European Online Journal of Natural and Social Sciences 2013; www.european-science.com
Vol.2, No.3 Special Issue on Accounting and Management.
ISSN 1805-3602

Using Data Mining Techniques for Improving Customer Relationship Management

Golchia Jenabi[1], Seyed Abolghasem Mirroshandel[2]

[1]Information Technology Engineering, Electronic Trade, University of Guilan, University campus 2
[2]Department of software Engineering, University of Guilan
*Email: Gol.jenabi@gmail.com

Abstract

Customer relationship management (CRM) refers to the managerial efforts to technologies and processes that helped to understand firms' customers. For this reason data mining techniques have an important role to extract the hidden knowledge and information which is inherit in the data used by researchers. This investigation focuses on the current automotive maintenance industry in Iran and applies various data mining technologies to partitioning customers. Its purpose is to determine the group of potential customers who are more likely to purchase optional services. Whereas the dataset used in this study is the real data of company, many steps of preprocess were applied and dataset records have been divided into two categories by attributing labels to the records. After preprocess steps, CAID and C5.0 methods of decision tree have been applied to classify customers and help the desired organization to make decision. By the results of two decision tree methods, there are some more important features for the firm to making decision.

Keywords: customer relationship management, loyalty, customer behavior, Data mining, decision tree.

.

Introduction

Since 1980s, customer relationship management (CRM) has become an important concept in business and marketing. There is no general definition of CRM (Lin & Yen, 2001; Ngai, 2005). It can be described as a strategic approach of using information, processes, technology and people to manage the customers' relationship with the company which means marketing, sales, services, and support across the whole customer life cycle (kincaid 2003). Swift (2001) defined CRM as the following:

> *"Enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability"*.

CRM consists four dimensions named as customer identification, customer attraction, customer retention, and customer development. Customer identification referred to customer equation in some research and it involves customer segmentation to analyze target customer and understanding potential and lost customers. (Ngai et. al, 2009) Customer attraction is the second dimension, which follows customer identification, and it can lead the organizations to attract target customers. The central concern for CRM is customer retention. Customer satisfaction which refers to an attitude or evaluation that is formed by the customer comparing their pre-purchase expectations of what they would receive from the product to their subjective perceptions of the performance they actually received, is the essential condition for retaining customers. (Jenabi & Ghanadan 2013) High customer satisfaction has many benefits for firms, such as enhancing firm reputation, attracting and retaining customers, increasing customer loyalty, reducing price lower
\

Table 1. Literature of scope studies

Article subject	Methods	Researchers
Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty	RFM, k-means, DT, NN	Seyed Hosseini et. al (2010)
Optimized K-means algorithm and application in CRM system	RFM, FRAT, K-means	Qin et. al (2010)
Apply of data Ming Technology in CRM	K-means	Nanm & Shao (2010)
Towards an optimal classification model against	`v rvcdb n cdcede	Tu et.al, (2011)
The analysis on the customers churn of charge email based on data mining	DT, See5	Nie (2006)
A customer intelligence system based on improvement LTV model and data mining	SOM. Fuzzy Decision Tree	Chen et. al (2006)
A new model for assessment fast food customer behavior case study an Iranian fast food restaurant	RFM, K-means	Jafari Momtaz (2013)
Mining customer knowledge for exploring online group buying behavior	Apriori, k-means	Liao et al (2012)
Classifying the segmentation of customer value via RFM model and RS theory	K-means, RFM, RS	Cheng & Chen (2009)
Identifying patients in target customer using a two-stage clustering-classification approach: A hospital-based assessment	K-means, RFM, Rough Set	Chen (2012)
Mining customer knowledge for tourism new product development and customer relationship management	Apriori, K-means, association rule	Liao et. al (2010)
Mining the text information to optimizing the customer	OLAP, DT	Chang et.al (2009)
A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry	OLAP, HQMS	Lee et.al (2013)
Predicting customer profitability during acquisition	LR, DT, Bagged tree	D'Hean et.al (2013)

costs of future transactions, and higher employee efficiency. Customer development is the final dimension and it conducts expansion or transaction intensity, transaction value and individual customer profitability. Customer life-time value analysis, up/cross selling, and market basket analysis have been described as elements of customer.

Analyzing the market, market segmentation and determine the target segment accurately are the most important factors for firms to promote their product or services to distribute their resources on the most efficient and effective ways (Jang et.al, 2002). As a result, it can decrease the firms cost and increase the probability of retaining customers and life-time value. For this reason they can take advantage of data mining methods to find valuable and useful information or knowledge which cannot be discovered directly, and organization can also predict their customer behaviors. Therefore, it can provide the basis for decision making and market planning (Liang, 2010).

Data mining can help the organization to understand and interact with their customers by appropriate market strategies, improve their competitive advantages, and increase value to the customers (Cerny, 2001). These issues supposed that data mining technology has an important role in customer relationship management. It has positive role in learning customer needs to develop strategy, evaluating the effectiveness of advertisement and promotions, increasing competitive advantages, responding the expectation of customer and service quality which due to customer satisfaction.

Many studies have employed data mining to analyze customer data up to now, but few studies have tried to implement various data mining methods within the same field in CRM scope. For instance a research had been has focused on analyzing customer value for the automotive maintenance industry in Taiwan thus K-means and SOM methods had been used to establish a customer value analysis model for analyzing customer value so the model divided customers into three value groups and decision tree has utilized to mine the characteristics of each customer segment. Finally researchers develop different strategies for groups of customers with different values (Liang, 2010). Further research had conducted an optimal classification model against imbalanced data for CRM to present great flexibility and outperforms with highest sensitivity, lowest classification cost and shortest modeling time. Researchers had used Bayesian Network and TAN algorithm in WEKA software package (Tu et. al, 2011). The following table shows some other studies in selected scope.

The general propose of this study is to analyze customer data for Iranian automotive maintenance industry by integrating data mining approaches to improve customer relationship management and the profit for the organization.

Materials and Methodology

The main function of data mining technology generally divided into five kinds including classification, estimate, clustering, association rule and prediction (cheng et al, 2005). Classification is one of the most common learning models in data mining (Ahmed, 2004; Berry & Linoff, 2004) It helps to build a model to predict future customer behaviors by classifying database records into a number of classes (Ahmed, 2004). Neural network, Bayesian, SVM, decision trees and if-then rules are the tools that used for classification.

Decision tree is one of the most important methods of unsupervised classification techniques in data mining. It divides a data set in subsets and in comparison of other methods, decision tree can produce understandable rules, perform tasks without much computing, learn features that are more important for classification (Chen et al, 2003). This technique use recursive division to assess the effect of specific variables and generating groups of customers with similar characteristic features. The division of customers into groups by using specific variables, generates a tree-structured model that can be analyzed to extract rules for organizational usage. Decision tree methods such as Chi-squared automatic interaction detector (CHAID), classification and regression tree (C&RT), quick, unbiased, efficient statistical tree (QUEST), Commercial version C5.0 (C5.0) are more efficient than classical statistical methods. (Ture, 2009)

The CHAID method calculate x^2 and ρ-value of node categories in every division by using a Chi-square test. The CART method is a binary splitting algorithm and it makes a tree by dividing subsets of a dataset. Processing continuous attributes of data is one of the advantage of this algorithm. C5.0 is the later version of C4.5 which is supervised learning classification algorithm which can use continuous data, information theory and learning method to build a decision tree.

CHAID tree is a decision tree that is constructed by continues division subsets of the space into child nodes and it based on the x^2-test (Michael & Gordon, 1997) and only receives nominal or ordinal categorical variables. So when variables are continuous, they must be converted into ordinal predictors. Any acceptable pair of predictor variable categories are merged until there is no difference between the pair according to target variable to define the best division at any node. For this reason, the ρ-value calculated in merging step which is used in division step.

As it mentioned earlier, C5.0 is a supervised learning classification algorithm which construct decision tree from a data set (Quinlan, 1993) by using Gain and Gain Ratio parameters. This algorithm like most empirical learning systems, gives a set of pre-classification cases and learns decision tree classifiers. It uses divide-and-conquer algorithm for growing decision trees, (Benjamin et.al, 200) and punning procedure for decreasing the overall tree size and compute error rate of the tree (Quinlan, 1993).

Data analysis

The data set used in this research has been collected from 2011 to 2014 in SaipaYadak for automotive industry in Iran. The company was founded in 1891 and it is supporting the companies which prepares equipment and car accessories for some car factories and provides after sale services for the customers. SaipaYadak defines the business scope of automotive maintenance industry including cleaning checking, maintaining, adjusting and changing part of automobiles. It offers many different services as name of orange packages for various type of automobiles. This dataset is a good case study to achieve high quality CRM data mining project in real life.

In this paper the classification framework has been used to portion the dataset to different groups, identify target customers and compare the promotion and advertisement costs with the revenue of customers for mentioned organization in the multivariate analysis models. For this reason decision tree is used to mining the characteristics of customers. This technique help to analyze the outcome drawn from the target and make it easy to classifying customers and decision tree algorithms are pervasive in researches in CRM scope so C5.0 and CHAID decision tree methods were used in this research.

As it mentioned, the data set used in this research is the original data of a maintenance organization that has 126 features. This huge data presented many challenges for data mining process so many steps of preprocess was needed to prepare usable dataset as an input for data mining methods. For decreasing number of the features, interview with expert people were done which helped to select features. Then casting variable into numeric type, set to flag, filtering and aggregation were done to prepare dataset for data mining. Since the supervised classification methods had been used in this study, records have been allocated by labels. After all preprocess steps, dataset had 2849 records and 23 features for data mining such as chassis number, engine number, vehicle type, automobile production date, admission mileage of the automobile and some characteristics features of car owners such like gender and age. This usable dataset was highly imbalanced so it divided into test and train sets which contain 70% and 30% of dataset respectively then it became a good input for CHAID and C5.0 methods of data mining so the new dataset were present the characteristics of population. C5.0 decision tree algorithm in data imbalance studies to set on different resampling strategies to counter data imbalance problem (Nathalie, 2002)

The propose of this paper is to keep the cost in lower level and increase the revenue for the firm so the cost matrix of C5.0 method used for the experiment of organization to set so that the misclassification cost for false negative (FN) is 10 times more than the false positive (FP), in the other word in formula FN:FP=10:1.

Results

As it mentioned earlier, expert feature selection and preprocess steps reduced number of features to 23 attributes and these are the input of the methods. Due to the following figures, it’s clear that applied CHAID algorithm has more than 82% accuracy and it is about 74% for C5.0 algorithm.

Since used data in this research were imbalanced, in addition to accuracy, other criteria for evaluating models is required such as precision, recall, f-measure and cost. One of the most important purpose of the research is reducing the service presentation cost and increase the revenue for the organization. The applied CHAID and C5.0 algorithms have 3434 and 2956 cost respectively which is not too much versus the revenue for the organization.

Analysis criteria of two applied algorithm for positive and negative instances has been shown in table 1. Due to the table, negative instances have much higher precision, recall and f-measure that positive instances.

Table 2. The classification results

Method	Cost	Accuracy	Precision P	f-measure P	Recall P	Precision N	f-measure N	Recall N
CHAID	3434	82%	0.19	0.29	0.65	0.96	0.89	0.83
C5.0	2956	73%	0.16	0.26	0.73	0.97	0.83	0.73

Propose of using decision tree in this study is to provide classification rules for predicting customer behaviors according to various customers features. The extracted rules illustrate what improvements in the marketing strategy should be taken to encourage and attract them to purchase optional services. For this analysis, customers should be divided into two categories; first category is customers who do not buy any optional services and second one is for ones who buy optional services.

The mining on data indicate that some kind of automobiles owners don’t interest to purchase any optional services for their cars so the organization have to use special strategies to encourage them to buy optional services and make them loyal to the firm. In addition created decision tree shows that some features (nodes) like mileage, city, chassis number, automobile type and model are more important factors to classify customers. For instance, decision tree shows that if less than one year from the date of manufacture of the car, it is more possible that car owners purchase optional services. In addition it’s more possible for owners of Tiba model to buy additional services. It means that these groups are potential customers that organization can focused on to catch more revenue for the organization. Classification rules extracted from the decision tree providing results to predict the effect of customer behaviors.

Conclusion

This study offers a business a simple and effective approach for obtaining accurate information regarding potential customers and focuses on sailing optional services which have most revenue for desired organization. For this reason, this study examined two methods for customer relationship management and adopted C5.0 and CHAID methods to classify customers. Results show that tested algorithms have acceptable classifying result and extracted rules can prepare important strategies for the firm.

The proposed research advocated that the selected organization should conduct a market survey and segmentation to develop proper market strategies to understand customer characteristic features and achieve their needs in order to make customer satisfy and loyal and finally gain more profit for the company.

The customer data used by this study was limited and it's only contained three useful features. So it's better for the firm to collect more customer information for future researches to achieve more complete analysis and make better decisions in future.

References

Ahmed, S.R. (2004). Applications of data mining in retail business. Information technology: coding and computing, 2, 455-459.

Berry, M.J.A., & Linoff, G.S. (2004). Data mining techniques second edition-for marketing sales and customer relationship management. Wiley.

Cerny, J.Z. (2001). Data mining and neural networks form a commercial perspective. In the 36th annual ORSNZ conference, Christchurch, NZ.

Chang, C.W., Lin, C.T. & Wang, L.Q. (2009). Mining the text information to optimizing the customer. Expert system with application, 36, 1433-1443.

Cheng, B.W., Chang, C.L., & Liu, I.S. (2005). Enhancing care services quality of nursing homes using data mining. Total quality management, 16, 575-596.

Cheng, C.H. & Chen, Y.S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert system with applications, 36, 4176-4184.

Chen, Y.L., Hsu,C.L., & Chou, D.C. (2003). Contructing a multi-value and multi-labled decision tree. Expert system with application, 25, 199-209.

Chen, Y.S., Cheng, C.h., Lai, C.J., Hsu, C.J & Syu H.J, (2012). Identifying patients in target customer using a two-stage clustering-classification approach: A hospital-based assessment. Computer in biology and medicine, 42, 213-221.

Chen, Y.Z., Zhao, M.H., Zhao, S.L. & Wang, Y.J (2006). A customer intelligence system based on improvement LTV model and data mining. Fifth international conference on machine learning and cybernetic, Dalian, 13-16 august.

D'Hean J., Poel, D.V & Thorleuchter, D. (2013). Predicting customer profitability acquisition: Finding the optimal combination of data source and data mining technique. Expert system with applications, 40, 2007-22012.

Jafari Momtaz, N., Alizadeh, S. & Sharif Vaghefi, M. (2010). A new model for assessment fast food customer behavior case study an Iranian fast food restaurant. British food journal, 115, 4, 601-613.

Jang, S.C., Morrison, A.M.T., & O'Leary, J.T. (2002). Benefit segmentation of Japanese Pleasure Travels to the USA and Canada: Selecting target markets based on the profitability and the risk of individual market segment. Tourism Management, 23, 367-378.

Kincaid, J.W. (2003). Customer relationship management: Getting it right. Upper Saddle River. N.J: Prentice Hall PTR.

Lee, C.K.H, Choy, K.L., Ho., G.T.S., Chin K.S., Law, K.M.Y & Tse, Y.K. (2013). A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry. Expert systems with application,40, 2435-2446.

Liang, Y.H (2010), Integration of data mining technologies to analyze customer value for the automotive maintenance industry. Expert system with application, 37, 7489-7496.

Ling, R, & Yen, D.C. (2001) Customer relationship management: An analysis framework and implementation strategies. Journal of computer information systems, 41, 82-97.

Liao, S.H., Chu, P.H., Chen Y.J. & Chang C.C., (2012). Mining customer knowledge for exploring online group buying behavior. Expert systems with application, 39, 3708-3716.

Liao, S.H, Chen, Y.J & Deng, M.Y, (2010). Mining customer knowledge for tourism new product development and customer relationship management. Expert system with application, 37, 4212-4223.

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

Ngai, E.W.T. (2005). Customer relationship management research (1992-2002): Am academic literature review and classification. Marketing Intelligence, Planning, 23, 582-605.

Ngai, E.W.T, Li Xiu & Chau D.C.K (2009). Application of data mining techniques in customer relationship management: a literature review and classification. Expert system with application, 36, 2592-2602.

Nathalie, J. & Shaju, S. (2002). The class imbalanced problem: a systematic study, Ottawa, Ontario Canada.

Nanm W.K. & Shao Q. (2010). Apply of data Ming Technology in CRM. IEEE.

Nie, G., Zhang, L., Li, X. & Shi, Y. (2006). The analysis on the customers churn of charge email based on data mining. Sixth IEEE international conference on data mining.

Qin, X., Zheng, S., He, T., Zou, M. & Huang, Y. (2010). Optimized K-means algorithm and application in CRM system. International symposium on computer, communication, control and automation.

Quinlan, W., Hubner, K., Schmoor, C., & Schumacher, M. (1997). Validation of existing and development of new prognostic classification schemes in node negative breast cancer. Breast cancer research and Treatment, 42, 249-163.

Seyed Hosseini, S.M, Maleki, A. & Gholamian, M.R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert systems with application, 37, 5259-5264.

Swift, R.S.(2001). Accelerating customer relationships: Using CRM and relationship technologies. Upper Saddle River. N.J: Prentice Hall PTR.

Tu, y., Yang, Z. & Benslimane, Y (2011). Towards an optimal classification model against imbalanced data for customer relationship management. Seventh international conference on natural computation.

Ture, M., Tokatli, F. & Kurt, I. (2009). Using Kapalan-Meier analysis together with decision tree methods in determining recurrence-free survival of breast cancer patient. Expert systems with application, 36, 2017-2026.