Expert Systems with Applications 38 (2011) 9274–9280

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Product development with data mining techniques: A case on design of digital camera

Jae Kwon Bae [a,*], Jinhwa Kim [b]

[a] Department of Railroad Management Information, Dongyang University, #1 Gyochon-dong, Punggi, Yeongju, Gyeongbuk 750-711, Republic of Korea
[b] School of Business, Sogang University, #1 Sinsu-dong, Mapo-gu, Seoul 121-742, Republic of Korea

ARTICLE INFO

Keywords:
New product development
Data mining based methodology
Association rule
Decision tree based models

ABSTRACT

Many enterprises have been devoting a significant portion of their budget to product development in order to distinguish their products from those of their competitors and to make them better fit the needs and wants of customers. Hence, businesses should develop product designing that could satisfy the customers' requirements since this will increase the enterprise's competitiveness and it is an essential criterion to earning higher loyalties and profits. This paper investigates the following research issues in the development of new digital camera products: (1) What exactly are the customers' "needs" and "wants" for digital camera products? (2) What features is more importance than others? (3) Can product design and planning for product lines/product collection be integrated with the knowledge of customers? (4) How can the rules help us to make a strategy during we design new digital camera? To investigate these research issues, the *Apriori* and *C5.0* algorithms are methodologies of association rules and decision trees for data mining, which is implemented to mine customer's needs. Knowledge extracted from data mining results is illustrated as knowledge patterns and rules on a product map in order to propose possible suggestions and solutions for product design and marketing.

1. Introduction

With the ever-changing information technology and the current consumption patterns change, product life cycle becomes shorter and shorter. Enterprises must master the ever-changing market trends, and create high value business activities continuing to develop of new products designed to enhance the competitiveness of enterprises. To satisfy customers' needs, customer-specific products should be produced. However, the latter increases production costs and the product market price. Manufacturing cost can be reduced by standardizing products to realize the benefits of the economy of scale.

Concurrent engineering is a management procedure for the traditional sequential engineering arising out of the product development loss. The concept which in its product design stage can be considered as thinking the problems may faced before the product life cycle processes, the problem such as manufacturing, assembly, cost and reliability other factors, and then reached the purpose of shortening the design time and reducing development costs. Concurrent engineering is a systematic approach to integrate product development that emphasizes the response to customer expectations. It embodies team values of co-operation, trust and sharing in such a manner that decision making is by consensus, involving all perspectives, from the beginning of the product life cycle. Accordingly, the entire product life cycle related activities can all be fully taken into account early in product development, not only to reduce development costs and shorten the time to market but also to increase product and process quality, lower costs and enhance the competitiveness of the new product. At present, the development and research of concurrent engineering in many areas of integration have many good results; for example, with design for manufacturing, with design for assembly, with design for reliability, with design for quality, with design for cost and so on (Boothroyd, Knight, & Dewhurst, 2001; Parsaei & Sullivan, 1993). However, with the design for customer on the integration of the design, there is not much written.

A new product development cannot only be pursuant to the business of the design and manufacturing capability one also has to consider the customer's needs and preferences and translate then into the design map. Cooper and Kleinschmidt (1993) also pointed out that with customer-oriented enterprises, when developing new products, one must be fully aware of the needs of customers, market competition and the nature of the market as these are critical success factor to new any product. The model of product development driven by sales has been gradually replaced by the customer and market orientation. If an enterprise can exactly understand what the customer wants, preferences and buying

* Corresponding author. Tel.: +82 54 630 1298; fax: +82 54 630 1275.
E-mail address: jkbae@dyu.ac.kr (J.K. Bae).

0957-4174/$ - see front matter Crown Copyright © 2011 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.01.030

behavior will provide clues to the development of new products. This study applies association rule and decision tree techniques to analyze customer preferences portfolio information and make a new product to customers. This will bring fast and accurate feed back to the product designers; the enterprises can make a quick response for short-lived product life cycle, and grasp the real needs of customers.

This paper investigates the following research issues in the development of new digital camera products: (1) What exactly are the customers' "needs" and "wants" for digital camera products? (2) What features is more importance than others? (3) Can product design and planning for product lines/product collection be integrated with the knowledge of customers? (4) How can the rules help us to make a strategy during we design new digital camera? To investigate these research issues, the *Apriori* and *C5.0* algorithms are methodologies of association rules and decision trees for data mining, which is implemented to mine customer's needs. Knowledge extracted from data mining results is illustrated as knowledge patterns and rules on a product map in order to propose possible suggestions and solutions for product design and marketing.

The remainder of this paper is structured as follows. Section 2 presents a research background review focused on the new product development using data mining techniques. Section 3 presents a research framework and analysis procedure. Section 4 presents data preparation and analysis. Some experimental results are presented and analyzed in Section 5, and finally our concluding remarks are provided in Section 6.

2. Literature review

2.1. New product development using data mining techniques

Before a product is designed, most companies perform marketing studies. The goal of these studies is to understand the customers' expectations. Online analytical processing tools are used to extract relevant customer information from multi-dimensional database. Classical statistical tools are used to compute various models (e.g. regression models) and parameters (e.g. mean, confidence intervals) based on the collected data. Hypotheses can be validated in support of decision-making (Agard & Kusiak, 2004). The goal of the research discussed in this paper is to extract unknown information and knowledge from databases rather than validate a hypothesis.

Anand and Buchner (1998) defined data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets. They classified data mining tasks as predictive and descriptive. Predictive tasks are those that produce models that can be used for classification. Descriptive tasks produce understandable and useful patterns and relationships describing a complex data set. Westphal and Blaxton (1998) identified four functions of data mining: classification, estimation, segmentation, and description. Classification involves assigning labels to previously unseen data records based on the knowledge extracted from historical data. Estimation is the task of filling in missing values in the fields of an incoming record as a function of fields in other records. Segmentation (called also clustering) divides a population into smaller subpopulations with similar behavior. Clustering methods maximize homogeneity within a group and maximize heterogeneity between the groups. The description task focuses on explaining the relationships among the data. Moreover, Fayyad (1996) defined a data mining process for the extraction of knowledge from a data set. Several steps are considered with frequent iterations aimed at the extraction of valuable knowledge. They begin with the development of an understanding of the application domain, the relevant prior knowledge and the goals of the end user. The next steps deal with the creation and preparation of the data to be mined (selection, cleaning, preprocessing, reduction, and projection of the data). Then, the most suitable data mining algorithm is selected to search for patterns in a particular representation form or a set of such representations. Knowledge is then extracted, interpreted, and validated.

This section outlines methodology for the application of data mining in new product development as shown in Table 1. Moore, Louviere, and Verma (1999) introduced how one can combine different conjoint analysis studies, each containing a core of common attributes, to help design product platforms that serve as the foundation for multiple derivative products. The illustration is based on actual, but disguised, data from a small company that makes electronic test equipment. Steiner and Hruschka (2003) have proposed the use of genetic algorithms to solve the problem of identifying an optimal single new product using conjoint data set. Tsai, Chang, and Wang (2003) describe the concepts of data mining and their application with product development. This research applied association rule technique to analyze the customer's preference from different product combination of current market. Agard and Kusiak (2004) introduced a methodology for using data mining algorithms in the design of product families. An analysis of the requirements for the product design was performed and association rules extracted. Shahbaz, Srinivas, Harding, and Turner (2006) applied data mining to extract knowledge from a fan blade manufacturer database. This paper examines the application of association rules to extract useful information about a manufacturing system's capabilities and its constraints. The quality of each identified rule is tested and, from numerous rules, only those that are statistically very strong and contain substantial design information are selected. In manufacturing engineering, Jiao, Zhang, Zhang, and Pokharel (2007) introduced a data mining approach for dealing with product and process variety mapping. The mapping relationships are embodied in association rules, which can be deployed to support production planning of product families within exiting production processes. Liao, Hsieh, and Huang (2008) introduced the product map obtained from data mining results, which investigates the relationships among customer demands, product characteristics, and transaction records, using the *Apriori* algorithm as a methodology of association rules for data mining. The product map shows that different knowledge patterns and rules can be extracted from customers to develop new products and possible marketing solutions. Chen (2009) introduced a new approach for problem solving using decision tree induction based on intuitionistic fuzzy sets to develop the problem formulation for the symptoms and causes of the problem based on intuitionistic fuzzy sets. And then provide the approach to find the optimal cause of the problem for the consideration of product design.

2.2. Association rules and decision tree based models

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Extraction of frequent patterns (called association rules mining) leads to data patterns with some predefined level of regularity. Measures such as *support* and *confidence* permit the evaluation of the quality of the extracted rules (patterns). For example, for the association rule $A \Rightarrow B$, *support* expresses the number of times A occurs as a fraction of the total number of examples. *Confidence* is the fraction of the number of times B exists in the data when A is present. An association with high *confidence* and *support* is called strong and could be potentially useful. Association rule mining technique finds all collections of items in a database whose *confidence* and *support* meet or exceed pre-specified threshold values. *Apriori* algorithm is one of the pre-

Table 1
Data mining based methodology for new product development.

Application domain	Authors	Methods
Using conjoint analysis to help design product platforms	Moore, Louviere, and Verma (1999)	Conjoint analysis
Genetic algorithms for product design	Steiner and Hruschka (2003)	Genetic algorithms, Conjoint analysis
Applying association rule and neural networks to product development	Tsai, Chang, and Wang (2003)	Association rules, Neural networks
Data mining based methodology for the design of product families	Agard and Kusiak (2004)	Association rules
Product design and manufacturing process improvement using association rules	Shahbaz, Srinivas, Harding, and Turner (2006)	Association rules
Association rule mining for product and process variety mapping	Jiao, Zhang, Zhang, and Pokharel (2007)	Association rules
Mining product maps for new product development	Liao, Hsieh, and Huang (2008)	Association rules
Product design using decision tree induction	Chen (2009)	Fuzzy decision trees

valent techniques used to find association rules (Agrawal, Imielinski, & Swami, 1993). *Apriori* operates in two phases. In the first phase, all large item sets are generated. This phase utilizes the downward closure property of *support*. The second phase of the algorithm generates rules from the set of all large item sets. Association rule first was applied to the analysis of supermarket shopping basket (market basket analysis) from the customer's transaction records to identify the relationship between goods as a reference of supermarket display of goods and stock purchase.

Decision tree learning is one of the most widely used and practical methods for inductive learning. Rule induction refers to the rules derived from the decision tree techniques in data mining. The data set is separated into many partitions in a way to increase the purity, which is the degree to which the dependent variable belongs to a certain class. The rules that are applied for splitting the data are called the inducted rules. Decision tree is a non-parametric method and suitable for figuring out interaction effect or non-linearity. In many cases, decision tree is used for the sake of interpretation of the analysis results. A decision tree have four types of method such as *CHAID* (Kass, 1980), *CART* (Breiman, Friedman, Olshen, & Stone, 1984), *QUEST* (Loh & Shih, 1997), and *C5.0* (Quinlan, 1993). *CHAID* (Chi-squared automatic interaction detection) method is based on the *chi-square* test of association. A *CHAID* tree is a decision tree that is constructed by repeatedly splitting subsets of the space into two or more child nodes, beginning with the entire data set (Michael & Gordon, 1997). To determine the best split at any node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. This *CHAID* method naturally deals with interactions between the independent variables that are directly available from an examination of the tree. The final nodes identify subgroups defined by different sets of independent variables (Magidson & SPSS Inc., 1993). *CART* (Classification and regression tree) is a recursive partitioning method to be used both for regression and classification. *CART* is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of impurity or diversity measures (Gini, towing, ordered towing, and least-squared deviation). The goal is to produce subsets of the data which are as homogeneous as possible with respect to the target variable (Breiman, Friedman, Olshen, & Stone, 1984). *QUEST* (Quick, unbiased, efficient statistical tree) is a binary-split decision tree algorithm for classification and data mining. *QUEST* can be used with univariate or linear combination splits. A unique feature is that its attribute selection method has negligible bias. If all the attributes are uninformative with respect to the class attribute, then each has approximately the same change of being selected to split a node (Loh & Shih, 1997). *C5.0* (Commercial version 5.0) is a supervised learning classification algorithm used to construct decision trees from the data (Quinlan, 1993). Most empirical learning systems are given a set of pre-classified cases, each described by a vector of attribute values, and construct from them a mapping from attribute values to classes. *C5.0* is one such system that learns decision tree classifiers. It uses a divide-and-conquer approach to growing decision trees. The main difference between *C5.0* and other similar decision tree building algorithms is in the test selection and evaluation process. In this study, we used measure of *entropy index* that used for categorical target variables.

3. Research framework and analysis procedure

In this paper, our research is consisted of three steps see the Fig. 1 as follow: In step 1, pre-test with digital camera manufacturers, vendors, and relevant researchers was conducted before the main test to derive functional attributes that influence the purchase of digital cameras. The questionnaire and interview methods were used together for the pre-test. And then, we design our questionnaire and choose sample make the survey and collect the data. A questionnaire is a data collection method that a respondent support completes in written format. Questionnaire surveys are an important part of marketing and customer relationship management. The types of questions in the questionnaires can be roughly classified into two categories, open-ended and closed-ended questions. Step 2, using *SAS Enterprise Miner 5.3* and *Clementine 9.0* to analyze the data input, and get association rules and decision tree rules. These rules contain useful information will be used for next step. Step 3, using association rules and decision trees we will analyze respondents' demands and needs and extract useful information for new product development.

Fig. 2 displays the analysis procedure used in this study. Step 1 shows sampling stage. The quality of model depends largely on the quality of data collected in many cases. This study uses a simple random sampling method. Exploration in step 2 collects useful data through data exploration. Modification in step 3 transforms the collected data to enhance the performance of the model through processes such as transformation, quantification, and grouping. Modeling in step 4 builds models using data mining techniques of association rules and tree-based models. These models are consolidated to build integrated rules. Integrated rules were extracted from the association rules and *C5.0* algorithm, which is implemented for mining product knowledge from customers. Assessment in step 5 tests the reliability, validity, and usability of the integrated rules for new digital camera development.

4. Data preparation and analysis

In this study, a pre-test with digital camera manufacturers, vendors, and relevant researchers was conducted before the main test to derive functional attributes that influence the purchase of digital cameras. The questionnaire and interview methods were used together for the pre-test. As shown in Table 2, nine functional attributes (i.e., price, size, resolution, functions, colors, weights, LED

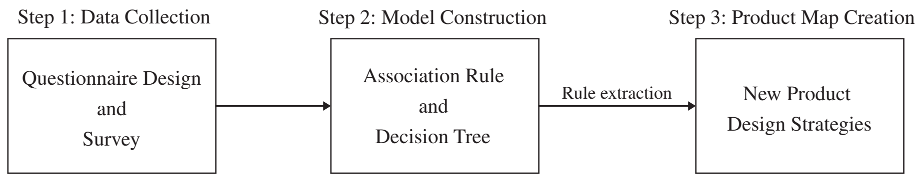

Fig. 1. Research framework.

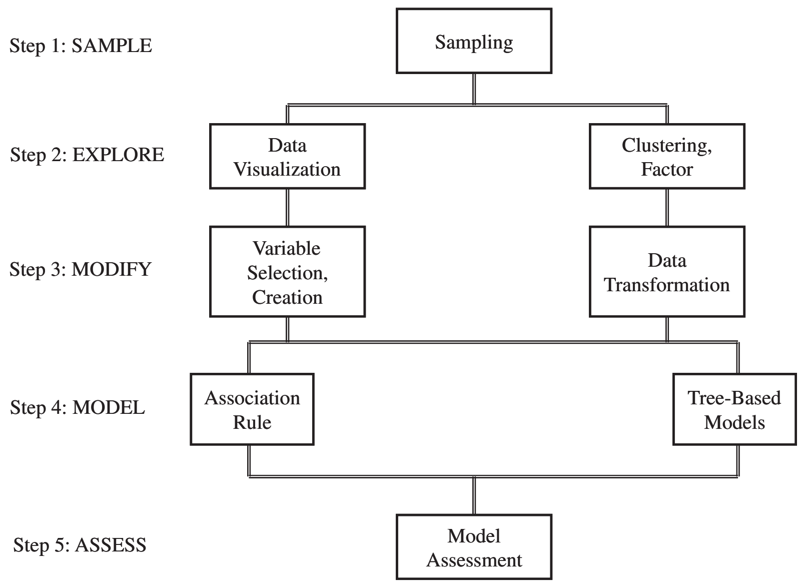

Fig. 2. Analysis procedure.

screen size, battery category, and ease of use) of digital cameras that influence the digital camera purchase were derived from the pre-test.

The second questionnaire survey was performed with the questions that were prepared based on the pre-test results. Data collection was conducted between September and December 2008 at the Business School of Sogang University, Republic of Korea. A total of 350 questionnaires were sent, and 272 completed questionnaires were returned. Excluding incomplete ones, there were 234 valid responses, for a response rate of about 66.86%, and the relational database construction was completed in March 2009.

The population of interest is person who already had or wants to buy a digital camera. The respondents were requested to complete the questionnaire by answering questions regarding two parts: personal information and digital camera features. The first part questionnaires were consisted of questions relating to personal information such as gender, age, job title, marital status, education degree, annual income, characteristic, favorite sports, time of internet using per day, and so on. Of the respondents, 149 (63.7%) were men and 85 (36.3%) were women. Respondents were in their early 20s (151, 64.5%), late 20s (51, 21.8%), and over 30s (32, 13.7%). The most job title of respondents was undergraduate students (201, 85.9%) and full-time MBA students (33, 14.1%). Marital status is consisted of single (203, 86.8%) and marriage (31, 13.2%). Annual income is distributed as $0–$30000 (203, 86.8%), $30001–$60000 (11, 4.7%), over $60000 (20, 8.5%). Characteristic is consisted of introvert (101, 43.2%) and extrovert (133, 56.8%). The detailed information is shown in Table 3.

The second part questionnaires were consider of questions relating to digital camera features, such as price, size, resolution, functions, colors, weights, LED screen size, battery category, and ease of use. In this part the respondents were asked to rank the nine digital camera features according to the importance by the number from 1 to 9. Number 1 stands for most important; number 9 stands for least important. We required respondents choose the most important three features (top 3) and then chose one camera style (see Fig. 3 – the camera photos A, B, and C). Moreover, a group that thinks nothing of the camera style if all functional attributes are satisfactory was defined as "Camera Style D". We try to analyze the relationship between these three features and camera style by the association rule. And then choose one attribute from each of features. As one example, the feature color has six choices: (1) black, (2) gray, (3) argent, (4) red, (5) white, and (6) does not matter. Further, we defined six colors as follows: (1) CO1, (2) CO2, (3) CO3, and so on (see Tables 4 and 5).

We developed the decision table as discussed in Table 5. The decision table includes 234 samples reflecting the respondents' favorite for each digital camera. For each record nine conditional attributes are registered and the style was defined as "Target". Further, we try to get decision tree rule like: What kind of attributes the respondents chose will make them choose relevant style of digital camera? A decision tree can be got by splitting the source data set into subsets based on an attribute-value test. Rows with attribute values above a certain threshold are placed in one partition,

Table 2
Digital camera features from pre-test.

No.	Functional attributes	Abbreviation	No.	Functional attributes	Abbreviation
1	Price	PR	6	Weights	WE
2	Size	SZ	7	LED screen size	LED
3	Resolution	RE	8	Battery category	BAT
4	Functions	FU	9	Ease of use	EOU
5	Colors	CO	–		

Table 3
Descriptive statistic of respondent characteristics.

Measure	Item	Frequency/(%)	Measure	Item	Frequency/(%)
Gender	Male	149 (63.7%)	Annual income	\$0–\$30000	203 (86.8%)
	Female	85 (36.3%)		\$30001–\$60000	11 (4.7%)
Age	20–25	151 (64.5%)		Over \$60000	20 (8.5%)
	26–30	51 (21.8%)	Characteristic	Introvert	101 (43.2%)
	Over 30	32 (13.7%)		Extrovert	133 (56.8%)
Marital status	Single	203 (86.8%)	Usage	Fun	227 (97.0%)
	Marriage	31 (13.2%)		Professional	7 (3.0%)
Education degree	Undergraduate students	201 (85.9%)	Favorite car style	Sedan	154 (65.8%)
	MBA students	33 (14.1%)		SUV	75 (32.1%)
Job title	Student	201 (85.9%)		Van	5 (2.1%)
	Employee	33 (14.1%)			

Fig. 3. Digital camera styles.

Table 4
Association rule data set.

ID	Feature	Abbreviation	ID	Feature	Abbreviation
1	Ease of use	EOU	3	Style C	STC
1	Price	PR	4	Price	PR
1	LED size	LED	4	Size	SZ
1	Style A	STA	4	Ease of use	EOU
2	Resolution	RE	4	Style B	STB
2	Functions	FU	:	:	:
2	Ease of use	EOU			
2	Style B	STB	234	Price	PR
3	Price	PR	234	Colors	CO
3	Functions	FU	234	LED size	LED
3	Weights	WE	234	Style A	STA

Table 5
Decision tree data set.

Feature	Abbreviation (Range)	1	2	3	...	234
Price	PR (1–4)	PR3	PR2	PR2	...	PR1
Size	SZ (1–4)	SZ1	SZ2	SZ4	...	SZ1
Resolution	RE (1–4)	RE4	RE3	RE2	...	RE3
Functions	FU (1–4)	FU1	FU2	FU3	...	FU1
Colors	CO (1–6)	CO3	CO1	CO6	...	CO4
Weights	WE (1–4)	WE1	WE2	WE2	...	WE4
LED size	LED (1–4)	LED4	LED2	LED2	...	LED3
Battery	BAT (1–3)	BAT3	BAT2	BAT2	...	BAT3
Ease of use	EOU (1–3)	EOU1	EOU2	EOU3	...	EOU1
Style	ST (A–D)	STA	STB	STA	...	STD*

Note: STD (Camera Style D) is a group that thinks nothing of the digital camera style (style A, B or C).

and the remaining rows are placed in another. This process is repeated on each derived subset in a recursive manner.

5. Experimental results

Tables 6–8 show the results of association rules and decision trees for new digital camera design. The new product development analysis is developed by choosing various decision variables and integrated consumers' purchasing tendencies to obtain knowledge about customer's product preferences. Using association rule, the analysis *lift* value should be set at greater than 1; while the minimum *support* and *confidence* values initially are set at least 5% and 10%, respectively, and then adjusted accordingly if necessary during the analysis process. Further, the study finds that the optimal depth of *C5.0* is 5 from the analysis of the purity of parent nodes and child nodes in each depth.

As shown in Table 6, we can easily know that style *A* digital camera has *confidence* (62.07%–69.89%) and *support* (7.69%–27.78%). When we design the style *A* digital camera we could consider all the rules as follow: (1) *size* => *style A*, (2) *size* & *price* => *style A*, (3) *size* & *resolution* => *style A*, (4) *weight* => *style A*, (5) *ease of use* & *size* => *style A*, (6) *LED screen size* => *style A*, and (7) *resolution* & *function* => *style A*. From the rules we can see that the *confidences* are very high which stand for that the rules of style *A* have big significance. Further, for style *A* digital camera the size is very important. Small size this key point should be considered first when designing other features.

Basic on this we discuss the rules from decision tree as follow: (1) If color = argent and function = digital zoom or does not matter and weight = less than 125 g or 125–320 g Then style A, (2) If LED screen size = 3.0 in. or 1.8 in. and color = gray or white and function = digital zoom or does not matter and weight = less than 125 g or 125–320 g Then style A, (3) If price = less than \$300 or \$300–\$500 and weight = less than 125 g and color = black Then style A, (4) If function = fuzzy or still image recording and weight = less than 125 g or 125–320 g and color = gray or argent or white Then style A, (5) If price = less than \$300 or does not matter and function = digital zoom and weight = 125–320 g or over 320 g or does not matter and color = black Then style A, (6) If LED screen size = 1.8 in. or 3.0 in. and color = red or does not matter Then style A. Basic on these rules we could design digital camera for customers in detail to satisfy different needs for style A digital camera. Combining the association rules and C5.0 rules we could make a clear detailed digital camera. Basic on the important characteristics small size style A digital camera, some customers focus on the size and resolution and some of them focus on the

Table 6
An example of style A (STA) rules.

AR rules for STA	(*Support*, *Confidence*)	*Lift*	DT rules for STA	(*N*, %)
SZ => STA	(27.78, 69.89)	1.19	IF (Colors = CO3) and (Functions = FU1 or FU4) and (Weights = WE2 or WE1) THEN STA	(43, 93%)
SZ & PR => STA	(15.81, 67.27)	1.15	IF (LED size = LED3 or LED1) and (Colors = CO2 or CO5) and (Functions = FU1 or FU4) and (Weights = WE2 or WE1) THEN STA	(8, 87.5%)
SZ & RE => STA	(11.11, 74.29)	1.27	IF (Price = PR2 or PR1) and (Weights = WE1) and (Colors = CO1) THEN STA	(26, 80.8%)
WE => STA	(10.68, 71.43)	1.22	IF (Functions = FU2 or FU3) and (Weights = WE1 or WE2) and (Colors = CO3 or CO5 or CO2) THEN STA	(15, 80%)
EOU & SZ => STA	(10.26, 66.67)	1.14	IF (Price = PR1 or PR4) and (Functions = FU1) and (Weights = WE2 or WE4 or WE3) and (Colors = CO1) THEN STA	(15, 66.7%)
LED => STA	(7.69, 62.07)	1.06	IF (LED size = LED2 or LED3) and (Colors = CO6 or CO4) THEN STA	(47, 63.8%)
RE & FU => STA	(7.26, 73.91)	1.26	–	

Table 7
An example of style B (STB) rules.

AR rules for STB	(*Support*, *Confidence*)	*Lift*	DT rules for STB	(*N*, %)
RE => STB	(12.82, 28.04)	1.40	IF (Functions = FU2 or FU3 or FU4) and (Weights = WE2 or WE3 or WE4) and (Colors = CO1) THEN STB	(23, 79.3%)
FU => STB	(11.97, 25.45)	1.21	IF (Price = PR3 or PR2) and (Functions = FU1) and (Weights = WE2 or WE4 or WE3) and (Colors = CO1) THEN STB	(16, 62.5%)
EOU => STB	(9.83, 25.27)	1.10	IF (Size = SZ1 or SZ2) and (LED size = LED4 or LED1) and (Colors = CO6 or CO4) THEN STB	(8, 50.0%)
RE & PR => STB	(8.12, 32.20)	1.10	–	

Table 8
An example of styles C and D (STC and STD) rules.

AR rules for STC and STD*	(*Support*, *Confidence*)	*Lift*	DT rules for STC and STD*	(*N*, %)
PR => STC	(7.26, 11.89)	1.07	IF (LED size = LED4 or LED2) and (Colors = CO2 or CO5) and (Functions = FU1 or FU4) and (Weights = WE2 or WE1) THEN STC	(5, 60.0%)
RE => STC	(5.13, 11.21)	1.01	IF (Weights = WE4) and (Colors = CO3 or CO5 or CO2) THEN STC	(9, 50.0%)
EOU => STD*	(5.27, 10.99)	1.51	IF (Size = SZ4) and (LED size = LED4 or LED1) and (Colors = CO6 or CO4) THEN STD*	(6, 66.7%)

Note: STD (Camera Style D) is a group that thinks nothing of the digital camera style (style A, B or C).

design and function some of them just focus on weight of the digital camera. As one example, if the customers focus on the weight of style A digital camera, they will consider what kind of weight of the digital camera they want, at here most of them chosen the weight of less than 125 g or 125–320 g. Accordingly, we follow the C5.0 rules, we suggest the other characteristics combination as follow: (1) Argent color, digital zoom function or no digital zoom function, weight less than 125 g or 125–320 g, (2) LED screen size 1.8 in. or 3.0 in., gray or white color, digital zoom function or no digital zoom function, weight less than 125 g or 125–320 g, (3) Price less than \$300 or \$300–\$500, black color and weight less than 125 g, (4) Fuzzy function or still image recording function, gray or white color, weight less than 125 g or 125–320 g, (5) Price less than \$300 or does not matter, digital zoom function, black color, weight 125–320 g or over 320 g or does not matter. These rules could satisfy most of different customers' different requirements which could help the company to narrow down the target and improve the effectiveness.

As shown in Table 7, we can easily know that style *B* digital camera has *confidence* (25.27%–32.20%) and *support* (8.12%–12.82%). Actually, styles *B* and *C* digital cameras have less significance than style *A*. Even though, when we design the style *B* digital camera we could consider all the rules as follows: (1) *resolution => style B*, (2) *functions => style B*, (3) *ease of use => style B*, and (4) *resolution* & *price => style B*. From these rules we get some useful information that the people chose this kind of style they put more attention on its resolution, functions, and easy of use. They require more about its performance and high technology, most of time this kind of digital cameras are not easy to operate. Therefore, the operation procedures should be simplified when we designing them.

Basic on this we can design the digital camera using the rules from decision tree as follow: (1) If function = fuzzy or still image recording or does not matter and weight = 125–320 g or over 320 g or does not matter and color = black Then style B, (2) If price = \$300–\$500 or over \$500 and function = digital zoom and weight = 125–320 g or over 320 g or does not matter and color = black Then style B, (3) If size = less than $86.8 \times 54.8 \times 22.0$ mm or $86.8 \times 54.8 \times 22.0$–$117 \times 80 \times 78$ mm and LED screen size = 1.8 in. or does not matter and color = red or does not matter Then style B. Accordingly, when the company develops their digital camera, they should improve the resolution as high as possible, increase the style or design, and make the use of the digital camera easier. After that, for other characteristics combination they should follow the decision tree rules to develop as about.

As shown in Table 8, there are two association rules for style *C* digital camera design. Actually, the *confidences* of those two rules are very low than styles *A* and *B*. When we design the style *C* digital camera we could consider all the rules as follow: (1) *price => style C* and (2) *resolution => style C*. For style *C* the customers pay more attention on the price and resolution basic on this information we can careful design the follow detail rules from decision tree: (1) *If LED screen size = 2.5 in. or does not matter and color = gray or white and function = digital fuzzy or does not matter and weight = less than 125 g or 125–320 g Then style C* and (2) *If weight = does not matter and color = argent or white or gray Then style C*. Accordingly, the company should focus on the lower or right price and high level of resolution, and then they should follow the *C5.0* rules to develop as about.

Some customers do not think the type of the style is important for they so they could buy any style of them (style *D* = style *A* or style *B* or style *C*), that sounds not important to the enterprise. However, if the enterprise could consider them stand by the opposition side and put more attention on them, they will get some useful information about their needs to get more competitiveness in the market: (1) *ease of use => style D* and (2) *If size = does not matter and LED screen size = does not matter or 1.8 in. and colors = does not matter or red Then style D*. From the first rule we can see that these people just focus on the convenience of using, so any style of the digital camera with this function it would be commented to the customers. Accordingly, when the enterprise designing the product they could reduce the complexity of the using and by the way they will attract these customers. From the second rule we also can see that if the customers they do not care the size and LED screen size or just choose 1.8 in. and do not care the color or just choose red color, they also do not care the style of the digital camera. These customers could buy any style of the digital camera. Therefore, we do not need to make a special design for these customers.

Since there is a growing trend towards industry globalization, industry leaders have to focus on diverse product designing and effective marketing strategies to meet the specific customers' needs and preferences, and to therefore, reach the enterprise's goals. Therefore, this study suggests that the digital camera industry should pay more attention, not only on experiments, but also to consumer factors for new product development. It suggests that the digital camera industry should extract customer information from the demand and use it as resource on its designing. Because more sophisticated information on digital camera product production and design are necessary factors and variables for consideration on data analysis. In addition, these actions not only offer better capabilities for understanding its market, but also enhance its manufacturing and product innovation capabilities by extending its product lines.

6. Conclusions

Customer needs and wants are sensitive and complex. If a firm can understand them and make efforts to fulfill customer demands and provide friendly service, then customers will be more supportive and loyal to the enterprise. During the process of development from the product concept to the actual product, the customer can only passively receive new information, and can only select from the products that are currently on sale in the market. No matter which type of product, the customer cannot individually come up with a product concept and then develop it.

Furthermore, buying what is available on the market does not mean that customers are satisfied with the current product, because the customer's experiences and preferences were not considered in developing the product so they can only accept the product as it is. As a result, a business should develop products that fulfill the customer's needs and wants, since this will increase the enterprise's competitiveness and it is an essential criterion to earning higher loyalties and profits.

Data mining technology can make dramatic changes to business practice which gathering of information about the customer to analyze and integrate, provide the engine to realize the knowledge of customer's requirements. In this paper, mining customer information for new product development is an example of implementing a data mining approach for analyzing and providing decision supports. Data mining techniques should be implemented using the on data mining process in order to enhance data analysis capabilities for classification, clustering, and prediction analysis.

In this study, the functional attributes of digital cameras that influence the digital camera purchase were found and emphasized to increase the digital camera repurchase rate and to present a product sales strategy for digital camera manufacturers and relevant researchers. This paper suggested that integrated rules were extracted from the association rules and *C5.0* algorithm, which is implemented for mining product knowledge from customers. Knowledge extraction from data mining results is shown as rules in order to propose suggestions and solutions for new product development and possible marketing solutions.

Despite the many findings from this study, it has some limitations. Firstly, the results from the study should be generalized. It would be better to investigate more products in order to generalize the results of this study. Secondly, the sample of this study is mainly on undergraduate and full-time MBA students. Actually, we should extend the range and amount of the sample to get more data if that we can get more detail information, the rules will be better than now. The result of the study also should be analyzed with the engineers to research the feasibility of the rules. For future study, we could focus on combine the rules and the customers in details and use other data mining techniques such as neural networks, genetic algorithms, and support vector machines by analyzing past years data to predict future new product design direction.

References

Agard, B., & Kusiak, A. (2004). Data-mining-based methodology for the design of product families. *International Journal of Production Research, 42*(15), 2955–2969.

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on management of data* (pp. 207–216).

Anand, S. S., & Buchner, A. G. (1998). *Decision support using data mining*. London: Financial Times Pitman.

Boothroyd, G., Knight, W., & Dewhurst, P. (2001). *Product design for manufacture and assembly* (2nd ed.). CRC Publisher.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman and Hall.

Chen, R. Y. (2009). A problem-solving approach to product design using decision tree induction based on intuitionistic fuzzy. *European Journal of Operational Research, 196*(1), 266–272.

Cooper, R. G., & Kleinschmidt, E. J. (1993). Uncovering the keys to new product success. *IEEE Engineering Management Review, 21*(4), 5–18.

Fayyad, U. M. (1996). Data mining and knowledge discovery: Making sense out of data. *IEEE Expert: Intelligent Systems and their Applications, 11*(5), 20–25.

Jiao, J., Zhang, L., Zhang, Y., & Pokharel, S. (2007). Association rule mining for product and process variety mapping. *International Journal of Computer Integrated Manufacturing, 21*(1), 111–124.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.

Liao, S.-H., Hsieh, C.-L., & Huang, S.-P. (2008). Mining product maps for new product development. *Expert Systems with Applications, 34*(1), 50–62.

Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815–840.

Magidson, J.SPSS Inc.. (1993). *SPSS for windows CHAID release 6.0*. Chicago: SPSS Inc..

Michael, J. A., & Gordon, S. L. (1997). *Data mining technique: For marketing, sales and customer support*. Wiley.

Moore, W. L., Louviere, J. J., & Verma, R. (1999). Using conjoint analysis to help design product platforms. *Journal of Production and Innovation Management, 16*(1), 27–39.

Parsaei, H. R., & Sullivan, W. G. (1993). *Concurrent engineering: Contemporary issues and modern design tools (Design and manufacturing)* (1st ed.). Springer Publisher.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA: Morgan Kaufmann.

Shahbaz, M., Srinivas, M., Harding, J. A. & Turner, M. (2006). Product design and manufacturing process improvement using association rules. In *Proceedings of the institution of mechanical engineers. Part B: Journal of engineering manufacture* (pp. 243–254).

Steiner, W., & Hruschka, H. (2003). Genetic algorithms for product design: How well do they really work? *International Journal of Market Research, 45*(2), 229–240.

Tsai, C.-Y., Chang, P.-C., & Wang, S.-J. (2003). Applying association-rule techniques and artificial neural network to product development. *Journal of the Chinese Institute of Industrial Engineers, 20*(2), 101–112.

Westphal, C., & Blaxton, T. (1998). *Data mining solutions*. New York: Wiley.