Available online at www.sciencedirect.com
ScienceDirect
Computers & Operations Research 34 (2005) 2902–2917
computers & operations research
www.elsevier.com/locate/cor

Computer assisted customer churn management: State-of-the-art and future trends

John Hadden[a,*], Ashutosh Tiwari[a], Rajkumar Roy[a], Dymitr Ruta[b]

[a]Enterprise Integration, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK
[b]British Telecom, Adastal Park, Ipswich, Suffolk IP5 3RE, UK

Available online 19 December 2005

Abstract

A business incurs much higher charges when attempting to win new customers than to retain existing ones. As a result, much research has been invested into new ways of identifying those customers who have a high risk of churning. However, customer retention efforts have also been costing organisations large amounts of resource. In response to these issues, the next generation of churn management should focus on accuracy. A variety of churn management techniques have been developed as a response to the above requirements. The focus of this paper is to review some of the most popular technologies that have been identified in the literature for the development of a customer churn management platform. The advantages and disadvantages of the identified technologies are discussed, and a discussion on the future research directions is offered.

Keywords: Customer churn management; Soft computing; Churn prediction

1. Introduction

As markets become increasingly saturated, companies have acknowledged that their business strategies should focus on identifying those customers who are likely to churn.

Predictions of behaviour, customer value, customer satisfaction and customer loyalty are examples of some of the information that can be extracted from the data that should already be stored within a company's database. However, to perform such a complex analysis of the information it is necessary to either purchase commercial software or implement a solution based on one of the many data mining techniques that have been developed for this purpose.

This paper is aimed towards individuals who wish to develop their own churn management solution, or use an existing one.

Research is described on the most common techniques for predicting customer churn, indicating some of the current research trends, and directions for future research.

Section 2 provides a brief introduction to customer churn management. A five stage model is also presented to illustrate the steps that should be addressed for creating a churn management system.

* Corresponding author. Tel.: +44 (0) 1234 750111x2858; fax: +44 (0) 1234 750852.
E-mail addresses: j.hadden.2003@cranfield.ac.uk (J. Hadden), a.tiwari@cranfield.ac.uk (A. Tiwari), r.roy@cranfield.ac.uk (R. Roy), dymitr.ruta@bt.com (D. Ruta).

0305-0548/$ - see front matter
doi:10.1016/j.cor.2005.11.007

Section 3 discusses the techniques that are commonly used for the first stage of the churn management model, 'identifying the best data'. This section also discusses the variables that have been identified as being the most useful for churn prediction.

Section 4 provides details of the second stage of the churn management model, details of the current lack of research in this area, and the problems that maybe encountered by the users of the data are discussed.

Section 5 details the feature selection phase, which is the third phase of the churn management model. It provides details of the steps that should be performed for feature selection and the current research in this area. Examples of the feature selection methods that are currently available are also provided.

Section 6 is focused around the churn management predictive models. The section is sub-divided into traditional methods and soft computing techniques. The reader is provided with knowledge of techniques for predicting customer churn, including the most popular and accurate methods.

Validation for performance and accuracy is discussed in Section 7, which details the fifth phase of the churn management tool. Section 8 provides a discussion on all five stages of the model. It also identifies the areas of further research for customer churn management research.

2. Customer churn management

It is becoming common knowledge in business, that retaining existing customers is the best core marketing strategy to survive in industry [1]. Churn management is the term that has been adopted to define customer turnover. More specifically, churn management is the concept of identifying those customers who are intending to move their custom to a competing service provider. Once identified, these customers can be targeted with proactive marketing campaigns for retention efforts. Customer retention becomes an important aspect of everyday business strategy. When the number of customers belonging to a business reaches its peak, finding and securing new customers becomes increasingly difficult and costly. At this point of the businesses lifecycle it should be higher priority to retain the most valuable, existing customers, than trying to win new ones.

Churning customers can be divided into two main groups, voluntary and non-voluntary churners. Non-voluntary churners are the easiest to identify, as these are the customers who have had their service withdrawn by the company. There are several reasons why a company could revoke a customer's service, including abuse of service and non-payment of service. Voluntary churn is more difficult to determine, because this type of churn occurs when a customer makes a conscious decision to terminate his/her service with the provider. Voluntary churn can be sub-divided into two main categories, incidental churn and deliberate churn.

Incidental churn happens when changes in circumstances prevent the customer from further requiring the provided service. Examples of incidental churn include changes in the customer's financial circumstances, so that the customer can no longer afford the service, or a move to a different geographical location where the company's service is unavailable. Incidental churn usually only explains a small percentage of a company's voluntary churn. Deliberate churn is the problem that most churn management solutions try to battle. This type of churn occurs when a customer decides to move his/her custom to a competing company. Reasons that could lead to a customer's deliberate churn include technology-based reasons, when a customer discovers that a competitor is offering the latest products, while their existing supplier cannot provide them. Economical reasons include finding the product at a better price from a competing company. Examples of other reasons for deliberate churn include quality factors such as poor coverage, or possibly bad experiences with call centres, etc. [2].

Churn management efforts should not focus across the entire customer base because (i) not all customers are worth retaining, and (ii) customer retention costs money; attempting to retain customers that have no intention of churning will waste resource.

Companies need to understand their customers. Liu and Shih [3] reinforce this statement by suggesting that intense competition is forcing organisations to develop novel marketing strategies to capture customer needs, and improving satisfaction and retention. Canning [4] states that selling more to everyone is no longer a profitable sales strategy and a market place that continually grows more competitive requires an approach that focuses on the most efficient use of sales resources.

A number of products exist for customer relationship management (CRM), which aims at analysing a company's customer base. Organisations should take great care if they decide to purchase a CRM product 'off-the-shelf'. Chen

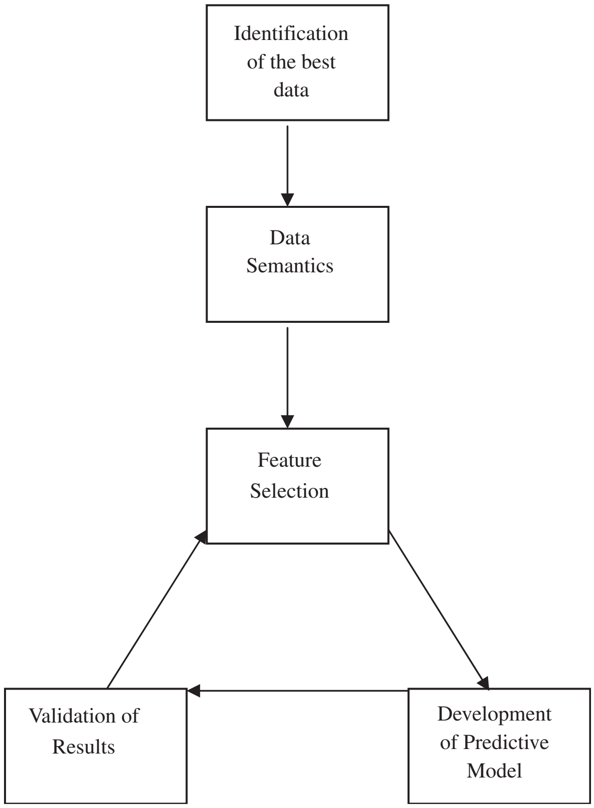

Fig. 1. The stages a churn management framework [6].

and Popovich [5] states “CRM vendors might entice organisations with promises of all powerful applications, to date there is no 100% solution”. The authors agree with the quote from Chen and Popovich [5]; however, we believe that no 100% solution is available, due to the uncertainty involved in churn prediction.

A five stage model for developing a customer churn management framework has been established [6]. These stages are illustrated in Fig. 1.

The techniques that are used in each of the five stages illustrated in Fig. 1 are discussed in detail through the following sections.

3. Techniques for identification of the best data

Fig. 1 shows that ‘identification of the best data’ is the initial step in developing a customer churn management framework. Different combinations of data hold different analytical powers. It is necessary to identify the data that best suits the type of analysis that is being performed. Different sets of data provide better indicators for different problems and the service sectors. For example, Ng and Liu [7] suggest that usage data should be mined for identifying customer churn in the ISP (Internet Service Provider) and telecommunications industry. Usage data has also been used for understanding e-customer behaviour of website users [8] and predicting mail order repeat buying [9]. Verhoef and Donkers [10] state that the purchasing of products and services is best predicted using the historical purchasing data. This claim is supported by Hsieh [11] who propose that the analysis of transaction data, along with account data and customer data could provide clues about the best incentives to offer bank customers for a better marketing strategy.

These examples reinforce the importance of the first stage identification the best data of developing a churn management framework. It is the authors understanding that care, thought and research should be committed to this initial phase, because the quality of the data will determine the power and accuracy of the overall model.

Customer data basically consists of many variables. There is a large amount of research that suggests recency, frequency, monetary (RFM) variables appear to be a good source for predicting customer behaviour [3,9–13]. The following definitions of all three RFM variables have been taken from Liu and Shih [12].

The recency variables are related to the time since the last purchase or use of the service. A lower value suggests a higher probability of the customer making repeat purchases.

Frequency variables are those connected to how often the service is used. In general it can be assumed that the higher the frequency, the more satisfied the customer is with the service. However, there are other reasons due to which a customer could be a heavy user, such as personal or business reasons.

A monetary variable for a customer would be the total money that customer has spent on services over a certain time period. Those customers with high monetary values are the ones the organisation should be most interested in retaining.

4. Establishment of data semantics

Data semantics is the process of understanding the context of the data in a database.

A definition provided by Ram [14] describes a semantic model as "objects, relationships amongst objects, and properties of objects".

The fields and data that are stored in a database can usually be viewed as a collection of words. These words could seem self-explanatory and unambiguous, however, it is common for data to be difficult to interpret. For example, the data stored could be abbreviations of company specific terms, in a numerical format, or having a dissimilar meaning to that which could be considered obvious. An example provided by Volz et al. [15] suggests that there are more than 500 free bio-informatics databases available over the internet, however, these databases can be difficult to understand due to inconsistent data definitions. According to one database the word "gene" is defined as "a DNA fragment that can be transcribed and translated into a protein". Another bio-informatics database describes a gene as "DNA region of biological interest with a name that carries a genetic trait or phenotype".

There are various models available for the capture of the meaning and structure of data in a database [16]. According to Ram and Khatri [16] "semantic models were developed to provide a precise and unambiguous representation of an organisation's information requirements".

The different types of tools that could be used for communication between the designer of a database and the end-users are as follows:

- *The entity relationship model* (ER)—the ER model represents data using entities, relationships, and attributes [17].
- *The relational model* (RM)—RM supports a data sub-language for data definition, but can be seen as a complete database model, supporting all aspects of data management [18].
- *The unifying semantic model* (USM)—USM is a formal specification for providing an accurate means of documentation and communication between users [14].

Despite claims by Datta et al. [6] that data semantics is one of the most difficult phases of dealing with large data warehouses, an analysis of 100 other research papers has discovered that the phase has not been documented in most research. There could be several reasons for this, including data sensitivity resulting in researchers being unable to publish any examples based on the companies data, or again due to sensitivity, researchers are forced to use what they have been given, and are not given the chance to explore a company's data warehouse for themselves.

Unforeseen circumstance could cause the fields of the data warehouse to change over time. Furthermore, the data warehouse is likely to use a range of regional databases as sources, and each region could have variances in the semantics of some fields. Depending on the methods to be used for churn analysis, null fields can cause problems. Data semantics is recognised as a continuous task, requiring close collaboration with the data semantic and model developers [6].

5. Feature selection

Feature selection is the process of identifying the fields which are the best for prediction, described by Sun et al. [19] as a critical process. It is an important stage because it helps with both data cleansing and data reduction, by including the important features and excluding the redundant, noisy and less informative ones [20]. There are two main

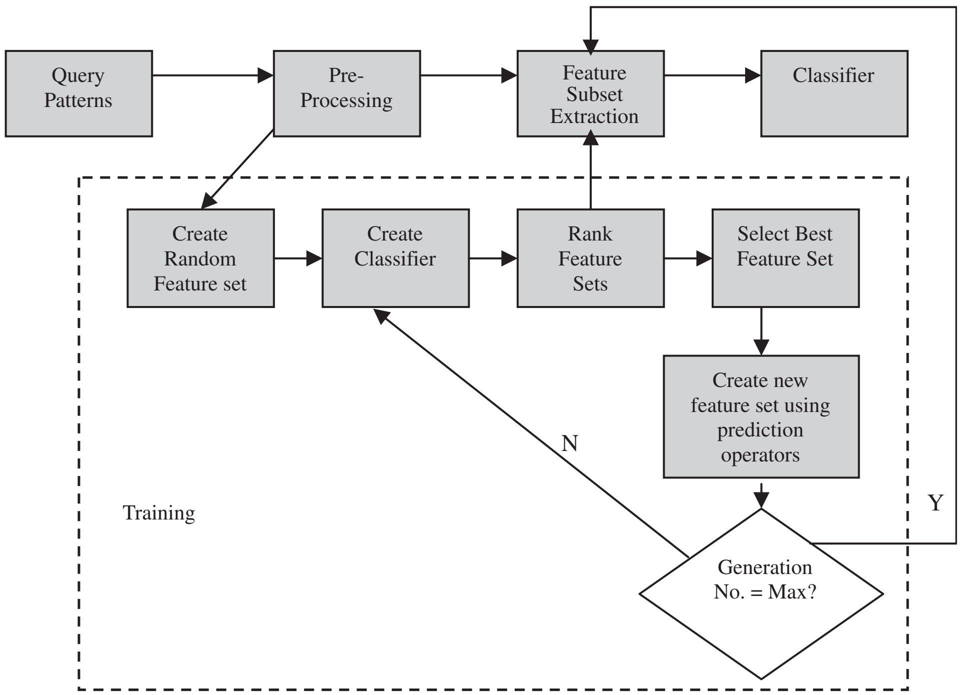

Fig. 2. The process of GA based feature selection, adapted from Sun et al. [19].

stages to feature selection. The first is a search strategy for the identification of the feature subsets, and the second is an evaluation method for testing their integrity, based on some criteria.

5.1. Search phase

The search phase of feature selection can be split into three categories, optimal, heuristic and randomised searches. The most straightforward optimal search method is an exhaustive search, however, with this method the number of possible subsets grows rapidly, making it unusable for even moderately sized feature sets. There are optimal search methods that avoid the exhaustive approach, using for example the branch and bound algorithm [19].

Two well-known heuristic methods are sequential forward selection (SFS) and sequential backward selection (SBS). SFS begins with an empty feature set and adds the best single feature to it. SBS works in the opposite way, by starting with the entire feature set, and at each step, drops the feature that least decreases the performance. SFS and SBS can be combined to create another search method, called the 'plus l − take away r' feature selection method. This method first enlarges the subset by adding l using the SFS method, and then deletes r using SBS. Two variations of this method have emerged called sequential forward floating search (SFFS) and sequential backward floating search (SBFS). These methods automatically and dynamically update the values of l and r, but they face limitations in a globally optimal solution [19].

Randomised search uses probabilistic steps, or sampling techniques. One such method is called the relief algorithm. This method assigns weights to the features. The features with weights exceeding a user-defined threshold are selected to train the classifier. Another randomised search method, which has been attracting more and popularity, is the genetic algorithm (GA) based approach. The research suggests that a GA can offer higher accuracy at the expense of greater computational effort. The main steps of feature selection using a GA are illustrated in Fig. 2 [19].

This paragraph outlines the feature selection methods that are recommended by various other researchers. Sun et al. [19] suggest a GA search approach, stating that a GA can provide a simple, general and powerful framework for selecting good subsets of features. Due to the characteristics of a GA, several feature subsets will be created and tested.

The best feature subset will then be evolved based on the fitness assignment. Yan et al. [20] suggest a receiver operating characteristic (ROC) curve for feature selection, or more specifically, they state the area under the curve (AUC) can be used for feature selection. The AUC is measured using the Wilcoxon–Mann–Whitney method of statistics

$$u = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x_i, y_j)}{mn},$$

where

$$I(x_i, y_j) = \begin{cases} 1, & x_i > y_j, \\ 0, & \text{otherwise.} \end{cases}$$

The values $[x_1, x_2, \ldots, x_n]$ are identified as the classifier output for positive samples, and the values $[y_1, y_2, \ldots, y_n]$ are identified as the classifier output for negative samples. The samples are used to perform a pairwise comparison. The AUC is then obtained for a linear classifier using only this single feature as the input. The feature is ranked by using a larger value between the AUC and an alternative value obtained by replacing $I(x_i, y_j)$ with $I(-x_i, -y_j)$.

A method suggested by Datta et al. [6] involves initially finding a subset of features from the data warehouse by manually selecting those that appear most suitable for the task.

It has been suggested by Meyer-Base and Watzel [21] that neural networks can be used for feature selection. Their research investigated the use of radial based neural networks (three-layered, feed-forward networks), however, it was necessary to include an additional layer to the traditional architecture in order to obtain a representation of relevant features. The additional layer converts the feature space linearly by multiplying the input vector and the centre of the nodes by the unit matrix B.

Ng and Liu [7] have performed feature selection by running an induction algorithm on the dataset. Datta et al. [6] have developed a technique to predict churn for the cellular mobile service sector. A two-step process is used for feature selection. The first step uses forward feature selection (FFS). A decision tree sorts the features according to their error rate, which is done by using a single feature to predict churn. The first 30–50 features with the lowest error are then presented to the modelling system. The second step is then added to the existing generation. A GA is used to find groups of features that are more predictive than using single features alone. The GA uses logistic regression to evaluate each possible set of features. Some of the features that tend to be valuable for churn prediction with the mobile telephone service sector include *minutes of use*, *balance brought forward from previous bills* and *the tenure of the customer*.

Datta et al. [6] use a decision tree to sort the features according to their error rate. Although the decision tree method was the one preferred by Datta et al. [6] they state that they also experimented with K-nearest-neighbour (KNN) and found no differences between the accuracy and performance of the models. Ng and Liu [7] also recommend the decision tree approach for feature selection suggesting that this method is highly accurate, and has a record for good performance.

Once features have been extracted from a dataset they need to be validated. It is important to note that the data the feature extraction was based on should be completely independent from the validation data; else there is a risk of over fitting.

5.2. The evaluation phase

Evaluation strategies can be divided into two categories, the first is called filter and the second is called wrapper. A wrapper evaluation method is where the feature subset evaluation is performed using a learning algorithm that is incorporated in the classification design, while the filter approaches uses q feature subset evaluation external to the classification design. Filter approaches are more efficient than wrapper approaches because they evaluate the fitness of features using criteria that can be tested quickly, although this approach could lead to non-optimal features, especially when the features are dependent on the classifier, leading to poor classifier performance [19,22].

Once features have been extracted from a dataset they need to be validated. It is important to note that the data the feature extraction was based on should be completely independent from the validation data; else there is a risk of over fitting.

6. Development of predictive model

A predictive model is defined as one that takes patterns that have been discovered in the database, and predicts the future [23]. According to Crespo and Weber [24] the most important predictive modelling techniques include decision trees and neural networks. The popularity of these technologies are reinforced by Baesens et al. [25] who claim neural networks and decision trees are typically traditional classification technologies. The following subsections provide an overview of both traditional and soft computing techniques used for predictive modelling.

6.1. Traditional methods

6.1.1. Decision trees

The most popular type of predictive model is the decision tree. Decision trees have become an important knowledge structure, used for the classification of future events [26]. Decision tree development usually consists of two phases, tree building and tree pruning. The tree-building phase consists of recursively partitioning the training sets according to the values of the attributes. The partitioning process continues until all, or most of the records in each of the partitions contain identical values. Certain branches may need to be removed because it is possible that they could consist of noisy data. The pruning phase involves selecting and removing the branches that contain the largest estimated error rate. Tree pruning is known to enhance the predictive accuracy of the decision tree, while reducing the complexity [27]. Pruning should be regarded as a process of experimentation because it is possible that pruning the tree could decrease the accuracy of the output rather than enhance it.

The C5.0 classification tree assembles classification trees by recursively splitting the instance space into smaller sub-groups until only instances from the same class remain known as a pure node, or a sub-group containing occurrences from different classes known as impure nodes. The tree is allowed to grow to its full potential before it is pruned back in order to increase its power of generalisation on unseen data.

A classification and regression tree (CART) is constructed by recursively splitting the instance space into smaller sub-groups until a specified criterion has been met. The decrease in impurity of the parent node against the child nodes defines the goodness of the split. The tree is only allowed to grow until the decrease in impurity falls below a user-defined threshold. At this time the node becomes a terminal, or leaf node [28].

Kitayama et al. [29] used a decision tree based approach to propose a model for customer profile analysis. The customer base was first segmented into groups of customers that were labelled preferred and regular, the preferred customers being those most valuable to the company. The decision tree was then applied to the segments in order to determine the necessary measures to take for both the preferred and regular divisions aim to prevent customers from switching to new companies. Dividing a population of data and generating nodes using optional explanatory variables creates the decision tree.

Experiments performed by Hwang and Euiiho Suh [47] involved a decision tree, neural network and logistic regression. The decision tree showed slightly better accuracy over the other technologies, however, Hwang and Euiiho Suh [47] state that these results do not prove decision trees to be the best choice in all cases. This is supported by Mozer et al. [30].

Ng and Liu [7] suggest that for the purpose of customer retention, the feature selection process has to be very accurate. For the purpose of identifying potential defectors, they chose C4.5, a decision tree induction method for classification, because it has shown proven good performance and it automatically generates classification rules.

Decision trees have been used for a variety of tasks. Fig. 3 demonstrates how classification rules can be applied to data, to perform decision-making tasks.

Jobless	Bought	Sex	...	Age	Savings	Granted
No	Car	Male	...	38	$150K	Yes
Yes	Jewel	Female	...	26	$60K	Yes
Yes	Stereo	Male	...	20	$10K	No

Fig. 3. Example of classification rules [7].

Fig. 3 shows an example of how classification rules can be used in a decision-making process. Examples of possible rules that could be generated from Fig. 3 include

If (Jobless = "No" and Bought = "Car" and Savings > 100k)
Granted = "yes"
If (Jobless = "Yes" and Bought = "Jewel" and Savings > 50k and Sex = "Female")
Granted = "Yes"
If (Jobless = "Yes" and Bought = "Stereo" and Sex = "male" and Age < 25)
Granted = "No"

Many variations of rules can be generated from the set of attributes, however, in a real case scenario not all attributes will be valuable to the decision-making process. Those attributes that are influential are known as the objective indicators. It should be noted that care must be taken when establishing the best classification rules because although a classification rule set can be highly accurate if correctly defined, poorly defined rules will result in a poor, unreliable system.

Classification rules appear to be common practice as a data mining technique. Liao and Chen [31] use classification rules to perform customer segmentation, establishing the rules from a relational database.

Datta et al. [6] also carried out research in the area of churn prediction and developed a model that they called Churn Analysis Modelling and Prediction (CHAMP). CHAMP also uses decision trees to predict customer churn in the telecommunications industry.

6.1.2. Regression analysis

Regression analysis is a popular technique used by the researchers dealing with predicting customer satisfaction. This provides a first step in model development. Mihelis et al. [32] developed a method to determine customer satisfaction, using an ordinal regression based approach. Another model for assessing the value of customer satisfaction was developed by Rust and Zahorik [33]. They used logistic regression to link satisfaction with attributes of customer retention. They claim that the logistic function can be interpreted as providing the retention probability.

Kim and Yoon [2] use a logit model to determine subscriber churn in the telecommunications industry, based on discrete choice theory. Discrete choice theory is the study of behaviour in situations where decision makers must select from a finite set of alternatives. According to Au et al. [27] regression analysis is fine for determining a probability for a prediction, however, it is unable to explicitly express the hidden patterns in a symbolic and easily understandable form. Hwang and Euhiio Suh [47] discovered that logistic regression performed best for predicting customer churn, when compared with neural networks and decision tree. It should be noted that Hwang and Euhiio Suh [47] were investigating a prediction of the customer lifetime value (CLV), with the intent of including customer churn. Hwang and Euhiio Suh [47] suggest that logistic regression is the best model for their purpose. The authors believe that many factors could influence these results, such as, the neural network parameters which were chosen and the data that the experiment was based on. It is fair to assume that the data may have been more suited for a logistic regression analysis than for a neural network or decision tree.

Datta et al. [6] used simple regression to initially predict churn but later experimented with nearest neighbour, decision trees and neural networks. The overall model used to develop the churn prediction platform presented by Datta et al. [6] was done using neural networks; however, they also experimented with KNN and decision trees. Their research could not establish a best method, and they have stated future directions as including an explanation of customer behaviour, because their model could predict customer churn, but it was unable to provide an explanation as to why a customer might churn. They also suggest that information stored externally to the organisations database should be included, such as the state of the telecommunications market, and competing offers, etc. The model suggested by Datta et al. [6] fails to distinguish between loyal customers, valuable customers and less profitable customers. They suggest that future research should include a more financial orientated approach, by optimising payoff. They suggest that by concentrating on payoff rather than churn, the developed model would weight those customers bringing in higher profits, more than those bringing in lesser profits. Baesens et al. [25] used Bayesian network classifiers for identifying the slope of the CLV for long-life customers, but used simple linear regression on the historical contributions of each customer to capture their individual lifecycles. The slope was then separated into either positive or negative classes to represent increased or decreased spending, and then this variable was used as the dependent variable in their study.

6.1.3. Other traditional methods

Other techniques dealing with predicting customer behaviour include semi-Markov processes, used by Jenamani et al. [8] to propose a model that considers e-customer behaviour. The discrete-time semi-Markov process was designed as a probabilistic model, for use in analysing complex dynamic systems. Prinzie and Van Den Poel [34] introduce a mixture transition distribution (MTD) to investigate purchase-sequence patterns. The MTD is designed to allow estimations of high order Markov chains, providing a smaller transition matrix facilitating managerial interpretation. Auh and Johnson [35] use five firm-level variables named as customer satisfaction index scores (SAT), customer loyalty index scores (LOY), year of data (YEAR), relative quality importance (RQI) and the average ease of comparing quality differences (EQ), to estimate a general linear model. Although Prinzie and Van Den Poel [34] do not offer a precise explanation of how these variables were established, the authors have determined the following definitions from the literature. SAT has been defined as a cumulative evaluation of a customer's purchase and consumption habits. YEAR is the year that the data is relevant to. RQI determines the relative impact that the perceived quality and value have on customer satisfaction. Estimations were established using a model based on quality-versus-price, as defined by the ASCI model (American Customer Satisfaction Index). EQ is defined as a measure of the ease of judging and comparing quality. It is basically the average of the response argued by asking the following question: “Thinking about the quality of product (e.g. mobile telephones) do you consider it easy or difficult to judge what is high versus low quality?” This question would have a score ranging between 1 and 10. A score of 1 would represent very easy while a score of 10 would represent very difficult. LOY is an indication of the customer's intentions to repurchase services or goods. Regression was used as an alternative to test for linear equality restrictions.

Chiang et al. [36] introduce their own algorithm for identifying potential churners, which they have named goal-oriented sequential pattern. This work uses association rules, which are defined as a technique that identifies relationships amongst variables. A simple example given by their research suggests that people, who purchase milk at a supermarket, will also buy bread at the same time. The paper defines two steps for finding out association rules. The first step is to detect the large itemset and the second is to establish the association rules by exploiting the large itemset. To accomplish the first step, the following conditions have to be met:

$$\begin{aligned}&\mathit{Support}(X, \cup, Y, D) \geqslant \text{Minsup}\\&\mathit{Confidence}(X \rightarrow Y) \geqslant \text{Minconf}\end{aligned}$$

Minsup and Minconf are user defined. The number of transactions that are contained in X is called the support of X and this determines the minimum number of support (Minsup). If the support of X meets the condition $\sigma x \geqslant$ Minsup, X is the large itemset.

For the second step, the a priori algorithm is used for the exploration of association rules. The a priori algorithm determines rules by a straightforward sequential process to determine relationships in the database.

Tao and Yeh [37] document two database marketing tools named USC (Usage Segment Code) and NRE (Net Revenue Equation) for the purpose of understanding customer behaviour. It is claimed that customer retention is one of the non-marketing activities that these two tools can be used for. Although it appears that these tools are not used for predicting potential defectors, but to help in making a decision on the marketing strategies to be used if a customer calls the company wishing to cancel the subscription.

Madden and Savage [38] document a churn model, for retaining customers in the Australian ISP (Internet Service Provider) industry. It could be argued that this work is no longer relevant, because it was carried out when ISPs charged a monthly fee for a certain amount of online dial-up time. The development of broadband and flat-rate dial-up offers could see this early work as obsolete. However, it is still of interest to examine the techniques used for identifying customer churn and the variables used. The variables were grouped into four categories, economic variables, usage variables, ISP choice variables and demographic variables. It is stated in the research that the demographic variables are the ones that had been found important in other research. As stated by Madden and Savage [38] “binominal probit model was then used to relate the probability of a subscriber leaving their current ISP”.

Support vector machine (SVM) is another technology that is worth investigating for its suitability for use with customer churn management; however, a thorough search of the literature has revealed little investigation into this method. The closest research to CRM identified by the authors is the mining of customer credit scores [39]. This research used regression trees, however, SVM was also investigated. The paper states that an SVM approach has emerged as a new and promising approach for the process of data classification. The technology is described as methodical and inspired by statistical learning theory. The experiments performed in the research took advantage of

MATLAB's SVM toolbox. The experiments showed similar results to neural networks. It is documented that on a standard PC, it took roughly 20 h to analyse a dataset of 8000, which is a major drawback of the technology [39].

Naïve Bayes is another classification technique which has been over-looked for its potential suitability for customer churn management. This is surprising since the Naïve Bayes is a popular technique for text categorisation. The Naïve Bayes consists of a training dataset and a set of possible features. For each category a few features are chosen which contain the largest mutual information with respect to that category. All the features are extracted from new information and estimated for every category to determine the probability of which category the new information belongs to. Drawbacks of the Naïve Bayes approach are the first approximation only considers the features that are present in the new data a second approximation assumes that the presence of features is independent (consequently is where the Naïve in Naïve Bayes comes in) [40].

The final technology that should be mentioned but has received very little attention within the CRM research community is KNN. A KNN is another classification technique that is widely used within the text categorisation community. To classify unknown information, the KNN ranks the training information using class labels of k most similar neighbours to predict the class of the information [41].

6.2. Soft computing

Soft computing is a consortium of methodologies (such as fuzzy logic, neural networks, and genetic algorithms) that work synergistically and provides, in one form or another, flexible information processing capabilities for handling real-life problems. Exploiting the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human-like decision making is the aim of soft computing [42]. Technologies that fall into the category of soft computing are evolutionary computation (EC), artificial neural networks (ANN), fuzzy logic (FL), probabilistic computing and their combinations, for example, neuro-fuzzy systems.

Artificial neural networks have been successfully used to estimate intricate non-linear functions. A neural network is an analogous data processing structure that possesses the ability to learn. The concept is loosely based on a biological brain and has successfully been applied to many types of problems, such as classification, control, and prediction [43]. Neural networks are different from decision trees and other classification techniques because they can provide a prediction with its likelihood. Various neural network approaches have emerged over time, each with varying advantages and disadvantages [44], however, greater detail into these variances is beyond the scope of this paper. Research suggests that neural networks outperform decision trees and logit regression for churn prediction [27].

According to Rygielski et al. [23] neural networks provide a more powerful and accurate means of prediction, however, there is a potential risk of finding sub-optimal solutions, and over fitting when compared with a decision tree. Another important factor to be aware of when considering the use of neural networks is that they do not uncover patterns in an easily understandable form [27].

Many of the technologies mentioned in this section are examples of data mining techniques. The most popular being decision trees, primarily because they uncover classification rules for classifying records in the form of *if-then* rules, which are easy to understand. Soft computing techniques do not provide this easy to understand classification, and do not clearly convey the underlying patterns in an easily understandable form. Due to these limitations Au et al. [27] proposed a new algorithm called 'data mining by evolutionary learning' (DMEL). DMEL uses non-random initial population based on first order rules. Higher order rules are then obtained iteratively using a GA type process. When identifying interesting rules, DMEL uses a measurement of the object, which does not require user interaction. The fitness value of a chromosome uses a function that defines the probability that the attribute value is correctly determined using the rules it encodes. The likelihood of prediction is estimated and the algorithm handles missing values. DMEL is used to predict churn in the telecommunications industry.

Baesens et al. [25] attempt to estimate whether a new customer will increase or decrease future spending. The paper recognises this problem as a classification task and proposes a Bayesian network for the prediction. A Bayesian network is defined as probabilistic white box, which represents a joint probability distribution over a set of discrete stochastic variables.

Investigations by Au et al. [27] have discovered that there are two GA approaches for rule discovery. These approaches are known as the Pittsburgh approach and the Michigan approach. The difference between these approaches is the Michigan approach represents a rule set by the entire population, while the Pittsburgh approach represents a rule set

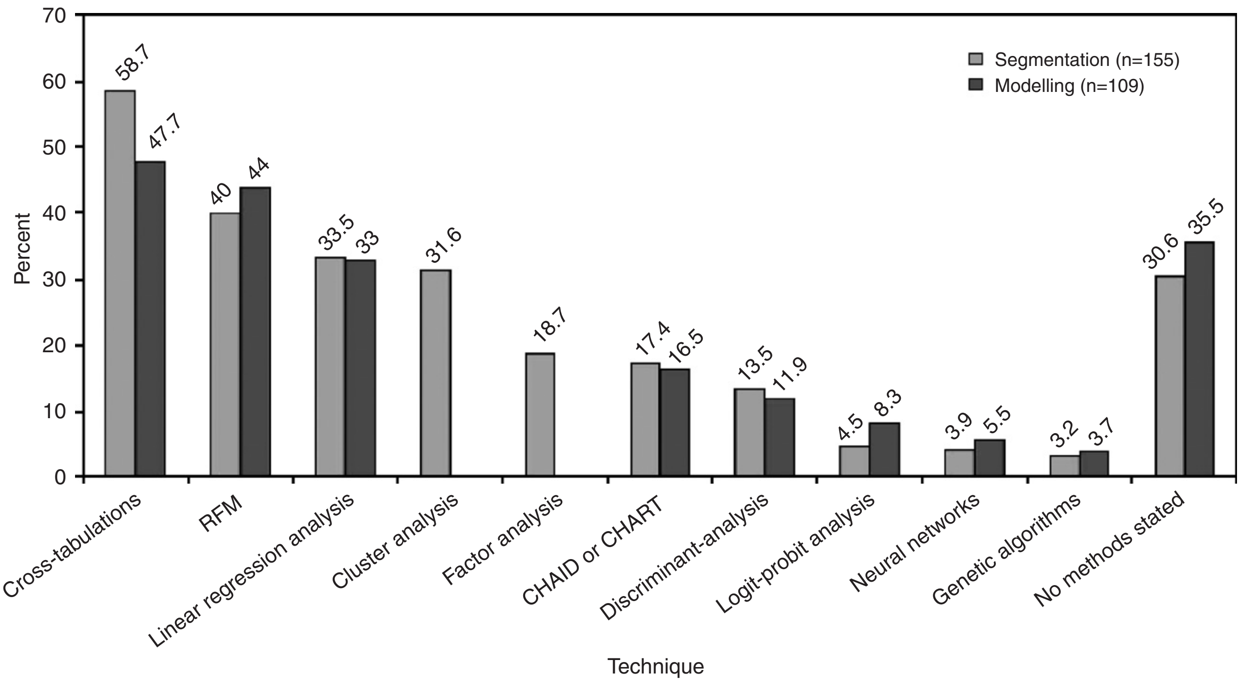

Fig. 4. Predictive modelling and segmentation techniques [10].

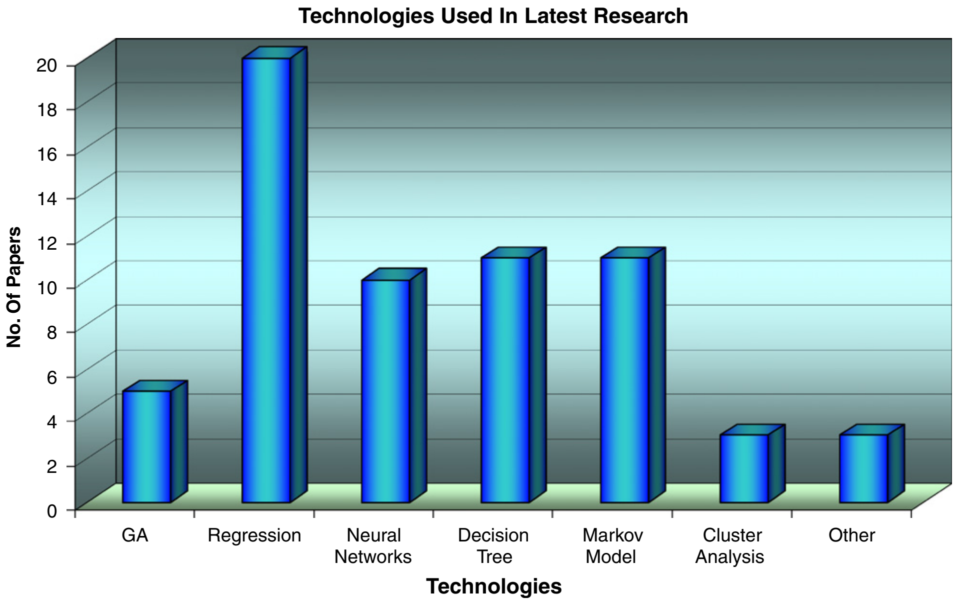

Fig. 5. An analysis of latest research papers according to technologies used.

by an individual chromosome. The research then states, “although GA based rule discovery can produce accurate predictive models, they cannot determine the likelihood associated with their predictions, preventing these techniques from being applicable to the task of predicting churn”.

Rygielski et al. [23] discuss neural networks as data mining technique for CRM. According to his work, neural networks provide a more powerful and predictive model than other techniques. They are also documented to be

Table 1
Reference of papers and the technologies

Reference	Genetic algorithm	Regression	Neural networks	Decision tree	Markov model	Cluster analysis	Other
Au et al. [27]			x	x			
Jenamani et al. [8]					x		
Hsieh [11]			x				
Hwang and Euiiho Suh [47]		x	x	x			
Kitayama et al. [29]				x			
Datta et al. [6]	x	x	x	x			
Baesens et al. [25]		x			x		
Bloemer et al. [28]		x					
Auh and Johnson [35]		x					
Ng and Liu [7]				x			
Mihelis et al. [32]		x					
Rust and Zahorik [33]		x					
Rygielski et al. [23]		x					
Kim and Yoon [2]		x					
Chiang et al. [48]							x
Liu and Shih [12]						x	
Liu and Shih [3]						x	
Prinzie and Van Den Poel [34]					x		
Jonker et al. [13]	x				x		
Viaene et al. [49]		x					
Sun et al. [19]	x						
Verhoef and Donkers [10]		x					
Van Den Poel [9]		x					
Mozer et al. [30]		x	x	x			
Morgan [50]						x	
Boone and Roehm [51]			x				
Shin and Sohn[46]				x			
Vellido et al. [45]			x				
Ranaweera and Neeley [52]		x					
Madden and Savage [38]		x					
Verhoef et al. [53]		x	x				
Berne et al. [54]							x
Kim et al. [1]							x
Kavzoglu and Mather [22]	x		x				
Meyer-Base and Watzel [21]	x		x				
Wei and Chiu [55]				x			
Liao et al. [44]		x					
Athanassopoulos [56]		x					
Slotnick and Sobel [57]					x		
Siskos et al. [58]		x					
Van Den Poel and Lariviere [59]				x			
Yan et al. [20]		x					
Jenamani et al. [8]					x		
Baesens et al. [25]				x			
Wei et al. [60]					x		
Ho Ha et al. [61]				x			
Chen and Rosenthal [62]					x		
Avrachenkov and Sanchez [63]					x		
Jonker et al. [13]					x		
Fleming [64]					x		

applicable to a wider area of applications. However, other disadvantages should be taken into account, like clarity of output, implementation and construction of model.

Boone and Roehm [51] have applied neural networks to segmentation of customer databases in the retail service sector. Vellido et al. [45] use neural networks to segment the online shopping market; more specifically they use a self organising map (SOM) which is an unsupervised neural network. A SOM has also been analysed by Shin and Sohn [46] for segmenting stock trading customers according to potential value. An investigation by Datta et al. [6] revealed that neural networks were only being used by a few companies. They state that a possible reason for this could be lack of clarity of output.

6.3. Overview of techniques for predictive modelling

This section provides a discussion on the various predictive modelling techniques and provides an overview of the research gap in this area. Much work has been done in the area of CRM and predicting customer behaviour, although it appears that to date customer churn management has not received the attention that it requires. Fig. 4 shows a chart of the known predictive and segmentation methods, compiled by Verhoef and Donkers [10]. It can be observed from Fig. 4 that cross-tabulation has been the most popular segmentation and predictive modelling technique in the past. The next column shows RFM as segmentation and modelling technique. The authors argue that RFM defines the variables for use with segmentation and modelling techniques rather than a technology of its. Methods such as weighting RFM variables have been reported in the literature [12], and we agree that weighting the variables could provide better results. It is assumed that this is the type of modelling classified as RFM by Verhoef and Donkers [10]. Linear regression is another popular technique, for performing both prediction and modelling, while cluster analysis and factor analysis have been shown for specifically segmentation. Interestingly, according to Verhoef and Donkers [10], the three main technologies for prediction, CART, logit analysis and neural networks have been least used methods. However, the figure shows that CART has been preferred over neural networks and logit regression.

Based on a review of current state-of-the-art literature, the authors compiled Fig. 5. There are clear differences between Figs. 4 and 5. The research conducted by the authors has focused mainly on publications during the last five years. Therefore, it updates the finding of Verhoef and Donkers [10]. Cross-tabulation is an older modelling technique, which has been overtaken by more advanced and accurate methods. Furthermore, the differences between Figs. 4 and 5 are also because the focus of Fig. 4 is on segmentation and predictive techniques in general, and the focus of Fig. 5 is on technologies for predicting customer behaviour. One of the major differences is that regression analysis has taken the lead for the preferred method, followed by decision trees and Markov models, and then neural networks. This means that four of the least used methods reported by Fig. 4 have become four of the most popular as reported by Fig. 5. Following the charts is Table 1 which offers a reference to which research papers are connected with which technologies.

7. Validation of results

There are several methods documented for validating a customer churn model. Some popular methods are discussed below:

- Cross-fold validation: Hwang and Euhiio Suh [47] performed validation by creating a 70/30 divide of the data. The 70% divide created the training set, and the 30% divide created the validation set. Cross-fold validation is based on the principle of using the available for both training and validation. Several cross-validation methods have been proposed in the literature, two examples follow [65]:
  - V-fold cross validation—the learning set is randomly partitioned into limited datasets of equal size. Each set is then used as a validation set.
  - Monte Carlo cross validation—the learning set is repeatedly divided into two random sets for training and validation.
- Cross-fold validation is most suitable in those cases in which there is a scarcity of data.
- Using a separate validation dataset: Datta et al. [6] validated their model by comparing their results against a simple regression model, decision trees and KNN classifiers. A validation dataset, containing 17,000 records, was used for validation. Bloemer et al. [28] again used a set of data that was not used for training or testing to validate their model. Prinzie and Van Den Poel [34] also used a validation dataset to validate their work. This method of validation is more suitable in those cases in which data availability is not an issue.

8. Discussion and future research directions

Despite the apparent benefits of using feature selection for data mining purposes, this step appears to have been neglected by much of the previous research performed in the area of customer churn management. Most efforts in the literature have largely ignored the feature selection problem and have focused mainly on developing feature extraction methods. Feature selection techniques offer three major advantages. First there might be a significant improvement over the performance of a classification by reducing the number of bands to a smaller set of informative features. Second, with a smaller number of bands, the processing time is greatly reduced. Third, in certain cases lower-dimensional datasets would be more appropriate, where a limited amount of training data is available (relevant for neural networks).

The main trends for developing models to predict customer behaviour include regression analysis, neural networks, decision trees and Markov models. Fig. 5 portrays similar findings for these technologies. Decision trees appear to be the most popular choice and neural networks have received slightly more interest than regression analysis. Focusing specifically on churn prediction, it can be observed that neural networks, decision trees, genetic algorithms, statistical methods and some new techniques have already been investigated. No work has been carried out specifically for the purpose of customer churn management, using certain other powerful predictive techniques such as Bayesian neural networks. The amount of work that has been done using GAs is very small, which suggests that research into these technologies has not yet been exhausted. Fuzzy logic is one other area of soft computing, that has never been investigated in either churn management or CRM research. Due to the fuzziness of the problem to be solved fuzzy logic could prove to be a valuable technique for identifying customer churn. More research is required to fully exploit the potential of fuzzy logic in this area.

Five areas that require further research have been identified from the literature.

The first area involves exploring RFM variable types for predicting customer churn. RFM variables are typically used for predicting purchase behaviour, but the authors believe it could be beneficial to use billing data from company databases in a similar way, for predicting a loss of customer.

The second area concerns data semantics. Only a small amount of the literature even mentions data semantics; however, it is argued that a customer churn management framework cannot be confidently developed without the knowledge and clear understanding of the data. The authors propose the development of a framework that could be followed by researchers and developers to achieve competence at the data semantics stage of development.

The third area concerns feature selection. It is argued that the data to be extracted from large data warehouses containing would be more accurately analysed by a feature selection method. The current literature suggests that previous work done for predicting customer churn has either neglected a feature selection phase, or failed to document one. Different variables hold different predictive powers for solving different problems. A powerful feature selection method, based around neural networks, is proposed to ensure that the best variables are extracted from the data warehouse for predicting customer churn. A common future direction suggested by churn management researchers is investigating the best variables for prediction. It is assumed that most of the presented research would have benefited from the inclusion of a stage that could identify the best variables for predicting customer churn. The literature has advised of two different GA approaches for rule discovery, the Michigan approach and the Pittsburgh approach. These approaches where briefly discussed in Section 6.2, soft computing techniques. Further investigations into GAs have to be carried out in order to establish their suitability to churn prediction.

The fourth area is the actual predictive model. To date, decision trees have been used because of their usefulness with prediction and classification. Neural networks could be potentially very useful for predicting customer churn. It is argued that the previous lack of interest in neural networks has come about because of the fact that classification rules are not output in an easily understandable form. Neural networks, combined with a powerful rule discovery method (GA) as proposed in the previous paragraph, could provide a customer churn prediction model with very powerful predictive capabilities. There is also a need to develop a framework that includes an analysis of the causes of churn.

Finally, the fifth area is the inclusion of a model that distinguishes between loyal and non-loyal customers. Various researchers have documented the current limitations of a churn management solution. Therefore, there is a need to develop a framework for calculating ‘loyalty index’.

It has become apparent from this research that customer retention efforts have been explored in many different ways. It is the purpose of this research to take the current churn management efforts to the next level. The authors believe that addressing the research gaps outlined above will create a powerful, accurate and robust churn identification system that will benefit and advance current churn management research. The future customer churn management systems will exploit the synergy between soft computing based modelling techniques (such as neural networks) and rule discovery methods (such as genetic algorithms) to provide an accurate churn prediction model that is based on the identification of the best variables extracted from the data warehouse for predicting customer churn. In addition, the future systems will provide a framework for calculating 'loyalty index'. They will also include an analysis of the causes of churn. This will provide the companies with a real time measure of customer loyalty that could be used for observing customer behaviour and predicting churn. This will help the companies to make right interactions at the right time for retaining the existing customers.

References

[1] Kim M, Park M, Jeong D. The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunications services. Telecommunications Policy 2004;28:145–59.

[2] Kim H, Yoon C. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy 2004;28:751–65.

[3] Liu D, Shih Y. Integrating AHP and data mining for product recommendation based on customer lifetime value. Information and Management 2004;42:387–400.

[4] Canning G. Do a value analysis of your customer base. Industrial Marketing Management 1982;11:89–93.

[5] Chen IJ, Popovich K. Understanding customer relationship management (CRM). Business Process Management Journal 2003;9:672–88.

[6] Datta P, Masand B, Mani DR, Li B. Automated cellular modeling and prediction on a large scale. Issues on the Application of Data Mining 2001; 485–502.

[7] Ng K, Liu H. Customer retention via data mining. Issues on the Application of Data Mining 2001; 569–90.

[8] Jenamani M, Mohapatra PKJ, Ghose S. A stochastic model of e-customer behaviour. Electronic Commerce Research and Applications 2003;2: 81–94.

[9] Van Den Poel D. Predicting mail-order repeat buying: which variables matter? Tijdschrift voor Economie and Management 2003;48(3):371–403.

[10] Verhoef PC, Donkers B. Predicting customer potential value an application in the insurance industry. Decision Support Systems 2001;32: 189–99.

[11] Hsieh N. An integrated data mining and behavioural scoring model for analysing bank customers. Expert systems with Applications 2004;27: 623–33.

[12] Liu D, Shih Y. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. The Journal of Systems and Software 2004;77:181–91.

[13] Jonker J, Piersma N, Van Den Poel D. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. Expert Systems with Applications 2004;27:159–68.

[14] Ram S. Intelligent database design using the unifying semantic model. Information and Management 1995;29:191–206.

[15] Volz R, Handschuh S, Staab S, Stojanovic L, Stojanovic N. Unveiling the hidden bride: deep annotation for mapping and migrating legacy data to the semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 2003;1:187–206.

[16] Ram S, Khatri V. A comprehensive framework for modeling set based business rules during conceptual database design. Information Systems 2003;30:89–118.

[17] Lee H. Semantics of recursive relationships in entity-relationship model. Information and Software Technology 1999;41:877–86.

[18] Apenyo K. Using the entity-relationship model to teach the relational model. SIGCSE Bulletin 1999;31:78–80.

[19] Sun Z, Bebis G, Miller R. Object detection using feature subset selection. Pattern Recognition 2004;37:2165–76.

[20] Yan L, Wolniewicz R, Dodier R. Predicting customer behaviour in telecommunications. IEE Intelligent Systems 2004;19:50–8.

[21] Meyer-Base A, Watzel R. Transformation radial basis neural network for relevant feature selection. Pattern Recognition Letters 1998;19: 1301–6.

[22] Kavzoglu T, Mather PM. The role of feature selection in artificial neural network applications. International Journal of Remote Sensing 2001;23:2919–37.

[23] Rygielski C, Wang J, Yen DC. Data mining techniques for customer relationship management. Technology in Society 2002;24:483–502.

[24] Crespo F, Weber R. A methodology for dynamic data mining based on fuzzy clustering. Fuzzy Sets and Systems 2004;150:267–84.

[25] Baesens B, Verstraeten G, Van Den Poel D, Egmont-Peterson M, Van Kenhove P, Vanthienen J. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. European Journal of Operational Research 2004;156:508–23.

[26] Muata K, Bryson O. Evaluation of decision trees: a multi criteria approach. Computers and Operational Research 2004;31:1933–45.

[27] Au W, Chan CC, Yao X. A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation 2003;7:532–45.

[28] Bloemer J, Brijis T, Vanhoof K, Swinnen G. Comparing complete and partial classification for identifying customers at risk. International Journal of Research in Marketing 2002;20:117–31.

[29] Kitayama M, Matsubara R, Izui Y. In: Power engineering society winter meeting, vol. 1. 2002; p. 632–4.

[30] Mozer MC, Wolniewicz R, Grimes DB, Johnson E, Kaushansky H. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. 2000;11:690–6.

[31] Liao S, Chen Y. Mining customer knowledge for electronic catalog marketing. Expert Systems with Applications 2004;27:521–32.

[32] Mihelis G, Grigoroudis E, Siskos Y, Politis Y, Malandrakis Y. Customer satisfaction measurement in the private bank sector. European Journal of Operational Research 2001;130:347–60.

[33] Rust RT, Zahorik AJ. Customer satisfaction, customer retention, and market share. Journal of Retailing 1993;69:193–215.

[34] Prinzie A, Van Den Poel D. Investigating purchasing-sequence patterns for financial services using Markov, MTG and MTGg models. European Journal of Operational Research 2004;170:710–34.

[35] Auh S, Johnson MD. Compatibility effects in evaluations of satisfaction and loyalty. Journal of Economic Psychology 2004;26:35–57.

[36] Chiang D, Wang Y, Lee S, Lin C. Goal-oriented sequential pattern for network banking and churn analysis. Expert Systems with Applications 2003;25:293–302.

[37] Tao YH, Yeh CR. Simple database marketing tools in customer analysis and retention. International Journal of Information Management 2003;23:291–301.

[38] Madden G, Savage G. Subscriber churn in the Australian ISP market. Information Economics and Policy 1999;11:195–207.

[39] Lee T, Chiu C, Chou Y, Lu C. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. Journal of Computational Statistics and Data Analysis 2004;50(4):1113–30.

[40] Keren D. Recognizing image "style" and activities in video using local features and naive bayes. Pattern Recognition Letters 2003;24: 2913–22.

[41] Tan S. An effective refinement strategy for KNN text classifiers. Expert Systems with Applications; in press, Uncorrected Proof.

[42] Pal S, Ghosh A. Soft computing data mining. Information Sciences 2004;163:1–3.

[43] Behara RS, Fisher WW, Lemmink J. Modelling and evaluating service quality measurement using neural networks. International Journal of Operations and Production Management 2002;22:1162–85.

[44] Liao Y, Fang S, Nuttle H. A neural network model with bound weights and for pattern classification. Computers and Operational Research 2004;31:1411–26.

[45] Vellido A, Lisboa PJG, Meehan K. Segmentation of the online shopping market using neural networks. Expert Systems with Applications 1999;17:303–14.

[46] Shin HW, Sohn SY. Segmentation of stock trading customers according to potential value. Expert Systems with Applications 2004;27:27–33.

[47] Hwang H, Euiiho Suh T. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunications industry. Expert Systems with Applications 2004;26:181–8.

[48] Chiang D, Wang Y, Lee S, Lin C. Goal orientated sequential pattern for network banking churn analysis. Expert Systems with Applications 2003;25:293–302.

[49] Viaene S, Baesens B, Gestel V, Suykens JAK, Van den Poel D, Vanthienen J. et al. Knowledge discovery in a direct marketing case using least squares support vector machines. International Journal of Intelligent Systems 2001;16:1023–36.

[50] Morgan FW. Profitability market segmentation: identifying the heavy users of overdraft checking. Journal of Business Research 1978;6: 99–110.

[51] Boone DS, Roehm M. Retail segmentation using artificial neural networks. International Journal of Research in Marketing 2002;19:287–301.

[52] Ranaweera C, Neeley A. Some moderate effects on the service quality-customer retention link. International Journal of Operations and Production Management 2003;23:230–48.

[53] Verhoef PC, Spring PN, Hoekstra JC, Leeflang PSH. The commercial use of segmentation and predictive modeling techniques for database marketing in The Netherlands. Decision Support Systems 2002;34:471–81.

[54] Berne C, Mugica JM, Yague MJ. The effect of variety seeking on customer retention in services. Journal of Retailing and Consumer Services 2001;8:335–45.

[55] Wei C, Chiu I. Turning telecommunications call details to churn prediction: a data mining approach. Expert Systems with Applications 2002;23:103–12.

[56] Athanassopoulos AD. Customer satisfaction cues to support market segmentation and explain switching behaviour. Journal of Business Research 2000;47:191–207.

[57] Slotnick SA, Sobel MJ. Manufacturing lead-time rules: customer retention versus tardiness costs. European Journal of Operational Research 2003;163:825–56.

[58] Siskos Y, Grigoroudis E, Zopounidis C, Saurais O. Measuring customer satisfaction using a collective preference disaggregation model. Journal of Global Optimization 1998;12:175–95.

[59] Van Den Poel D, Lariviere B. Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research 2003;157:196–217.

[60] Wei W, Wang B, Towsley D. Continuous-time hidden Markov models for network performance evaluation. Performance Evaluation 2002;49: 129–46.

[61] Ho Ha S, Min Bae S, Chan Park S. Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. Computers and Industrial Engineering 2002;43:801–20.

[62] Chen Y, Rosenthal RW. Dynamic duopoly with slowly changing customer loyalties. International Journal of Industrial Organization 1996;14: 269–96.

[63] Avrachenkov KE, Sanchez E. Fuzzy Markov chains and decision-making. Fuzzy Optimization and Decision Making 2002;1:143–59.

[64] Fleming KN. Markov models for evaluating risk-informed in-service inspection strategies for nuclear power plant piping systems. Reliability Engineering and System Safety 2003;83:27–45.

[65] Dudoit S, Van Der Laan MJ. Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. Statistical Methodology 2003;2:131–54.