theories. Information Sciences, 57–58, 135–144 (1991)

Smets, P.: What is Demspeter-Shafer's model? In: Yager, R.R., Fedrizzi, M., Kacprzyk, J. (eds.) Advances in the Dempster-Shafer theory of evidence, pp. 5–34. Wiley, New York (1994)

Talvinen, J.M.: Information systems in marketing: Identifying opportunities for new applications. European Journal of Marketing 29(1), 8–26 (1995)

Verwey, W.B.: Effect of Sequence Length on the Execution of Familiar Keying Sequences: Lasting Segmentation and Preparation? Journal of Motor Behavior 35(4), 343–354 (2003)

Wedel, M., Pieters, R.: Eye Fixations on Advertisements and Memory for Brands: A Model and Findings. Marketing Science 19(4), 297–312 (2000)

Decision Making in Multiagent Web Services Based on Soft Computing

Zhaohao Sun[1,3], Minhong Wang[2], and Dong Dong[3]

[1] Graduate School of Information Technology and Mathematical Sciences University of Ballarat, Mt Helen, Victoria 3353 Australia
e-mail: z.sun@ballarat.edu.au, zhsun@ieee.org
[2] Division of Information & Technology Studies, The University of Hong Kong, HK
e-mail: magwang@hku.hk
[3] School of Computer Science and Technology, College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang, 050016 China
e-mail: donald.ddong@gmail.com

Abstract. Web services are playing an important role in successful business integration and other application fields such as e-commerce and e-business. Multiagent systems and soft computing are intelligent technologies that have drawn an increasing attention in web services. This chapter examines decision making in multiagent web services based on soft computing. More specifically, it proposes a unified multilayer architecture, SESS, and an intelligent system architecture, WUDS. The SESS unifies e-services, web services and infrastructure services into an integrated hierarchical framework. The WUDS aims at implementing decision making in multiagent web services based on soft computing and implementation strategies. Both architectures tie together methodologies such as multiagent system, soft computing techniques such as case-based reasoning (CBR), fuzzy logic and their applications in web services into a unified framework that includes both logical and intelligent embodiment of decision making in web services. The chapter also proposes demand-driven web service lifecycle for service providers, brokers and requesters taking into account their decision making. Finally, the chapter explores unified case-based web services for discovery, composition and recommendation based on fuzzy logic. The proposed approach will facilitate the research and development of web services, e-services, intelligent systems and soft computing.

Keywords: Web services, soft computing, multiagent systems, intelligent systems, decision making, case-based reasoning.

1 Introduction

Decision making is essential not only for traditional commerce but also for web services. Web services are playing an increasingly important role in successful

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 389–415.
springerlink.com

business integration and other application fields such as e-commerce and e-business [54]. Web services are the provision of services over electronic networks such as the Internet and wireless networks [36]. The key motive for rapid development of web services is the ability to discover services that fulfil users' demands, negotiate service contracts and have the services delivered where and when the users request them [51]. With dramatic development of the Internet and the web in the past decade, web services have been flourishing in e-commerce, artificial intelligence (AI), soft computing because they offer a number of strategic advantages such as mobility, flexibility, interactivity and interchangeability in comparison with traditional services [19].

The fundamental philosophy of web services is to meet the needs of users precisely and thereby increase the market share and revenue [36]. Web services have helped users to reduce the cost of information technology (IT) operations and allow them to closely focus on their own core competencies [19]. At the same time, for business marketers, web services are very useful for improving interorganizational relationships and generating new revenue streams [45]. Furthermore, web services can be considered as a further development of e-commerce, because they are service-focused paradigms that use two-way dialogues to build customized service offerings, based on knowledge and experience about users to build strong customer relationships [36]. It implies, however, that one of the intriguing aspects of web services is that any web service cannot avoid similar challenges encountered in traditional services such as how to meet the customer's demands in order to attract more customers.

The current computer programs can do little to reason and infer knowledge about web services [54]. Current research trend is to add intelligent techniques to web services to facilitate discovery, invocation, composition, and recommendation of web services. This is why intelligent web services have been drawing an increasing attention [26]. However, there are still less intelligent techniques for facilitating main stages of the entire web service lifecycle.

Soft computing has found many successful applications in multiagent web services [26]. However, it is still a major issue for decision making in web services although the customer can obtain web services through the web. For example, web services providers need to address rational and irrational customer concerns regarding the adoption of new web services, and improve support for customers who wish to customize web service applications [19]. Further, there is no unified treatment for decision making in web services using soft computing taking into account web services lifecycle although there have been a great number of researches on web service discovery and composition [60].

This chapter will address the above mentioned issues by examining decision making in multiagent web services based on soft computing. More specifically, it proposes a unified multilayer architecture, SESS, and an intelligent system architecture, WUDS. The SESS unifies e-services, web services and infrastructure services into an integrated hierarchical framework. The WUDS is a system architecture for implementing decision making in multiagent web services based on soft computing. Both architectures tie together methodologies such as multiagent system, soft computing techniques such as case-based reasoning

(CBR), fuzzy logic and their applications in web services into a unified framework that includes both logical and intelligent embodiment of decision making in web services. The chapter also proposes a demand-driven web service lifecycle taking into account decision making. Finally, the chapter explores a unified case-based web services for discovery, composition and recommendation. To this end, the remainder of this chapter is organized as follows: Section 2 looks at the fundamental of web services. Section 3 proposes SESS: a unified multilayer architecture for integrating e-services, web services and infrastructure services into a hierarchical framework. Section 4 examines web service lifecycle by proposing demand-driven web service lifecycle for service providers, requesters and brokers. Section 5 discusses decision making in web services. Section 6 looks at soft computing for web services. Section 7 proposes WUDS, a unified decision support system architecture for decision making in web services based on soft computing. Section 8 explores case-based web services. The final section ends the chapter with some concluding remarks and future work.

2 Fundamentals for Web Services

This section examines e-services, web services and their relationships. It also looks at the parties involved in web services and corresponding architectures.

2.1 E-Services and Web Services

E-services are "electronic offerings for rent" made available via the Internet that complete tasks, solve problems, or conduct transactions [19]. Song states that e-services have the following features: integration, interaction, customization, self-services, flexibility and automatic response [43]. E-services allow customers to review accounts, monitor shipments, edit profiles, schedule pick-ups, adjust invoices, return merchandises and so on. HP defines e-services as the means by which an enterprise offers its products, services, resources, and know-hows via the Internet [9].

Some e-services, e.g. Amazon.com, are integrated with e-commerce applications such as shipping information, package tracking and rate inquiries [43]. E-services are seamlessly integrated with e-commerce applications to make the shipping experience simple and convenient. Therefore, e-services can be considered as a further development of e-commerce [36]. Services play a vital role in industry in many developed countries, service-oriented businesses made up about two-thirds of the economy [37] (p. 36). Therefore, any e-business or e-commerce activity is a kind of e-services.

Web services have drawn an increasing attention in building distributed software systems across networks such as the Internet, and also in business process reengineering [55]. Web services are defined from an e-commerce viewpoint at one extreme, in this case, web services are the same as e-services. At another extreme, web services are defined from a computer science viewpoint. For example, web services are defined as the network enabled reusable components that conform to an interface with standard description format and access protocols [61]. Between these two extremes, many different definitions have been proposed by different authors. For example, web services are self-contained, modular applications, accessible via the web, that provide a set of functionalities to businesses or individuals [51]. Further, there are also different levels for defining web services from a methodological viewpoint. For example, a web service is a way of publishing an explicit, machine-readable, common standard description of how to use a service and access it via another program using some standard message transports [33]. Others are at an intermediary level, for example, a web service is an operation typically addressed via a URI (Uniform Resource Identifier), declaratively described using widely accepted standards, and accessed via platform-independent XML-based messages [1] (p. 124). A more technical definition of web services is as follows: A web service [1] (p. 125) is "a standardized way of integrating web-based applications using the XML, SOAP (Simple Object Access Protocol), WSDL (Web Services Description Language), and UDDI (Universal Description Discovery and Integration) open standards over an Internet protocol backbone. XML is used to tag the data. SOAP is a protocol for exchanging XML messages over the web. WSDL is used for describing the e-service available. UDDI is used for listing what services are available."

This chapter considers web services as simple, self contained applications that perform functions, from simple requests to complicated business processes, taking into account the hierarchical relationship between e-services and web services, which is detailed in Section 3.

2.2 Parties in Web Services

There are mainly three parties related to web services: web service requester, web service broker, and web service provider [41][45]. Web service requester also denotes web service user, buyer, customer, consumer, receiver, and its intelligent agent. Web service broker denotes web service intermediary, middle agent and its intelligent agent. Web service provider denotes web service owner, seller, sender and its intelligent agent. A simple service oriented architecture (SOA) for web services was introduced by [41] (p. 20). In this architecture, web service providers create web services and advertise them to potential web service requesters by registering the web services with web service brokers, or simply offers web services [9]. The web service provider also needs to describe the web service in a standard format, and publish it in a central service registry. The service registry contains additional information about the service provider, such as address and contact of the providing company, and technical details about the service. Web service providers may integrate or compose existing services [25] using intelligent techniques. They may also register descriptions of services they offer, and monitor and manage service execution [9]. Web service requesters retrieve the information from the registry and use the service description obtained to bind to and invoke the web service. Web service brokers maintain a registry of published web

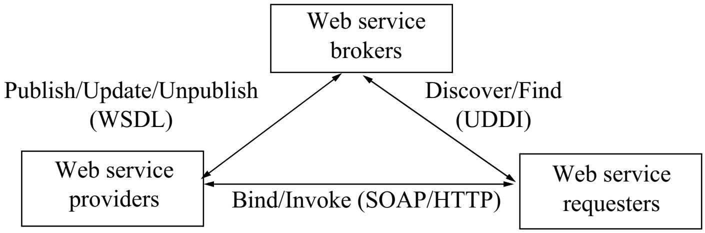


Fig. 1 A SOA for web services based on [41] and [50]

services and might introduce web service providers to web service requesters. They use UDDI to find the requested web services, because UDDI specifies a registry or "yellow pages" of services [41] (p. 20). They also provide a searchable repository of service descriptions where service providers publish their services, service requesters find services and obtain binding information for these services.

This SOA is simple because it only includes three parties (as mentioned above) and three basic operations: publish, find and bind. In fact, some behaviors of agents are also fundamentally important to make e-services successful. These fundamental behaviors at least include communication [41][46], interaction [41][46], collaboration [41][46], cooperation [41][46], coordination [41][46], negotiation [41][46], trust [41] and deception [46].

Papazoglou [27] proposes an extended SOA. The parties involved in this architecture are more than that in the simple SOA, because it includes service provider, service aggregator, service client, market maker, and service operator.

A service aggregator is a service agent that consolidates services provided by other service providers into a distinct value-added service [27]. It develops specifications and/or codes that permit the composite service to perform functions such as coordination, monitoring quality of service (QoS) and composition.

Web market makers aim to establish an efficient service-oriented market to facilitate the business activities among service providers to service brokers and service requesters. In the traditional market, the service broker is working in the market, while the market maker makes the market running.

The web service operator is responsible for performing operation management functions such as operation, assurance and support [27].

From a viewpoint of multiagent system [45], there are still other parties involved in web services, such as web service advisor, web service manager, and web service composer and so on. Some of them will be mentioned in the later sections.

3 SESS: A Unified Multilayer Architecture for E-Services

This section will propose a multilayer system architecture for integrating e-services, web services and infrastructure services.

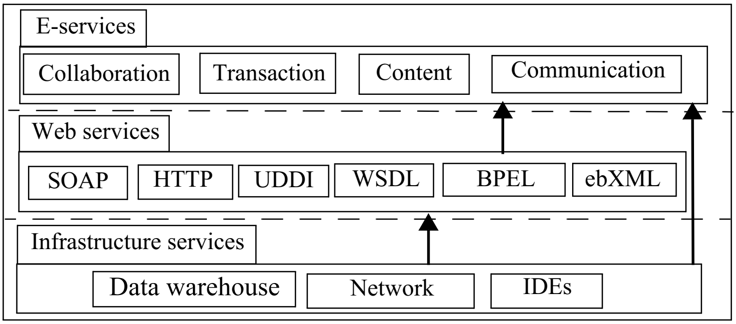


Fig. 2 A unified system architecture for e-services (SESS)

There are many system architectures available for e-services and web services [32][41][45]. These architectures provide a high level system design free of implementation details. For example, Vissers et al [53] propose a reference model for e-services taking into account open systems interconnection, which consists of seven layers: application, presentation, session, transport, network, data link and physical layer. Kreger [20] proposes a three-layer architecture for web services comprised of wire layer, description layer, and discovery agency layer. He demonstrates that web services technologies are being developed as the foundation of a new generation of B2B e-commerce. However, the relationship between e-services and web services has not been examined in a unified way based on different perspectives that include commerce, IT and information systems (IS), and ordinary customers. In order to resolve the above mentioned issue, we propose a unified multilayer system architecture for e-services (SESS), as shown in Fig. 2. The SESS consists of three layers: infrastructure services layer, web services layer and e-services layer.

The infrastructure services layer supports the realization of web services. The web services layer consists of software or software components that support the realization of e-services. The e-services layer is directly interacting with the e-service customer [45]. Therefore, these three distinct services are on three different layers, and constitute an integrated system. The infrastructure services are resided either on servers of internet service providers (ISP) [38]. Web services are resided on the server side that is normally managed by the e-service providers. E-services are accessed by customers on the client side. The following subsections will discuss each of these three layers in some detail.

3.1 First Layer: Infrastructure Services

The infrastructure services refer to the basic technology platform and features needed to implement web services, which at least consist of data warehouse, network and integrated development environments (IDEs).

A data warehouse or knowledge repository is typically defined by its content and structure [7]. A data warehouse consists of many databases related to implementation of web services. A knowledge repository could either be populated with data or documents. Both data warehouse and knowledge repository have been designed to capture text and multimedia information such as media documents.

Network provides communication services such as communication between different web services, between servers, between clients, between servers and clients. ISP is one of the most important providers of such communications.

IDEs provide an integrated environment for developing web services. IDEs include programming languages and special development tools used to implement web services such as Java, C++, Javascript, PHP, and XML.

It should be noted that the infrastructure services layer in the SESS is not specific for e-services and web services but also for other systems and applications such as knowledge management systems [7]. Further, the infrastructure services are basically studied by computer science/engineering scientists and electrical engineering scientists, while few IS researchers get involved in this layer.

3.2 Second Layer: Web Services

A web service is a server that listens for and replies with SOAP [23], generally via HTTP [29]. In practice, a web service is described by WSDL and is also listed/published in a UDDI registry [8][50]. WSDL descriptions are retrieved from the UDDI directory and allow the software systems of one business to extend to be used by others directly. If a web service comes in, it begins to work by publishing itself to the UDDI registry [22]. BPEL (Business Process Execution Language) supports automated business processes [64]. The e-services are invoked over the web using the SOAP/XMLP protocol [63]. ebXML (electronic business XML) provides a registry similar to UDDI for web service discovery [51], it also supports web service negotiation and transactions. SOAP, HTTP, WSDL, UDDI, ebXML are all considered as web services technologies, standards and protocols [23][54].

Based on above discussion, web services are not only "services" in a traditional sense but also application enablers of e-services. The final goal of web services is to realize e-services. Therefore, a necessary condition for successful e-services is the efficient and effective support of web services. However, this condition is not sufficient, because we have encountered many unsuccessful e-services although they have the efficient and effective support of web services. One of the reasons is that they have not looked into the non-technical aspects in e-services such as customer relationship management [45].

The web services layer is at a technical level. This service layer aims to ensure that different components of web services are operating with acceptable performances [6]. Middleware infrastructure, service deployment and service management [6] are some issues of this layer which are basically studied by computer science/engineering scientists. The IS researchers have been involved in this layer to some extent, whereas business researchers do not get involved in this layer.

3.3 Third Layer: E-Services

The e-services layer is a business application layer that directly interacts with end users such as e-service customers. E-services can be classified into four categories: collaborative services, transaction services, content and communication services [45].

Collaborative services are designed to support groups of interacting people in their cooperative tasks [53], for example, teachers tele-teach students.

Transaction services support formal and often traceable transaction between parties by organizing the exchange of predefined messages or goods in predefined orders [53]. In practice, these services do not cover huge data per transaction. They only exchange simple messages such as orders, bills, contracts and payments.

Content services allow people to access and manipulate e-content such as accessing e-libraries [53]. These kinds of services are very popular in universities to access large amounts of e-journal papers.

Communication services provide at least services such as communication between users, collaboration among users and workflow management [7]. The first communication services are implemented through utilities such as file sharing and e-mailing. The second communication services can be implemented through synchronous meeting and asynchronous discussion forums. The last one allows service users to manage workflow processes by supporting online execution and control of workflow.

In practice, e-services are comprised of only two major categories: free e-services and pay e-services from a business viewpoint.

Free e-services are the services provided by the e-services provider freely to potential e-service receivers [45]. For example, a website for detecting fraud and deception in e-commerce is free for anybody to visit. Free e-services have played an important role in promoting human culture development.

Pay e-services are the services provided by e-services providers to e-service receivers with application and usage fees. These services are a new kind of business activity and have played a vital role for e-services development. The traditional commerce and services are dramatically transferred to e-services to make effective use of the strategic advantages over the traditional services [19]. Collaborative services, transaction services and content services all belong to pay e-services [53]. Further, any pay e-service shares some common business features of traditional business activities: fraud, deception, pursuing maximum profit through bargaining between service providers and service receivers, etc. The service intermediaries such as bargainers, consultants, and brokers facilitate the success of these service activities [45][46].

The main research and development of the e-services layer is on customer satisfaction, customer relationship management and customer experience management, which are basically studied by IS and business researchers [45].

3.4 SESS: A Service Centered System Architecture

The proposed SESS is a service-centered system architecture. The services in the SESS have been in a hierarchical structure, as shown in Fig. 2. The receiver or customer of the infrastructure services is the web service provider and e-service provider, while the receiver of the web services is the e-service provider. The receiver or customer of e-services are ordinary customers such as university students.

It should be noted that some agents play a double role in the SESS [45]. In other words, they are service providers on one layer, they are service receivers on the other layer. For example, the web services provider is the receiver of the infrastructure services and the provider of e-services. Therefore, the services in the SESS constitute a service chain towards e-services.

Based on the above discussion, the competition in e-services not only occurs at the boundary between the customer market and e-services layer but also at the boundary between web services and e-services. It can also occur at the boundary between web services layer and infrastructure services layer in order to obtain the advance of acquiring different kinds of service customers [45]. Therefore, how to make decision to satisfy customer's demand becomes significant in web services, which will be examined in more detail in the later sections.

4 Web Service Lifecycle

This section reviews web service lifecycles, and proposes demand-driven web service lifecycle for web service provider, requester and broker respectively.

4.1 Introduction

There have been a number of attempts to address web service lifecycle (hereafter, WSLC) in the web service community [5]. For example, Leymann [24] discusses a WSLC based on explicit factory-based approach, in which a client uses a factory to create "an instance" of a particular kind of service, the client can then explicitly manage the destruction of such an instance, or it can be left to the grid environment. Sheth [39] proposes a semantic web process lifecycle that consists of web description (annotation), discovery, composition and execution or orchestration. Zhang and Jeckle propose a WSLC that consists of web service modeling, development, publishing, discovery, composition, collaboration, monitoring and analytical control from a perspective of web service developers [61]. Kwon proposes a WSLC consisting of web service identification, creation, use and maintenance [22]. Narendra and Orriens [30] consider the WSLC consisting of web service composition, execution, adaptation, and re-execution, etc. Tsalgatidou and Pilioura [50] propose a WSLC, which consists of two different layers: basic layer and value-added layer. The former contains web service creation, description, publishing, discovery, invocation and unpublishing, all of these activities are necessary to be supported by every web service environment. The latter contains the value-added activities of composition, security, brokering, reliability, billing, monitoring, transaction handling and contracting. These activities bring a better performance to any web service environment. They acknowledge that some of these activities take place at the web service requester's site, while others take place at the web service broker's or provider's site. They also explore technical challenges related to each activity in the WSLC. However, they have not classified the proposed activities of stages in their lifecycle based on web service requester, provider, and broker in detail. Some companies and organizations also propose their own WSLC. For example, W3C proposes a service lifecycle for web service management, which is expressed in the state transition diagrams [59]. Sun considers the WSLC consisting of four stages: design/build, test, deploy/execute, and manage [44], which can be considered as a model for web service developers. However, as mentioned in Section 2.2, web services consist of three main parties: Service providers, service requesters and service brokers [51], different parties require different web service lifecycles. Therefore, what is a web service lifecycle from the viewpoint of a web service provider, broker and requester respectively? How many stages (or activities) does a web service lifecycle consist of? The following subsections will resolve these issues by examining the web service lifecycle from a demand viewpoint.

From a perspective of computer science, lifecycle originated from software engineering [35]. It describes the life of a software product from its conception, to its implementation, delivery, use, and maintenance [34]. A traditional software development lifecycle mainly consists of seven phases: planning, requirements analysis, systems design, coding, testing, delivery and maintenance. Based on this, a web service lifecycle consists of the start of web service and the end of web service and its evolutionary stages that transform web service from the start to the end. Further, demand chain management is an important factor for market and economy development [52]. The decrease of demand is an implication for economic recession. Different parties generally have different demands of web services and then different web service life-cycles. Therefore, we will examine web service lifecycle from a demand perspective of service provider, broker and requester respectively.

4.2 Provider's Demand Driven Web Service Lifecycle

From a web service provider's demand perspective, a web service lifecycle mainly consists of web service identification [22][23][51], description/representation, creation (design/build, test, deploy) [22][51], publishing [51], unpublishing, composition [25][51], invocation, use and reuse [22], execution or orchestration, management and monitoring [9][51], maintenance [22], billing and security [51]. In what follows, we only examine some of the above mentioned activities owing to space limitation. This is also true for the following two subsections.

Web service identification is to identify appropriate services [23]. Web service invocation is to invoke the discovered web service interface [23]. Web services are published to intranet or the Internet repositories for potential users to locate [51]. Web service unpublishing is sometimes no longer available or needed, or it has to be updated to satisfy new requirements [51].

Web service composition primarily concerns requests of web service users that cannot be satisfied with any available web service [30]. It combines a set of available web services to obtain a composite service. Therefore, web service composition refers to the process of creating customised services from existing services by a process of dynamic discovery, integration and execution of those services in a deliberate order to satisfy user requirements [25][57]. It refers to intelligent techniques and efficient mechanisms of composing arbitrarily complex services from relatively simpler web services. Service composition can be either performed by composing elementary or composite services. Composite services in turn are recursively defined as an aggregation of elementary and composite services [9].

There are many techniques existing for web service composition. For example, Tang et al [51] propose an automatic web service composition method taking into account both services' input/output type compatibility and behavioral constraint compatibility. Further, Dustdar and Schreiner [9] discuss the urgent need for service composition and the required technologies to perform service composition as well as present several different composition strategies.

4.3 Requester's Demand Driven Web Service Lifecycle

From a web service requester's demand perspective, a web service lifecycle mainly consists of web service consultation, search [23], matching [23], discovery [51], composition, mediation [23], negotiation, evaluation and recommendation.

Web service discovery is a process of finding most appropriate web services needed by a web services requester [17]. It consists of web service identification (identifies a new web service) and detects an update to a previously discovered web service [23]. Services may be searched, matched, and discovered by service requesters by specifying search criteria and then be invoked [9][51]. Service invocation is restricted to authorised users [9].

Web service mediation is to mediate the request of web service from the web service requester. Web service negotiation consists of a sequence of proposal exchanges between the two or more parties with the goal of establishing a formal contract to specify agreed terms on the service [60]. Through negotiation, web service requesters can continuously customize their needs, and web service providers can tailor their offers. In particular, multiple web service providers can collaborate and coordinate with each other in order to satisfy a request that they can't process alone.

However, a web service requester might not need to know how the web services are retrieved, discovered and composed internally. Therefore, search, matching, and composition might less important for a web service requesters.

4.4 Broker's Demand Driven Web Service Lifecycle

Brokering is the general act of mediating between requesters and providers to match requester's needs and providers' offerings. It is a more complete activity than discovery [51]. A broker should enable universal service-to-service interaction, negotiation, bidding and selection of the highest quality of service (QoS) [41] (p.345-46). Brokering is supported by HP web services platform as a HP web intelligent broker [50]. After discovering web service providers that can respond to a user's service request, HP web services platform negotiates between them to weed out those that offer services outside the criteria of the request.

From a web service broker's demand perspective, a web service lifecycle mainly consists of web service consultation, personalization, search, matching, discovery, adaptation, composition, negotiation, recommendation, contracting and billing.

Web service consultation can be considered as the start of the web service lifecycle, because the web service broker begins to consultation as soon as a web service customer provides a request for a web service. In order to provide a service consultation, the web service broker has to conduct web service search, like Google does. During the web service search, the web service broker uses any techniques of web service matching such as CBR [46]. After discovering a number of web services, the web service broker can select one of them to recommend it to the web service customer. If the customer accepts the recommended web service, then the web service can be considered as a web service reuse; that is, the existing web service has been reused by customers.

Web service recommendation can be conducted through optimization, analysis, forecasting, reasoning and simulation, for example, an inference engine is a solver for making decisions through reasoning [22].

Different web service customers have different preferences. Therefore, a web service broker has to personalize web services in order to meet the requirement of the web service customer satisfactorily. It is necessary to compose web services based on the requirement of customers to personalize the web service. At the same time, web service composition allows web service broker to create a composite web service for customers rapidly [51].

Billing concerns service brokers and service providers [51]. Service brokers create and manage taxonomies, register services and offer rapid lookup for services and companies. They might also offer value-added information for services, such as statistical information for the service usage and QoS data.

4.5 Summary of Demand Driven Web Service Lifecycle

Based on the above discussion, the activities involved in the demand-driven web service lifecycle for web service provider, requester and broker can be summarized in

Table 1. Some of the detailed activities have not been listed in Table 1 owing to space limitation.

Table 1 Demand driven web service lifecycle

Parties in web services	identification	representation	search/matching	discovery	consultation	personalization	composition	recommendation	adaptation	mediation	negotiation	invocation	billing	contract
Provider	X	X	X				X					X	X	X
Requester			X	X	X	X	X	X	X	X	X	X	X	X
Broker			X	X		X	X	X	X	X	X		X	X

As shown in Table 1, some activities in web services are common demands of the main players: service providers, brokers, and requestors. This means that they share the some web service activities. However, different parties in web services demand the same activity in a different way. For example, the service provider demanding "web services search" means that s/he asks web services developers or her/his technology agents to provide efficient web services search function for his or her business. On the other hand, the service requestor demanding "web services search" means that s/he requires a fast search function from the service provider or broker in order to obtain the most satisfactory web service as soon as possible.

Search and matching are not unique activities related to web services, they are also involved in database and case based reasoning (CBR). Google uses search and matching to provide services. Adaptation, retrieval, classification [23], use/reuse [22], retention or feedback are not only in web services but also in CBR cycle [46]. Web service invocation, binding, billing, contract [51] can be considered as common features for any business activities. Therefore, we need not discuss each of them in detail in the context of web services.

Based on the above discussion, the service requestors demand the service providers and brokers for web services discovery and recommendation; the service brokers demand the service providers for web services discovery and composition; the service providers demand up-to-date techniques and tools for web service discovery, composition and recommendation. Therefore, the most important activities in web services can be web service discovery, composition and recommendation. We will examine these activities based on soft computing and its corresponding decision making in the following sections.

5 Decision Making in Web Services

This section first reviews decision making and then looks at decision making of web service provider, requester and broker respectively.

5.1 Decision Making

The term "decision" can have many different meanings, depending on who uses [62] (pp. 241). If a computer scientist uses it, then decision is a special kind of information processing and problem solving; if a statistician uses it, decision might be a mathematical model. Some decisions are formal, whereas other decisions are described in natural language.

In classical (normative, statistical) decision theory [62] (pp. 241), a decision can be characterized by a set of decision alternatives (decision space); a set of states of nature (state space); a relation assigning a result to each pair of a decision and state; and finally, the utility function orders the results based on the desirability of the decision supporter. When decision making under uncertainty, the decision maker does not know exactly which state will occur. In this case, the decision making becomes more difficult. For brevity, we will restrict our attention to decision making in the following sense: The decision space can be described either by enumeration or by a number of constraints. The utility function orders the decision space via a one-to-one relationship of results to decision alternatives. Hence, we can only have one utility function applying to the order, and we may have several constraints defining the decision space.

The following is a model for decision making in a fuzzy environment proposed by Bellman and Zadeh [3][62]. Suppose that there are *n* fuzzy goals $G_1, G_2, \ldots, G_n$ and *m* fuzzy constraints $C_1, C_2, \ldots, C_m$. Then, the resultant decision *D* is the intersection of the given goals $G_1, G_2, \ldots, G_n$ and the given constraints $C_1, C_2, \ldots, C_m$; that is,

$$D = G_1 \cap G_2 \cap \ldots \cap G_n \cap C_1 \cap C_2 \cap \ldots \cap C_m \tag{1}$$

where $\cap$ is an operator of fuzzy intersection. It should be noted that fuzzy decision making at least consists of fuzzy linear programming, fuzzy dynamic programming, fuzzy multicriteria analysis, multiobjective decision making, multiattributive decision making [62] (pp. 241-282). From the viewpoint of fuzzy linear programming, we can propose a model for decision making in a fuzzy environment as follows:

$$maximize\ \ G_1, G_2, \ldots, G_n such\ that\ C_1, C_2, \ldots, C_m \tag{2}$$

5.2 Decision Making in Web Services

Decision making in web services has been drawn an increasing attention in the web service community. For example, Yao et al [60] discuss flexible decision making strategies for web service negotiation. Decision making in web services refers to as a decision maker chooses one of the web services from the alternative recommended web services provided by a web service support agent. Although one may think of decision making in web services as only happening in one stage of the lifecycle such as recommendation stage, it is actually a basic activity in every stage of the web service lifecycle. Consequently, a chain of decision making is formed in web services. In what follows, we will not look into decision making in every stage of the mentioned web service lifecycle, but examine who are decision makers in web services, and what decision the decision makers should make.

As mentioned in Section 2.2, there are three main parties in web services: web service providers, requesters and brokers. Therefore, the decision makers in web services are mainly web service requesters, providers and brokers.

A web service requester makes decisions to acquire the most satisfactory web service with minimal cost throughout demand-driven web service lifecycle. For example, the service requester should make decision for web service use, which consists of web service identification, alternative generation, and execution [22].

The decision in web services made by the service provider is to provide the highest qualtity web services with maximal expected benefit throughout the demand-driven web service lifecycle. For example, the service provider should make decision for selecting the updated standards, technologies for representation, orchestration, publication, registry of web services.

The decision in web services made by the service broker is to provide the highest quality web services with maximal expected benefit from the service provider and service requester throughout the demand-driven web service lifecycle. For example, the service broker should make decision for web service personalization (retrieval, search and discovery), composition, negotiation and recommendation.

More generally, all the involved decision makers in web services should have useful information and services from web services and the database and knowledge base technologies to make decisions [22], they also share open decision modules on the Internet using standardized protocols such as HTTP, SOAP, and WSDL and so on. Various intelligent techniques such as soft computing and multiagent system technology are widely used to facilitate decision making in web services [16][46][58][62].

It should be noted that fuzzy decision making in web services is still an open problem, to our knowledge. In the future work, we will provide fuzzy models for decision making of service providers, requesters, and brokers respectively and also illustrate each of the models with examples.

6 Soft Computing for Web Services

This section will briefly review soft computing for web services and examine which stage in the above mentioned web service lifecycle requires soft computing.

The principal constituents of soft computing are fuzzy logic, neural networks, and generic algorithms [40][54]. Soft computing is also composed of rough sets, knowledge based systems, CBR and probabilistic reasoning [15][23]. Fuzzy logic is an intelligent technology that is primarily concerned with handling imprecision and uncertainty [40]. Neural networks focus on simulating human being's learning process, and genetic algorithms simulate the natural selection and evolutionary processes. Each component of soft computing is complementary to each other. Using combinations of several technologies such as fuzzy logic and neural networks will generally obtain better solutions. Soft computing is adequate to cope with real world problems and systems with fuzziness, imprecision and uncertainty [15][40].

Soft computing has been applied to web services, especially in matchmaking, discovery, brokering and composition of web services. For example, Fenza et al [11] discuss how to improve fuzzy service matchmaking through concept matching discovery. Wang et al [54] examine how to provide semantic web services with high QoS based on fuzzy logic, fuzzy neural networks with genetic algorithms. They apply a fuzzy neural network that is tuned by a genetic algorithm to evaluate QoS metrics. Ladner et al [23] use a case-based classifier for web services discovery which includes the application of rough set techniques in the feature selection component of the classifier. They also examine soft computing techniques for web service brokering by describing an integrated end-to-end brokering system that performs automated discovery, mediation and transformation of web service requests and responses.

7 WUDS: A Unified Decision Support System for Web Services

WUDS is a unified decision support system for web services which is being developed under the project supported by the Ministry of Education, Hebei China. It examines the decision support for decision making of service provider, requester and broker taking into account their corresponding demand-driven web service lifecycles (see Section 5). WUDS searches a variety of web services from the UDDI using the demand of the web service requesters. Based on the characteristics of the discovered web services, the WUDS uses the web services base (WSB) to support adaptation, composition, mediation, negotiation, and recommendation of web services. This section will examine the system architecture of the WUDS, as shown in Fig. 3. More specifically, it will look atagents within the WUDS. It also proposes a system model for the WS (web service) decision supporter. Finally, it discusses the workflowing of agents within the WUDS.

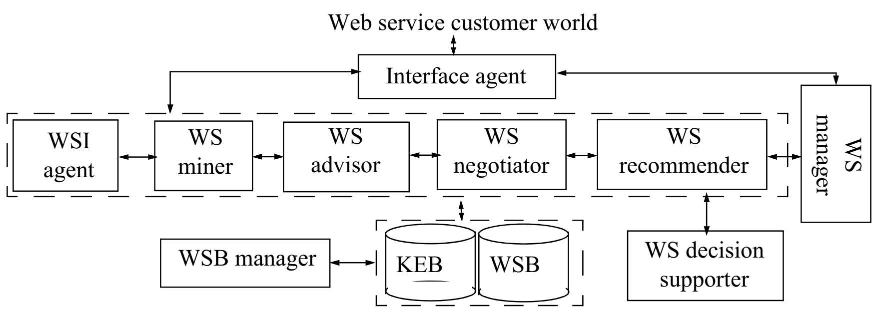


Fig. 3 WUDS: A unified decision support system architecture for web service

WUDS is a multilayer system architecture consisting of view layer, business logic layer, and database access layer [38]. The view layer consists of an intelligent interface agent. The business logic layer is mainly comprised of a WSI agent, web service service advisor, negotiator, and recommender. The database access layer mainly consists of a knowledge/experience base (KEB) and web service base (WSB). KEB stores the generated experience and knowledge discovered from web service users. WSB consists of information of web services.

7.1 Agents within WUDS

From the viewpoint of soft computing, a web service customer *i* creates uncertainty on three temporary levels: uncertainty about past experiences and behaviors, current behaviors and preferences, future needs in KEB. Brohman et al propose the following four strategies to resolve these uncertainties [4]:

1. Transaction strategy. Data in the KEB is analysed to identify profit concerns.
2. Data strategy. Data for the KEB is captured and used to service customers by providing access to a detailed record of past transactions.
3. Inference strategy. An inference strategy or decision making is used to infer future customer needs based on data in KEB and WSB.
4. Advice strategy. A recommendation strategy is used to provide accurate and reliable advice to customers based on their future needs.

Based on these strategies, the WUDS is comprised of nine intelligent agents: WSI agent, WS miner, WS advisor, WS negotiator, WS recommender, WS decision supporter, WS manager, WSB manager, and Interface agent. In what follows, we look at each of them in some detail:

- WSI agent is a web service information gathering agent. It is a mobile agent that proactively roams around the main search engines in the Internet such as Google and Yahoo. It interacts and collaborates with them in order to search and analyze the required web service information and then puts it in KEB [46].
- WS miner is an autonomous agent that discovers web services based on web service discovery algorithms, similar to knowledge discovery algorithms [2]. The discovered web service will be stored in WSB. The discovered web services are in a machine-readable form to facilitate the WS advisor, WS negotiator and WS recommender to make decisions. Fuzzy reasoning, CBR (see Section 8.3) and fuzzy inductive reasoning can be used in web service discovery.
- WS advisor is an autonomous agent that makes consultation of web services to WS requesters.
- WS recommender is a proactive agent that makes recommendation of web services based on the information or data available in WSB and KEB and soft computing technology. The proposed recommendation will be forwarded to the WS requester by the interface agent. There are many intelligent strategies for

recommendation. Because of uncertainty, incompleteness and inconsistency in customer experiences and web services, WS recommender has to use soft computing to make decisions for web services [45]. Case-based web service recommendation is one of them, which will be discussed in some detail in Section 8.3.

- WS negotiator is an autonomous [41], mobile and proactive agent that performs both integrative and distributive negotiation strategies during negotiation with the web service requester [46]. Because business negotiation is complicated in some cases, the intelligence of the WS negotiator lies in that it can change its negotiation strategies timely according to the changing resources or cases. It prepares a necessary compromise under bargaining. Thus, the WS negotiator may use all available inference methods such as CBR, fuzzy reasoning, soft computing [47] in different cases, if necessary. The WS negotiator sometimes works as service broker [46].
- WS manager is an intelligent agent that plays a leading role in the WUDS. Its main task is to decide which agent should do what and how to deal with a web service transaction.
- The interface agent is an intelligent agent consisting of the dynamic interactive exchange of information and service that occurs between the customer and the web services [37] (p. 141). It proactively interacts, cooperates with the web service requester and obtains the supply-demand information. At the same time, it obtains special information about the web service requesters and then stores it in KEB. The interface agent also interacts with the WS manager and transfers transaction messages to the web service requester.
- WSB manager is responsible for administering KEB and WSB. Its main tasks are creation and maintenance of KEB and WSB, web service evaluation, reuse, revision, and retention. The functions of the WSB manager are an extended form of those of a case base manager because case creation, retrieval, reuse, revision and retention are the main tasks of a CBR system [46][48].

It should be noted that soft computing has not yet been applied in every activity of the mentioned architecture. This implies that there is still a long way to go for soft computing to enable intelligent web services.

7.2 WS Decision Supporter

Decision making is an important part for the WUDS based on discovered web services, existing web services and data in KEB and WSB. The WS decision supporter is an intelligent agent within the above-mentioned WUDS.

The system architecture of the WS decision supporter consists of WS decision users (requesters, providers and brokers), *U,* which are either human users or intelligent agents, as shown in Fig. 4. The WS decision supporter, as a system, mainly consists of a user interface, a global web service base (GWB) and a multi-inference engine (MIE). The user interface consists of some kinds of natural language processing systems that

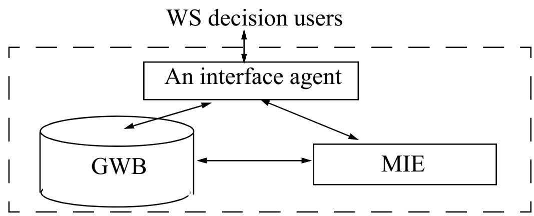


Fig. 4 A system architecture of the WS decision supporter [48]

allow the user to interact with the soft computing strategies [31] (p. 282). The GWB consists of all the web services that the system collects regularly and the new web services discovered when the system is running. GWB can be considered as a collection of KEB and WSB in the WUDS. The MIE consisting of the mechanism for implementing any inference paradigms based on soft computing for web services to manipulate the GWB to infer web services *X (X* denotes discovery, composition, recommendation, etc) requested by the user [48]. The remarkable difference between the mentioned WS decision supporter and the traditional knowledge-based systems lies in that the latter's inference engine is based on a unique reasoning paradigm (or inference rule), whereas the MIE is based on many different reasoning paradigms.

7.3 Agents Workflowing in WUDS

Now let us have a look at how the WUDS works. The web service customer, C, inquires the interface agent about certain web services. Then the interface agent asks *C* to *login* and fill in information about C's preferences for web services which will be stored in KEB. Then the interface agent forwards the information from *C* to the WS advisor. The WS advisor and recommender can use the data of WSB and certain recommendation strategies to recommend web services that are then forwarded to *C*. If *C* does not agree to the recommendation from the WS advisor and recommender, he likes to negotiate over the price of the provided services. In this case, WS negotiator uses a few negotiation strategies [46] to negotiate with *C* over the price or related items.

The negotiation should be helped by the WS decision supporter, because the WS negotiator might not know which negotiation strategies or reasoning paradigms that *C* has used at the negotiation. In this case, the WS decision supporter will first recognize the reasoning paradigms that *C* has used and then selected one of other reasoning paradigms to make decisions under the deceptional environment, because any negotiation usually hides some truths in order to get advantages in the interest of conflicts.

If *C* accepts one of the web services after recommendation and negotiation, then the WUDS completes this web service transaction. Otherwise, the WUDS will ask the WS composer to conduct web service composition and then obtain a composite web service based on the requirement and then forward it to C as a recommended web service. This will go to the recommendation and negotiation process of the WUDS.

Finally, if C accepts the recommended web service, the WSB manager will look at whether this web service is a new one. If yes, then the manager will add it to the WSB. Otherwise, it will keep some routine records to update all the related bases. If C does not accept the recommended e-service, the interface agent will ask C to adjust some attributes of her/his requirement, and then further forward the revised requirement to the related agents within the WUDS for further processing.

WSI agent collects the information in the web services world and saves the acquired information into WSB. WS miner discovers the WS models from WSB and KEB. These models will be used for web service mediation, recommendation and negotiation. The WS manager coordinates the activities of agents within the WUDS.

8 Case Based Web Services

This section will provide a CBR model for unifying the processes in web services such as service discovery, composition and recommendation.

8.1 Introduction

Case-based reasoning (CBR) is a reasoning paradigm based on previous experiences or cases; that is, a CBR system solves new problems by adapting solutions that were used to successfully solve old problems [23][42]. Ladner et al [23] use case-based classification for web service discovery by applying CBR to supervised classification tasks. Kwon [22] examines how to find the most similar web service case among cases using CBR. Limthanmaphon and Zhang [25] examine composition of e-services using CBR and present a model of web services composition. CBR has been successful in making recommendation of business activities such as in e-commerce to recommend different e-services with high QoS [54][46][49][45]. However, how to unify discovery, matching, composition and recommendation of web services remains open for CBR research. The following subsections will fill this gap by providing a unified treatment of web services based on CBR.

8.2 Web Services vs. CBR

A case in case base in the context of CBR is denoted as $c = (p, q)$, where p is the structured problem description and q is the solution description [42]. In web services, the service case base stores the collection of service cases [25]. A service case, $w = (d, s)$, consists of the service description d and its service solution (or functions) s as well as other information including functionally dependency among web services [22]. The service description corresponds to the requirement of the service user, while the service solution corresponds to the answer to the requirement. In this way, a web service case in web services corresponds to a case in a case base [42].

When service definitions change or new providers and services are registered within the web services platform such as the WUDS, the services need to be adaptive to the change in the environment with minimal user intervention, in order to manage and even take advantage of the frequent changes in the service environment [9]. In other words, web service adaptation is necessary for web services. In fact, case retrieval (search), reuse, revise (adaptation) and retention constitutes the basic activities of a CBR cycle [40][42][46]. Web service retrieval (search), reuse, adaptation, and retention in web services can then correspond to the activities of CBR. Therefore, at a general level, CBR can be used for processing web service retrieval, reuse, adaptation, and retention, implying that CBR is naturally applicable to web services. This is the reason why CBR has been successfully applied to web service discovery (including search and matching) [23][25]. It is significant to apply CBR to the special activities of web services such as web service discovery, composition and recommendation.

8.3 A Unified Treatment of Case Based Web Services

This subsection provides a unified treatment for case based web services, and examines case-based web service search, matching, retrieval, discovery, adaptation, composition and recommendation in a context of fuzzy case based web service reasoner (FCWSR), which is used by the WS miner, WS recommender, WS negotiator and WSB manager within the WUDS.

The web service user's demand is normalized into a structured service description p'. Then the FCWSR uses its similarity metric mechanism to retrieve its service case base, which consists of service cases, each of which is denoted as $c = (p, q)$, where p is the structured service description and q is the service solution description. The inference engine of the FCWSR performs similarity-based reasoning that can be formalized as [12][42]:

$$\frac{P',\ P' \sim P,\ P \rightarrow Q,\ Q \sim Q'}{\therefore\ \ Q'} \tag{3}$$

where P, P', Q, and Q' represent fuzzy compound propositions, P' ~ P (Q' ~ Q) means that P and P' (Q and Q') are similar in terms of fuzz logic.

From a fuzzy CBR viewpoint, the service case retrieval process from web service search and matching is used to discover the following service cases from the web service case base in the FCWSR [42][46]:

$$C(p') = \{c_i | c_i = (p_i, q_i), p \sim p'\} = \{c_1, c_2, \ldots, c_n\} \tag{4}$$

This is the result of *case based web service discovery*, where n is a positive integer, c_i, $i = 1, 2, \ldots, n$ are all service cases with their demand description p similar to the current demand description p'. Usually, $C(p') = \{c_1, c_2, \ldots, c_n\}$ satisfies the following property: for any integer i, $1 \le i < n$ and $c_i = (p_i, q_i)$,

$$s(p_i, p') \ge s(p_{i+1}, p') \tag{5}$$

where $s(\cdot)$ is a similarity metric, which measures the similarity between one service demand and another.

If n is small, then the FCWSR will directly recommend the web service solutions of $c_1, c_2, \ldots, c_n$, $q_1, q_2, \ldots, q_n$, to the WS requester through the interface agent. If n is very large, the FCWSR has to recommend the web service descriptions of the first m cases of $c_1, c_2, \ldots, c_n$; that is, $q_1, q_2, \ldots, q_m$, to the requester, in order to meet the demand of the WS requester, where $1 < m < n$. This process can be called *case-based web service recommendation.*

After obtaining the recommended web services from the FCWSR, the WS requester will evaluate them and then select one of the following:

1. Accept one of the recommended web services, ^{q}k, and contract it, where $1 < k < m$.
2. Adjust her/his demand descriptions p' and then send them to the FCWSR.
3. Reject the recommended e-services and leave the FCWSR.

It is obvious that only the first two among these three choices require further discussion. For the first choice, the deal was successfully done and the FCWSR routinely updates the successful service case $c_k = (p_k, q_k)$ in the WSB. At the same time, the FCWSR has reused the service case successfully; that is, FCWSR completes the process of *case-based web service use* and *reuse.* For the second choice, the demand adjustment is the process of demand adaptation that corresponds to problem adaptation. After having adjusted the demand, the requester then submits it to the FCWSR. The FCWSR will conduct web service retrieval, recommendation and reuse again. Therefore, the web service demand submission, retrieval, recommendation, and adaptation constitute a cycle.

Further, if the web service adaptation is unsuccessful, the FCWSR has to conduct *case based web service composition.* Assume that the web service requester's demand is normalized into a structured service description and service solution description $c = (p', q')$, and the FCWSR has discovered m web services $c_1, c_2, \ldots, c_m$ (where m is the least positive number) such that

$$p' \subseteq p_1 \cup p_2 \cup \ldots \cup p_m \text{ and } q' \subseteq q_1 \cup q_2 \cup \ldots \cup q_m \tag{6}$$

where $\cup$ is the union operation of the set theory. This is a necessary condition for case based web service composition. Based on Eq.(6), the composite web service case $c = (p, q)$ is obtained through *case based web service composition* of the FCWSR:

$$p = p_1 \oplus p_2 \oplus \ldots \oplus p_{\mathrm{m}} \text{ and } q = q_1 \otimes q_2 \otimes \ldots \otimes q_{\mathrm{m}} \tag{7}$$

where $\oplus$ and $\otimes$ are composition operations for web services. For example, when they are replaced by the ordinary (or fuzzy) union operation of set theory, the composite web service is the same as that discussed in DIANE [21] or similar to the composite web service in [22]. When they are replaced by the "independence" operation taking into account interindependent relationships among the services, the composite service is similar to that discussed in [25]. However, it is still a big issue for case based web service composition to use a more sophisticated composition operation to obtain a composite service case.

After obtaining a composite service case, the FCWSR will recommend it to the service requester for acceptance. This goes to the early mentioned process for acceptance, adaptation or rejection.

So far, we have provided a unified treatment of case-based web service retrieval, discovery, adaptation, reuse, composition and recommendation. We will not illustrate the above models with examples any more owing to space limitation.

9 Conclusions and Future Work

The chapter examined decision making in web services based on soft computing. More specifically, it proposed a unified architecture, SESS, and an intelligent system architecture, WUDS. The SESS unifies e-services, web services and infrastructure services into a hierarchical framework. The WUDS is a system model for implementing multi-agent web services such as discovery, composition and recommendation. Both architectures tie together methodologies, techniques, and applications into a unified framework that includes both logical and intelligent embodiment of decision making in web services. The chapter also proposed a demand-driven model for web service lifecycle taking into account decision making. The chapter finally explored a unified treatment for case-based discovery, composition and recommendation of web services. The proposed approach will facilitate the development of web services, intelligent systems and soft computing.

There have been a variety of techniques and approaches developed for web service discovery. For example, OWL-S (of W3C) provides classes that describe what the service does, how to ask for the service, what happens when the service is carried out, and how the service can be accessed [23]. With the development of web service technology, web service discovery will be easier.

Web service composition is at its early stage for research and development. However, web service composition will become an important topic for research and development in the near future, because composing web services to meet the requirement of the service requester is the most important issue for web service providers and brokers.

Web service recommendation is a significant challenge for web service industry, in particular for web service brokers. The next generation of web services is web service composition and recommendation. Intelligent techniques such as soft computing, multiagent systems and CBR will play an important role in these aspects, as they have done in e-commerce and e-business [46].

Applying intelligent techniques to decision making in web services is still a new topic for soft computing, AI and web services. However, intelligent decision making in web services will significantly alter web services lifecycle and represent major competition for parties involved in web services. In future work, we will develop a system prototype based on the proposed WUDS and soft computing. We will also integrate web service discovery, composition and recommendation using the FCWSR based on soft case based reasoning.

Acknowledgements. This research is supported by the Ministry of Education Hebei, China under Grant No. ZH200815. The authors would like to thank Ms Yanxia Wang and Mr. Zhengke Kang from Hebei Normal University, China for providing many useful materials and discussions which contributed in the elaboration of this chapter.

References

[1] Alonso, G., Casati, F., Kuno, H., Machiraju, V.: Web Services: Concepts, Architectures and Applications. Springer, Heidelberg (2004)
[2] Becerra-Fernandez, I., Gonzalez, A., Sabherwal, R.: Knowledge Management: Challenges, Solutions and Technologies. Prentice-Hall, Upper Saddle River (2004)
[3] Bellman, R.E., Zadeh, L.A.: Decision-Making in a Fuzzy Environment. Management Science 17(4), B-141–B-164 (1970)
[4] Brohman, M.K., Watson, R.T., Piccoli, G., Parasuraman, A.: Data completeness: A key to effective net-based customer service systems. Comm. of the ACM 46(6), 47–51 (2003)
[5] Bucchiarone, A., Gnesi, S.: A Survey on Services Composition Languages and Models. In: Proceedings of the International Workshop on Web Services - Modeling and Testing (Web service-MaTe 2006), Palermo, Sicily, June 9, pp. 51–63 (2006)
[6] Casati, F., Shan, E., Dayal, U., Shan, M.-C.: Business-oriented management of web services. Comm. ofthe ACM 46(10), 55–59 (2003)
[7] Chua, A.: Knowledge management systems architecture: A bridge between KM consultants and technologies. Intl. JInformation Management 24, 87–98 (2004)
[8] Chung, J.Y., Lin, K.J., Mathieu, R.G.: web services computing: Advancing software interoperability. Computer 36(10), 35–37 (2003)
[9] Dustdar, S., Schreiner, W.: A survey on web services composition. Intl. J. Web and Grid Services 1(1), 1–30 (2005)
[10] Erl, T.: Service-Oriented Architecture (SOA): Concepts, Technology and Design. Prentice Hall, Englewood Cliffs (2006)
[11] Fenza, G., Loia, V., Senatore, S.: Improving fuzzy service matchmaking through concept matching discovery. In: Proceedings FUZZ-IEEE, pp. 1–6 (2007)
[12] Finnie, G., Sun, Z.: A logical foundation for the CBR cycle. Intl. J. Intell. Syst. 18(4), 367–382 (2003)
[13] Ferris, C., Farrell, J.: What are web services. Comm. of the ACM 46(6), 33–35 (2003)
[14] Guruge, A.: Web Services: Theory and Practice. Elsevier Inc., Amsterdam (2004)

[15] Groumpos, P.P., et al.: Heterogeneous process and modelling, http://www.sim-serv.com/wg_docAVG7_Heterogenous.pdf (accessed on 05 01 09)

[16] Henderson-Sellers, B., Giorgini, P. (eds.): Agent-Oriented Methodologies. Idea Group Publishing, Hershey (2005)

[17] Hung, P.C.K., Li, H., Jeng, J.: Web service negotiation: An overview of research issues. In: Proc. of 37th Hawaii Intl. Conf. on System Sciences, pp. 1–10 (2004)

[18] Huhns, M.N., Singh, M.P. (eds.): Readings in Agents. Morgen Kaufmann Publishers, Inc., San Francisco (1998)

[19] Hoffman, K.D.: Marketing + MIS = E-Services. Comm. of the ACM 46(6), 53–55 (2003)

[20] Kreger, H.: Fulfilling the web services promise. Comm. of the ACM 46(6), 29–34 (2005)

[21] Kuster, U., Konig-Ries, B., Stern, M., Klein, M.: DIANE: An integrated approach to automated service discovery, matchmaking and composition. In: Intl. Conf. on World Wide Web (WWW 2007), Alberta, Canada, May 8-12, pp. 1033–1041 (2007)

[22] Kwon, O.B.: Meta web service: building web-based open decision support system based on web services. Expert Systems with Applications 24, 375–389 (2003)

[23] Ladner, R., et al.: Soft computing techniques for web service brokering. Soft Computing 12, 1089–1098 (2008)

[24] Leymann, F.: Web Services: Distributed Applications without Limits: An Outline. In: Proceedings Database Systems For Business, Technology and Web BTW, Leipzig, Germany, Feburary 26-28, Springer, Heidelberg (2003), http://doesen0.informatik.uni-leipzig.de/proceedings/pa-per/keynote-leymann.pdf

[25] Limthanmaphon, B., Zhang, Y.: Web service composition with case-based reasoning. In: ACM Intl. Conf. Proc. Series, Proc. 14th Australasian Database Conf., Adelaide, Australia, pp. 201–208 (2003)

[26] Lu, J., Ruan, D., Zhang, G.: E-service intelligence: An introduction. In: Lu, J., et al. (eds.) E-ServiceIntelligence: Methodologies, Technologies and Applications, pp. 1–33. Springer, Berlin (2007)

[27] Mike, P.: Papazoglou Service -Oriented Computing: Concepts, Characteristics and Directions. In: Proc. of 4th Intl. Conf. on Web Information Systems Engineering (WIS E2003), December 10-12, pp. 3–12 (2003)

[28] Negnevitsky, M.: Artificial Intelligence: A Guide to Intelligent Systems, 2nd edn. Addison-Wesley, Harlow (2002)

[29] Miller, G.: .NET vs. J2EE. Comm. of the ACM 48(7), 64–67 (2005)

[30] Narendra, N.C., Orriens, B.: Requirements-driven modeling of the web service execution and adaptation lifecycle. In: Madria, S.K., Claypool, K.T., Kannan, R., Uppuluri, P., Gore, M.M. (eds.) ICDCIT 2006. LNCS, vol. 4317, pp. 314–324. Springer, Heidelberg (2006)

[31] Nilsson, N.J.: Artificial Intelligence. A New Synthesis. Morgan Kaufmann Publishers Inc., San Francisco (1998)

[32] Papazoglou, M.P., Georgakopoulos, D.: Service-oriented computing. Comm. of the ACM 46(10), 25–28 (2003)

[33] Petrie, C., Genesereth, M., et al.: Adding AI to web services. In: van Elst, L., Dignum, V., Abecker, A. (eds.) AMKM 2003. LNCS (LNAI), vol. 2926, pp. 322–338. Springer, Heidelberg (2004)

[34] Pfleeger, S.L., Atlee, J.: Software Engineering: Theory and Practice, 3rd edn. Pearson Education, Inc., London (2006)
[35] Pressman, R.S.: Software Engineering: A Practitioner's Approach, 5th edn. McGraw-Hill, Boston (2001)
[36] Rust, R.T., Kannan, P.K.: E-service: A new paradigm for business in the electronic environment. Comm. of the ACM 46(6), 37–42 (2003)
[37] Schmitt, B.H.: Customer Experience Management: A revolutionary approach to connecting with your customers. John Wiley & Sons, Hoboken (2003)
[38] Satzinger, J.W., Jackson, R.B., Burd, S.D.: Systems Analysis and Design in a Changing World, 3rd edn. Thompson Learning, Boston (2004)
[39] Sheth, A.: Semantic web process lifecycle: Role of Semantics in Annotation, Discovery, Composition and Orchestration, Invited Talk. In: WWW 2003 Workshop on E-Services and the Semantic Web, Budapest, Hungary, May 20 (2003)
[40] Shiu, S.C.K., Pal, S.K.: Case-based reasoning: Concepts, features and soft computing. Applied Intelligence 21, 233–238 (2004)
[41] Singh, M.P., Huhns, M.N.: Service-oriented computing: Semantics, Processes, and Agents. John Wiley & Sons, Ltd., Chichester (2005)
[42] Sun, Z., Finnie, G.: A unified logical model for CBR-based e-commerce systems. Intl. J. Intelligent Syst. 20(1), 29–46 (2005)
[43] Song, H.: E-services at Fed Ex. Comm. of the ACM 46(6), 45–46 (2003)
[44] Sun Microsystem, Web services life cycle: Managing enterprise Web services, White Paper (October 2003), http://www.sun.com/software(accessed on December 24, 2008)
[45] Sun, Z., Lau, S.: Customer experience management in e-services. In: Lu, J., Ruan, D., Zhang, G. (eds.) E-Service Intelligence: Methodologies, Technologies and applications, pp. 365–388. Springer, Berlin (2007)
[46] Sun, Z., Finnie, G.: Intelligent Techniques in E-Commerce: A Case-based Reasoning Perspective. Springer, Heidelberg (2004)
[47] Sun, Z., Finnie, G.: Experience based reasoning for recognising fraud and deception. In: Proc. Intl. Conf. on Hybrid Intelligent Systems (HIS 2004), Kitakyushu, Japan, December 6-8, pp. 80–85. IEEE Press, Los Alamitos (2004)
[48] Sun, Z., Finnie, G.: MEBRS: A multiagent architecture for an experience based reasoning system. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3681, pp. 972–978. Springer, Heidelberg (2005)
[49] Sun, Z., Finnie, G.: A unified logical model for CBR-based e-commerce systems. Intl. J. Intelligent Syst. 20(1), 26–29 (2005)
[50] Tsalgatidou, A., Pilioura, T.: An Overview of Standards and Related Technology in web Services. Distributed and Parallel Databases 12, 135–162 (2002)
[51] Tang, X.F., Jiang, C.J., Ding, Z.J., Wang, C.: A Petri net-based semantic web services automatic composition method. Journal of Software 18(12), 2991–3000 (2007) (in Chinese)
[52] Turban, E., Volonino, L.: Information Technology for Management: Improving Performance in the Digital Economy, 7th edn. John Wiley & Sons, Inc., Chichester (2009)
[53] Vissers, C.A., Lankhorst, M.M., Slagter, R.J.: Reference models for advanced e-services. In: Mendes, M.J., Suomi, R., Passons, C. (eds.) Digital Communities in A Networked Society, pp. 369–393. Kluwer Academic Publishers, Dordrecht (2004)
[54] Wang, H., Zhang, Y.Q., Sunderraman, R.: Extensible soft semantic web services agent. Soft Compting 10, 1021–1029 (2006)

[55] Wang, M., Cheung, W.K., Liu, J., Xie, X., Lou, Z.: E-service/process composition through multi-agent constraint management. In: Dustdar, S., Fiadeiro, J.L., Sheth, A.P. (eds.) BPM 2006. LNCS, vol. 4102, pp. 274–289. Springer, Heidelberg (2006)

[56] Wang, M., Wang, H., Xu, D., Wan, K.K., Vogel, D.: A web service agent-based decision support system for securities exception management. Expert Systems with Applications 27(3), 439–450 (2004)

[57] Wang, M., Liu, J., Wang, H., Cheung, W.K., Xie, X.: On-demand e-supply chain integration: A multi-agent constraint-based approach. Expert Systems with Applications 34(4), 2683–2692 (2008)

[58] Weiss, G. (ed.): Multiagent Systems: A modern approach to Distributed Artificial Intelligence. MIT Press, Cambridge (1999)

[59] W3C, Web Service Management: Service life cycle (2004), `http://www.w3.org/TR/2004/NOTE-webservicelc-20040211/` (accessed on December 24, 2008)

[60] Yao, Y., Yang, F., Su, S.: Flexible decision making in web services negotiation. In: Euzenat, J., Domingue, J. (eds.) AIMSA 2006. LNCS (LNAI), vol. 4183, pp. 108–117. Springer, Heidelberg (2006)

[61] Zhang, L.J., Jeckle, M.: The next big thing: Web services composition. In: Jeckle, M. (LJ) Zhang, L.-J. (eds.) ICWS-Europe 2003. LNCS, vol. 2853, pp. 1–10. Springer, Heidelberg (2003)

[62] Zimmermann, H.J.: Fuzzy Set Theory and its Application. Kluwer Academic Publishers, Dordrecht (1991)

[63] `http://www.google.com.au/search?hl=en&lr=&oi=defmore&defl=en&q=define:web+Services`

[64] `http://www-128.ibm.com/developerworks/library/specification/webservice-bpel/`

Dynamic Price Forecasting in Simultaneous Online Art Auctions

Mayukh Dass[1], Wolfgang Jank[2], and Galit Shmueli[2]

[1] Rawls College of Business, Texas Tech University
[2] Robert H. Smith School of Business, University of Maryland

Abstract. In recent years, the simultaneous online auction (SOA) has become a popular mechanism for selling heterogeneous items such as antiques, art, furniture, and collectibles. These auctions sell multiple items concurrently to a selected group of bidders who often participate in multiple auctions simultaneously. Such bidder behavior creates a unique competitive environment where bidders compete against each other both within the same auction as well as across different auctions. In this chapter, we present a novel dynamic forecasting approach for predicting price in ongoing SOAs. Our proposed model generates a price forecast from the time of prediction until auction close. It updates its forecasts in real-time as the auction progresses based on newly arriving information, price dynamics and competition intensity. Applying this method to a dataset of contemporary Indian art SOAs, we find high predictive accuracy of the dynamic model in comparison to more traditional approaches. We further investigate the source of the predictive power of our model and find that price dynamics capture bidder competition within and across auctions. The importance of this finding is both conceptual and practical: price dynamics are simple to compute at high accuracy, as they require information only from the focal auction and are therefore a parsimonious representation of different forms of within-auction and between-auction competition.

1 Introduction

With the growing popularity of online auctions and the increasing number of items sold through them, price prediction has become a vital research topic in recent years. In contrast to earlier studies (Ghani and Simmons 2004; Gneezy 2005) who rely solely on “static” characteristics that are known at the start of the auction (e.g., the opening price, product characteristics, and seller reputation), more recent approaches attempt to capture the dynamic aspects of the auction process along with the more traditional static components. To that end, Wang et al. (2008) consider the price velocity (or rate of change in price), and Jap and Naik (2008) and Bajari and Hortascu (2003) incorporate the underlying bid distribution into the forecasting process. A common aspect across all of these studies is that they focus on individual auctions where auctions for a particular item are held independently of each other and bidders typically bid only on one auction at a time.

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 417–445.
springerlink.com

Recently, there has been increasing interest in an alternative auction format, commonly known as the Simultaneous Online Auction (SOA), which has become very popular for selling high-priced complementary items such as fine art and collectibles. These auctions are held by specialized auction houses dedicated to selling only one type of item (e.g., SaffronArt.com sells only Indian contemporary art, Attinghouse.com sells only Chinese art and Southeast Asian art) and with a price tag ranging from a few thousands to a few millions of dollars. With such high stakes, forecasting prices in SOAs is crucial to both auction house managers and bidders. A method that can provide price projections can help auction house managers make real-time decisions and take actions and adjustments during the ongoing auction such as inviting additional bidders or running promotions to attract more bidding activity. Typically, online art auctions are held for three days and, as a consequence, auction house managers have an opportunity to actively intervene in the auction process by running promotions or by calling additional bidders who may have an interest in the art and who are likely to influence the auction process based on their behavior in previous auctions (Dass et al. 2007). Individual bidders can also benefit from real-time price forecasts. For example, auctioned items can be dynamically ranked based on their forecasted price, from lowest to highest surplus for the bidder (Ghani and Simmons 2004; Wang et al. 2008). Such rankings can help bidders to focus on items that are within their budget and/or maximize their expected surplus.

In this chapter, we present a dynamic forecasting model based on Functional Data Analysis to forecast the price of an ongoing auction. Prior studies (Wang et al. 2008) have provided some evidence that price dynamics in online auctions matter, and that capturing dynamics leads to improved real-time forecasting. By price dynamics we mean the speed at which the price changes throughout the auction (price velocity), and perhaps even the rate at which this speed changes (price acceleration). Our current study investigates two questions: The first question addresses the role of price dynamics in forecasting SOA prices and how it differs from the individual-auction case. The second question investigates why the incorporation of price dynamics results in superior prediction, and in particular, we examine the role of bidder competition and its relation to price dynamics. This is done in the context of simultaneous online fine art auctions, where two types of bidder competition are salient: competition within a single auction and competition across different auctions.

Price forecasting in online auctions and particularly in SOAs is challenging due to their dynamic environment. One aspect of this environment is the changing bid density, where the number of bids per unit time changes dramatically throughout the auction (Roth and Ockenfels 2002; Russo et al. 2008). The resulting unequally-spaced time-series of bids make traditional forecasting models such as ARIMA or its variants such as ARCH, GARCH and so on (which assume evenly spaced measurements) inadequate. Furthermore, price dynamics across different auctions follow different paths. In other words, the speed at which the price travels during the auction and the rate at which this speed changes varies across auctions. Therefore, traditional forecasting models, which do not account for such instantaneous change, fail to accurately forecast auction prices. To incorporate the dynamic nature into a forecasting model, we take a functional data modeling approach.

Functional Data Analysis (FDA) is an emerging statistical methodology that operates on functional observations such as the price curves in online auctions. While FDA has received a lot of enthusiasm within the statistics literature, it is only slowly entering the marketing and information systems literature. Only recently, Sood, James and Tellis (Sood et al. 2007) have proposed an FDA-based model as an alternative to the Bass model for predicting market penetration of new products. Foutz and Jank (Foutz and Jank 2007) use an FDA-based method to early and dynamically forecast box-office success by analyzing the trading shapes from online virtual stock markets. Although FDA is less frequently mentioned as a soft computing method, it fits well with the definition of Soft Computing as it is tolerant to imprecision, uncertainty, partial truth, and approximation[1]. Most computational intelligence techniques mimic some aspect of the human mind. For example, genetic algorithms mimic evolutionary theory; neural networks mimic the neural system of a human body, and fuzzy logic mimics fuzzy approximate reasoning of the human mind. FDA belongs to the same family, as it mimics the continuity processing performed by the human mind. Although many events are discrete, the human mind has the ability to combine them into a continuous scheme (e.g., seeing motion in cartoons that are essentially a set of discrete pictures, or hearing music from a set of discrete tones). FDA attempts to represent such continuity, obtained from discrete events, and it supplies tools for studying the continuous realms and investigates its characteristics and determining factors.

In online auctions, FDA has been shown to be useful as a graphic tool for advanced data visualization of electronic commerce data (Jank et al. 2008) and as a mechanism to capture price dynamics in online auctions (Bapna et al. 2008; Jank and Shmueli 2006; Reddy and Dass 2006). In this chapter, we employ FDA to capture the dynamic components of an SOA and to build a real-time forecasting model for price. We follow the approach in Wang et al. (2008) and incorporate price dynamics in addition to other available information into the forecasting model. The underlying idea is to represent the price path during an auction as a continuous curve that describes the price formation process. Then, following functional principles, we “recover” (i.e., estimate) the price curves of individual auctions using smoothing techniques (Ramsay and Silverman 2005). From the price curves, we can then obtain estimates of the price dynamics such as price velocity and price acceleration via first and second derivatives of the price curves, respectively. The price curves and price dynamics are subsequently incorporated into the forecasting model to produce the real-time price forecast.

In this study, we develop two major Dynamic Forecasting Models, DFM-I and DFM-II, to forecast price in an SOA. The first model (DFM-I), incorporates both the price path and the price velocity information until the time of prediction, whereas the second model (DFM-II) considers only the price path until the time of prediction. Both models also include static pre-auction information, which does not change throughout the auction (e.g., the opening price or the item characteristics), but neither directly incorporates bidder competition information. We later

[1] `http://www.soft-computing.de/def.html`

supplement these two models with more direct measures of bidder competition to create DFM-III and DFM-IV, in order to study the relationship between price dynamics and competition. We investigate the predictive performance of all models to uncover the source of price dynamics. We compare the dynamic models with two additional models: one that forecass the final price based only on static pre-auction information (STATIC) and another that is a simple dynamic model (DFM-0) as it includes only price at the time of prediction in addition to static information. Comparing the mean absolute percentage error (MAPE) of all models on a holdout set, we find that DFM-I outperforms all competing models in terms of predictive accuracy. Thus, our first conclusion is that, like in individual-auction forecasting, price dynamics, and in particular the price velocity, has a major impact on forecasting price in SOAs. We also conclude that the dynamic forecasting model DFM-I is sufficiently flexible and powerful to capture very different types of price dynamics, and that it can be used in a wide range of auction formats.

Our second goal is to investigate possible sources of the predictive power of price dynamics. Prior studies (Ariely and Simonson 2003; Heyman et al. 2004; Ku et al. 2005) suggest that bidder emotions play a significant role in the formation of auction dynamics. Such emotions result from rivalry (or competition) among bidders to acquire the item (Ariely and Simonson 2003) and thus affect auction dynamics. Therefore, we expand our study to analyze the relationship between bidder competition and auction dynamics. In SOAs, bidders compete both within an auction as well as across multiple simultaneous auctions. This results in two types of bidder-competition, namely, within-auction competition and between-auction competition. Using new metrics for measuring within- and between-auction bidder competition we examine the performance of the resulting forecasters (DFM-III and DFM-IV) in the presence of directly observed bidder competition information. Compared to the advantage of DFM-I over DFM-II, we find that the difference in predictive performance between DFM-III and DFM-IV vanishes, suggesting that dynamics essentially proxy for competition. That is, both DFM-III and DFM-IV predict price equally well compared to DFM-I, which implies that a forecaster with direct bidder competition information is equivalent to its counterpart using only a proxy based on price dynamics. The practical implication of this finding is that price dynamics can provide a simple and parsimonious measure for the competitive nature of online marketplaces. It is simple because it only requires the information from within the focal auction; it is parsimonious because it summarizes many different forms of competition between auction participants in one single measure.

We further examine this relationship using another formulation, where we condition the forecasting model on the level of competition. This is done by splitting auctions into 4 segments based on different levels of within-auction and between-auction competition. Once again, we observe that the effect of dynamics vanishes after controlling for bidder competition, suggesting that dynamics effectively capture bidder competition in SOAs.

The rest of the chapter is organized as follows. First, we describe the mechanism of simultaneous online art auctions and in particular that on SaffronArt.com. We also describe and explore our available data. Second, we derive and estimate the dynamic forecasting models and discuss the results. Third, we define and incorporate bidder competition information into the forecasting models and study their relationship with price dynamics. We conclude with answers to our two research questions as well as managerial implications and future directions.

2 Simultaneous Online Auctions

SOAs are different from the auctions held on popular auction sites such as eBay, both in terms of the auction design and the types of items that they sell. SOAs sell multiple objects simultaneously in a first-price ascending auction format. This means that auctions start and end at the same time for all items, and that the highest bidder wins the auction and pays the amount that s/he bid. Since many items are highly complementary, bidders are typically interested in purchasing more than one item at a time. As a result, bidders frequently compete against each other, not only within the same auction (i.e., for the same item), but also across other auctions that sell complementary items, which leads to unique bidding dynamics (Rothkopf 1977). In contrast, items on eBay are sold in a variant of the second-price[2] sealed-bid auction (Krishna 2002) and are held independently of each other. On eBay bidders are rarely observed to consciously compete against each other across different auctions that take place simultaneously, as it is unlikely that bidders will have similar product demand. Even if they do, it is highly implausible that they will compete for the same item, as products auctioned on eBay typically have multiple listings. Moreover, eBay now masks bidder identities thereby eliminating the ability of bidders to identify competitors across auctions. Another difference between eBay auctions and SOAs is the type of closing rule. While eBay has mostly fixed hard-closing times, SOAs tend to have soft-closing times where the auction closing time automatically extends after a late bid. A soft-close auction format not only encourages bidders to bid early (Roth and Ockenfels 2002), but also discourages sniping[3] in the last moments. Finally, SOAs are organized by only one seller, i.e. the auction house, whereas eBay provides an auction platform for many different sellers.

3 Data Used in This Study

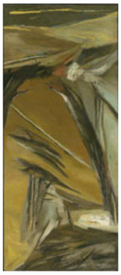
Like other online auction houses, SaffronArt.com posts detailed bid histories of items for sale on their website during periods when auctions are in progress. For this study, we collected the bid histories of auctions held on SaffronArt.com in December 2005. The auctions lasted three days and 196 art items were sold. A snapshot of a bid history is shown in Figure 1. The bid history includes information on each submitted bid, its time and amount, and the bidder's ID. Apart from the bidding activity information, each bid history also includes information about the item: name of the artist, physical characteristics of the item (size and media),

---

[2] In a second-price auction the highest bidder wins the item and pays the second highest bid.

[3] Sniping is a strategic bidding activity where bids are submitted in the last moments of the auction to allow minimal time to other bidders to react to this bid. Such behavior is prominent in eBay auctions as the auction closes promptly at a specific time.

The following are the bids that have been placed on this Lot : COUNTDOWN : 1 days 22:29:05



❖36
Ram Kumar

Untitled

Signed in Devnagari and dated in English (lower right)
1969
Oil on canvas
70 x 30 in (177.8 x 76.2 cm)

$100,000 - 150,000
Rs 4,400,000 - 6,600,000

Next Valid Bid : $ 157,500 (Rs 6,930,000)

Rank	Nick Name	Amount($)	Amount(Rs)	Date & Time (US EST)
Current Highest Bid 1	Anonymous 118	147,500	6,490,000	Dec 6 2005 10:59:25 AM
2	Poker	137,500	6,050,000	Dec 6 2005 4:20:08 AM
3	Kyozaan	127,500	5,610,000	Dec 6 2005 2:07:14 AM
4	Poker	117,500	5,170,000	Dec 6 2005 1:25:20 AM
5	Kyozaan	107,500	4,730,000	Dec 6 2005 1:25:16 AM
6	Anonymous 3	100,000	4,400,000	Dec 6 2005 1:25:04 AM
7	Kyozaan	95,000	4,180,000	Dec 6 2005 1:25:04 AM
8	Anonymous 3	87,500	3,850,000	Dec 5 2005 10:30:00 PM
	Start price	80,000	3,520,000	Dec 5 2005 10:30:00 PM

Fig. 1 Snapshot of a Bid History

Painting	Description	Bidding
 View Bigger Image add to my auction gallery Send to a friend	3 M.F. Husain (b.1915) Untitled Signed in English (lower left) Circa 1980's Oil on canvas 53.5 x 81 in (135.9 x 205.7 cm) **$232,600 - 290,700** Rs 10,000,000 - 12,500,000 Click here to read more	Comparables

http://www.saffronart.com - Painting Description - Mozilla Firefox

Shortly following the formation of the Progressive Artists' Group, its founding members M.F. Husain, and F.N. Souza visited an exhibition in New Delhi in 1948, which was to have a profound effect on their work; allowing them to imbibe India's Classical aesthetic tradition. The show was to affect the manner in which Husain presented the human form and its movement within the confines of two-dimensional pictorial space. This trip was a "turning point" in his career. "It was at this juncture that he conceived the essential form that is pivotal to his work.

He states, 'One reason why I went back to the Gupta period of sculpture was to study the human form – when the British ruled we were taught to draw a figure with the proportions from Greek and Roman sculpture...That was what I thought was wrong...In the east the human form is an entirely different structure... the way a woman walks in the village there are three breaks...from the feet, the hips, the shoulder...they move in rhythm...the walk of a European is erect and archaic.' " (Yashodhara Dalmia, The Making of Modern Indian Art: The Progressives, OUP, 2001, p. 102)

[Close]

Comparable prices

Artist Information

Fig. 2 Snapshot of Item Information

pre-auction price estimates, the item's expected value based on analysis by auction house art experts, and provenance of the item. The auction house also provides information about results of previously auctioned comparable items by the same artist. A snapshot of the item listing is provided in Figure 2. Since the items are of high value, the auction house tries to provide as much information about the items as possible in order to help bidders make informed bidding decisions. Additional information about the auction format, general bidding rules, and the closing schedules is also provided by the auction house.

Our data set includes sales of 196 art items (lots) from 70 different artists. In these auctions 256 bidders participated, posting 3042 bids. The average number of bids per lot is 15 and the average number of bidders participating in an auction is 6. The average value realized for all 196 items is $56,233, ranging between $3,135 and $1,486,100. Other descriptive statistics of the data are shown in Table 1.

Table 1 Summary Data Description

	Mean (SD)	**Median**	**Min.**	**Max.**
No. of Unique Bidders/ Lot	6.38 (2.47)	6	2	14
No. of Unique Lots Bid / Bidder	4.89 (7.76)	3	1	63
No. of Bids/lot	15.52 (7.49)	15	2	48
Opening Bid in $	$19,145 ($36,830)	$6,400	$650	$300,000
Pre-Auction Low Estimates of the Lots	$23,880 (45,954)	$8,000	$795	$375,000
Pre-Auction High Estimates of the Lots	$30,816 (60,676)	$10,230	$1,025	$475,000
Realized Value of the Lots in USD($)	$61,845 (134,109)	$21,400	$3,135	$1,486,100
Realized Sq. Inch Price of the Lots in USD($)/ Sq. Inch	$109.39 (227.13)	$45.06	$1.40	$1,865.42

4 Bidder Competition in Simultaneous Online Auctions

Bidder competition in art SOAs is different from eBay auctions as bidders compete against each other not only within an auction, but also across auctions. Therefore, bidder competition in simultaneous online auctions can be defined in two ways. The first type of competition is the rivalry-intensity level between two specific bidders for the same item, and thus is termed within-auction competition. The second type of bidder competition is the level of rivalry between bidders across different auctions, and is thus termed between-auction competition (Dass et al. 2007).

Operationally, we compute within-auction competition between two bidders as the maximum number of sequential pairs of bids between two bidders. For every auction, we first determine the unique pairs of bidders *j, k* participating in the auction. Then for each of these bidder pairs, we count the number of times the two bidders bid sequentially n_{jk} (e.g., A→B→A). The maximum number of sequential bid pairs denotes the within-auction competition. Therefore, within-auction competition (wa) for auction i is given by

$$wa_i = \max\left(n_{jk}\right) \text{ for } j = 1\cdots B_i - 1, \text{ and } k = j + 1\cdots B_i \tag{1}$$

where B_i denotes the number of bidders in auction *i*, n_{jk} the number of sequential bids between bidder *j* and bidder *k*. Consider the example shown in Figure 1: There are four bidders participating in the auction. Therefore, we have $\binom{4}{2}$=6 unique bidder pairs. For each of the 6 pairs, we compute the number of times two bidders bid sequentially and compute the maximum value of all pairs. In the case of Figure 1, the within-auction competition equals 3. We use the maximum value as our measure because heated rivalry between a specific pair of bidders can induce higher bidder dynamics in the entire auction (Ariely and Simonson 2003; Heyman et al. 2004)[4].

In contrast to within-auction competition, between-auction competition measures the competitive reach of a bidder pair across several auctions. Like for the previous measure, we first determine the number of unique bidder pairs. Then, for all bidder pairs, we count the number of auctions in which the pair is competing simultaneously. The between-auction competition for a certain auction is the average of this number across all pairs. Therefore, between-auction competition (ba) for auction i is given by

$$ba_i = \sum_{j=1}^{B_i-1} \sum_{k=j+1}^{B_i} cl_{jk} \bigg/ N_i \tag{2}$$

where B_i denotes the number of bidders in auction *i*, cl_{jk} the number of common auctions bid by bidders *j* and *k*, N_i the number of bidder pairs in auction *i*.

For example, consider auction #36 in Figure 3. There are 4 bidders participating in the auction leading to $\binom{4}{2}$=6 unique pairs of bidders. Considering only the

[4] We also analyzed within-auction competition as the average value across all the bidder pairs. Results from that analysis are similar to those obtained using the maximum value.

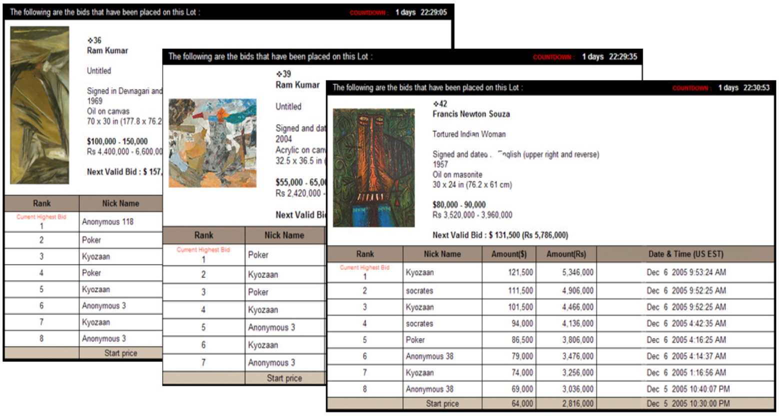


Fig. 3 Between-Auction Competition

three auctions displayed in that Figure (#36, #39 and #43), we find that the total number of auctions bid simultaneously by the bidder pairs Anonymous118-Anonymous3, Anonymous11-Poker, Anonymous118-Kyozaan, Poker-Kyozaan, Poker-Anonymous3 and Kyozaan-Anonymous3 are 1,1,1,2,1, and 1 respectively. Therefore, the between-auction competition for auction #36 is 1.167 (=7/6).

Earlier studies on online auction competition have considered bidder rivalry as a component that influences bidding dynamics during the auction (Ariely and Simonson 2003; Heyman et al. 2004). They showed that such rivalry increases bidders' quasi-endowment feeling and escalates their commitment towards the item, thereby leading to a phenomenon called "auction fever.[5]" Since these studies focus on eBay and eBay-like individual online auctions, they do not consider between-auction rivalry. Our chapter extends this literature by considering bidder competition which goes beyond the rivalry within a single individual auction and looks in addition at competition across simultaneous auctions.

5 Dynamic Price Forecasting

Prior research on price forecasting in online auctions is limited and has mostly focused on predicting the final price of items using static or pre-auction information. For example, Ghani and Simmons (2004) use data-mining techniques to predict the final price in eBay auctions using only information available at the outset of the auction such as the opening price, product characteristics, and seller reputation. Their model therefore does not account for new information arriving during the ongoing auction. Bajari and Hortascu (2003) recover the bid distribution using

[5] Auction fever is an emotional phenomenon where bidders become irrational in their bidding decision and bid higher than what they would normally pay for the item.

a structural modeling technique, but they too only predict the final price. And finally, Gneezy (2005) uses step-level models of reasoning to predict the auction outcome, but like others, does not account for the dynamics during the auction. Only two recent studies dynamically forecast price in online auctions. Jap and Naik (2008) develop a method to estimate dynamic bidding models in online corporate procurement reverse auctions. Wang et al. (2008) (referred to as WJS from hereon) build a dynamic forecasting model using FDA. In both studies, the models are designed for individual auctions such as those on eBay. Our forecasting model builds upon the WJS approach and adapts it to the SOA setting.

5.1 Model Formulation

Following WJS, our model consists of an initial step of recovering (or estimating) for each auction the underlying price curve and the corresponding price dynamics from the observed bid histories. Since bids arrive at unevenly spaced time intervals, we need the flexible FDA approach to approximate a continuous underlying price curve and the rate of change in price, or price-velocity, which is estimated via the first derivative of the price curve[6]. Recovering the price curve is done using monotone smoothing splines (Ramsay and Silverman 2005; Simonoff 1996) in order to guarantee price curves that are continuous and monotonically non-decreasing. See Appendix A for further details on the curve recovery step.

The smoothed price curves and their first derivatives (i.e., price velocity) for our 196 auctions are shown in Figure 4. The average price and price velocity plots show that the price formation in a typical auction is fast at the beginning and near the end of the auction. Also note that the average price velocity (i.e., the rate of change in price) nearly doubles towards the end of the auction compared to the beginning of the auction.

After creating smooth price curves and their dynamics, we use these components as the basis for our dynamic forecasting model. Our model contains 3 conceptually different pieces of information: static pre-auction and time-varying information, the price path, and the price dynamics information. We later supplement it with a fourth piece, which is bidder competition. The dynamic forecasting model of price at time t $(y(t))$ is given by Wang et al. (2008) as:

$$y(t) = \alpha + \sum_{i=1}^{Q} \beta_i x_i(t) + \sum_{j=1}^{J} \gamma_j D^{(j)} y(t) + \sum_{l=1}^{L} \eta_l y(t-l) + \varepsilon(t) \tag{3}$$

where $x_1(t), \ldots, x_Q(t)$ is a set of static (pre-auction) and time-varying predictors, $D^{(j)} y(t)$ denotes the j^{th} derivative of price at time t, and $y(t-l)$ is the l^{th} price

[6] In general, we can estimate further price dynamics by taking higher order derivatives. For instance, the second derivative estimates price acceleration. See Shmueli & Jank (2008) for further details.

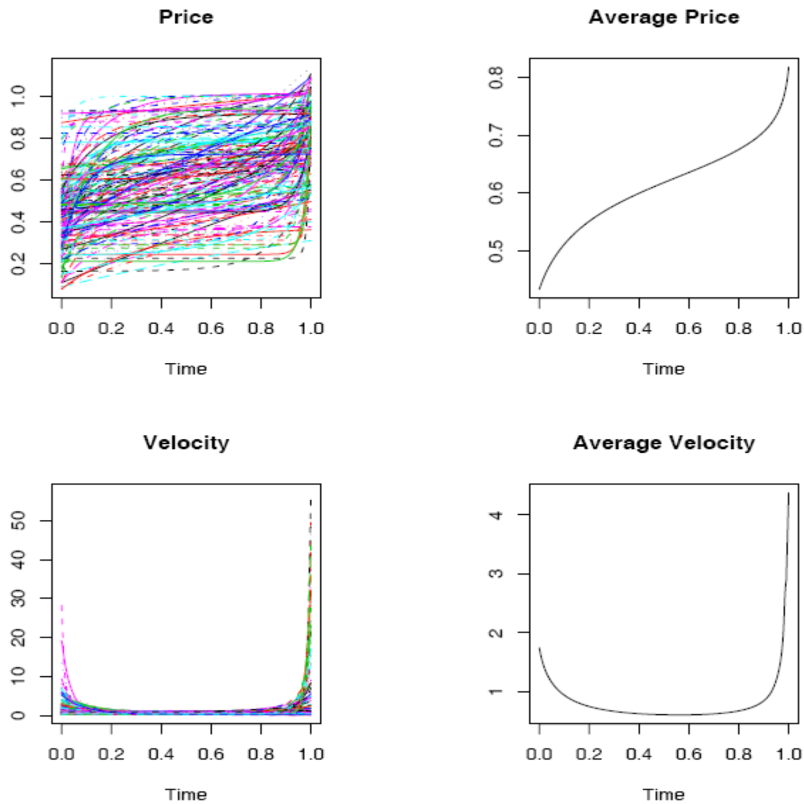


Fig. 4 Price Dynamics for 196 Lots Sold in the Online Art Auction

lag. In our case the static predictors, which do not change over the course of the auction, include the opening bid, item characteristics (size of the item and type of art work), and artist characteristics (artist type, average price per sq. inch of the artist's sold items in the previous year's auction); see also category 1a in Table 2. Time-varying predictors, which do change as the auction progresses, include the number of bids (see category 1b in Table 2). Note that although WSJ did not directly include bidder competition information in their model, the model in (3) is flexible enough to incorporate any type of static, time-varying, or dynamic information. A brief description of all model-components is given in Table 2. We perform two sets of analysis with the above predictors. In the first set, our model follows WJS closely and does not use the competition covariates; in the second set, we introduce the two competition variables into the DFM-I, DFM-II models to create DFM-III and DFM-IV.

Table 2 Predictors Used in the Model

Category	Covariates	Description
Static Predictors		
1a	Opening Bid	Opening bid is the first bid in the auction.
	Size of the Item	It is the dimension of the artwork in square area.
	Type of Artwork	Artworks can be categorized into works on paper and works in canvas. We used an indicator variable in our model to indicate whether the item is a canvas work or not.
	Artist Reputation	The artists are categorized into established artists and emerging artists.
	Previous Auction History	The price/sq. inch of the artworks of the artists in the previous years
Time-varying Predictors		
1b	Current number of Bids	This is a time-varying predictor indicating the current number of bids placed in the auction.
Current Price		
2	Current and previous price (price path)	Price lags at time t, t-1…
Price Dynamics		
3	Price Velocity	First derivative of price at time t
Competition		
4	Within-auction competition	This indicates the current level of within-auction competition in the auction.
	Between-auction competition	This indicates the current level of between-auction competition in the auction.

Using equation (3), the resulting h-step ahead forecast, given information until time T, is given by

$$\tilde{y}(T+h\,|\,T)=\hat{\alpha}+\sum_{i=i}^{Q}\hat{\beta}_i x_i(T+h\,|\,T)+\sum_{j=1}^{J}\hat{\gamma}_j\bar{D}^{(j)}y(T+h\,|\,T)+\sum_{l=1}^{L}\hat{\eta}_l\tilde{y}(T+h-1\,|\,T) \quad (4)$$

As explained in WSJ, equation (3) faces two challenges that need to be addressed. First, the price dynamic components $D^{(j)}y(t)$ are coincident indicators, and therefore must be forecasted prior to their use in equation (4). The solution in WSJ is to forecast price dynamics using a polynomial-trended linear regression model with static and time-varying predictors and autoregressive (AR) residuals. It is of the form:

$$D^{(j)}y(t)=\sum_{k=0}^{K}a_k t^k+\sum_{i=1}^{P}b_i x_i(t)+u(t) \quad (5)$$

where $t=1,2,...T$ and u(t) follows an AR model of order R.

Once this model is estimated from a training set, it can be used to forecast price dynamics of a new ongoing auction. See Appendix B for further details.

The second challenge with the forecasting model in equation (3) is that the static predictors do not change during the auction, i.e. they are independent of time *t*, and therefore their estimated coefficients are confounded with the price function. The solution in WSJ is to transform the static variables into time varying predictors by considering each static variable's impact on the price evolution. This is done by fitting a functional regression model of price on each of the static predictors and then using their resulting time-varying estimated coefficient as a time-varying predictor in equation (3). See Wang et al. (2008) for further details.

With the above general approach, we build two Dynamic Forecasting Models (DFMs). DFM-I consists of the static, time-varying, current price and dynamic components 1-3 in Table 2 and is therefore equivalent to eq. (4). The two price components that we include are the price curve and its first derivative (i.e. price velocity). For the purpose of investigating the specific role of price dynamics, we also consider DFM-II, which uses the same information as above, except for the price velocity (i.e. only components 1 & 2 from Table 2). Comparing DFM-I and DFM-II allows us to assess the importance of price dynamics in the SOA price prediction process.

5.2 Benchmark Models

In order to benchmark the performance of our dynamic models DFM-I and DFM-II, we consider a competing static model (STATIC) that includes only pre-auction information (i.e. only component 1a from Table 2) via a linear regression model on price (e.g. Lucking-Reiley 1999), and a simple dynamic model (DFM-0) that includes the price at the time of prediction in addition to the static information (i.e. components 1a & 2 from Table 2).

5.3 Model Estimation and Evaluation

In order to test and compare the performance of the different models, we randomly partition our data into a training set (70% or 137 auctions) and a holdout set (30% or 59 auctions), where the training set is used to estimate the model, and the holdout set is used to measure predictive accuracy. For the DFM-I and DFM-II models, the training set is used for fitting price curves, for estimating the dynamics prediction model (equation 1.5), and for estimating the final forecasting model. The STATIC and DFM-0 models are estimated using the same training set. Note that the training set consists of auctions that are fully observed between the start and end; in contrast, auctions in the validation set are only partially observed, i.e. information is only available until time T, the time at which a forecast is desired. T is flexible and can be set by the user.

Since our art auctions are 3-day long, our intention is to forecast the price of an ongoing auction early enough so that the auction house managers can take action and potentially intervene, and that bidders can decide which items to concentrate on in each auction. Therefore, we forecast the price during the last T=18 hours prior to the closing of the auction.

6 Results

6.1 Estimated Models

The estimated coefficients for the STATIC and the DFM-0 models are given in Table 3. As the primary goal of this chapter is predictive, our main emphasis is on

Table 3 Parameter Estimates of the Models in the Training Set

Covariates	STATIC (Std. Error)	DFM-0 (Std. Error)
Opening Bid	0.1894* (0.0230)	0.0810* (0.0249)
Previous Auction History of the artist	-0.2127* (0.0248)	-0.1239* (0.0258)
Size	-0.2018* (0.0254)	-0.1199* (0.0259)
Current Price		0.4836* (0.0693)

* Significant at 0.01 level

the forecasting performance of the different models, rather than on inference. From Table 3, we see that three static predictors are significant both in the STATIC and DFM-0 models (at the 1% level): opening bid, previous auction history of the artist, and size of the art item. The effect of these static predictors on the final price has already been shown in prior research. In particular, the positive effect of the opening bid reflects the direct relation between an item's value and the choice of the starting price. This is in accordance with findings from prior studies (Bajari and Hortacsu 2003; Czujack et al. 1996). The negative effect of an artist's previous year's values on this year's price could imply that bidders are looking for "bargains" (i.e. artists that had low values in the previous year) and are willing to bid rather aggressively for them. Finally, the negative effect of the size of the art works on price is in accordance with the findings of Czujack et al. (1996). We also find the current price to have a positive and significant impact in the DFM-0 model.

In the dynamic models we have time-varying coefficients and static variables that were transformed into time varying predictors via functional regression weighting. This impact of each of the predictors can now be assessed at different times of the auction. Figure 5 displays the time-varying coefficient-curves (together with associated confidence bands). The results show that size has a significant negative effect only at the end of the auction. This indicates that smaller art objects are more expensive than the bigger ones in accordance with the findings of Czujack et al. (1996) and that bidders take this into consideration more consciously towards the auction-end. We also observe that the current number of bidders has an initial significant positive effect, suggesting that high initial bidder participation is a strong signal for price. Previous auction history is found to have an initial negative effect, but becomes positive as the auction progresses. This indicates that while bidders may be shopping for "bargains" early on, as the auction comes to a close they trust more artists that have previously done well. Finally, we find that opening

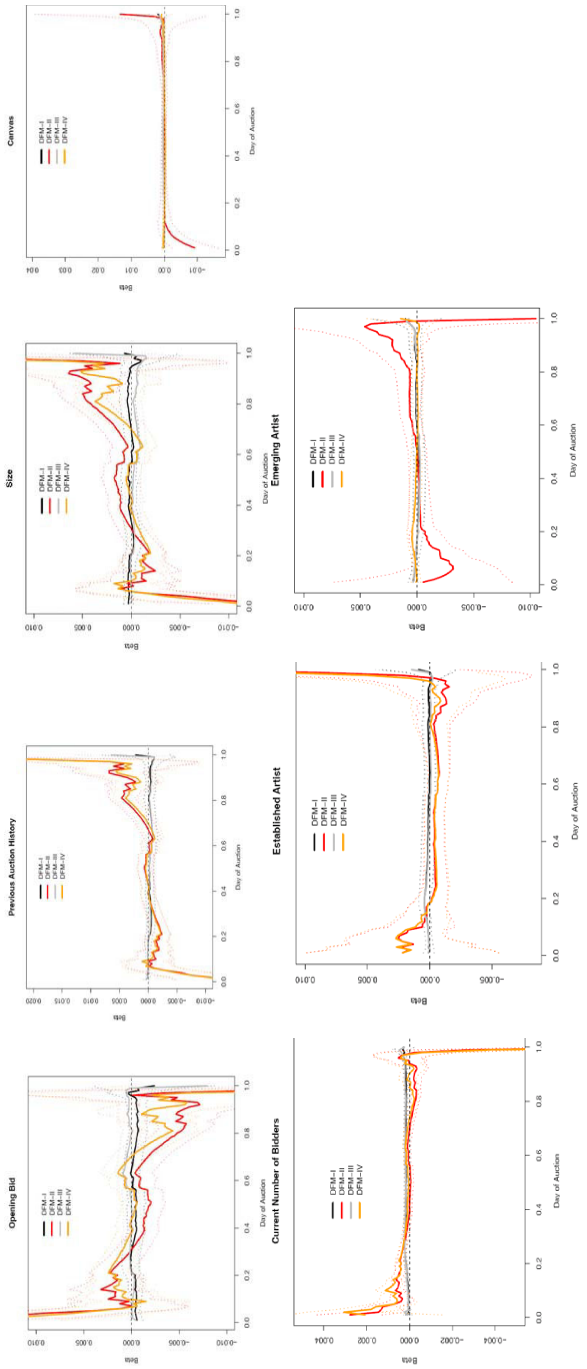


Fig. 5a Parameter Estimates of Time-varying Covariates in the Model with Training Set

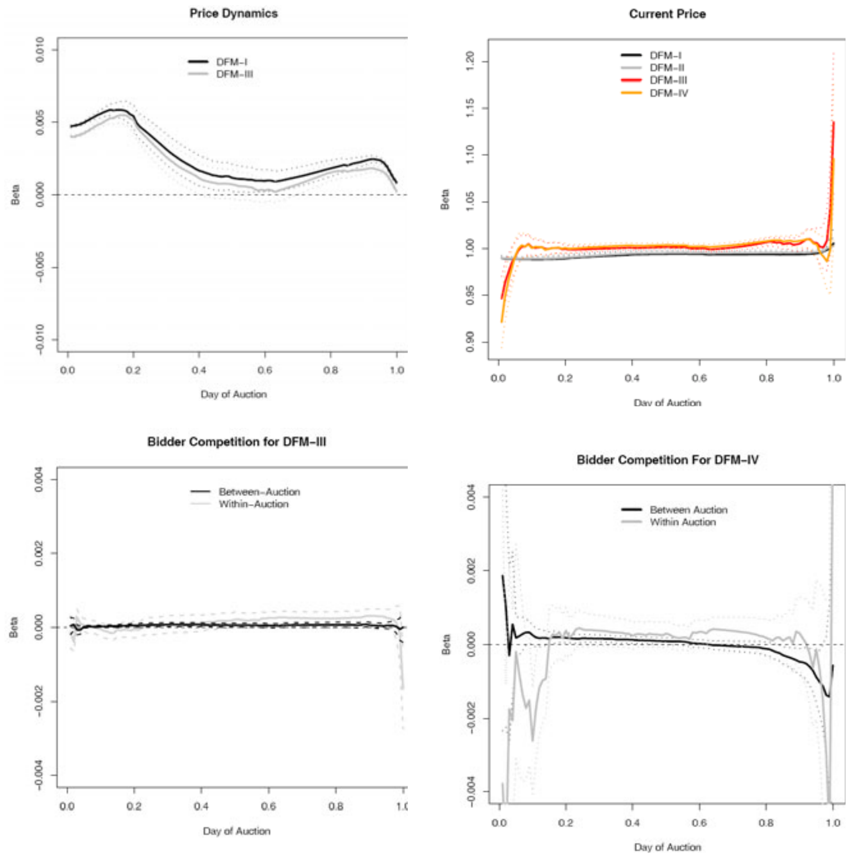


Fig. 5b Parameter Estimates of Time-varying Covariates in the Model with Training Set

bid has an initial positive effect that later switches to a negative effect at the auction end. This finding is similar to that of Reddy and Dass (2006) and suggests that bidders initially draw information from the opening bid, but then gradually discount this information as more signals come in from competing bidders.

In what follows, we discuss the impact of the other variables in the dynamic forecasting model. Recall that the model for price velocity is a linear regression model with quadratic trend (K=2) and three static predictors. Figure 6 illustrates the accuracy of the price velocity prediction for four randomly selected auctions (#12, #26, #39 and #54). For each of these auctions we see that the true and forecasted price curves are close for all but auction #39, where the velocity is under-predicted.

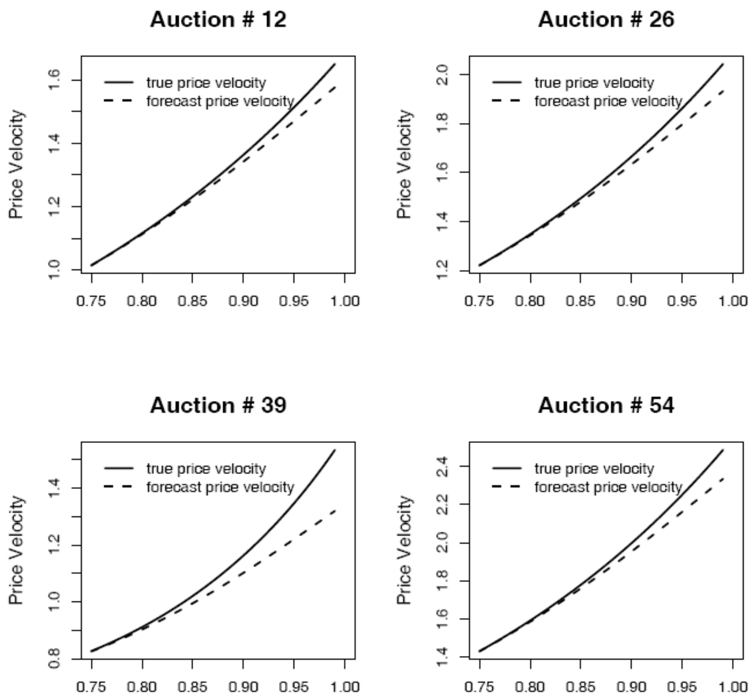


Fig 6 Performance of Forecasting Price Dynamics of the Last 18 Hours for Four Sample Lots using DFM-I

6.2 Forecasting Performance

Since our main goal for modeling is generating accurate price forecasts, we evaluate and compare the different models in terms of their predictive accuracy. To that end we use the mean absolute percentage error (MAPE) computed on the holdout set. MAPE is computed here as the difference between the forecasted curve and the true curve. The MAPE values for all four models (DFM-0, DFM-I, DFM-II, STATIC) are shown in Figure 7. Recall that we observe the first 75% (or 54 hours) of the 3-day auction, and forecast the last 25% (or 18 hours). We see that DFM-I and DFM-II outperform DFM-0 and STATIC, but that DFM-I (which includes price velocity) is by far the best of all four forecasters. Figure 8 shows price forecasts for the same four auctions as in Figure 6, using only DFM-I. We see that while the velocity (Figure 6) is forecasted very accurately for auction #12, its price is not which suggests that for #12, our model does not perform as well it does for other items. More importantly, we can see that DFM-I accomplishes dynamic and real-time forecasts, and “customizes” its forecast for each individual auction, depending on that auction’s dynamic environment.

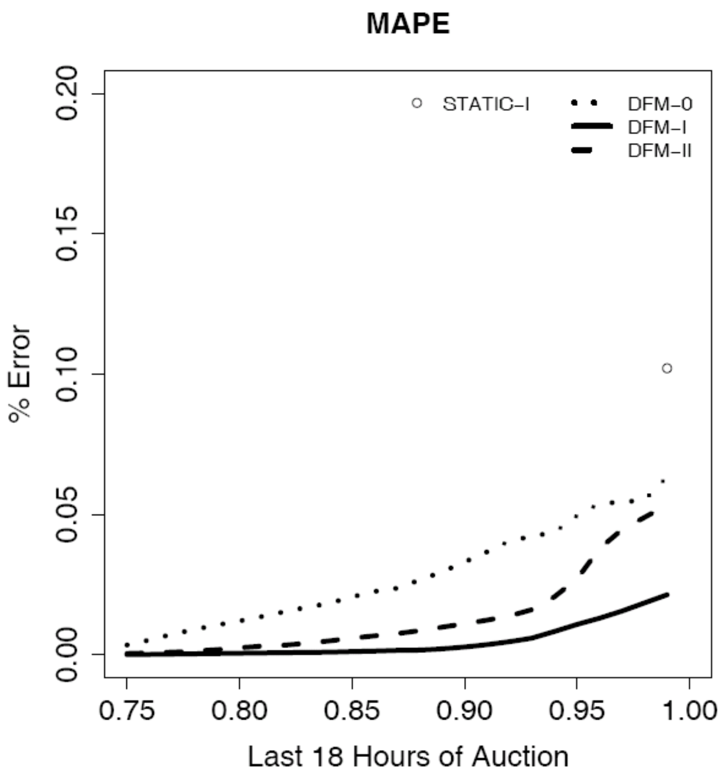


Fig. 7 Mean Absolute Percentage Errors (MAPE) of Models without Bidder Competition

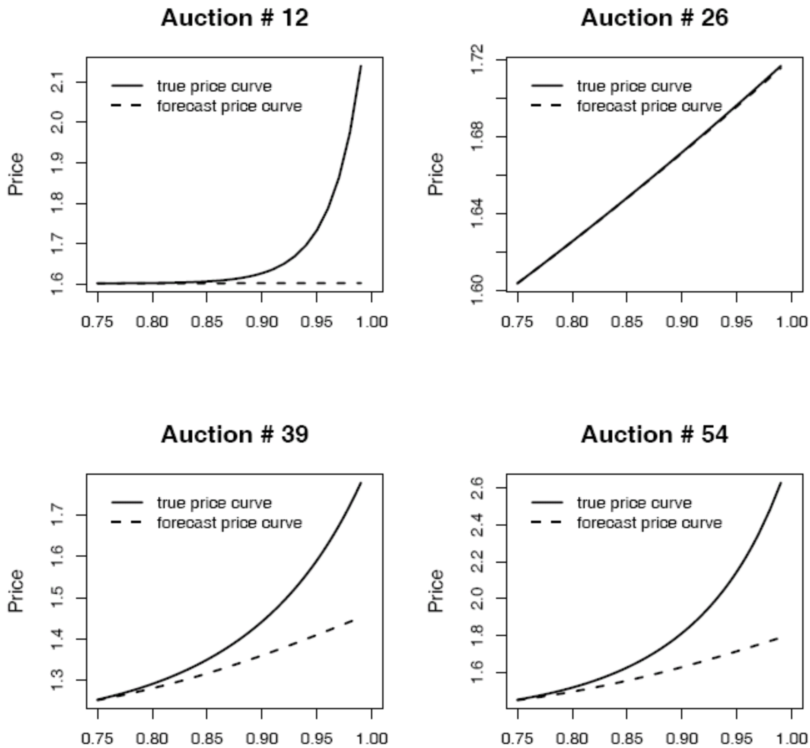


Fig. 8 Dynamic Forecasting of the Last 18 Hours for Four Sample Auctions using DFM-I

7 Bidder Competition and Price Forecasting

The previous results show that dynamics matter and that they lead to superior forecasts. An important yet unanswered question is why dynamics lead to better predictive performance. To study this question and in particular the relationship between price dynamics and bidder competition, we incorporate direct measures of bidder competition into the dynamic models DFM-I and DFM-II. This results in two new models DFM-III and DFM-IV, respectively[7]. Under the hypothesis that bidder competition is one of the main forces behind price dynamics, we expect that the inclusion of direct bidder competition information will mitigate the effect of price dynamics. In other words, DFM-III will lose its advantage over DFM-IV. We incorporate bidder competition into the forecasting model in two different ways: One approach is to incorporate the within- and between-auction competition measures directly into the models as additional time-varying predictors. The

[7] Correlation (Within-auction, Between-auction) = 0.1909, Correlation (Within-auction, Price Dynamics) = -0.1081 and Correlation (Between-auction, Price Dynamics) = -0.0380.

second approach is to segment the auctions by competition levels and then estimate models separately within each segment. The results from both approaches are consistent, indicating that price dynamics proxy for bidder competition. We describe the results for each of these approaches next.

7.1 Bidder Competition as Time-Varying Predictors

We repeat the estimation process described earlier and now include the two bidder competition measures (category 4 in Table 2) in the dynamic models DFM-I and DFM-II to create DFM-III and DFM-IV. Qualitatively, the resulting estimated coefficients are all very similar except for one important difference: comparing DFM-I and DFM-III in Figure 5b (top left panel), we see that dynamics, which are highly significant in DFM-I, become less significant in the presence of the direct competition predictors (DFM-III). In other words, price dynamics appear to carry the same information about price as bidder competition, i.e. they act as a proxy. Moreover, examining the estimated coefficients of bidder competition (bottom panels of Figure 5b), we find that the inclusion of dynamics (DFM-III) reduces the size of the competition-effect by a factor of almost 10, which is another indicator that both components carry similar information.

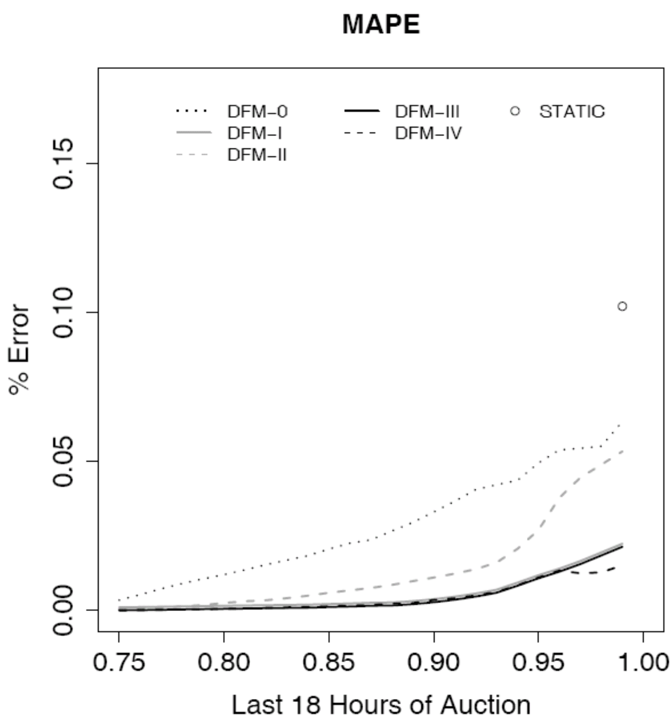


Fig. 9 Mean Absolute Percentage Errors (MAPE) of Models with (DFM-III, DFM-IV) and without (DFM-I, DFM-II)

Figure 9 (similar to Figure 7) compares the predictive accuracy for all models (now including DFM-III and DFM-IV). We see that the initial advantage of DFM-I over DFM-II due to price dynamics (Figure 7) vanishes when including direct measures for competition (i.e. DFM-III vs. DFM-IV). In other words, the inclusion of competition mitigates the impact of price dynamics. Further, we find that both DFM-I, DFM-III and DFM-IV perform equally well. However, note that DFM-I is conceptually simpler and more parsimonious compared DFM-III and DFM-IV: it is conceptually simpler because it only requires information on the focal auction (i.e., an estimate of the price dynamics); in contrast, the other two models require information on the focal auction (i.e. within-auction bidder competition) as well as on all other simultaneous auctions (i.e. between-auction bidder competition). While conceptually simpler, it is also operationally easier to compute measures only from within the focal auction, compared to monitoring and measuring all other, simultaneous auctions. Another advantage of the use of dynamics is parsimony. In order to capture price dynamics, we only need one additional predictor. In contrast, bidder competition requires two predictors, and might even require further predictors in other types of auctions, such as eBay, which sells a much wider variety of items, over a much longer period of time.

7.2 Bidder Competition as a Conditioning Variable

To evaluate the relation between bidder competition and price dynamics from another angle, we extend our investigation by splitting the 196 auctions into 4 competitive segments based on their level of within-auction and between-auction competition. To that end, we split our dataset into four parts based on low-high[8] values of the within-auction and between-auction competition and then estimate the forecasting models separately within each segment. Summary statistics for each of the segments are given in Figure 10. The average price-path and price-dynamics (Figure 10) support prior empirical findings that suggest that high between-auction competition has a negative effect on the auction outcome as they result in a slower price growth than other types of auctions (Dass et al. 2007)[9]. In contrast, the average price curve for low between-auction competition steadily increases throughout the auction with a dramatic increase near auction-end.

In each of the four segments we randomly partition the auctions into a training set (70%) and a validation set (30%) and estimate the models DFM-0, DFM-I, DFM-II and STATIC as described earlier. The predictive accuracy for each of the four segments (Figure 11) shows that, as in the combined dataset (Figure 7), DFM-I and II outperform DFM-0 and STATIC. But unlike the results for the combined dataset, for each segment the advantage of DFM-I over DFM-II vanishes. This further supports our above conclusion that, when explicitly controlling for competition, the predictive power of price dynamics are mitigated.

[8] To classify each item as low-high level of bidder competition, we first create the scatter plot of the two types of competition and took the natural separation line for the measures. The separation value for within-auction competition is 2.94 and for between-auction competition is 6.

[9] Dass et. al attributed this phenomenon to tacit collusion among bidders.

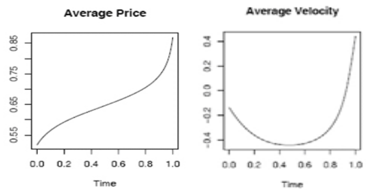
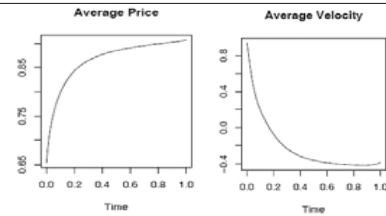
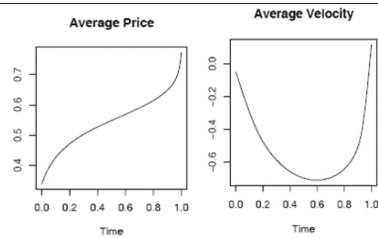
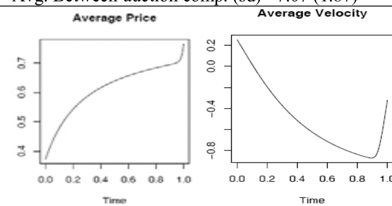
Within-auction dyadic bidder competition mean (sd) =2.96(1.44)	**Between-auction dyadic bidder competition** mean (sd) = 3.18 (2.63)	
	Low	**High**
Low	 Sample Size = 62 Avg. Within-auction comp. (sd) = 1.76(0.43) Avg. Between-auction comp. (sd) = 2.07(0.93)	 Sample Size = 25 Avg. Within-auction comp. (sd)= 1.69 (0.48) Avg. Between-auction comp. (sd)= 7.07 (1.87)
High	 Sample Size = 67 Avg. Within-auction comp. (sd)= 3.94 (1.2) Avg. Between-auction comp. (sd)= 2.27 (0.9)	 Sample Size = 42 Avg. Within-auction comp.(sd)= 3.86 (1.29) Avg. Between-auction comp.(sd)= 8.25 (3.46)

Fig. 10 Average Price and Velocity Curves by Bidder Competition

In summary, we control for competition in two different ways: by including competition predictors directly into the forecasting model and by segmenting auctions based on their competitive landscape. In both cases, we observe that the power of price dynamics is mitigated in the presence of competition information. We therefore conclude that price dynamics effectively proxy for competition in art SOAs.

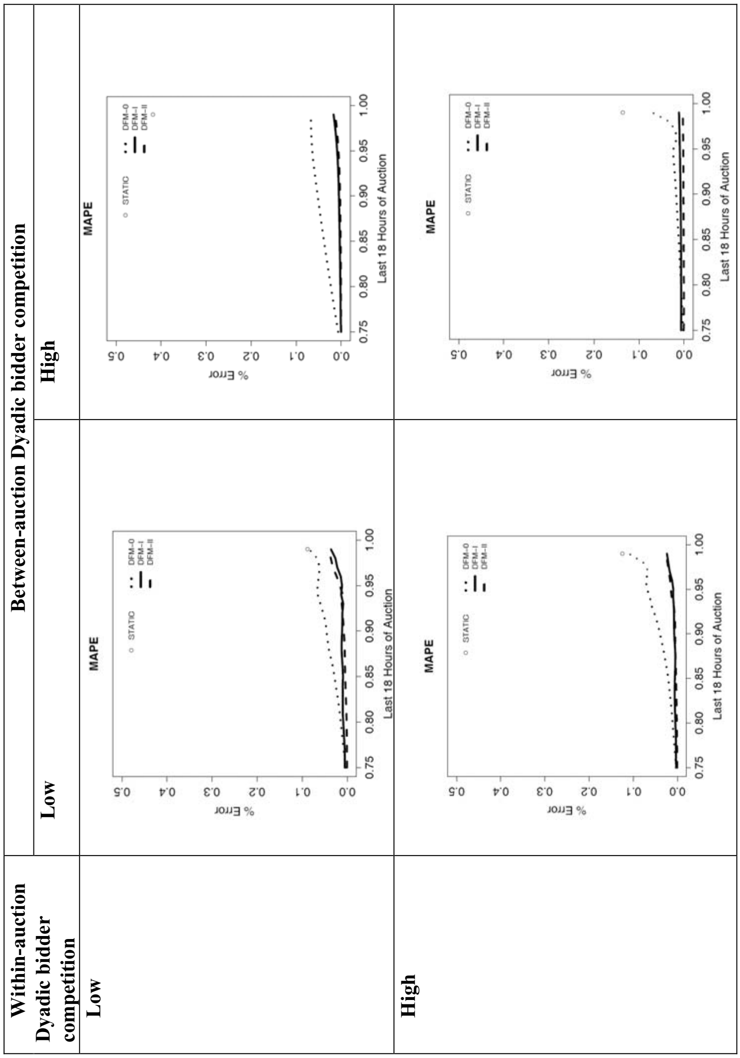


Fig. 11 Mean Absolute Percentage Errors of Competing Models by Bidder Competition

8 Conclusion and Future Direction

In this chapter, we present an innovative forecasting model for ongoing simultaneous auctions and use it to predict auction prices in real-time. We also compare our dynamic model with competing approaches and show that price dynamics matter and result in superior predictive capabilities. We also investigate the source for the predictive power of price dynamics and find that dynamics essentially proxy for bidder competition. One practical implication of this finding is that dynamics, which are conceptually simpler to capture than direct measures of competition, result in a more parsimonious model for competitive marketplaces.

Considering the higher stakes associated with online art auctions, our forecasting model gives added power to both auction house managers and bidders. For auction house managers, knowing the expected final price early enough gives them sufficient time to run promotions or call specific bidders to participate. It also provides them valuable insights regarding what combinations of items generate higher dynamics in simultaneous art auctions. For bidders with budget constraints, our models provide vital information regarding the items which are within their budget or provide them with the largest surplus. In these art auctions, most bidders participate with a desire to purchase more than one art object. Therefore, our forecasting model provides a tool for selecting complementary art items that are more affordable. Finally, our model can be used to build a dynamic price estimation system. Auction houses provide pre-auction estimates for their auctioned items. These values are computed by experts, curators, etc. Using our model, they can supplement the expert estimates with data-driven estimates, thereby providing richer pre-auction information. Moreover, using our dynamic forecaster, the initial estimates can be dynamically updated during the auction, thereby providing bidders with more up-to-date information. This can make the auction process more transparent.

Although our initial model is based on the approach developed by Wang, Jank and Shmueli (2008), we improve upon their work in two important ways. First, we apply their dynamic forecasting framework to the simultaneous online auction context where high-end art items are sold, ranging from a few thousands to a few millions of dollars. Finding that the dynamic forecasting framework is useful beyond individual-auction eBay-type auctions is important, as it shows the strength and flexibility of this dynamic forecaster. Second, we show that price dynamics that are represented by functional objects, in fact capture bidder competition. This provides an explanation to why dynamics matter and how to interpret them in an online auction context.

References

Ariely, D., Simonson, I.: Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions. Journal of Consumer Psychology 13(1&2), 113–123 (2003)

Bajari, P., Hortacsu, A.: The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions. RAND Journal of Economics 34(2), 329–355 (2003)

Bapna, R., Jank, W., Shmueli, G.: Price Formation and Its Dynamics in Online Auctions. Decision Support Systems 44(3), 641–656 (2008)

Czujack, C., Flores Jr., R., Ginsburgh, V. (eds.): On Long Run Co-Movements Between Paintings and Prints. Amsterdam (1996)

Dass, M., Reddy, S.K., Du, R.: Dyadic Bidder Interactions and Key Bidders in Simultaneous Online Auctions. University of Georgia (2007)

Foutz, N., Jank, W.: Prerelease Forecasting via Functional Shape Analysis of the Online Virtual Stock Market. In: MSI Report (2007)

Ghani, R., Simmons, H.: Predicting the End-Price of Online Auctions. In: International Workshop on Data Mining and Adaptive Modeling Methods for Economics and Management, Pisa, Italy (2004)

Gneezy, U.: Step-Level Reasoning and Bidding in Auctions. Management Science 51(11), 1633–1642 (2005)

Heyman, J.E., Orhun, Y., Ariely, D.: Auction Fever: The Effect of Opponents and Quasi-Endowment on Product Valuations. Journal of Interactive Marketing 18(4), 8–21 (2004)

Jank, W., Shmueli, G., Plaisant, C., Shneiderman, B.: Visualizing Functional Data With An Application to eBay's Online Auction. In: Haerdle, Chen, U. (eds.) Handbook on Computational Statistics on Data Visualization, Academic Press, San Diego (2008)

Jank, W., Shmueli, G.: Functional Data Analysis in Electronic Commerce Research. Statistical Science 21(2), 155–166 (2006)

Jap, S., Naik, P.: Bid Analyzer: A Method For Estimation and Selection of Dynamic Bidding Models. Marketing Science (2008)

Krishna, V.: Auction Theory. Academic Press, San Diego (2002)

Ku, G., Malhotra, D., Murnighan, J.K.: Towards a Competitive Arousal Model of Decision-making: A Study of Auction Fever in Live and Internet Auctions. Organizational Behavior & Human Decision Processes 96(2), 89–103 (2005)

Lucking-Reiley, D.: Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet. American Economic Review 89(5), 1063–1080 (1999)

Ramsay, J.O.: Matlab, R, and S-PLUS functions for functional data analysis (2003), `ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns`

Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005)

Reddy, S.K., Dass, M.: Modeling Online Art Auction Dynamics Using Functional Data Analysis. Statistical Science 21(2), 179–193 (2006)

Roth, A.E., Ockenfels, A.: Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet. American Economic Review 92(4), 1093–1103 (2002)

Rothkopf, M.H.: Bidding in Simultaneous Auctions with a Constraint on Exposure. Operations Research 25(4), 620 (1977)

Russo, R.P., Shmueli, G., Shyamalkumar, N.D.: Models of Bidder Activity Consistent with Self-Similar Bid Arrivals. In: Jank, Shmueli (eds.) Statistical Methods in eCommerce Research, Wiley, Chichester (2008)

Shmueli, G., Jank, W.: Modeling Dynamics in Online Auctions: A Modern Statistical Approach. In: Kauffman, R., Tallon, P. (eds.) Economics, Information Systems and Electronic Commerce: Empirical Research, M.E. Sharpe, Armonk (2008) ISBN: 978-0-7656-1532-9

Simonoff, J.S.: Smoothing Methods in Statistics, 1st edn. Springer, New York (1996)

Sood, A., James, G.M., Tellis, G.J.: Functional Regression: A New Model and Approach for Predicting Market Penetration of New Products. In: Marketing Science (2007) (forthcoming)

Wang, S., Jank, W., Shmueli, G.: Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis. Journal of Business and Economic Statistics (2008) (forthcoming)

Appendix A: ***Estimating Price Curves from Observed Bid Histories***

Data Pre-Processing: The first step is to recover a smooth price curve from the observed bid history. This recovery stage often needs to be preceded by data pre-processing. Let $y_i^{(j)}$ denote the bid placed at time t_{ij}. To better capture bidding activity at the beginning and near the end of the auction, we transform the bids into their log score. Next, we linearly interpolate the raw data and sample it at a common set of time points $t_i, 0 \leq t_i \leq 3,$ where $i = 1,\ldots n$ in order to account for the irregular spacing of the bid arrival. Thus, each auction can be represented by a vector of equal length

$$y^{(j)} = (y_1^{(j)},\ldots, y_n^{(j)}) \tag{6}$$

which forms the basis for the smooth price curves.

Recovering the Underlying Price Function: To recover the underlying price curves, we use penalized monotone curves (Ramsay and Silverman 2005; Simonoff 1996), which provide both small local variation and overall smoothness. They also readily yield higher-ordered derivatives of the target price curve as desired in our case. We first start with selecting an appropriate basis function for the price dynamics. We decided on using b-spline basis function as it is commonly used in cases when the data is not periodic. Next, for every auction, we express a price function *w(t)* as a linear combination of a basis function $\phi_k(t)$. Therefore,

$$w(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{7}$$

where c_k is a constant and k ranges from 1 to K basis functions. Then, we fit the data by minimizing the error sum of squares by

$$SSE = \sum_{j=1}^{n} [y_j - f^{(j)}(t_j)]^2 \tag{8}$$

where y_j is the price of item j observed in time t_j. j is 1…100 in our case and *f(t)* is the price function that fits the observed values. A roughness penalty function is imposed to measure the degree of departure from the straight line

$$PEN_m = \int \left[D^m f(t) \right]^2 dt \tag{9}$$

where $D^m f$, m =1, 2, 3 …., is the m^{th} derivative of the function f. The goal is to find a function $f^{(j)}$ that minimizes the penalized residual sum of squares

$$PENSS_{\lambda,m}^{(j)} = \sum^{n} \left(y_i^{(j)} - f^{(j)}(t_i) \right)^2 + \lambda \times PEN_m^{(j)} \tag{10}$$

where the smoothing parameter λ provides the trade-off between fit $[(y_i^{(j)} - f^{(j)}(t_i))^2]$ and variability of the function (roughness) as measured by PEN_m.[10] We use the monospline module developed by (Ramsay 2003) for minimizing $PENSS_{\lambda,m}^{(j)}$.

Appendix B: ***Forecasting Price Dynamics***

Since the forecasting model (equation 1.6) uses the price dynamics component of the same time-period, we must predict this component before forecasting price. To do so, we model $D^{(j)}y(t)$ as a polynomial in t with autoregressive (AR) residuals along with other covariates x_i as they also play a significant role in affecting the price dynamics (Figure 4). This leads to the following model for predicting price dynamics

$$D^{(j)}y(t) = \sum_{k=0}^{K} a_k t^k + \sum_{i=1}^{P} b_i x_i(t) + u(t) \tag{11}$$

where $t = 1, 2, ...T$ and $u(t)$ follows an AR model of order R

$$u(t) = \sum_{i=1}^{R} \phi_i u(t-i) + \varepsilon(t), \text{ where } \varepsilon(\mathrm{t}) \sim \text{iid N}(0, \sigma^2) \tag{12}$$

This results in a two-step forecasting procedure as we first estimate the parameters $a_0, a_1, ..., a_K, b_1, ..., b_p$ and the residuals $\hat{u}(t)$. Then, using the residuals, we estimate $\phi_1, ..., \phi_R$. Therefore with the information until time T, we first forecast the next residual by

[10] Sensitivity tests were performed with different values of p (4, 5, 6 were used) and λ (14 different values between 0.001 to 100 were used). We found the model fit to be insensitive to different values of p and λ. However, the RMSE for the model was the lowest with p=4 and $\lambda = 0.1$. Thus, we use these smoothing parameters in recovering the price curves.

$$\tilde{u}(T+1 \mid T) = \sum_{i=1}^{R} \tilde{\phi}_i u(T-i+1) \tag{13}$$

and then use it to predict the corresponding price derivative

$$D^{(j)}\tilde{y}(T+1 \mid T) = \sum_{k=0}^{K} \hat{a}_k (T+1)^k + \sum_{i=1}^{P} \hat{b}_i x_i (T+1 \mid T) + \tilde{u}(T+1 \mid T) \tag{14}$$

We can rewrite equation 1.10 to predict $D^{(j)}y(t)$ with h steps ahead by

$$D^{(j)}\tilde{y}(T+h \mid T) = \sum_{k=0}^{K} \hat{a}_k (T+h)^k + \sum_{i=1}^{P} \hat{b}_i x_i (T+h \mid T) + \tilde{u}(T+h \mid T) \tag{15}$$

Analysing Incomplete Consumer Web Data Using the Classification and Ranking Belief Simplex (Probabilistic Reasoning and Evolutionary Computation)

Malcolm J. Beynon* and Kelly Page

Cardiff Business School, Cardiff University, Colum Drive,
Cardiff, CF10 3EU, Wales, UK
e-mail: BeynonMJ@Cardiff.ac.uk, pagekl@Cardiff.ac.uk

Abstract. Consumer attitudes, involvement and motives have long been identified as important determinates of decision making in classic models of consumer behaviour. Online consumer attitudes may differ depending on the level of web experience of the intended consumer. This chapter considers Classification and Ranking Belief Simplex (CaRBS) analyses of consumer web data, considering attitudes from consumers with different levels of web experience. The CaRBS technique is based on Probabilistic Reasoning (Dempster-Shafer theory) and Evolutionary Computation (Trignometric-Differential Evolution), two known components of soft computing. An important facet of the presented analyses is the ability of the CaRBS technique to analyse incomplete data, without the need for the missing values present to be managed in anyway. The chapter allows a pertinent demonstration of how soft computing, here using CaRBS, can offer the opportunity for realistic analysis, more realistic than traditional techniques.

1 Introduction

Industry experts predict that online retailing sales in Europe will grow to an estimated €263 billion by 2011, as the number of online shoppers grows to 174 million (Forrester 2006). Interesting the economic crisis is not dampening but accelerating the shift in retailing to the Internet, as many e-retailers see a big opportunity to appeal to increasingly price-conscious consumers. Sales, in the UK online retailing sector is forecast to more than double by the end of 2011 to £21.3 billion, whilst high street sales will plummet by up to £8.3 billion (Experian 2009). This increase in choice in online retailers means a need for increased understanding of consumer online shopping behaviour for effective segmentation. Consumer reports of dissatisfied site experiences and the rising number of baskets left empty or 'drop-out rates' begs the question what is it about online retailing that interests or inhibits some consumers from purchasing? (Page-Thomas et al. 2006). Consequently a rise

J. Casillas & F.J. Martínez-López (Eds.): Marketing Intelligence Systems, STUDFUZZ 258, pp. 447–473.
springerlink.com

in the need for marketers to develop valid and reliable methods for modeling and segmenting consumers on their online shopping attitudes, motives and behaviour is evident (Liao and Tow-Cheung 2001; Lee and Lin 2005).

Consumer attitudes, involvement and motives have long been identified as important determinates of decision making in classic models of consumer behaviour. However, in an online context this research is still in its infancy (Chang 1998), dominated by traditional demographic and behavioural bases for segmentation, with limited research exploring online shopping motives (Rohm and Saminathan 2004) and Web-usage-lifestyle (Brengman et al. 2005). Online consumer attitudes may also differ depending on the level of web experience of the intended consumer (Donthu and Garcia 1999). Given the reported differences in consumers based on shopping experience (Smith and Carsky 1996); technology experience (Swoboda, 1998); and web shopping experience (Donthu and Garcia 1999), how consumers with differing past technology experiences differ in their level of perceived web purchase attitudes and involvement is important (*ibid.*). With continued use of information technology users evolve in both experience and skill (Davis et al. 1989). Web experience as a segmentation variable is important as current evidence has identified that the length or degree of prior experience has been found to be an important driver of the adoption and perceptions of electronic technologies (Moore and Benbasat 1991; Taylor and Todd 1995; Handzic and Low 1999). Further, with respect to the web, the degree of use has been found to be an important moderator of attitude toward the web and the success of the web-activity undertaken (Diaz et al. 1997). As such profiling user attitudes toward online shopping, partitioned by their level of web experience (low and high) can identify important market segments for targeting and strategic development.

However, unlike behavioural measures of usage which can be measured through observation with log-file and cookie data, variables measuring consumer attitudes, involvement and past usage experience, require measurement through self-report methods of data collection such as Internet Surveys. With the rise in the number of consumers online has resulted in the rise in the use of online data collection methods such as Internet surveys in marketing (Christian et al. 2007; Manfreda et al. 2008). According to the ESOMAR (2006) the number of online research studies increased by 80% in 2005 with collection methods such as Internet surveys accounting for 20% of the total data collection expenditure worldwide. However, Internet surveying is not without its limitations. Given participant self-selection and item-non-response, missing or incomplete data is a common occurrence through internet-surveying methods (Smith 1997; Dillman and Christian 2005).

The whole point of conducting segmentation analyses in marketing is to be able to provide marketers with useful information about why some segments are similar whilst others differ (Hansen 2005). However, with the presence of incomplete data (with missing values), the ability to develop reliable segment profiles with confidence decreases. This chapter discusses and applies the use of a technique for the realistic analysis of incomplete data collected through an Internet survey about on consumer attitudes towards online shopping and past usage experience to aid reliable and valid segmentation analysis.

One of the most critical issues in model formulation and marketing analytics is the treatment of missing data (Koslowsky 2002), and subsequently, their management in marketing intelligent systems. For example, Roth (1994) summarises that missing data causes two specific problems: the loss of statistical power and bias in parameter estimates. As such, the standard/traditional solutions, in the marketing literature, have been their external management, either through case-wide deletion or mean imputation (Huisman 2000). However, the range of choice on how to manage the presence of missing values can seem as confusing as the issues for their presence. An overriding concern is that some thought should be made, indeed, the effect of a lack of thought is well expressed by Huang and Zhu (2002, p. 1613):

> "Inappropriate treatment of missing data may cause large errors or false results."

Once such management is undertaken, the ability to realize the associated marketing insights desired is reduced, since the richness of the original incomplete data is ignored. The importance of this issue is that most data analysis techniques adopted in marketing are not designed for the presence of missing data (Schafer and Graham 2002).

The analysis technique employed in this chapter is the Classification and Ranking Belief Simplex (CaRBS). Introduced in Beynon (2005*a*, 2005*b*), see also Beynon (2008), its rudiments are based on the Dempster-Shafer theory of evidence - DST (Dempster 1967; Shafer 1976), subsequently its analysis is performed in the presence of 'mathematical' ignorance (a facet of the more general term uncertain reasoning). DST, through its association with probabilistic reasoning, is considered one of the three key mathematical approaches to soft computing (Roesmer 2000), along with the evolutionary computing and fuzzy logic approaches (see Mantores 1990; Zadeh 1975; Yang et al. 2006). Indeed, the CaRBS technique also employs evolutionary computing in the inherent optimization undertaken in its segmentation (the terms used here for classification), namely using the nascent trigonometric differential evolution (Fan and Lampinen 2003).

In a marketing context, the explicit representation of a level of ignorance in the evidence from survey question responses and final segmentation of respondents, for example, is a novel feature associated with the employment of the CaRBS technique. Moreover, it allows the realistic analysis of incomplete data (the presence of non-responses to survey questions by respondents), with the retention of the missing responses rather than having to facilitate their external management (see for example, Schafer and Graham 2002). The role exposited here for the CaRBS technique is the facilitation of a marketing intelligent system based on uncertain reasoning, applied to a relevant marketing problem.

The specific marketing problem considered in this chapter relates to the segmentation of online consumers' web experience through their related consumer attitudes to online shopping. One feature of the associated web experience data set investigated here, part of a larger survey, is its incompleteness. Two CaRBS analyses are undertaken on the web experience data set, to facilitate respective marketing intelligent systems. The first analysis is on the original incomplete data set, with the second analysis undertaken on a completed form of the data set (completed here using mean imputation of the missing values - see later). These two CaRBS based analyses enable the clear demonstration of the impact of the management of missing values in incomplete data analysis, and the beneficial consequence of employing soft computing, here uncertain reasoning using DST, within a marketing context.

A feature of the CaRBS technique is the emphasis on the graphical representation of its results, primarily through the domain of the simplex plot (Beynon 2005a). Findings are reported at the individual respondent (consumer) level, as well as the overall segmentation of all the considered respondents (in both cases using the simplex form of data representation). Further, findings are reported (using simplex plots), on the relevance of the individual questions to the segmentation of the respondents, in respect of their levels of web experience. The intention of this chapter is to offer a benchmarkable understanding to the issues of consumer web purchasing analysis using CaRBS, and the ability to work with incomplete data (with missing values retained), in such segmentation based analysis, all facilitated through the soft computing associated methodologies of probabilistic reasoning and evolutionary computing.

2 Background

This section offers a description of the CaRBS technique, employed in the analysis of a survey-based web experience data set (later introduced). The reader is encouraged to read through the research studies Beynon (2005a, 2005b), for a fuller understanding of this technique, here the description will include the exposition of Dempster-Shafer theory of evidence - DST (Dempster 1967; Shafer 1976) and trigonometric differential evolution - TDE (Fan and Lampinen 2003). Where appropriate, the introduced terminology will be described using the language later used in the subsequent analysis of the referred to web experience data set. In general, DST is regularly considered a generalisation of the well-known probability theory (Lucas and Araabi 1999).

Formally, DST is based on a finite set of p elements $\Theta = \{s_1, s_2, ..., s_p\}$, collectively called a frame of discernment (Θ). A *mass value* is a function m: $2^\Theta \to [0, 1]$ such that $m(\varnothing) = 0$ ($\varnothing$ - the empty set) and $\sum_{s \in 2^\Theta} m(s) = 1$ (2^Θ - the power set of Θ). Any proper subset s of the frame of discernment Θ, for which $m(s)$ is non-zero, is called a focal element and the concomitant mass value represents the exact belief in the proposition depicted by s. The notion of a proposition here, being the collection of the hypotheses represented by the elements in a focal element. The collection of mass values (and the focal elements) associated with a piece of evidence is called a body of evidence (BOE).

In the context of the CaRBS technique, it is concerned with the binary segmentation of objects (respondents R_j, $1 \le j \le n_R$) to a hypothesis x (high web experience - see later) and not-the-hypothesis $\neg x$ (low web experience), and a level of concomitant ignorance, based on the respondents' responses to a number of survey questions (P_i, $1 \le i \le n_P$). For a single respondent (R_j) and their response to a survey question (P_i), in CaRBS the associated evidence is formulated in a *response* BOE, defined $m_{j,i}(\cdot)$, and is made up of the mass values, $m_{j,i}(\{x\})$ and $m_{j,i}(\{\neg x\})$, which denote the levels of exact belief in the association of the object to x and $\neg x$, and $m_{j,i}(\{x, \neg x\})$ the level of concomitant ignorance. In the case of $m_{j,i}(\{x, \neg x\})$, its association with the term ignorance is because this mass value is unable to be assigned specifically to either x or $\neg x$ (an unknown distribution exists in the allocation of this mass value to x and $\neg x$).

From Safranek et al. (1990), and used in CaRBS, the described triplet of mass values in a response BOE, are given by the expressions (for one of a respondent's survey question response values v);

$$m_{j,i}(\{x\}) = \frac{B_i}{1-A_i} cf_i(v) - \frac{A_i B_i}{1-A_i}, \; m_{j,i}(\{\neg x\}) = \frac{-B_i}{1-A_i} cf_i(v) + B_i$$

$$\text{and } m_{j,i}(\{x, \neg x\}) = 1 - m_{j,i}(\{x\}) - m_{j,i}(\{\neg x\}),$$

where $cf_i(v) = 1/(1 + \exp(-k_i(v - \theta_i)))$, with k_i, θ_i, A_i and B_i ($i = 1, .., n_P$) the control variables incumbent in CaRBS, which require value assignment for its configuration (optimum configuration). Importantly, if either $m_{j,i}(\{x\})$ or $m_{j,i}(\{\neg x\})$ are negative they are set to zero, and the respective $m_{j,i}(\{x, \neg x\})$ then calculated. Further exposition of this mathematical process for the construction of a response BOE is given in Figure 1, and the later representation of a response BOE as a simplex coordinate in a simplex plot also shown.

Figure 1(*a* to *b*) shows the process by which a response value v (from survey question P_i) is re-scaled into a confidence value $cf_i(v)$ (over the domain 0 to 1), and then transformed into a body of evidence, here a response BOE $m_{j,i}(\cdot)$, made up of a triplet of mass values; $m_{j,i}(\{x\})$, $m_{j,i}(\{\neg x\})$ and $m_{j,i}(\{x, \neg x\})$. The response BOE $m_{j,i}(\cdot)$ is then able to be represented as a single simplex coordinate $p_{j,i,v}$ in a simplex plot (Figure 1*c*). That is, a point $p_{j,i,v}$ exists within an equilateral triangle such that the least distance from $p_{j,i,v}$ to each of the sides of the equilateral triangle are in the same proportions (ratios) to the values, $v_{j,i,1}$ ($m_{j,i}(\{x\})$), $v_{j,i,2}$ ($m_{j,i}(\{\neg x\})$) and $v_{j,i,3}$ ($m_{j,i}(\{x, \neg x\})$).

The set of response BOEs $\{m_{j,i}(\cdot), i = 1, \ldots, n_P\}$, associated with an individual respondent R_j's survey question responses, can be combined using Dempster's combination rule into a *respondent* BOE, denoted $m_j(\cdot)$. This combination rule, denoted by $[m_{j,i_1} \oplus m_{j,i_2}]$ (on two BOEs $m_{j,i_1}(\cdot)$ and $m_{j,i_2}(\cdot)$), is defined by:

$$[m_{j,i_1} \oplus m_{j,i_2}](y) = \begin{cases} 0 & y = \varnothing \\ \dfrac{\sum_{F_1 \cap F_2 = y} m_{j,i_1}(F_1) m_{j,i_2}(F_2)}{1 - \sum_{F_1 \cap F_2 = \varnothing} m_{j,i_1}(F_1) m_{j,i_2}(F_2)} & y \neq \varnothing \end{cases}$$

where F_1 and F_2 are focal elements from the independent BOEs $m_{j,i_1}(\cdot)$ and $m_{j,i_2}(\cdot)$, respectively. This combination rule can then be used iteratively to combine any number of BOEs. In Figure 1*c*, the results of combining two potential response BOEs, $m_{j,1}(\cdot)$ and $m_{j,2}(\cdot)$, is graphically shown to produce a BOE $m_C(\cdot)$, for the technical details of this combination process, see Table 1.

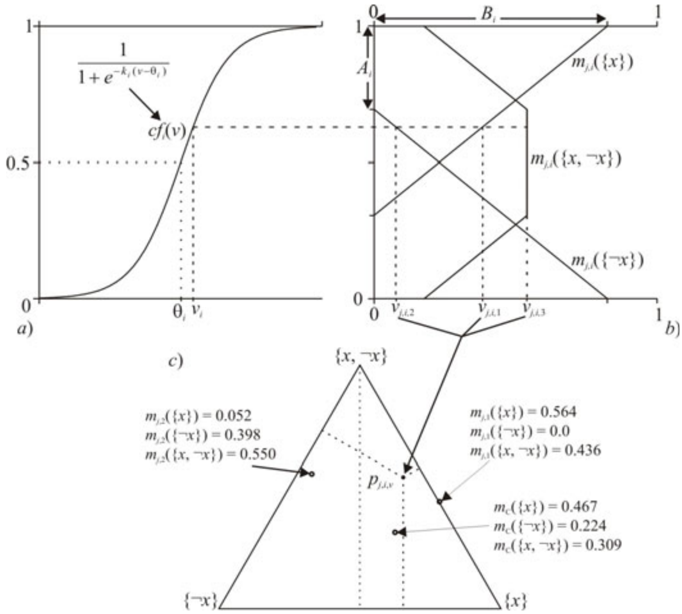


Fig. 1 Graphical representation of an intermediate stage in CaRBS, for a response value v to be transformed into a BOE and subsequent representation as a simplex coordinate in a simplex plot

Table 1 Intermediate stage of the combination of the two BOEs, mj,1(·) and mj,2(·)

$m_{j,2}(\cdot) \setminus m_{j,1}(\cdot)$	$m_{j,1}(\{x\})$, 0.564	$m_{j,1}(\{\neg x\})$, 0.000	$m_{j,1}(\{x, \neg x\})$, 0.436
$m_{j,2}(\{x\})$, 0.052	$\{x\}$, 0.029	$\varnothing$, 0.000	$\{x\}$, 0.023
$m_{j,2}(\{\neg x\})$, 0.398	$\varnothing$, 0.224	$\{\neg x\}$, 0.000	$\{\neg x\}$, 0.174
$m_{j,2}(\{x, \neg x\})$, 0.550	$\{x\}$, 0.310	$\{\neg x\}$, 0.000	$\{x, \neg x\}$, 0.240

Table 1 shows an intermediate stage of the combination of the BOEs, $m_{j,1}(\cdot)$ and $m_{j,2}(\cdot)$, namely the intersection and multiplication of the respective focal elements and mass values in the BOEs (see the definition of the combination rule). For example, the intermediate combination of $m_{j,1}(\{x\})$ and $m_{j,2}(\{x\})$, produces the focal element $\{x\} \cap \{x\} = \{x\}$ and mass value $0.564 \times 0.052 = 0.029$. Amongst the findings, a number of the resultant focal elements found are empty ($\varnothing$), it follows, $\sum_{F_1 \cap F_2 = \varnothing} m_{j,1}(F_1) m_{j,2}(F_2) = 0.224$ (part of the denominator of the combination rule), then the resultant BOE, defined $m_C(\cdot)$, can be taken from the summing of the values associated with the same focal elements in Table 1, and then divided by $1 - 0.224 = 0.776$ (the denominator in Dempster's combination rule). The subsequent, newly formed BOE $m_C(\cdot)$, is found to be;

$$m_C(\{x\}) = (0.029 + 0.023 + 0.310)/0.776 = 0.467,$$
$$m_C(\{\neg x\}) = 0.224 \text{ and } m_C(\{x, \neg x\}) = 0.309.$$

This newly created BOE is also shown in Figure 1*c*, its simplex coordinate position is further away from the $\{x, \neg x\}$ vertex than the two BOEs it was created from (through their combination). This is reduction in concomitant ignorance associated with $m_C(\cdot)$, compared to either of $m_{j,1}(\cdot)$ and $m_{j,2}(\cdot)$, is a consequence of the combination of their evidence.

Returning to the CaRBS technique, with the objects (respondents) known to be associated to either x or $\neg x$, a configured CaRBS system can be constructed, with respect to the intended optimization of the required segmentation (to x or $\neg x$). The effectiveness of such a configured CaRBS system is governed by the values assigned to the incumbent control variables k_i, θ_i, A_i, and B_i, $i = 1, \ldots, n_P$ (see Figure 1). This configuration process is defined a constrained optimization problem (using standardized response values), solved here using trigonometric differential evolution (TDE - Fan and Lampinen 2003), using an objective function (OB), which, from Beynon (2005b) is defined by:

$$\text{OB} = \frac{1}{4}\left(\frac{1}{|E(x)|}\sum_{R_j \in E(x)}(1 - m_j(\{x\}) + m_j(\{\neg x\})) + \frac{1}{|E(\neg x)|}\sum_{R_j \in E(\neg x)}(1 + m_j(\{x\}) - m_j(\{\neg x\}))\right)$$

where $E(\cdot)$ represents an equivalence class of respondents, to either x or $\neg x$ in this case (it can be shown the value of the objective function lies within the domain $0 \leq \text{OB} \leq 1$). It is noted, within the definition of the OB, maximising a difference value such as $(m_j(\{x\}) - m_j(\{\neg x\}))$ only indirectly affects the associated ignorance, rather than making it a direct issue (the mass value $m_j(\{x, \neg x\})$ is not included in the definition of the OB).

The TDE approach to effect this stated optimization is next briefly described, being a development on the nascent differential evolution (DE) algorithm, introduced in Storn and Price (1997). The domain of DE (and TDE) is the continuous space made up of the number of control variables considered. For a series of control variable values they are represented as a point in this continuous space (member vector). In DE, a population of vectors is considered at each generation of the progression to an optimum solution, measured through a defined objective function (such as OB).

Starting with an initial population, TDE generates new vectors by adding to a third member the difference between two other members (this change subject to a crossover operator). If the resulting vector yields a lower OB value then a predetermined population member it takes its place. This construction of a resultant vector is elucidated in Figure 2, where an example two dimensional (X_1, X_2) case is presented.

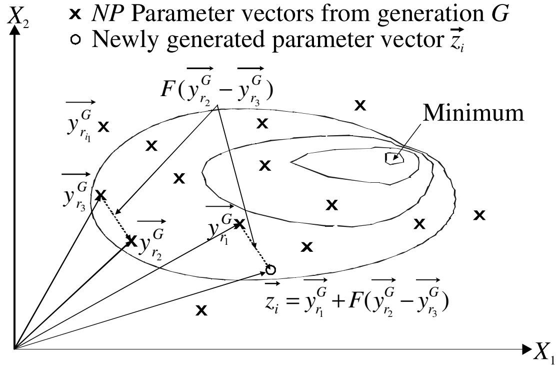


Fig. 2 Example of an OB with contour lines and process for generation of the resultant vector

In Figure 2, the effect of the 'vector' difference between two vectors $\overrightarrow{y^G_{r_2}}$ and $\overrightarrow{y^G_{r_3}}$ on the resultant vector $\vec{z}_i$ from another vector $\overrightarrow{y^G_{r_1}}$ is elucidated (F amplification parameter). The TDE development, presented in Fan and Lampinen (2003), takes account of the associated object function values of potential solutions, to hasten the convergence to an optimum solution (the optimum assignment of values to the control variables incumbent in CaRBS).

Pertinent to the analysis in this chapter, the ability of the CaRBS technique to analyse an incomplete data set is next described. Put simply, this is through the assignment of total ignorance to any missing response values present amongst their responses (the corresponding response BOE would be $m_{j,i}(\{x\}) = 0.000$, $m_{j,i}(\{\neg x\}) = 0.000$ and $m_{j,i}(\{x, \neg x\}) = 1.000$). Thus, there is no requirement for any external management of the missing values present, instead they are retained (modelled as ignorance as described), allowing the fullest opportunity to accrue insights from all the available/original data.

To gauge the relevance (quality) of the individual survey questions in the optimum segmentation of objects (respondents), the *average response* BOEs is introduced. As its name suggests, this BOE is simply the average of the response BOEs associated with an equivalence class of objects ($E(x)$ or $E(\neg x)$ in this case). The general average response BOE, in the marketing context, associated with a survey question P_i and equivalence class of respondents $E(R)$, defined $am_{i,R}(\cdot)$, is given by;

$$am_{i,R}(\{x\}) = \sum_{R_j \in E(R)} \frac{m_{j,i}(\{x\})}{|E(R)|},\ am_{i,R}(\{\neg x\}) = \sum_{R_j \in E(R)} \frac{m_{j,i}(\{\neg x\})}{|E(R)|} \text{ and}$$

$$am_{i,R}(\{x, \neg x\}) = \sum_{R_j \in E(R)} \frac{m_{j,i}(\{x, \neg x\})}{|E(R)|},$$

where R_j is a respondent. Further, the response BOEs from respondents which have missing values may or may not be included in the evaluation of their average response BOEs, depending on whether the level of missingness of responses to a survey question (over an equivalence class of objects), is of interest (or not). Since the average response BOE is a BOE, it can be represented by a simplex co-ordinate in a simplex plot; hence a graphical exposition of the relevance of survey questions can be given.

3 Measuring Consumer Web Purchasing Attitudes and CaRBS Analyses

This section provides an overview of the conceptualization and operationalization of the variables of interest in this study: online purchasing involvement and past web usage experience and the web-based internet survey method used for data collection. This is followed by two CaRBS analyses, the first on an original incomplete web experience data set, the second on completed version of the data set (missing values managed).

4 Conceptualisation and Operationalisation of Web Purchasing Involvement

Online retailers are generating consumer demand for services that provide increased consumer value through the provision of a satisfying online shopping experience (Ballantine 2005; Francis 2007). Consumer attitudes and perceptions of online retailing, especially their involvement with online purchasing is therefore becoming an increasingly important factor for marketing, strategy and technology development (Lee and Lin 2005). Transactional involvement is ‘the self-relevance of the purchasing activity or purchase-decision to the individual’ (Slama and Tashchian 1985; Mittal 1989). The level of involvement a consumer has with the purchase activity suggests that some consumers are more interested, concerned and/or involved in the purchase process and that this involvement influences the actual decision making and purchase behaviour (Slama and Tashchian 1985). So what is involvement derived from? From a channel perspective, past literature denotes that a consumer’s level of personal relevance in the web for transactional behaviour, maybe derived from a consumers percpetions of purchase risk, importance and entertainment (Babin et al. 1994; Korgaonkar and Wolin 1999; Liao and Tow Cheung 2001; Forsythe 2006).

From this discussion, we can thus define Perceived Web Purchase Involvement here as the level of personal relevance a consumer has with the conduct of transactional activity through the web and propose three underlying factors; risk, importance, and pleasure. *Perceived risk*: is the probability of a negative consequence occurring from transactional activity conducted through the web; *perceived importance*: the perceived ego-importance placed on the conduct of transactional activities through the web; and *perceived pleasure*: the perceived pleasure and excitement consumers derive from conducting transaction through the web.

To measure perceived web purchase involvement, content analysis of existing scale items measuring involvement and the factors of perceived risk, perceived hedonic value or pleasure and perceived importance or interest was conducted (Jain and Srinivasan 1990; Laurent and Kapferer 1985; McQuarrie and Munson 1986). In addition, existing scales measuring purchase decision involvement (Mittal 1989; Slama and Tashchian 1985) were also examined to assess existing item structure, content and design. A number of items were derived from these existing scales, with the content tailored for the context of interest in this study, the World Wide Web. To aid in this process and establish initial content validity, the item generation process also involved 6 in-depth interviews using a purchase orientated task-analysis and verbal elicitation technique. This process generated 12-items, reduced to 10-items (survey questions P_i - see Table 2) following exploratory factor analysis.

Table 2 Description of the ten survey questions (P1, ..., P10), on web purchasing attitudes

	Description	Number Missing	Mean	Standard Deviation
P1	Purchasing on the web seems exciting	17	3.668	1.403
P2*	Purchasing on the web is boring	33	4.315	1.271
P3	Purchasing goods on the web is very important to me	28	4.482	1.453
P4	Pleasure may be derived from the act of purchasing on the web	11	3.792	1.228
P5	One is never certain of ones choice to use the web to shop	26	4.230	1.349
P6	Purchasing and booking goods on the web is rather complicated	36	4.443	1.163
P7	I never know if I should purchase on the web	10	4.059	1.493
P8	Purchasing on the web is appealing	40	3.681	1.379
P9	I feel a little bit at a loss with knowing how to purchase goods on the web	12	4.799	1.581
P10*	I am not interested in purchasing on the web	11	4.360	1.693

5 Internet Survey Design

The survey was administered using a single cross-sectional web based survey design to a representative sample of the web population, numbering 300 in this case (Page 2003).

In Table 2, along with their description, descriptive statistics are presented on the ten survey questions measuring the three antecedents of web purchasing attitudes (risk, pleasure and importance). This includes the number of missing responses from the 300 considered respondents, the mean response and standard deviation values for each survey question (mean and standard deviation values used later). In the case of the missing values present, they total 224, which indicates 7.467% of the values are missing in this defined incomplete web experience data set. The two survey questions marked with *s in Table 2, P2 and P10, it is suggested have a negative relationship with the level of web experience, the others possibly having positive relationship.

One of the most critical issues in model formulation and marketing analytics is the treatment of missing data (Koslowsky 2002). As such the standard/traditional solution has been their external management, either through whole case deletion or imputation (Huisman 2000). The relevant need to concern oneself with this issue is that most data analysis techniques were not designed for their presence (Schafer and Graham 2002). However, once such management is undertaken the ability to realize the insights desired is reduced, since the richness of the original incomplete data is ignored (including possible differences between respondents and non-respondents). Furthermore, since the management of missing values is a pre-processing step, it may take a considerable amount of pre-processing time to allow for its completion (Pyle 1999; Huang and Zhu 2002).

The remainder of this section reports two CaRBS analyses of the web experience data set (previously described). The first analysis, on the incomplete web experience data set, retains the missing values present, with their associated evidence given as total ignorance (see description of the CaRBS technique given previously). The second analysis, is on a completed form of the web experience data set, where the missing values are managed through their replacement by the mean of the present values of the respective survey question (mean imputation, see Huisman 2000).

5.1 CaRBS Analysis of 'Incomplete' Web Experience Data Set

This section undertakes a CaRBS analysis of the incomplete web experience data set. Moreover, following the description of the CaRBS technique given previously, this section considers an analysis of the individual respondents' responses to the questions, P1, P2, …, P10, which model the evidence on the belief that a respondent (in their concomitant response $m_{j,i}(\cdot)$ and respondent $m_j(\cdot)$ BOEs), is more associated with high ({H} - $m_{j,i}(\{H\})$ and $m_j(\{H\})$) or low ({L} - $m_{j,i}(\{L\})$ and $m_j(\{L\})$) web experience, and a level of concomitant ignorance ({H, L} - $m_{j,i}(\{H, L\})$ and $m_j(\{H, L\})$).

This is considered here as a segmentation problem in the presence of ignorance, with a resultant marketing intelligent system constructed. As discussed earlier in this chapter, web experience as a segmentation variable is pertinent as current evidence has identified that prior experience has been found to be an important driver of the adoption and perceptions of usefulness of electronic technologies, like online retail store design (Moore and Benbasat 1991; Taylor and Todd 1995; Handzic and Low 1999).

To configure a CaRBS system, through the minimization of the objective function OB (defined previously), the respondents' survey question response values were standardized prior to the employment of TDE (using details given in Table 2), allowing consistent domains to be considered over the control variables incumbent in CaRBS, set as; $-2 \leq k_i \leq 2$, $-1 \leq \theta_i \leq 1$, $0 \leq A_i < 1$ and $B_i = 0.4$ (see Beynon 2005b). The explicit value for the B_i control variables ensured a predominance of ignorance in the evidence from the individual response values (in the concomitant response BOEs), so reducing over-conflict during the combination of the independent pieces of evidence (the combination of the response BOEs).

The TDE method was employed to produce an optimally configured CaRBS system, based on the previously defined TDE-based parameters, and run five times, each time converging to an optimum OB value, the best out of the five runs being OB = 0.3702. A reason for this value being away from its lower bound of zero is related to the imposed minimum levels of ignorance associated with each response BOE (due to the fixing of the B_i control variables), possibly also due to the presence of conflicting evidence from the response values. The resultant control variables, k_i, θ_i and A_i (B_i = 0.4), found from the best TDE run, are reported in Table 3.

Table 3 Control variables values associated with the configuration of CaRBS system, in the analysis of incomplete web experience data set

Parameter	P1	P2	P3	P4	P5
k_i	2.0000	−2.0000	2.0000	2.0000	2.0000
θ_i	0.5857	0.1395	−1.0000	0.5682	−0.1762
A_i	0.3323	0.6898	0.9423	0.6966	0.5029
Parameter	P6	P7	P8	P9	P10
k_i	2.0000	2.0000	2.0000	2.0000	−2.0000
θ_i	0.0519	0.2956	−1.0000	−0.0653	0.0970
A_i	0.2987	0.6616	0.5601	0.2933	0.9517

Amongst the values reported in Table 3, is the consistent absolute value of the k_i control variables, which from Figure 1 shows the evaluated confidence factor values are trying to be most discerning (producing values near the 0 and 1 limits). Only the two survey questions, P2 and P10, have negative k_i values (negative direction of association to low and high web experience), as expected from their introduction and concomitant description. Little consistency is found in the values assigned to the θ_i and A_i control variables. These control variable values are used in the construction of the response BOEs, modeling the evidence from the respondents' response values to the survey questions, and subsequent segmentation of the respondents' levels of web experience (0 - Low and 1 - High).

The construction of a response BOE is next demonstrated, considering the respondent R_{31} and the survey question P1. Starting with the evaluation of the confidence factor $cf_{\mathrm{P1}}(\cdot)$ (see Figure 1*a*), for the respondent R_{31}, P1 = 2.0000, when standardised it is $v = -1.1891$ (see Table 4 presented later), then;

$$cf_{\mathrm{P1}}(-1.1891) = \frac{1}{1+e^{-2.0000(-1.1891-0.5857)}} = \frac{1}{1+34.8005} = 0.0279,$$

using the control variables reported in Table 3. This confidence value is used in the expressions making up the triplet of mass values in the response BOE $m_{31,\mathrm{P1}}(\cdot)$, namely; $m_{31,\mathrm{P1}}(\{\mathrm{H}\})$, $m_{31,\mathrm{P1}}(\{\mathrm{L}\})$ and $m_{31,\mathrm{P1}}(\{\mathrm{H, L}\})$, found to be;

$$m_{31,\mathrm{P1}}(\{\mathrm{H}\}) = \frac{0.4}{1-0.5857}0.0279 - \frac{0.5857\times 0.4}{1-0.5857} = 0.0167 - 0.1991$$
$$= -0.1823 < 0.0000 \text{ so } = 0.0000,$$
$$m_{31,\mathrm{P1}}(\{\mathrm{L}\}) = \frac{-0.4}{1-0.5857}0.0279 + 0.4 = -0.0167 + 0.4 = 0.3833,$$
$$m_{31,\mathrm{P1}}(\{\mathrm{H, L}\}) = 1 - 0.0000 - 0.3833 = 0.6167.$$

For the respondent R_{31}, this response BOE is representative of all the associated response BOEs $m_{31,i}(\cdot)$ exhibiting the evidence in the respondent's response values to the survey questions, P1, …, P10, presented in Table 4 (using their standardised response values). These response BOEs, describe the evidential support from all the perceived survey question response values, from a respondent, to their association with low or high level of web experience (R_{31} - known to have a low level of web experience).

Table 4 Response values and response BOEs for the respondent R_{31}, from CaRBS analysis of incomplete web experience data set

Survey question	P1	P2	P3	P4	P5
R_{31} (actual)	2	7	4	4	7
R_{31} (standardised)	−1.1891	2.1122	−0.3316	0.1691	2.0537
$m_{31,i}(\{\mathrm{H}\})$	0.0000	0.0000	0.0000	0.0000	0.3908
$m_{31,i}(\{\mathrm{L}\})$	0.3833	0.3755	0.0000	0.0000	0.0000
$m_{31,i}(\{\mathrm{H, L}\})$	0.6167	0.6245	1.0000	1.0000	0.6092
Survey question	P6	P7	P8	P9	P10
R_{31} (actual)	6	5	2	6	7
R_{31} (standardised)	1.3385	0.6307	−1.2187	0.7596	1.5592
$m_{31,i}(\{\mathrm{H}\})$	0.3596	0.0000	0.0000	0.3088	0.0000
$m_{31,i}(\{\mathrm{L}\})$	0.0000	0.0000	0.0433	0.0000	0.0000
$m_{31,i}(\{\mathrm{H, L}\})$	0.6404	1.0000	0.9567	0.6912	1.0000

In Table 4, for the response values to support correct segmentation of the respondent R_{31}, in this case to low level of web experience (L), it would be expected for concomitant response BOEs to include $m_{31,i}(\{\mathrm{L}\})$ mass values to be larger than their respective $m_{31,i}(\{\mathrm{H}\})$ mass values, which is the case for the characteristics, P1, P2 and P8. Whereas, P5, P6 and P9, offer more evidence towards the respondent having a high level of web experience, and P3, P4, P7 and P10 only total ignorance.

The predominant strength of the response BOEs supporting incorrect segmentation - to {H} (of those giving evidence), is reflected in the final respondent BOE $m_{31}(\cdot)$ produced, through their combination (using Dempster's combination rule), which has mass values $m_{31}(\{\mathrm{H}\}) = 0.4995$, $m_{31}(\{\mathrm{L}\}) = 0.3161$ and $m_{31}(\{\mathrm{H, L}\}) = 0.1844$. This respondent BOE, with $m_{31}(\{\mathrm{H}\}) = 0.4995 > 0.3161 = m_{31}(\{\mathrm{L}\})$, suggests the respondent R_{31} is more associated with a high level of web experience, which is the incorrect segmentation in its case.

The results concerning another respondent R_{167} (known to have low level of web experience) are given in Table 5, with regard to the respective response BOEs $m_{167,i}(\cdot)$.

Table 5 Response values and response BOEs for the respondent R_{167}, from CaRBS analysis of incomplete web experience data set

Survey question	P1	P2	P3	P4	P5
R_{167} (actual)	3	-	-	6	-
R_{167} (standardised)	–0.4761	-	-	1.7983	-
$m_{167,i}(\{H\})$	0.0000	0.0000	0.0000	0.2962	0.0000
$m_{167,i}(\{L\})$	0.3360	0.0000	0.0000	0.0000	0.0000
$m_{167,i}(\{H, L\})$	0.6640	1.0000	1.0000	0.7038	1.0000
Survey question	P6	P7	P8	P9	P10
R_{167} (actual)	2	5	-	6	6
R_{167} (standardised)	–2.1006	0.6307	-	0.7596	0.9686
$m_{167,i}(\{H\})$	0.0000	0.0000	0.0000	0.3088	0.0000
$m_{167,i}(\{L\})$	0.3924	0.0000	0.0000	0.0000	0.0000
$m_{167,i}(\{H, L\})$	0.6076	1.0000	1.0000	0.6912	1.0000

In Table 5, a number of the response values, from the respondent R_{167}, are shown to be missing (denoted with '-'), namely the survey questions, P2, P3, P5 and P8. For each of these questions, their respective response BOEs offer only total ignorance (as stated earlier). Also assigned total ignorance are the response BOEs associated with the survey questions, P7 and P10, not because the value were missing, but that the concomitant control variable values have meant only total ignorance is assigned to them. The remaining response BOEs offer more than only ignorance to either $m_{167,i}(\{H\})$ or $m_{167,i}(\{L\})$, which from their combination, produce the respondent BOE $m_{167}(\cdot)$, found to be, $m_{167}(\{H\}) = 0.2986$, $m_{167}(\{L\}) = 0.4184$ and $m_{167}(\{H, L\}) = 0.2830$. This respondent BOE, with $m_{167}(\{H\}) = 0.2986 < 0.4184 = m_{167}(\{L\})$, suggests the respondent R_{167} is more associated with low level of web experience, which is the correct segmentation in this case.

The results concerning individual respondents, and their low or high level of web experience segmentation, are next graphically considered, see Figure 3, with response and respondent BOEs shown for four respondents, as simplex coordinates in simplex plots (simplex plot is the standard domain with CaRBS (see Figure 1c) - with each vertex identifying where there would be certainty in their segmentation to having low (L) or high (H) web experience, or ignorance (H, L), a vertical dashed line discerns between where there is majority belief of segmentation to L or H.

Figure 3 shows the evidence from survey question responses and final predicted segmentation of the respondents, R_{31} (3*a*), R_{167} (3*b*), R_{216} (3*c*) and R_{280} (3*d*). In each simplex plot shown, the grey shaded sub-domain of a simplex plot, at the top, defines where the respective response BOEs are able to exist (individually considered low level measurements whereby a level of ignorance is present in each response BOE - through the bounds on the B_i control variables, see Figure 1).

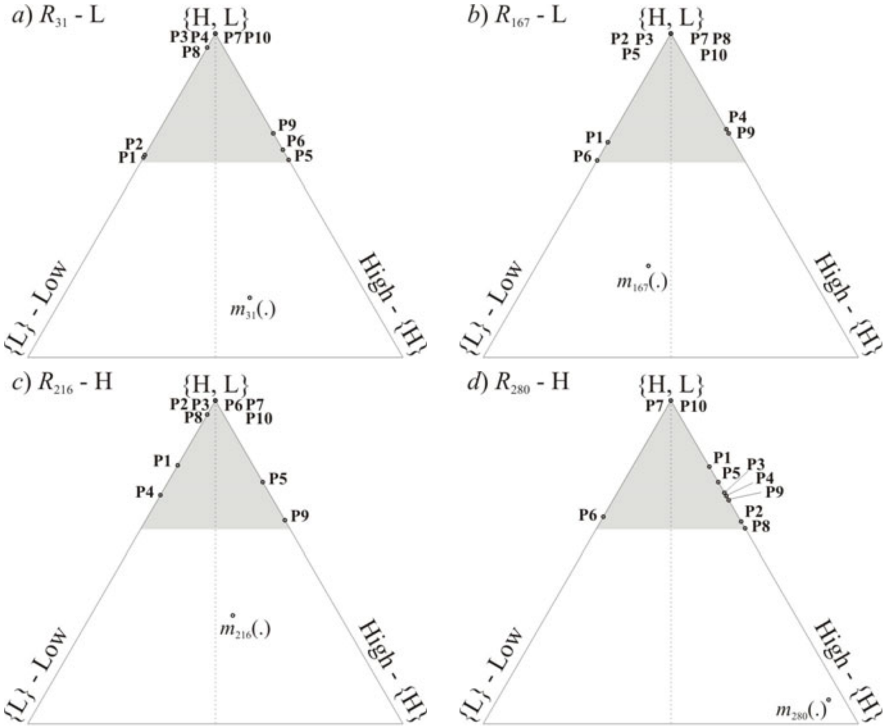


Fig. 3 Simplex plot based segmentation evidence of individual respondents, R_{31} (3*a*), R_{167} (3*b*), R_{216} (3*c*) and R_{280} (3*d*), from CaRBS analysis of incomplete web experience data set

With respondent R_{31} (Figure 3*a* - known to have low web experience), the ten circles in the shaded domain are the response BOEs representing the evidence from their responses to the survey questions P1, P2, …, P10 (labels for $m_{31,\mathrm{P}i}(\cdot)$ - see Table 4), with the lower circle shown the final respondent BOE $m_{31}(\cdot)$, found from the combination of the $m_{31,i}(\cdot)$ (compare with example in Figure 1). The discussion of this series of results on the respondent R_{31} follows the discussion surrounding Table 4, given previously. For example, the combination of the evidence in the response BOEs produces the respondent BOE $m_{31}(\cdot)$, shown to be to the right of the vertical dashed line in Figure 3*a*, indicating their predicted segmentation to being high experienced, incorrect in this case. The results for the respondent R_{167} (Figure 3*b*) similarly follow that surrounding the details reported in Table 5.

Two further respondents are considered in Figure 3, using the simplex plot representation, namely R_{216} and R_{280}, which are known to both have high levels of web experience. The results for the respondent r_{280} (Figure 3*c*), show evidence from their responses to the survey questions supporting its low (P1, P4 and P8) and high (P5 and P9) web experience segmentation, resulting in a respondent BOE offering weak evidence towards high level of web experience (correct segmentation). The results for the respondent r_{280} (Figure 3*d*), show strong supporting evidence from the majority of responses to the ten survey questions, with the implication being a very strong segmentation to them having high web experience, noticeably more certain in its predicted segmentation than for the other three respondents described.

The process of positioning the segmentation of a respondent, in a simplex plot, on their predicted level of web experience, can be undertaken for each of the 300 respondents considered, see Figure 4.

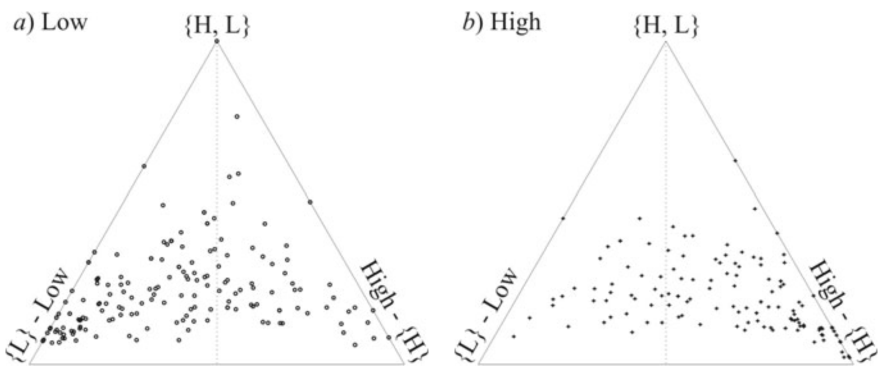


Fig. 4 Simplex plot based representation of segmentation of respondents with low (4*a*) and high (4*b*) levels of web experience, from CaRBS analysis of incomplete web experience data set

The simplex plots in Figure 4, contain the simplex coordinates, denoted by circles and crosses, representing the presented respondent BOEs of known low (4*a*) and high (4*b*) web experienced respondents, respectively. The different heights of the simplex coordinates (circles and crosses) in the simplex plots indicate variation in the levels of ignorance associated with the respondents' segmentations. One reason for the variation in the level of ignorance in a respondent's segmentation is the level of incompleteness in their responses to the questions considered, as well as conflicting responses. At the limit, in Figure 4*a*, there is a simplex coordinate at the {H, L} vertex, showing a respondent BOE for a respondent has only total ignorance in its predicted segmentation, due to a large number of missing values present amongst their responses, and total ignorance assigned to the responses they actually made.

The overall accuracy of this segmentation is based on whether the simplex coordinates representing respondent BOEs are on the correct side of the vertical dashed lines (low and high web experience to the left and right respectively). In summary, a total of 214 out of 300 (71.133%) respondents' have correct predicted segmentation (118 of 171 (69.006%) low and 96 of 129 (74.419%) high).

Throughout the descriptions of the predicted segmentation of the four respondents described in Figure 3, the evidence from their responses to the ten survey questions varied amongst the respondents. This variation potentially exists in the evidential relevance (quality) of the responses to the survey questions from all the respondents. The next results presented concern the relevance, discerning power, of each survey question used in segmenting the respondents' web experience. They have implications for how a website, or online retail presence is designed and communicated to differing segments of the target market categorized by high and low web usage. For example, if the item "purchasing and booking goods on the web is rather complicated" is reported as more important for users with low usage experience, targeting this segment of web users with specific communications to help educate and inform them about the booking process, could increase their overall satisfaction and positive attitudes with purchasing online. This information could also be used to inform web designers and marketing personnel on recommendations for improvement for effective transactional web design. As such the quality and accuracy of the data being analysed about not only a consumers perceptions, but also their web experience (i.e., unbiased), is of paramount importance to ensure correct and accurate results upon which electronic marketing decisions are made.

The elucidation of this relevance uses the average response BOEs defined previously, which accrue the level of evidence from survey questions to certain equivalence classes of respondents (known to have low and high web experience in this case), and as BOEs, they can be graphically reported using simplex plots, see Figure 5 (with only the grey shaded sub-domain shown - where response BOEs exist).

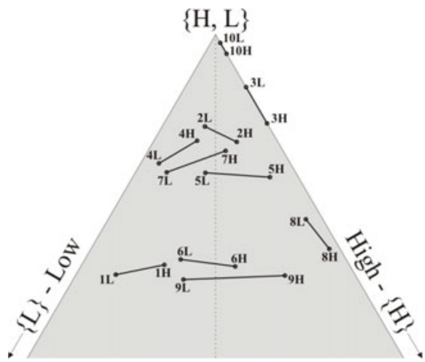


Fig. 5 Relevance of each survey question to the segmentation of respondents, from CaRBS analysis of incomplete web experience data set

In Figure 5, each simplex coordinate labelled '?L' and '?H' represents the average response BOEs associated with survey question P?, from known low and high web experienced respondents, respectively. For example, the average response BOE represented by 9H is found from the average of the response BOEs for survey question P9 of respondents with known high web experience (not including missing values). The further down the simplex plot sub-domain the ?L and ?H points appear the more relevance of the survey question (less ignorance) to the segmentation of low and high web experienced respondents, and the increased horizontal distance between respective points denotes the lesser ambiguity the survey question offers in the segmentation.

In the results in Figure 5, it follows, the survey questions, P1, P6 and P9 (also P8 to a lesser extent), are indicated to have the more relevance (than the others), based on the responses of the respondents, since they are lowest down the simplex plot sub-domain. In terms of ambiguity, the horizontal distance between the 9L and 9H simplex coordinates ,associated with the survey question P9, identify its limited associated ambiguity in what the survey question suggests with respect to low and high web experienced respondents, when compared with the other survey questions. Further interpretation can be given on the other survey questions considered, even for question P10, a rating an individuals of overall interest in the web for shopping, which from Figure 5, has little relevance to this study due to its position (10L and 10H) near the {H, L} vertex (large level of ignorance associated with it). This relevance could be attributed to the contextual nature of web purchase, in that interest could be moderated by product category, age or even gender, allowing for further segmentation.

5.2 CaRBS Analysis of 'Completed' Web Experience Data Set

This sub-section undertakes a further CaRBS analysis on the web experience data set, but here the missing values in the incomplete web experience data set, are now externally managed using mean imputation (Huisman 2000). It follows, all the respondents are retained in the data set (unlike when case deletion is employed), now termed a completed web experience data set, with the missing values replaced by the mean of the present values of the respective survey question.

To again configure a CaRBS system, to optimally segment the respondents to having low or high web experience, through the respondents' survey question response values, they were again standardized prior to the employment of TDE, allowing consistent domains over the control variables incumbent in CaRBS, set as; $-2 \le k_i \le 2$, $-1 \le \theta_i \le 1$, $0 \le A_i < 1$ and $B_i = 0.4$ (see Beynon 2005b). With standardized response values considered, and employing mean imputation, the missing values were now assigned the value 0.0000, since standardized data has mean zero (and unit standard deviation).

The TDE method was again employed to configure a CaRBS system, based on the previously defined TDE-based parameters, and run five times, each time converging to an optimum value, the best out of the five runs being OB = 0.3689. Like in the CaRBS analysis of the incomplete web experience data set, this value is noticeably above the lower bound, and is actually slightly lower than the previously found OB value (0.702). This OB value would suggest an improved level of segmentation has been achieved by the completing of the incomplete web experience dates set (using mean imputation). The resultant control variables, k_i, θ_i and A_i ($B_i = 0.4$), found from the best TDE run are reported in Table 6.

Table 6 Control variables values associated with the configuration of CaRBS system, in the analysis of completed web experience data set

Parameter	P1	P2	P3	P4	P5
k_i	2.0000	−2.0000	2.0000	2.0000	2.0000
θ_i	−0.1204	0.1453	−1.0000	0.4895	−0.1262
A_i	0.6714	0.6873	0.8902	0.7288	0.5628
Parameter	P6	P7	P8	P9	P10
k_i	2.0000	2.0000	2.0000	2.0000	−2.0000
θ_i	0.2410	0.2939	−0.4040	0.1092	0.6070
A_i	0.3818	0.6623	0.7801	0.2265	0.9820

The results in Table 6, concerning k_i, again show the same values and directions of association of the evidential contribution of the survey questions. The interesting feature here, is that the values found for the other control variables, θ_i and A_i, are mostly dissimilar to those found in the previous analysis (see Table 3). This is the first evidence on the impact of the external management of missing values, namely that the control variables found are different, so the configured CaRBS system here is different to that found in the previous analysis. Further evidence on this impact is shown by considering, in detail, the two respondents, R_{31} and R_{167}, first considered in the previous analysis.

Considering the respondent R_{31}, the construction of the response BOE associated with the question P1 is again described. For the respondent R_{31}, P1 = 2.000, when standardised, it is $v = -1.1891$ (see Table 7 presented later), then;

$$cf_{\mathrm{P1}}(-1.1891) = \frac{1}{1+e^{-2.0000(-1.1891+0.1204)}} = \frac{1}{1+8.4765} = 0.1055,$$

using the control variables in Table 6. This confidence value is used in the expressions making up the triplet of mass values in the response BOE $m_{31,\mathrm{P1}}(\cdot)$, namely; $m_{31,\mathrm{P1}}(\{\mathrm{H}\})$, $m_{31,\mathrm{P1}}(\{\mathrm{L}\})$ and $m_{31,\mathrm{P1}}(\{\mathrm{H, L}\})$, found to be;

$$m_{31,\mathrm{P1}}(\{\mathrm{H}\}) = \frac{0.4}{1-0.6714}0.1055 - \frac{0.6714\times0.4}{1-0.6714} = 0.1285 - 0.8173$$
$$= -0.6888 < 0.0000 \text{ so } = 0.0000,$$
$$m_{31,\mathrm{P1}}(\{\mathrm{L}\}) = \frac{-0.4}{1-0.6714}0.1055 + 0.4 = -0.1285 + 0.4 = 0.2715,$$
$$m_{31,\mathrm{P1}}(\{\mathrm{H, L}\}) = 1 - 0.0000 - 0.2715 = 0.7285.$$

For the respondent R_{31}, this response BOE is representative of all the associated response BOEs $m_{31,i}(\cdot)$, presented in Table 7 (using their standardised response values).

Table 7 Response values and response BOEs for the respondent R_{31}, from CaRBS analysis of completed web experience data set

Parameter	P1	P2	P3	P4	P5
R_{31} (actual)	2	7	4	4	7
R_{31} (standardised)	−1.1891	2.1122	−0.3316	0.1691	2.0537
$m_{31,i}(\{H\})$	0.0000	0.0000	0.0000	0.0000	0.3885
$m_{31,i}(\{L\})$	0.2715	0.3754	0.0000	0.0000	0.0000
$m_{31,i}(\{H, L\})$	0.7285	0.6246	1.0000	1.0000	0.6115
Parameter	P6	P7	P8	P9	P10
R_{31} (actual)	6	5	2	6	7
R_{31} (standardised)	1.3385	0.6307	−1.2187	0.7596	1.5592
$m_{31,i}(\{H\})$	0.3352	0.0000	0.0000	0.2893	0.0000
$m_{31,i}(\{L\})$	0.0000	0.0000	0.1018	0.0000	0.0000
$m_{31,i}(\{H, L\})$	0.6648	1.0000	0.8982	0.7107	1.0000

A consequence of the response BOEs shown in Table 7, through their combination, is the resultant respondent BOE, termed $m_{31}(\cdot)$, and found to be: $m_{31}(\{H\}) = 0.5014$, $m_{31}(\{L\}) = 0.2949$ and $m_{31}(\{H, L\}) = 0.2037$. This respondent BOE shows predominant association to high level of web experience, which is the incorrect segmentation in its case. Most interesting in these results concerning the respondent R_{31}, is what changes there is in its segmentation to that found in the previous analysis (when the missing values in the data set were not managed in any way - retained as missing). In terms of the respondent BOE, there is little difference in these BOEs between when the incomplete and completed web experience data sets were considered (there are limited differences in the individual response BOEs - compare the details presented in Table 4 and Table 7).

A further set of results are given with respect to the respondent R_{167}, see Table 8.

Table 8 Response values and response BOEs for the respondent R_{167}, from CaRBS analysis of completed web experience data set

Parameter	P1	P2	P3	P4	P5
R_{167} (actual)	3	-	-	6	-
R_{167} (standardised)	−0.4761	0.0000	0.0000	1.7983	0.0000
$m_{167,i}(\{H\})$	0.0000	0.0000	0.0000	0.2997	0.0000
$m_{167,i}(\{L\})$	0.0000	0.0000	0.0000	0.0000	0.0000
$m_{167,i}(\{H, L\})$	1.0000	1.0000	1.0000	0.7003	1.0000
Parameter	P6	P7	P8	P9	P10
R_{167} (actual)	2	5	-	6	6
R_{167} (standardised)	−2.1006	0.6307	0.0000	0.7597	0.9686
$m_{167,i}(\{H\})$	0.0000	0.0000	0.0000	0.2893	0.0000
$m_{167,i}(\{L\})$	0.3941	0.0000	0.0000	0.0000	0.0000
$m_{167,i}(\{H, L\})$	0.6059	1.0000	1.0000	0.7107	1.0000

Using the response BOEs reported in Table 8, the resultant respondent BOE, termed $m_{167}(\cdot)$, and found to be; $m_{167}(\{H\}) = 0.3795$, $m_{167}(\{L\}) = 0.2445$ and $m_{167}(\{H, L\}) = 0.3760$. This respondent BOE suggests association to high level of web experience, an incorrect segmentation in this case. The results here are in contrast to the findings from the analysis of the incomplete web experience data set, where a correct segmentation was found.

These findings on the respondent R_{167} demonstrate clearly the potential negative impact of externally managing the presence of missing values in a data set (in any way). Comparing the results in Tables 5 and 8, on the response BOEs used in constructing the respective respondent BOEs (from their combination), there is an impacting difference in the response BOEs $m_{167,\mathrm{P1}}(\cdot)$ found. That is, in the analysis of the completed data set, the response BOE $m_{167,\mathrm{P1}}(\cdot)$ offers only total ignorance, instead of the evidence towards their low level of web experience as offered by $m_{167,\mathrm{P1}}(\cdot)$ in the analysis of the incomplete data set.

A visual representation of the evidential support of the response BOEs and subsequent respondent BOEs are given in Figure 6, for the four respondents, R_{31} (6*a*), R_{167} (6*b*), R_{216} (6*c*) and R_{280} (6*d*).

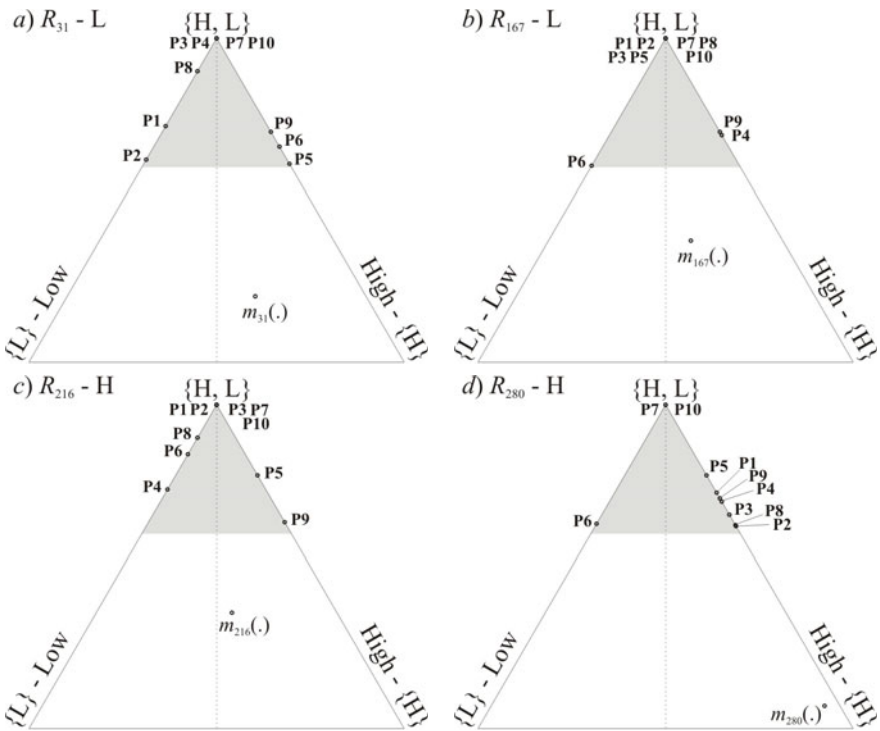


Fig. 6 Simplex plot based segmentation evidence of individual respondents, R_{31} (6*a*), R_{167} (6*b*), R_{216} (6*c*) and R_{280} (6*d*), from CaRBS analysis of completed web experience data set

The primary benefit of these simplex plots here, is in the ability to compare the segmentation results of these respondents with their segmentation in the previous analysis (see Figure 3). For the respondent R_{31}, as mentioned previously, there is limited change in the resultant respondent BOE, but some minor positional changes of the simplex coordinates representing the response BOEs. For the respondent R_{167}, the difference in the two analyses is more impacting than for R_{31}, where the positional change of the respondent BOE $m_{167}(\cdot)$ now to the right of the vertical dashed line is shown in Figure 6*b*, instead of being the left in Figure 3*b*. Inspection of the response BOEs, associated with the respondent R_{167}, shows the lack of evidential contribution now by their response to survey question P1 compared to in the previous analysis. The other two respondents R_{216} (6*c*) and R_{280} (6*c*), have similar results from both analyses, in terms of the positions of the respective respondent BOEs, but further inspection does show changes in the evidential contributions of the response BOEs.

The process of positioning the segmentation of a respondent, in a simplex plot, on their predicted level of web experience, can be again undertaken for each of the 300 respondents considered, see Figure 7.

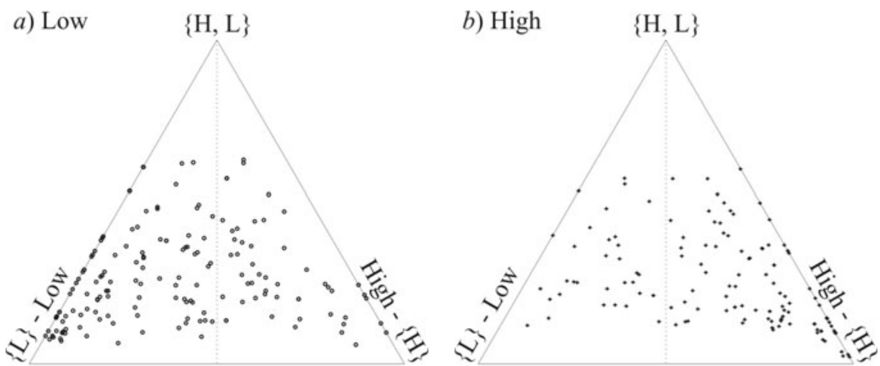


Fig. 7 Simplex plot based representation of segmentation of respondents with low (7*a*) and high (7*b*) levels of web experience, from CaRBS analysis of completed web experience data set

In Figure 7, the respondent BOEs are shown as circles and crosses (simplex coordinates), depending on whether the respondents' are known to have low or high levels of web experience. While the spread of the simplex coordinates shown appears similar to those reported in Figure 4 (from the CaRBS analysis of the incomplete web experience data set), there are changes. One noticeable change, is the non-presence of a simplex coordinate at the {H, L} vertex in Figure 7*a*, as there was in Figure 4*a*. This is due to the replacement of the missing response values of this respondent, which has meant some evidence has been assigned to its response BOEs, and so its movement away from the top vertex.

The overall accuracy of this segmentation is again based on whether the simplex coordinates representing respondent BOEs are on the correct side of the vertical dashed lines (low and high web experience to the left and right, respectively). In summary, a total of 214 out of 300 (71.133%) respondents' have correct predicted segmentation (119 of 171 (69.591%) low and 95 of 129 (73.643%) high). The overall segmentation accuracy is the same as in the CaRBS analysis of the incomplete web experience data set, but the separate accuracies of the low and high web experienced respondents do differ slightly.

To consider the relevance of the individual questions, the average response BOEs defined previously are again used, graphically reported in simplex plots in Figure 8 (grey shaded sub-domain shown only again).

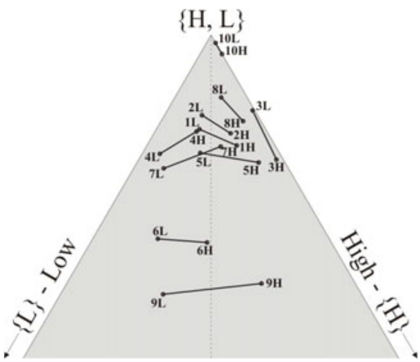


Fig. 8 Relevance of each survey question to the segmentation of respondents, from CaRBS analysis of completed web experience data set

The relevance results reported in Figure 8 indicate the survey questions, P6 and P9, are noticeably more relevant than the other survey questions (their positions lower down the simplex plot sub-domain shown). When comparing with the same results in Figure 5, from the CaRBS analysis of the incomplete web experience data set, the relevancies of the two survey questions, P1 and P8, have changed noticeably.

The changes in the relevancies of the survey questions, P1 and P8, between the two CaRBS analyses demonstrates most clearly the impact of managing the missing values present in a data set. Within the marketing context, here we can see that items P6 and P9 suffer less relevance inference from the presence of missing data than the other, previously found P1 and P8 more relevant items.

6 Future Trends

The potential future trends that can be considered, from this chapter, is the recognition that there is the possibility that the external management of missing values in incomplete data set can impact negatively on the inference that subsequent analysis allows.

The use of the CaRBS technique on the original incomplete and a completed version of the web experience data set, through the comparison of the results, clearly shows the variation in inference that could be the consequence of the management of missing values. It of course requires the existence of techniques, like CaRBS, that enable the analysis of incomplete data sets that a change of mind set towards the presence of missing values can take place in the future.

7 Conclusions

One of the most critical issues in model formulation and marketing analytics is the treatment of missing data, and subsequently, their management in marketing intelligent systems, causing problems through the loss of statistical power and quality in parameter estimates. As such, the standard/traditional solutions, in the marketing literature, have been their external management. The CaRBS technique employed throughout this chapter offers a novel analysis approach, with its inclusion of the notion of ignorance in the evidence and final segmentation of the respondents to their association with having low or high web experience.

A feature of the utilisation of the CaRBS technique is its ability to analyse incomplete data, in the case of the web experience data set, missing responses by respondents to certain survey questions. This is an important development, through the use of the soft computing associated methodology Dempster-Shafer theory in the CaRBS technique, since there has been limited ability to analyse such incomplete data sets, without having to externally manage the missing values present in some way. Indeed, this is a clear example of how soft computing approaches, in general, can offer new incites in how to undertake the pertinent analysis of marketing data, and creation of intelligent marketing intelligent systems.

The whole point of conducting segmentation analyses in marketing is to be able to provide marketers with useful information about why some segments are similar whilst others differ (Hansen 2005). However, with the presence of incomplete data (with missing values), the ability to develop reliable segment profiles with confidence decreases. By using a technique that enables researchers to analyse the relevance (quality) of the data, or level of bias in the dataset at either individual (respondent) level or variable item (question) level, it enables them to strategically discern the quality of the dataset for more informed and correct interpretation. This allows for more accurate marketing insight generation upon which strategic marketing decisions are made. This chapter has discussed and applied the use of a technique for the realistic analysis of incomplete data collected through an Internet survey about on consumer attitudes towards online shopping and past usage experience to aid reliable and valid segmentation analysis.

References

Babin, B.J., Darden, W.R., Griffin, M.: Work and/or fun: Measuring hedonic and utilitarian shopping value. Journal of Consumer Research *20, 644–656 (1994)*

Ballantine, P.W.: Effects of Interactivity and Product Information on Consumer Satisfaction in an Online Retailing Setting. International Journal of Retailing and Distribution Management 33(6), 461–471 (2005)

Bellenger, D.N., Korgaonkar, P.K.: Profiling the Recreational Shopper. Journal of Retailing 56, 77–91 (1980)

Beynon, M.J.: A Novel Technique of Object Ranking and Classification under Ignorance: An Application to the Corporate Failure Risk Problem. European Journal of Operational Research 167, 493–517 (2005a)

Beynon, M.J.: A Novel Approach to the Credit Rating Problem: Object Classification Under Ignorance. International Journal of Intelligent Systems in Accounting, Finance and Management 13, 113–130 (2005b)

Beynon, M.J.: Optimising Object Classification: Uncertain Reasoning based Analysis using CaRBS Systematic Search Algorithms. In: Vrakas, D., Vlahavas, I. (eds.) Artificial Intelligence for Advanced Problem Solving, pp. 234–253. IDEA Group Inc., PA (2008)

Brengman, M., Geuens, M., Weijters, B., et al.: Segmenting Internet Shoppers Based On Their Web-Usage-Related Lifestyle: A Cross-Cultural Validation. Journal of Business Research 58(1), 79–88 (2005)

Chang, S.: Internet segmentation: state-of-the-art marketing applications. Journal of Segmentation Marketing 2(1), 19–34 (1998)

Chen, Z.: Data Mining and Uncertain Reasoning: An Integrated Approach. John Wiley, New York (2001)

Christian, L.M., Dillman, D.A., Smyth, J.D.: Helping Respondents Get it Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. Public Opinion Quarterly 71(1), 113–125 (2007)

Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: A comparison of two theoretical models. Management Science 35(8), 982–1003 (1989)

Dempster, A.P.: Upper and lower probabilities induced by a multiple valued mapping. Ann. Math. Statistics 38, 325–339 (1967)

Diaz, A.N., Hammond, K., McWilliam, G.: A Study of Web Use and Attitudes Amongst Novices, Moderate Users and Heavy Users. In: Paper presented at the 25th EMAC Conference Proceedings (1997)

Dillman, D.A., Christian, L.M.: Survey Mode as a Source of Instability in Responses Across Surveys. Field Methods 17, 30–52 (2005)

Donthu, N., Garcia, A.: The Internet Shopper. Journal of Advertising Research, 52–58 (May/June 1999)

ESOMAR, Global MR Trends ESOMAR Report (2006), (September 5, 2009), `http://www.esomar.org/web/publication/paper.php?page=1&id=1416&keyword=industry%20trends`

Experian, Online to Rescue Britain's Retail Sector from Recession, Experian Report Accessed (June 24, 2009), `https://www.paypal-press.co.uk/Content/Detail.asp?ReleaseID=169&NewsAreaID=2`

Fan, H.-Y., Lampinen, J.A.: Trigonometric Mutation Operation to Differential Evolution. Journal of Global Optimization 27, 105–129 (2003)

Forrester, European Online Retail Consumer (December 20, 2004), `http://www.forrester.com/ER/Press/Release/0,1769,973,00.html` (retrieved September 20, 2005)

Forsythe, S., Liu, C., Shannon, D., et al.: Development of a Scale to Measure the Perceived Benefits and Risks of Online Shopping. Journal of Interactive Marketing 20(2), 55–75 (2006)

Francis, J.E.: Internet Retailing Quality: One Size Does Not Fit All. Managing Service Quality 17(3), 341–355 (2007)

Handzic, M., Low, G.C.: The role of experience in user perceptions of information technology: An empirical examination. South African Computer Journal 24, 194–200 (1999)
Hansen, T.: Consumer Adoption of Online Grocery Buying: A Discriminant Analysis. International Journal of Retail and Distribution Management 33, 101–121 (2005)
Huang, X., Zhu, Q.: A pseudo-nearest-neighbour approach for missing data on Gaussian random data sets. Pattern Recognition Letters 23, 613–1622 (2002)
Huisman, M.: Imputation of Missing Item Responses: Some Simple Techniques. Quality & Quantity 34, 331–351 (2000)
Jain, K., Srinivasan, N.: An Empirical Assessment of Multiple Operationalisation of Involvement. In: Paper presented at the Advances in Consumer Research, Provo, UT (1990)
Korgaonkar, P.K., Wolin, L.D.: A Multivariate Analysis of Web Usage. Journal of Advertising Research, 53–68 (March/April 1999)
Koslowsky, S.: The case of missing data. Journal of Database Marketing 9(4), 312–318 (2002)
Lee, G.-G., Lin, H.-F.: Customer Perceptions of e-Service Quality in Online Shopping. International Journal of Retailing and Distribution Management 33(2), 161–176 (2005)
Liao, Z., Tow-Cheung, M.: Internet-based E-shopping and consumer attitudes: An empirical study. Information and Management 38, 299–306 (2001)
Lucas, C., Araabi, B.N.: Generalisation of the Dempster-Shafer Theory: A Fuzzy-Valued Measure. IEEE Transactions on Fuzzy Systems 7(3), 255–270 (1999)
McQuarrie, E.F., Munson, J.M.: The Zaichkowsky Personal Involvement Inventory: Modification and Extension. In: Paper presented at the Advances in Consumer Research, Provo, UT (1986)
Manfreda, K.L., Bosnjak, M., Berzelak, J., et al.: Web Surveys versus Other Survey Modes: A Meta-analysis Comparing Response Rates. International Journal of Marketing Research 50(1), 79–104 (2008)
Mantores, R.L.: De Approximate Reasoning Models. Ellis Horwood, West Sussex (1990)
Mittal, B.: Measuring Purchase-Decision Involvement. Psychology & Marketing 6, 147–162 (Summer 1989)
Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Information Systems Research 2(3), 192–222 (1991)
Page-Thomas, K.L., Moss, G., Chelly, D., et al.: The provision of service delivery information prior to purchase: A missed opportunity. International Journal of Retailing & Distribution Management 34(4/5), 258–277 (2006)
Page, K.L.: World Wide Web Perceptions and Use: Investigating the Role of Web Knowledge. Unpublished Doctoral Dissertation, UNSW, Sydney (2003)
Pyle, D.: Data Preparation for Data Mining. Morgan Kaufmann, Los altos (1999)
Roesmer, C.: Nonstandard Analysis and Dempster-Shafer Theory. International Journal of Intelligent Systems 15, 117–127 (2000)
Rohm, A.J., Swaminathan, V.: A typology of online shoppers based on shopping motivations. Journal of Business Research 57(7), 748–757 (2004)
Roth, P.: Missing Data: A Conceptual Review for Applied Psychologists. Personnel Psychology 47, 537–560 (1994)
Safranek, R.J., Gottschlich, S., Kak, A.C.: Evidence Accumulation Using Binary Frames of Discernment for Verification Vision. IEEE Transactions on Robotics and Automation 6, 405–417 (1990)

Schafer, J.L., Graham, J.W.: Missing Data: Our View of the State of the Art. Psychological Methods 7(2), 147–177 (2002)

Shafer, G.A.: Mathematical theory of Evidence. Princeton University Press, Princeton (1976)

Slama, M.E., Tashchian, A.: Selected Socioeconomic and Demographic Characteristics Associated with Purchasing Involvement. Journal of Marketing 49, 72–82 (Winter 1985)

Smith, C.: Casting the Net: Surveying an Internet population. Journal of Computer Mediated Communication 3(1) (1997)

Smith, M.F., Carsky, M.L.: Grocery Shopping Behaviour. Journal of Retailing and Consumer Services 3(3), 73–80 (1996)

Storn, R., Price, K.: Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimisation 11, 341–359 (1997)

Swoboda, B.S.: Conditions of Consumer Information Seeking: Theoretical Foundations and Empirical Results of Using Interactive Multimedia Systems. The International Review of Retail, Distribution, and Consumer Research 8(4), 361–381 (1998)

Taylor, S., Todd, P.: Assessing IT usage: The role of prior experience. MIS Quarterly 19, 561–570 (1995)

Yang, J.-B., Liu, J., Wang, J., et al.: Belief Rule-Base Inference Methodology Using the Evidential Reasoning Approach—RIMER. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 36(2), 266–285 (2006)

Zadeh, L.A.: Fuzzy Logic and Approximate Reasoning (In Memory of Grigore Moisel). Synthese 30, 407–428 (1975)

Author Index