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-