A Hybrid Approach to Predict Churn

Conference Paper · January 2011

DOI: 10.1109/APSCC.2010.87 · Source: IEEE Xplore

2010 IEEE Asia-Pacific Services Computing Conference

A Hybrid Approach to Predict Churn

Javad Basiri
School of Electrical and Computer Engineering
University of Tehran, Tehran, Iran
basiri@ut.ac.ir

Fattaneh Taghiyareh
School of Electrical and Computer Engineering
University of Tehran, Tehran, Iran
ftaghiyar@ut.ac.ir

Behzad Moshiri
Control and Intelligent Processing Center of Excellence, School of ECE
University of Tehran, Tehran, Iran
moshiri@ut.ac.ir

***Abstract*— Acquiring new customers in any business is much more expensive than trying to keep the existing ones. As a result, many prediction algorithms have been proposed to detect churning customers. In this paper, the ordered weighted averaging (OWA) technique is brought to the attention of marketing researchers. We have applied OWA technique to improve the prediction accuracy of existing churn management systems. The decision lists of underlying prediction algorithms have been fused using OWA algorithm. Applied to the database of a telecommunication company, this method is found to significantly improve accuracy in predicting churn compared to the best existing result in the literature of the churn management. Our findings lead us to believe that using OWA technique could cause to increase profit for the companies.**

Keywords-churn management system; LOLIMOT; Bagging and Boosting; hybrid approach; OWA

I. INTRODUCTION

Customer retention has become a major issue for business in recent years. The importance of managing customer churn has been highlighted in many recent studies making it a critical subsystem in customer relationship management. As an example, a 1% increase in retention rate has been shown to increase firm value by 5%, on average [1]. To reduce customer churn, managers need to be able to predict this behavior accurately and establish links between customer attrition and factors under their control [2].

Churn prediction is a binary classification task which differentiates churners and non-churners. In the literature, many prediction algorithms have been applied to predict the customer churn, such as neural networks [3], decision trees [4], logistic regression [5] and genetic algorithm [4], but nevertheless, this problem remains a hot topic in marketing research [6]. About decision tree algorithms, as some nodes may have similar probabilities and the approach is vulnerable to noise, they may have a less optimal solution [6]. Also, neural network outperforms certain types of decision trees in churn prediction but potentially there is a risk of finding suboptimal solutions and over fitting [7]. Although genetic algorithms can produce accurate predictive models, there is a shortcoming for them: they cannot clarify the likelihood associated with their predictions. Finally, the regression algorithms do not normally lead to the best results and in the literature they have been used as a baseline method for comparison [7]. As a result, the current literature requires an applicable algorithm with a high performance in churn prediction so that companies would be able to detect as many churners as possible.

Two important prediction approaches in the churn management literature are bagging and boosting classification trees [8] and LOLIMOT [6]. To the best of our knowledge, bagging and boosting classification tree is the most accurate prediction algorithm in the churn management area; also the LOLIMOT algorithm has had better performance in comparison with many other algorithms [6].

In this paper, using the strengths of both bagging and boosting and LOLIMOT algorithms, we have proposed OWA approach to combine these algorithms. Also, the application of optimistic operator as weighting function has been investigated. These algorithms have been learned using different sets of extracted features and the outputs of all learned models have been combined. So, applying an ensemble method, this paper aims at improving the prediction accuracy in the churn management area.

To evaluate this approach, we apply it to churn prediction in the telecommunication industry. Our method can also be applied to many other service industries. Our approach significantly has improved the accuracy in predicting churn compared to the best existing result in the literature.

The remaining of the paper is arranged as follows. First, we review the churn management systems. Then the optimistic OWA and the proposed approach are discussed in details. After that, the proposed approach compared with several methods on the churn data. Finally, we conclude the paper.

978-0-7695-4305-5/10 $25.00 © 2010 IEEE
DOI 10.1109/APSCC.2010.87

II. CUSTOMER CHURN MANAGEMENT

"Churn" refers to the act of those customers who are intending to move their custom to a competing service provider. Churn management on the other hand, aims at identification of such churners and the carrying out of some proactive campaigns for retention efforts [9]. Geppert [12] defines customer churn as the following equation:

$$\text{Monthly Churn} = (C0 + A1 - C1)/C0$$

Where:

C0 = Number of customers at the start of the month
C1 = Number of customers at the end of the month
A1= Gross new customers during the month

There are two main groups of churning customers, voluntary and non-voluntary churners [9]. Identification of non-voluntary churners is fairly easy since they are the customers who have had their service withdrawn by the company. On the other hand, voluntary churners can be divided into two subgroups: incidental churn and deliberate churn. Incidental churn refers to the act of those customers who unintentionally leave the company because of some reasons like house movement and change in occupation [6]. Deliberate churn occurs when a customer decides to leave the company deliberately in search of better or cheaper products or services. Generally, churn management systems try to identify this specific group of customers and the carrying out of some proactive campaigns for retention efforts.

Many prediction algorithms have been used for predicting churn. Some well-known of these algorithm are neural networks [3], decision trees [4], logistic regression [5] and genetic algorithm [4]. But, as mentioned in the introduction, each of them has some shortcomings. Also two most accurate algorithms in the churn management area are bagging and boosting classification trees [8] and LOLIMOT [6]. In this study, these algorithms have been used as the input for the hybrid approach.

Fig. 1 illustrates the share of each algorithm in the churn management systems until 2005 [9].

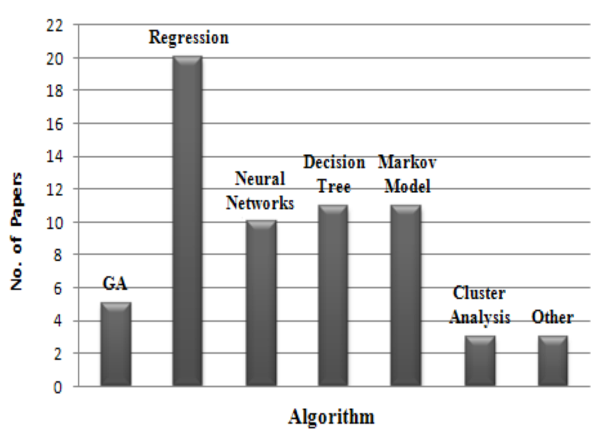

Figure 1. Share of each algorithm in the churn management [9].

III. PROPOSED APPROACH

In this section, to improve the prediction accuracy of existing churn management systems, we introduce a hybrid approach which unites the strengths of both bagging and boosting and LOLIMOT algorithms. Briefly, it has three main steps. First, using feature selection algorithm, most significant features are specified. Then using different size of features, some prediction algorithms are learned and finally, using OWA technique the decision lists of underlying prediction algorithms are fused. Also, the application of optimistic operator as weightening function for OWA has been investigated. Fig. 2 depicts the proposed ensemble approach. In the following, each step has been explained.

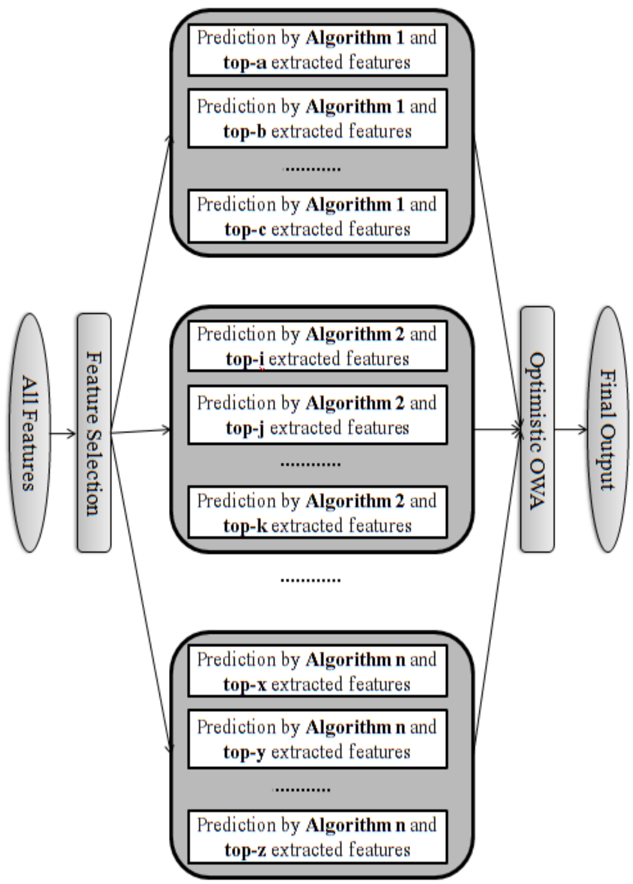

Figure 2. Diagram of the proposed hybrid approach

A. *Feature Selection*

Feature selection process aims at identification of features which are the best for the prediction [9]. "It is an important stage because it helps with both data cleansing and data reduction, by including the important features and excluding the redundant, noisy and less informative ones" [9]. In this paper, Chi-Square algorithm [20] has been used as the feature selection algorithm.

B. *Prediction Algorithm*

The prediction algorithm computes the churn probability for each customer. The output of the prediction algorithm is a number which shows how likely a customer may leaves the company. Generally, the output for a prediction algorithm is in the range of 0 to 1 [13].

In this paper, two different prediction algorithms have been used, including: bagging and boosting classification trees and LOLIMOT. These algorithms have been defined in the following. Although we have used two different algorithms, it is possible to use more than two algorithms in the proposed hybrid approach.

1) Bagging and Boosting

A. Lemmens and C. Croux have been applied bagging and boosting classification trees on the churn data [8]. To the best of our knowledge, this approach is the most accurate prediction algorithm in the literature.

Bagging [10] is a bootstrap ensemble method that trains each classifier on a randomly drawn training set. The training set of each classifier consists of the same number of instances which randomly drawn from the original training set. Probability of drawing for any given example is equal to the others. Samples are drawn with replacement. Combining these classifiers, the resulting ensemble leads to a lower error than a single classifier [11].

Boosting [12] trains a series of classifiers, while each training set is produced based on the diagnosis accuracy of the previous classifiers. New classifiers are constructed to better predict instances for which the current ensemble's diagnosis accuracy is poor. This procedure is carried out using adaptive resampling. It means, instances that are incorrectly classified by the previous classifiers are selected more frequently, or alternately, given a higher cost of misclassification [11].

While bagging is a simple method, there are some variants with more complexity. Stochastic gradient method is one of the recent developments, and includes weights in the resampling procedure [8]. A. Lemmens and C. Croux have been applied this specific type to predict churn. For more details about, see [8].

2) LOLIMOT

LOLIMOT algorithm has been introduced to the marketing and churn management literature by Ghorbabi et al. [6]. According to [6], this algorithm has outperformed the other algorithm, such as, artificial neural networks, decision trees, and logistic regression.

LOLIMOT unites the advantage of neural networks, tree model and fuzzy modeling. "This procedure which is a combination of tree models and neural networks eliminates the shortcomings of these approaches while making the best use of the advantages" [6]. The model is very flexible, and the optimal structure is selected automatically by the learning procedure. Based on an incremental tree based learning algorithm, LOLIMOT starts as an optimal linear least squares estimation, and the nonlinear neurons are added if they lead to an improvement in performance. Thus, the learning algorithm omits the need for linearity tests, and automatically builds the model to gain the highest generalization [6].

Churn prediction has a fuzzy nature which has not appropriately been considered [9] and the LOLIMOT algorithm covers this aspect as well.

After feature selection phase, all features are ordered based on their importance. Then, using different number of selected features, some classifiers are learned by each prediction algorithm. For example, we have learned bagging and boosting classification trees, using different feature size, such as 10, 12, 20 features.

3) OWA Technique

In this study, we use the optimistic exponential type of ordered weighted averaging (OWA) operator to fuse the output of each learned classifiers, that introduced by Filev and Yager [14]. OWA were introduced by Yager [16] (cf. also Yager and Kacprzyk [17]) and are defined as follows: The OWA operator of dimension n is a mapping such as:

F: $R^n \rightarrow R$ and is given by

OWA ($a_1, a_2, \dots a_k \dots, a_n$) = $\sum_{i=1}^{n} w_i b_i$

Where b_i is i[th] largest element among a_k's [18]. The weights are non-negative ($w_i \geq 0$) and their sum equals to one ($\sum_{i=1}^{n} w_i = 1$).

The OWA operators include many of the well-known operators such as the maximum, the minimum, the k-order statistics, the median and the arithmetic mean. The OWA operators are commutative, monotone, and they have a compensatory behavior [15, 18]. Base on last property the aggregation done by an OWA operator always is between the maximum and the minimum. It can be seen as a parameterized way to go from the min to the max. In [16] a degree of maxness (initially called orness) was introduced as follow:

Maxness $(w_1, w_2, \dots \dots, w_n) = \frac{\sum_{i=1}^{n}(n-i)w_i}{n-1}$

Where for the minimum, maxness (0, 0, . . . ,1) = 0 and for the maximum maxness (1, 0. . . ,0) = 1 [19].

A simple class of OWA operators as exponential class was introduced to generate the OWA weights satisfying a given degree of maxness. The optimistic and pessimistic exponential OWA operators were correspondingly introduced as follows [14]:

Optimistic:

$$w_i = \alpha \times (1-\alpha)^{i-1}, \forall\, i \neq n;\; w_n = (1-\alpha)^{n-1} \quad (1)$$

Pessimistic:

$$w_1 = \alpha^{n-1};\; w_i = (1-\alpha) \times \alpha^{n-i}, i \neq 1 \quad (2)$$

Where parameter α belongs to the unit interval [0 1] and is related to orness value regarding the n.

In this paper, we have used the optimistic exponential kind of OWA as the hybrid approach.

After constructing the classifiers, their outputs for each instance in the test data are used as inputs for OWA algorithm. Finally, the fused result of this algorithm for each instance is used as the final churn probability of that instance.

IV. METHODOLOGY

A. Dataset

The dataset is provided by the Teradata Center at Duke University. The database contains datasets of mature subscribers (i.e. customers who were with the company for at least six months) from a major U.S. Telecommunication Company. There is three different datasets in this database including: Calibration Data, Current Score Data and Future Score Data.

A total of 172 variables are included in each dataset, one for churn indication, and 171 variables for prediction. The prediction variables include three types of variables: behavioral data such as minutes of use, revenue, handset equipment; company interaction data such as customer calls into the customer service center, and customer household demographics. The churn response is coded as a dummy variable with churn = 1 if the customer churns, and churn = 0 otherwise.

There are 100,000 records in the Calibration dataset, 51,306 records in the Current Score and 100,462 records in the Future Score dataset. The actual average monthly churn rate is reported around 1.8%. Calibration sample is a balanced data set with 50-50 split between churners and non-churners. Also churn rate in the Current Score and the Future Score dataset is about 1.8%.

The Calibration dataset (a balanced one) is used as the training set and Future Score dataset is used as the test set in our experiments.

B. Data Preparation

All features with more than 30% missing value have been omitted. Also, using linear interpolation method, missing value replacement has been done for all numerical variables. This procedure has been carried out by SPSS 17 platform.

C. Feature Selection

As before mentioned, feature selection phase has been done by the Chi-Square algorithm. The top 20 features have been presented in table 1. This procedure has been done by Weka release 3.6.1.

D. Hybrid Approach

Here, we have used an optimistic exponential OWA technique with $\alpha = 0.46$ which was simply obtained by a try-and-error algorithm on the training data. We have discussed about sensitivity of optimistic exponential OWA to this parameter in the empirical analysis section.

E. Evaluation Criteria

The lift criterion is a performance measure which is the result of the ratio between the obtained outcomes using and without using the prediction model. The higher the lift means the model is more accurate, and instinctively, the more profitable a targeted proactive churn management program will be [21].

Table 1. Most Significant Features

Name of feature	Description of feature
Months	Total number of months in service
Eqpdays	Number of days of the current equipment
hnd_price	Current handset price
rev_Mean	Mean monthly revenue (charge amount)
totmrc_Mean	Mean total monthly recurring charge
mou_Mean	Mean number of monthly minutes of Use
change_mou	Percentage change in monthly minutes of use vs previous three month average
Adjrev	Billing adjusted total revenue over the life of the customer
mou_Range	Range of number of minutes of use
Totrev	Total revenue
avg3mou	Average monthly minutes of use over the previous three months
totmrc_Range	Range of total monthly recurring charge
avg3qty	Average monthly number of calls over the previous three months
mou_opkv_Mean	Mean unrounded minutes of use of off-peak voice calls
mou_cvce_Mean	Mean unrounded minutes of use of completed voice calls
opk_vce_Mean	Mean number of off-peak voice calls
complete_Mean	Mean number of completed data calls + Mean number of completed voice calls
mou_peav_Mean	Mean unrounded minutes of use of peak voice calls
comp_vce_Mean	Mean number of completed voice calls
peak_vce_Mean	Mean number of inbound and outbound peak voice calls

As the top-decile lift is the main criterion in the churn management area [21], this criterion has been used to evaluate the performance of applied algorithms. The top-decile lift focuses on the customers predicted most likely to churn. The top 10% critical customers (i.e. the customers that have the highest churn probability) is an ideal portion for targeting a retention marketing campaign [6]. The criterion equals the rate of churners in this segment, a10%, divided by the rate of churners in the whole set, A:

$$Top - decile = \frac{a_{10\%}}{A} \tag{3}$$

To compute the top-decile lift, we should sort the customers from predicted most likely to predicted least likely to churn. Next, we should take the top 10% and calculate the exact percentage that did in fact churn. Finally, we divide this value by the average churn rate across all customers. Repeating this procedure for the other top m% (i.e 20%, 30%, ….., 100%) sorted customers, the lift chart will be formed.

V. EMPIRICAL ANALYSIS

To evaluate the performance of the proposed hybrid approach, it has been applied to the churn data. Additionally, to demonstrate the efficiency of hybrid method, it has been compared with six other well-known prediction methods, including: neural networks, C5.0, logistic regression, baysian network, LOLIMOT and bagging and boosting classification tree.

In these experiments, neural networks, C5.0, logistic regression, baysian network and bagging and boosting classification trees have been carried out with the aid of the Clementine tool for Windows version 12.0. Also, LOLIMOT algorithm has been performed using MATLAB version 7.6.

Fig. 3 shows the performance of the hybrid method (OWA) and four other well-known prediction algorithms, including neural networks, C5.0, logistic regression, baysian network by a lift chart. The horizontal line is a random selection model with 50% chance of correct churn detection.

According to the fig. 3, in this experiment OWA approach has the highest accuracy among these five prediction methods and baysian network and regression have worst results.

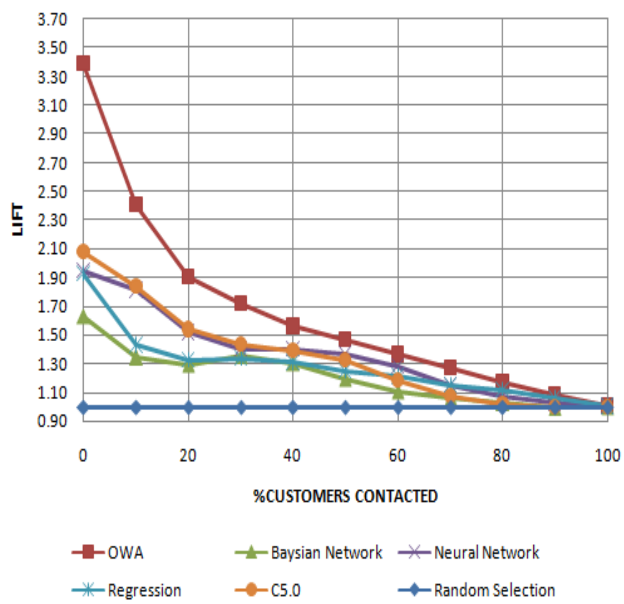

Figure 3. Lift chart for OWA and some other well-known prediction algorithm

Fig. 4 illustrates the performance of OWA and two other prediction algorithms which used as the inputs for OWA technique, including bagging and boosting classification tree and LOLIMOT by a lift chart. In this figure the results of four different bagging and boosting tree classifiers can be seen which have been learned using 10, 12, 13 and 24 most significant features. LOLIMOT algorithm has been learned using 5 and 32 most significant features too. As it can be seen, the OWA approach has the highest accuracy among these six classifiers.

Top-decile lift criterion for all applied prediction algorithms can be seen in table 2. These results also confirm the results of two lift charts which OWA performs very accurately in comparison with other prediction algorithms.

Table 2 Top-decile lift

Algorithm Name	Top-decile lift
OWA	**2.41**
C5.0	1.84
Neural Networks	1.80
Logistic Regression	1.44
Baysian Network	1.35
Bagging and Boosting-10 Features	2.14
Bagging and Boosting-12 Features	2.16
Bagging and Boosting-13 Features	2.25
Bagging and Boosting-24 Features	2.05
LILOMOT-5 Features	1.89
LOLIMOT- 32 Features	2.19

Four algorithms, LOLIMOT, C5.0, neural network and logistic regression, have been compared in [6] for this particular dataset. They have found the LOLIMOT algorithm to be the most accurate and logistic regression the worst one among these four algorithms, as we see in this experiment too.

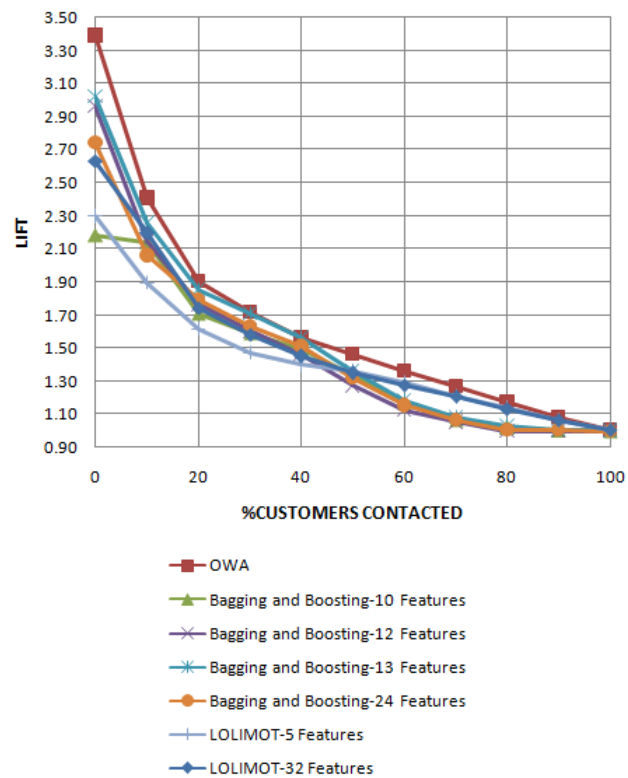

Figure 4. Lift chart for OWA and its inputs

To the best of our knowledge, bagging and boosting classification tree is the most accurate prediction algorithm in the churn management literature. This implementation has been done by A. Lemmens and C. Croux [8]. Table 3 illustrates this result and the result of our proposed hybrid approach.

Table 3 Top-decile lift of OWA and Bagging and Boosting

Algorithm Name	Reference	Top-decile lift
OWA	Our Proposed Approach	**2.41**
Bagging and Boosting	[8] The best result in the literature	2.29

As it can be seen, our proposed approach has a better performance in comparison with the bagging and boosting classification tree algorithm. Note that table 2 (our experiment) and table 3 [8] are different because all the used features and feature selection method have not been mentioned in [8].

To investigate the sensitivity of optimistic exponential OWA to the parameter α (equation 1) in our application, we have computed the top-decile lift of this algorithm for different values of α. We have changed this parameter from 0.3 to 0.6 increasing by 0.005. Fig. 5 shows the result of this experiment. The horizontal line is the result of bagging and boosting which is the best result in the literature.

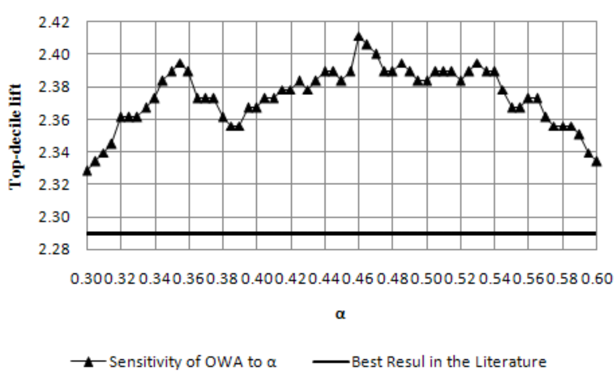

Figure 5. Sensitivity of OWA to the parameter α

From fig. 5, we can see, for all investigated values of parameter α in equation 1, the performance of OWA is better than bagging and boosting algorithm.

From the empirical results, we can conclude that uniting the capability of both bagging and boosting and LOLIMOT algorithms, the OWA has increased its performance compared to the other experimented algorithm. Also, our findings lead us to believe that using OWA technique could cause to increase profit for the companies.

VI. CONCLUSION AND FUTURE WORK

In this paper we have proposed a hybrid approach in order to improve the prediction accuracy of the existing churn management systems. In the proposed approach, two most important prediction algorithm in the churn management literature, bagging and boosting classification tree and LOLIMOT algorithm, are learned using different number of most significant features. For each instance, the decision lists of underlying learned classifiers have been fused using the optimistic exponential OWA technique. We have tested our hybrid approach by applying it to the database of a telecommunication company. Results gave us enough confident to be sure about its superiority to many other classifiers due to our comparison with some other well-known classifiers. Also, this method is found to significantly improve accuracy in predicting churn compared to the best existing result in the literature of the churn management systems.

In order to give more robustness to our hybrid method, the fusion technique may be developed. It can be improved by giving dynamicity to α, the currently static parameter in the equation (1). During the training phase it may be tuned instead of static setting.

ACKNOWLEDGEMENT

We thank the Iran Telecommunication Research Center for partially financial support. The authors thank Mr.Morteza Zihayat, Mr.Kaveh Kavousi and Mr. Amir Hosein Keyhanipour for friendly helping this research. We wish to thank the developers of Weka. We also express our gratitude to the Teradata Center for Customer Relationship Management at Duke University.

REFERENCES

[1] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram (2006), "Modeling customer lifetime value," Journal of Service Research, 9 (2), 139.

[2] Z. Jamal and RE Bucklin (2006), "Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach," Journal of Interactive Marketing, 20.

[3] B. Masand, P. Datta, Mani, DR., Li, B., "CHAMP: A Prototype for Automated Cellular Churn Prediction", Data Mining and Knowledge Discovery, vol. 3, no. 2, 1999, pp. 219-225.

[4] P. Datta, B. Masand, D. Mani, B. Li, "Automated cellular modeling and prediction on a large scale", Artificial Intelligence Review, vol. 14, no. 6, 2000, pp. 485-502.

[5] H. Noh, M. Kwak, and I. Han, Improving the prediction performance of customer behavior through multiple imputation. Intelligent Data Analysis, 2004. 8: p. 563-577.

[6] A. Ghorbani, F. Taghiyareh, and C. Lucas, "The Application of the Locally Linear Model Tree on Customer Churn Prediction." SoCPaR, 2009, pp. 472-477.

[7] C. Rygielski, J. Wang, and D. Yen, Data mining techniques for customer relationship management. Technology in Society, 2002. 24(4): p. 483-502.

[8] A. Lemmens, and C. Croux, "Bagging and boosting classification trees to predict churn," Journal of Marketing Research, vol. 43, no. 2, pp. 276-286, 2006.

[9] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends", Computers and Operations Research, vol. 34, no. 10, 2007, pp. 2902-2917.

[10] L. Breiman, Bagging predictors, Machine Learning, 24 (2) (1996) 123–140.

[11] Y. Kim, "Toward a successful CRM: variable selection, sampling, and ensemble," Decision Support Systems, vol. 41, no. 2, 2006, pp. 542-553.

[12] Y. Freund, R. Schapire, Experiments with a new boosting algorithm, Proceedings of the 13th Int'l Conf. on Machine Learning, Bari, Italy, 1996, pp. 148– 156.

[13] A. Ghorbani, and F. Taghiyareh, "CMF: A framework to improve the management of customer churn.", IEEE APSCC ,2009, pp. 457-462.

[14] D. Filev, and R. Yager, "On the issue of obtaining OWA operator weights," Fuzzy Sets and Systems, vol. 94, no. 2, 1998, pp. 157-169.

[15] H. Keyhanipour, B. Moshiri, M. Kazemian, M. Piroozmand, C. Lucas, "Aggregation of web search engines based on users' preferences in WebFusion", Knowledge-Based Systems 20 (2007) 321–328

[16] R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making", IEEE transactions on Systems, Man and Cybernetics 18 (1988) 183–190.

[17] R. Yager, & J. Kacprzyk, (Eds.), "The ordered weighted averaging operators: Theory and applications". Boston: Kluwer. (1997).

[18] J. Fodor, J. Marichal, and M. Roubens, "Characterization of the ordered weighted averaging operators," IEEE Transactions on Fuzzy Systems, vol. 3, no. 2, pp. 236-240, 1995.

[19] X. Liu, "A general model of parameterized OWA aggregation with given orness level," International journal of approximate reasoning, vol. 48, no. 2, 2008, pp. 598-627.

[20] X. Jin, A. Xu, R. Bie, P. Guo, Machine Learning Techniques and previous term Chi-Square Feature Selection next term for Cancer Classification Using {SAGE} Gene Expression Profiles, Data Mining for Biomedical Applications, {PAKDD} Workshop, Bio{DM} Proceedings, vol. 3916, Springer, 2006, pp. 106–115.

[21] S. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. Mason, "Defection detection: Measuring and understanding the predictive accuracy of customer churn models," Journal of Marketing Research, vol. 43, no. 2, 2006, pp. 204-211.