HANKEN
Svenska handelshögskolan

SOFT COMPUTING METHODS FOR CUSTOMER CHURN MANAGEMENT

LITERATURE REVIEW

Author: Triin Kadak

Helsinki, 2007

1. INTRODUCTION

There holds a common understanding in business, that retaining existing customers is the best core marketing strategy to survive in an industry. Churn management is the term that has been widely used to define customer turnover. In more detail, churn management is the concept of detecting those customers who are intending to move to a competing service provider. When such customers are identified, they can be targeted with proactive marketing campaigns for retention. Customer retention has become a vital aspect in everyday business strategy. At the point when the number of customers belonging to a business reaches its peak, finding and retaining new customers becomes increasingly difficult and costly. Thus, at this stage of the businesses lifecycle it should be of higher priority to secure the most valuable existing customers, than trying to win new ones. In addition, it has been a common practice that finding new customers can prove to be a more expensive business strategy than retaining existing ones.

The old product-oriented view of “design-build-sell” is being replaced by a customer-oriented view of “sell-build-redesign” and thus, the traditional process of mass-marketing is being challenged by the new approach of one-to-one marketing. The marketing goal in the traditional model is to reach as many customers as possible and expand the customer base. However, knowing the high cost of acquiring new customers, it is more reasonable to continue business with existing customers. As a consequence, the marketing focus shifts from the customer base to the needs of each individual customer and the performance metric from market share to “wallet share”. Businesses do not just deal with customers in order to make transactions but they aim to establish a long-term relationship with each customer.

In today’s business world company’s databases should store data that can be used to obtain information about customer value, customer satisfaction, customer loyalty and prediction of behaviour. 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.

As explained previously, it is known that a company faces higher expenses while attempting to win new customers than while retaining existing ones. However, customer retention efforts have as well been costing organizations large amount of resources. Thus, it is believed that the new generation of churn management should focus on accuracy. There exists a variety of churn management techniques that have been developed to meet business requirements. These predictive modelling techniques can be divided into traditional and soft computing methods. Much research has been made by applying traditional methods to churn management and two of the most popular of those are regression analysis and decision trees. The newest technologies in the area fall into the category of soft computing.

Soft computing refers to a collection of computational techniques used in computer science, artificial intelligence, machine learning and some engineering disciplines, which attempt to study, model, and analyze very complex phenomena: those for which more conventional methods have not yielded low cost, analytic, and complete solutions. Earlier computational approaches could model and precisely analyze only relatively simple systems. More complex systems arising in biology, medicine, the humanities, management sciences, and similar fields often remained intractable to conventional mathematical and analytical methods. [7]

Key areas of soft computing include neural networks (NN), fuzzy systems (FS), evolutionary computation (EC) (including evolutionary algorithms and swarm intelligence), probabilistic computing (inclusing Bayesian network), Chaos theory.

Soft computing techniques resemble more closely with biological processes than with traditional techniques, which are largely based on formal logical systems (such as sentential logic and predicate logic) or rely heavily on computer-aided numerical analysis (as in finite element analysis). Unlike hard computing schemes, which strive for exactness and full truth, soft computing techniques exploit the given tolerance of imprecision, partial truth, and uncertainty for a particular problem. Soft computing techniques often complement each other. Another common contrast comes from the

observation that inductive reasoning plays a larger role in soft computing than in hard computing. [7]

The current paper will provide an overview of the various soft computing technologies that have been identified in the literature for the development of customer churn management. In addition, areas, which could be further investigated, will be identified and thus, ideas for future research will be provided in the end of each paper reviewed. The first paper to be reviewed provides the state-of-art of the various kinds of churn management technologies and thus, served as the basis for the selection of the underlying topic as well as for finding the related scientific research papers.

2. LITERATURE REVIEW

2.1. PAPER ONE

COMPUTER ASSISTED CUSTOMER CHURN MANAGEMENT: STATE-OF-THE-ART AND FUTURE TRENDS ***by John Hadden, Ashutosh Tiwari, Rajkumar Roy and Dymitr Ruta*** [3]

Source: Computers and Operations Research, Volume 34, Issue 10
Date: October 2007

2.1.1. OVERVIEW

The focus of the paper is to review the most popular technologies that have been identified in the literature for the development of a customer churn management platform. In the paper the advantages and disadvantages of the identified technologies are discussed, and a discussion on the future research directions is offered.

The study begins with introducing the customer churn management. A five-stage model is presented to illustrate the steps that should be addressed for creating a churn management system. The steps in the model are the following: (1) Identification of the best data; (2) Data semantics; (3) Feature selection; (4) Development of predictive model and (5) Validation of results. After explanation of the model the techniques that

are commonly used for each of the stage of the churn management model are introduced.

The authors of the paper conclude the following:

- the feature selection step appears to have been neglected by the previous research performed in the area of customer churn management and have focused mainly on developing feature extraction methods, though being a highly significant step;
- the main trends for developing models to predict customer behaviour include regression analysis, neural networks, decision trees and Markov models. Decision trees appear to be the most popular choice and neural networks have received slightly more interest than regression analysis;
- 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 genetic algorithms is very small, which suggests that research into these technologies has not yet been exhausted;
- Fuzzy logic is also an area of soft computing, that has never been investigated in either churn management or CRM research. It 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.

2.1.2. FUTURE RESEARCH

The authors highlight five areas for future research: (1) exploring RFM (i.e. recency, frequency, monetary) 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; (2) only a small amount of the literature even mentions data semantics, thus, 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; (3) the current literature suggests that previous work done for predicting customer churn has either neglected a feature selection phase, or failed to

document one. 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; (4) neural networks, combined with a powerful rule discovery method (genetic algorithm) 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; (5) include a model that distinguishes between loyal and non-loyal customers, i.e. developing a 'loyalty index'.

2.2. PAPER TWO

DATA MINING TECHNIQUES FOR CUSTOMER RELATIONSHIP MANAGEMENT ***by Chris Rygielski, Jyun-Cheng Wang and David C. Yen*** [5]

Source: Technology in Society, Volume 24, Issue 4
Date: November 2002

2.2.1. OVERVIEW

This article examines the concepts of customer relationship management and one of its components, data mining. It begins with an overview of the concepts of data mining and CRM, followed by a discussion of evolution, characteristics, techniques, and applications of both concepts. It as well integrates the two concepts and illustrates the relationship, benefits, and approaches to implementation, and the limitations of the technologies. Through two studies, the authors offer a closer look at two data mining techniques, namely Chi-square Automatic Interaction Detection (CHAID) and Neural Networks. Based on those two case studies, CHAID and neural networks are compared and contrasted on the basis of their strengths and weaknesses.

The study examines neural networks and CHAID through two providers of data mining tools, NeoVista Solutions Inc. and Applied Metrix, Ltd. respectively. Both case studies are selected from the respective websites of the data mining tool providers and the identities of their clients are withheld at the client's request. The client of NeoVista Solutions Inc. was a large retailer who found its profits suffering

due to less-than-optimal seasonal product demand forecasting. The retailer was overstocked on its slow-moving products and under-stocked on its most popular items at critical selling periods. NeoVista designed and implemented a solution that combined elements of the clustering and neural network technology, which enabled the retailer to automatically review its point-of-sale history and equate store groupings to sales patterns.

The client of Applied Metrix was a home equity marketer that extended home equity lines of credit at the national level. The client's goal was to increase the efficiency of targeting current mortgage customers who might be interested in the client's service. Applied Metrix used CHAID for the initial segmentation modeling. After the segments were identified, logistic regression response probability modelling was applied, which allowed measurement of the unique influence of each predictor on a specific criterion and allowed scoring of individual customers or prospects in the marketing database.

As a result of the comparison of the two techniques, the authors concluded:

- CHAID model is understandable as a set of rules, whereas a neural network is obscure, with weights that have no intuitive meaning;
- it is much easier for a CHAID model to be implemented than a neural network. The performance of an implemented CHAID model will be significantly faster than an implemented neural network;
- neural networks provide more accurate (i.e., powerful and predictive) models, especially for complex problems. However, there is also a risk of finding sub-optimal solutions and over-fitting;
- CHAID is easier and quicker to construct, whereas neural networks have many parameters that must be set and require more skilled manipulation to ensure over-fitting does not occur;
- It is harder to apply background domain knowledge using neural networks, whereas it is easier to see mistakes and over-fitting in a CHAID tree;
- building a neural network is more costly than building a CHAID model. A neural network requires more time and a higher level of building skills than a CHAID model.

As a conclusion the results of Rygielski et al. suggest that 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. In addition, other disadvantages should be taken into account, like clarity of output, implementation and construction of model.

The authors as well highlight that businesses must bear in mind that they have to use technology responsibly in order to achieve a balance between privacy rights and economic benefits. In addition, while choosing a suitable technology for personalization of CRM, organizations must be aware of the tradeoffs when considering differing data mining software applications since different technologies vary in terms of effectiveness and ease of use. According to the authors, current CRM solutions focus primarily on analyzing consumer information for economic benefits, and very little touches on ensuring privacy.

2.2.2. FUTURE RESEARCH

As privacy issues become major concerns for consumers, an integrated solution that streamlines and enhances the entire process of managing customer relationships will become more necessary.

2.3. PAPER THREE

BAYESIAN NETWORK CLASSIFIERS FOR IDENTIFYING THE SLOPE OF THE CUSTOMER LYFECYCLE OF LONG-LIFE CUSTOMERS ***by Bart Baesens, Geert Verstraeten, Dirk Van den Poel, Michael Egmont-Petersen, Patrick Van Kenhove and Jan Vanthienen*** [1]

Source: European Journal of Operational Research, volume 156, Issue 2
Date: July 2004

2.3.1. OVERVIEW

The authors of the study indicate that recent findings in relationship marketing literature have shown large differences within the group of long-life customers in terms of spending and spending evolution. Therefore, they focus on introducing a measure of a customer's future spending evolution that might improve decision-making in relationship marketing.

From a marketing point of view Baesens et al. focus on predicting whether a newly acquired customer will increase or decrease his/her future spending from initial purchase information, which is an essential classification task. However, the main contribution of the study lies in comparing and evaluating several Bayesian network classifiers with statistical and other artificial intelligence techniques for the purpose of classifying customers in the binary classification problem at hand. A Bayesian network is defined as probabilistic white box, which represents a joint probability distribution over a set of discrete stochastic variables.

The authors discuss and evaluate several types of Bayesian network classifiers and their corresponding structure learning algorithms. The descriptive part is continued by a research on UPC scanner data of a large Belgian DIY (Do-It-Yourself) retail chain. The data were gathered by the customer loyalty cards and contain four complete years of information. In order to assess the quality of the models, the database was randomly divided into two parts: 2/3 of the observations were used for learning the classifiers and the remaining 1/3 was used as a test set for estimating the generalization behavior of the classifiers.

The study compared and contrasted the performance of the Naive Bayes, TAN, CL multinet, GBN, GBN multinet, C4.5, C4.5rules, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) classifiers on our marketing data set. The results of the study provide experimental evidence that:

- Bayesian network classifiers offer an interesting and viable alternative for customer lifecycle slope estimation problem;
- the Markov Blanket concept allows for a natural form of attribute selection that was very effective for the application at hand;

- the sign of the slope can be predicted with a powerful and parsimonious general, unrestricted Bayesian network classifier;
- a set of three variables measuring the volume of initial purchases and the degree to which customers originally buy in different categories, are powerful predictors for estimating the sign of the slope, and might therefore provide desirable additional information for relationship marketing decision making.

2.3.2. FUTURE RESEARCH

In the domain of marketing to create variables, which have better predictive capabilities for predicting the sign of the slope of the linear lifecycle. Another idea suggested by the underlying authors is to replicate the study over different customer bases in diverse industries and countries to deliver an insight into the stability of the findings or to investigate the power of other structure learning algorithms and the presence of hidden variables in the Bayesian network.

2.4. PAPER FOUR

SEGMENTATION OF STOCK TRADING CUSTOMERS ACCORDING TO POTENTIAL VALUE *by H.W. Shin and S.Y. Sohn* [6]

Source: Expert Systems with Applications, Volume 27, Issue 1
Date: July 2004

2.4.1. OVERVIEW

In the article the authors propose a robust clustering algorithm to classify the stock traders into several groups in terms of the three 3-month transactions in order to suggest the graded commission policy for each group.

Variables used for clustering criteria are transactions made on both representative assisted trading and online Home Trading System (HTS) in the Korean stock market. Clustering methods used are K-means clustering, self-organizing map (SOM), and

fuzzy K-means method. The cut-off value of each customer group is set based on classification and regression tree (CART).

The paper begins with introducing the three clustering algorithms: K-means, SOM and fuzzy K-means. They continue with the comparison of the algorithms and finally conduct a case study. The authors randomly selected 3000 customers who had transaction records from the middle of July to the middle of October in 1999 from a stock corporation ‘A’ and apply the three clustering methods. The stock transaction modes used are either representative assisted or online HTS. HTS customers directly buy and sell their stocks without the advice of the corporation's representatives.

One of the objectives of the study was to suggest a graded brokerage commission policy based on the three clusters of customers, which is effective enough to avoid the churning behavior of the existing customers and at the same time resulting in sufficient profit to the security corporation.

As a result, the study found a fuzzy K-means clustering being the most stable to group stock trading customers and used it to classify three tiers of customers (Normal, Best, and VIP level) based on the total trade amount over 3-month period. For each group, different brokerage commission rate was assigned as 0.4, 0.45, and 0.5% for the representative assisted mode while 0.06, 0.1, and 0.18% for HTS.

The authors suggest that this approach is different from the existing graded commission policy in that the proposed policy adopts the idea of the graded commission based on the historically accumulated transaction amount made by customer. This new approach is expected to bring more profit by treating loyal customers in a better manner and subsequently retain them in a longer term.

2.4.2. FUTURE RESEARCH

It is suggested to redo a similar study containing longer time-series data. In addition, since the policy derived in the study is mainly dependent on the cumulative transaction some other facts such as frequency of transaction could be included in the policy.

2.5. PAPER FIVE

APPLYING DATA MINING TO TELECOM CHURN MANAGEMENT ***by Shin-Yuan Hung, David C. Yen and Hsiu-Yu Wang*** [4]

Source: Expert Systems with Applications, Volume 31, Issue 3
Date: October 2006

2.5.1. OVERVIEW

The aim of the study is to compare various data mining techniques that can assign a `propensity-to-churn` score periodically to each subscriber of a mobile operator. It intends to illustrate how to apply IT technology to facilitate telecom churn management. The research uses various data mining techniques (decision tree and neural network) to find a best model of predictive churn from data warehouse to prevent the customers` turnover, further to enhance the competitive edge.

To begin with, the methodology of the study consisted of three parts: (1) creation of an IT infrastructure to facilitate the research, including a common customer base, attributes and transactions, modelling parameters, model results, etc.; (2) a model-independent knowledge discovery process to discover customer behaviour prior to churn, by using data mining techniques, and (3) a set of measurements to quantify the performance of models developed by different modelling tools, such as decision tree and neural network.

A wireless telecom company in Taiwan was used to obtain the customer-related data. To protect customer privacy, the data source included data of approximately 160,000 subscribers, including 14,000 churners, from July 2001 to June 2002, randomly selected based on their telephone numbers.

The authors developed possible variables from other research and interviews with telecom experts and then analyzed these variables with *z*-test from four dimensions

and listed significant variables of churn based on the analysis database. These variables included customer demography (age, tenure, gender), bill and payment analysis (monthly fee, billing amount), call detail records analysis (in-net call duration, call type) and customer care/service analysis (MSISDN change count, count of bar and suspend).

First of all they used K-Means to model the customers into five clusters and continued with assessment of how models built by decision tree (C5.0) and back propagation neural network (BPN) techniques perform. The decision tree technique was used to create many models under different scenarios: one of the scenario to create a decision tree model for all the customers and another scenario to create separate decision tree models in each of the customer segments.

The findings suggest that the performance of the decision tree model without segmentation is better than that with segmentation and the performance of BPN is better than the decision tree model without segmentation on both hit ratio and capture rate.

Overall, both decision tree and neural network techniques can deliver accurate churn prediction models by using customer demographics, billing information, contract/service status, call detail records, and service change log. Data mining techniques can effectively assist telecom service providers to make more accurate churner prediction.

2.5.2. FUTURE RESEARCH

The authors suggest developing attractive retention programs to satisfy the customers in addition to recognition of the customers who are about to leave. In addition, they propose integrating churn score with customer segment and applying customer value to help mobile service providers to design the right strategies to retain valuable customers. Another future research idea might include applying data mining techniques in various CRM fields, such as credit card fraud detection, credit score, affinity between churners and retention programs, response modelling, and customer purchase decision modelling.

2.6. PAPER SIX

TOWARD A HYBRID DATA MINING MODEL FOR CUSTOMER RETENTION ***by Bong-Horng Chu, Ming-Shian Tsai and Cheng-Seen Ho*** [2]

Source: Knowledge-Based Systems
Date: October 2006

2.6.1. OVERVIEW

The paper proposes a hybridized architecture to deal with customer retention problems. It aims to predict churn probability as well as proposes retention policies. The architecture works in two modes: learning and usage. In the learning mode, a churn model, which can analyze and predict the probability of churning is constructed. The authors also create a policy model, which clusters the churners by labelling each group with its most significant attributes. The policy model is then used to create appropriate policies for each customer cluster. In the usage mode, the churn model is used to determine whether a given customer is going to defect. If there is a high probability of defection the policy model is used to propose a specific policy to retain the customer. The most significant feature of the approach is that it not only predicts churning but also makes proactive attempts to reduce it.

The historical information was gathered from a subscriber database of a major wireless carrier in Taiwan. The database consisted of 65516 business subscribers. The subscribers came from all regions of Taiwan in July 2001. The carrier defined churn as 'the cancellation of all services received by a customer'. The database contained historical details for each individual subscriber and the details included: defection history, de-activation data, payment history, usage patterns, trends and changes. The subscriber database initially contained 253 data fields. From these fields, the domain expert, who works for a telecommunication company and majors in customer service identified 12 attributes to be most likely relevant to customer churning (gender, zip-code, industry code, dealer ID, site code, package code, discount type, tenure, times of

service suspended, times of service resumed, times of service terminated, average invoice).

In the learning mode the authors construct a churn and a policy model. The churn model is created by using a classification technique (based on a decision tree algorithm C5.0) to predict the conditions under which a subscriber may defect. The policy model, which analyses the churn model is constructed by using a clustering technique (based on GHSOM - Growing Hierarchical Self-Organizing Map) and it develops policies that can be associated with each cluster group.

The study suggested that according to the experiments the churn model has an evaluation accuracy of approximately eighty-five percent (no appropriate data was obtained to evaluate the constructed policy model), meaning that the policy model construction represents an interesting and important technique in investigating the characteristics of churner groups. Furthermore, the study indicated that understanding the relationships between churns is essential in creating effective retention policy models for dealing with 'churners'.

2.6.2. FUTURE RESEARCH

The policy model developed in the study is currently semi-automatically constructed, thus the authors suggest finding out whether the system can automatically construct a policy model according to the nature of churn groups, through the use of a set of primitive policies.

3. CURRENT STATE-OF-THE-ART

The six papers reviewed provided understanding about the current status of the soft computing techniques used for churn management. In addition, they provided several interesting ideas for future research and study. The current state-of-the-art indicates that neural networks have been widely used for developing churn management models. However, the amount of work that has been done using genetic algorithms is very small, which suggests that research into these technologies has not yet been exhausted. Fuzzy logic is another area of soft computing, that has not been much

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 [3].

REFERENCES

[1] B. Baesens, G. Verstraeten, D. Van Den Poel, M. Egmont-Peterson, P. Van Kenhove and J. Vanthienen, Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers, *European Journal of Operational Research* **156** (2004)

[2] B. H. Chu, M. S. Tsai and C. S. Ho, Toward a hybrid data mining model for customer retention, *Knowledge-Based Systems*, (2006)

[3] J. Hadden, A. Tiwari, R. Roy and D. Ruta, Computer assisted customer churn management: State-of-the-art and and future trends, *Computers and Operations Research* **34**, (2007)

[4] S. Y. Hung, D. C. Yen and H. Y. Wang, Applying data mining to telecom churn management, *Expert Systems with Applications* **31**, (2006)

[5] C. Rygielski, J. Wang and D.C. Yen, Data mining techniques for customer relationship management, *Technology in Society* **24** (2002)

[6] H.W. Shin and S.Y. Sohn, Segmentation of stock trading customers according to potential value, *Expert Systems with Applications* **27** (2004)

[7] Wikipedia. "Soft computing" <http://en.wikipedia.org/wiki/Soft_computing>