A decision-making framework for precision marketing

Zhen You [a], Yain-Whar Si [b], Defu Zhang [a,*], XiangXiang Zeng [a], Stephen C.H. Leung [c], Tao Li [a,d]

[a] *Department of Computer Science, Xiamen University, Xiamen 361005, China*
[b] *Department of Computer and Information Science, University of Macau, Macau, China*
[c] *Department of Management Sciences, City University of Hong Kong, Hong Kong*
[d] *School of Computer Science, Florida International University, Miami, FL, USA*

ARTICLE INFO

ABSTRACT

Precision marketing offers personalized customer service and is used to help enterprises increase their profits by means of high-efficiency marketing. This paper presents a novel decision-making framework for precision marking using data-mining techniques. First, this study presents a trend model to accurately predict monthly supply quantity; second, it uses a RFM (Recency, Frequency and Monetary) model to select attributes to cluster customers into different groups; third, it uses CHAID decision trees and Pareto values to identify important attribute values to distinguish different customer groups; and finally, it creates different supply strategies targeting each customer group. The objective of the proposed precision-making framework is to help managers identify the potential characteristics of different customer categories and put forward appropriate precision marketing strategies, which can greatly reduce inventory for every customer category. The real-world data from a company in China were collected and used in a case study to illustrate how to implement the proposed framework. This case study demonstrates that our proposed decision-making framework is efficient and capable of providing a very good precision marketing strategy for enterprises.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Due to the accelerated pace of economic globalization and increasing market competition, economic pressures and competition have led enterprise managers to face the problem of choosing the right strategic decision-making policies for selling the right products to the right customers at the right time, such that the companies can increase their profits. Recently, it has been recognized that precision marketing has become a key means of generating profit and is becoming increasingly important as customers become better informed about the products and their rights as consumers. The availability of customer data and transaction records provides better understanding of customers' consumption behaviors and preferences. In the increasingly competitive environment, enterprises have to create a decision-making model for precision marketing that provides appropriate strategies to manage the market positioning system for fulfilling their customers' needs. The research motivation of this paper is stemmed from a real-world project. This project considers a marketing problem where the supplier or manufacturer provides different products for retail customers, of which some may sell well in some customer segments and some may not. Products that are not sold will be returned back to the supplier. Therefore, the supplier needs to find a good marketing strategy that minimizes goods in stock and satisfies the supplier, retailers and consumers. In recent years, the decision-making problems have received much attention due to a wide range of real-world applications. Many decision-making techniques have been proposed in literature. Chen and Wang (2009) presented multi-criteria optimization and compromise solutions. Saen (2010) developed a technique for order performance via similarity to ideal solution. Hsu, Chiang, and Shu (2010) developed a nonlinear programming for decision-making, and Lin, Chen, and Ting (2011) presented a Linear programming for decision making.

When discussing the decision making in this context, it is important to pay attention to the role of artificial intelligence (AI). The decision support framework of AI is considered to be a major tool for obtaining information related to historical data collections using some artificial Intelligence technique such as genetic algorithm (GA), artificial colony optimization (Ghasab, Khamis, Mohammad, and Fariman, 2015), and other data-mining techniques (Chai, Liu, & Ngai, 2013). Nowadays, data-mining, which can extract useful customer information and discover the hidden customer's behaviors from big data, has a great influence on

guiding decision making and forecasting the effects of decisions. Guo, Yuan, and Tian (2009) used a hierarchical potential support vector machine (HPSVM) for supplier selection and improved the accuracy. Chang and Hung (2010) presented a rough set theory (RST) to analyze the rules of supplier selection and guide the decision making. Chai et al. (2013) conducted a systematic review of literature about the application of decision-making techniques in supplier selection. They classified these techniques into three categories: Multi-criteria decision making techniques (MCDM), mathematical programming techniques (MP), and artificial intelligence techniques (AI). MCDM is a methodological framework that aims to provide decision makers a knowledgeable recommendation amid a finite set of alternatives (Chai et al., 2013). Wang, Wu, Wang, Zhang, and Chen (2014) developed two kinds of prioritized aggregation operators of IVHFLNs for MCDM.

However, in practice, it is not an easy task to choose a good data mining tool since, each data-mining tool has its own advantages and disadvantages. For example, artificial neural networks (ANNs) involve too many hidden neurons and training parameters (Zhang, Jiang, & Li, 2005). Its disadvantages include the "black box" nature, greater computational burden and proneness to over-fitting. However, an advantage of ANN is that ANN has the ability to implicitly detect complex nonlinear relationships between dependent and independent variables. Decision trees are simple to use and easy to understand and they offer many advantages compared with other decision-making tools. However, the disadvantages of decision trees include their instability and relatively low performance. Hence, there is no single best model for all the cases (Akın, 2015). Recently, researchers attempt to combine different models together by considering their respective advantages. Guo, Wong, and Li (2013) presented a multivariate intelligent decision-making mode which combined three different model to complete every phase of the decision making process. Tadić, Zečević, and Krstić (2014) developed a novel hybrid MCDM model that combines fuzzy decision making trial and evaluation laboratory model to provide support to decision makers. Furthermore, Yan and Ma (2015) proposed a novel two-stage group decision-making approach to uncertain quality function deployment.

To the best of our knowledge, the marketing problem considered in this paper is still a new problem. The objective of this paper is to propose a decision-making framework for combining various data-mining algorithms to achieve precision marketing of real products. Real-world data that include historical monthly supply (quantity) and information of every customer were collected from a company in China. The goal is to find a model that can classify targeted customers and predict supply quantity and then provide a strategy for precision marketing, i.e., deciding the quantity of products that every store needs. Depending on the different characteristics and requirements of each phase of the marketing model, the decision-making framework uses four data-mining models/algorithms, which are K-means algorithm, decision tree, Pareto ratio method, and RFM model, for decision-making. Overall, the purpose of this study is to generate, using data-mining techniques, a decision-making model for products' precision marketing. The proposed decision-making framework is more accurate due to its integrated precision strategies which combine prediction models, clustering, and classifying model. Moreover, the proposed framework incorporates RFM model and Pareto ratio into the process of customer segmentation so that the generated strategies are more convincing.

The rest of this paper is organized as follows. Section 2 introduces related works on common data-mining models and algorithms. Section 3 describes the methodology in our proposed framework briefly. Section 4 presents a case study and results of analyses. Finally, Section 5 concludes the paper.

2. Related works

Since a single data-mining model may only be suitable for a specific problem, such as prediction or clustering or classification, in the proposed framework, we have combined four data-mining models or algorithms to derive a precision marketing strategy for enterprises. The literature on these four models or algorithms was reviewed below.

2.1. K-means algorithm

Clustering is the process of grouping a set of physical or abstract items into classes of similar items where the groups are either meaningful or useful, or both. A well-known clustering algorithm is K-means, which was first proposed by MacQueen (1967). The accuracy of this algorithm depends on the initialization and the number of clusters (Mesforoush & Tarokh, 2013). The basic idea of K-means is to discover k clusters, such that the records within each cluster are similar to each other and distinct from the records in other clusters. K-means is an iterative algorithm: an initial set of clusters is defined and the clusters are repeatedly updated until no further improvement is possible (or the number of iterations exceeds a specified threshold). The K-means algorithm is widely used to pre-process data or for clustering because of its simplicity and efficiency (Mesforoush & Tarokh, 2013). K-means has been widely used to effectively identify the valuable customers and develop the related marketing strategies (Wei, Lee, Chen, & Wu, 2013; Mesforoush & Tarokh, 2013). In particular, Cheng and Chen (2009) use the RFM model and K-means to perform customer relationship management, and the experimental results demonstrate that their proposed model is an effective method in customer value analysis.

2.2. Decision tree

A decision tree is an efficient data mining algorithm with a strong explanatory capability (Zhang, Zhou, Leung, & Zheng, 2010). This study uses the Chi-squared Automatic Interaction Detector (CHAID) decision tree, a decision tree algorithm that is a type of database segmentation. The concept of CHAID was first published in Kass (1980). CHAID is used for prediction and classification. Like other decision trees, CHAID's advantages are that its output is highly visual and easy to interpret. A recent study indicates that CHAID is superior to judgment RFM for identification of likely responders (McCarty & Hastak, 2007). CHAID is similar to the RFM approach because it can identify the terminal nodes that will break even with respect to expected profit and costs. Therefore, CHAID was widely adopted used for segmenting customers and extracting rules that show the associations between the input and output variables (McCarty & Hastak, 2007; Chen & Wang, 2009; Mistikoglu et al., 2015). In addition, Coussement, Van den Bossche, and De Bock (2014) demonstrate that, when performing customer segmentation at different levels of uncertainty, CHAID outperforms RFM and logistic regression. Furthermore, CHAID can also assess its robustness to data accuracy problems. So in our study, we use CHAID to segment customers.

2.3. Pareto ratio

The concept of Pareto ratio is from the "Pareto Principle" (also known as the 80–20 rule, the law of the vital few or the principle of factor sparsity) which states that for many events, roughly 80% of the effects come from 20% of the causes. This is a common rule of thumb in business. This distribution is relevant to entrepreneurs and business managers, stating, for example, that 80% of the