Cost Sensitive Feature Selection in Decision-Theoretic Rough Set Models for Customer Churn Prediction: The Case of Telecommunication Sector Customers

Emel Kızılkaya Aydogan, Mihrimah Ozmen, Yılmaz Delice

Abstract**—**In recent days, there is a change and the ongoing development of the telecommunications sector in the global market. In this sector, churn analysis techniques are commonly used for analysing why some customers terminate their service subscriptions prematurely. In addition, customer churn is utmost significant in this sector since it causes to important business loss. Many companies make various researches in order to prevent losses while increasing customer loyalty. Although a large quantity of accumulated data is available in this sector, their usefulness is limited by data quality and relevance. In this paper, a cost-sensitive feature selection framework is developed aiming to obtain the feature reducts to predict customer churn. The framework is a cost based optional pre-processing stage to remove redundant features for churn management. In addition, this cost-based feature selection algorithm is applied in a telecommunication company in Turkey and the results obtained with this algorithm.

Keywords**—**Churn prediction, data mining, decision-theoretic rough set, feature selection.

I. Introduction

NOWADAYS, on behalf of companies in telecommunications sector, to maintain market has become an important issue for the privatizations, market institutions and participation of new companies. Therefore, companies are obliged to adopt a customer-oriented approach in order to survive in the severe competition. Skill of making customers and keeping and controlling their loyalty depends on the speed and effectiveness of companies. To prevent the loss of customers, many companies are seeking ways to increase customer loyalty. Thus, telecommunication companies make surveys to customers and organize campaigns based on the results of these surveys not to lose customers.

"Customer churn" is a term used in the telecommunication industry to describe the customer movement from one provider to another, and "churn management" is a term that describes an operator's process to retain profitable customers [1]. In telecommunication sector, churn management has gained importance by increasing competition with the participation of new institutions into the market. As noted in Mattersion's study, handling loss of customers is the most important issue for the survival of the company for many telecommunications managers [2]. Therefore, for an effective churn management, companies should tend to hold customers instead of making new ones. Future success or failure of telecommunications companies depends on their knowledge of markets and customers. In addition, segmentation of customers allows developing marketing programs by grouping customers with their similar characteristics. Identifying segment of customers who tend to churn provides companies to develop customized marketing campaigns to those customers.

In this paper, we present a feature selection framework for customer churn management to the telecommunication company "TURKNET Communications Services Inc." in Turkey. The framework is based on minimum cost genetic approach in decision-theoretic rough set model (DTRS). This method ensures minimal subset of attributes with the minimum decision cost.

This paper is structured as follows. In the following section, we review the methods used in this study and some data mining approaches applied for predicting customer churn in telecommunication companies. The framework of case study is described and illustrated in Sections III and IV. In Section V, experimental results and their analysis are discussed. After that, we conclude in Section VI.

II. Literature

The telecommunication sector is turning into highly competitive market and needs a defensive marketing strategy. Developments and the restructuring bring forward new market opportunities and increase competition. Under these circumstances, managing customer churn is a crucial challenge in the telecommunications industry. Moreover, customer churn affects negatively these companies by decreasing profit levels. In this case, companies are directed to use marketing opportunities for adding value to products. With effective churn management, companies should be able to determine which customers are loyal and which are close to churn. If companies have this information, they develop strategies to

Emel Kızılkaya Aydoğan is with the Department of Industrial Engineering, Erciyes University, Talas 38039, Kayseri, Turkey (e-mail: emelkizilkaya@gmail.com).

Mihrimah Ozmen is with the Department of Industrial Engineering, Erciyes University, Talas 38039, Kayseri, Turkey.

Yılmaz Delice is with the Department of Logistics, Erciyes University, Develi Vocational High School, Talas 38039, Kayseri, Turkey.

prevent the loss of customers and reach the right customer at the right campaign. Therefore, to manage customer churn in a good way, companies should predict a customer's behavioural churn path and the factors.

Lately, data mining techniques have been used widely to cope with the customer churn challenging problem in telecommunication service field [3]-[10]. To develop the customer churn management in telecommunication service field, we present a new feature selection model. Our model is based on rough set model and the experimental results show that the presented feature reducts are effective.

In classical rough set models, most attribute reduction studies focus on generally positive or non-negative region maintains unchanged, while these regions do not decrease with additional features. Feature selection is an important subject of rough set theory and there are many studies on this subject [11]-[14]. In Pawlak rough sets and other generalized models, various quality degradation issues have been studied. Overall, feature reduction can be interpreted as the process of finding the minimum features set. The minimum set of attributes is called a reduct.

Reduct definitions can be classified in two categories: Qualitative and quantitative definitions. For qualitative descriptions, some qualitative criteria (such as dependency and the quality of the classification) are provided to describe a reduct. Pawlak defined the attribute reduct as keeping the attributes in the positive region unchanged [15]. For quantitative description, quantitative criteria are applied. They may be the corresponding criteria of the quantitative definitions.

In probabilistic model of rough sets, some qualitative properties in Pawlak rough set models cannot be kept longer. Therefore, some models were studied mostly with quantitative definitions. One of those, DTRS [16]-[19] is a probabilistic rough set model.

For attribute reduction in the DTRS models, Yao and Zhao [20] studied positive region, nonnegative region, confidence of rules, coverage of rules, cost of rules and other quantitative criteria. Also Jia et al. proposed a new attribute reduct definition with the objective of minimizing the cost of decisions for DTRS models [21]. Minimizing the decision cost is a very important concept for DTRS models, so it is necessary to take into account during feature selection algorithm. To define an attribute reduction, the decision cost can be regarded as heuristic objective criteria. Therefore, Jia et al. defined a minimum cost attribute reduct in DTRS models decreasing the cost of attribute reduct [21].

Feature selection is possible to improve problems with multi-objective algorithms. Therefore, the objective function of feature selection algorithm is applied to improve the results in the literature are as follows.

The objective function consists of classification of quality and the ratio of reduct cardinality [22], [23].

$$Fitness = \alpha * \gamma_R(D) + \beta * \frac{|C|-|R|}{|C|} \; \alpha\epsilon[0,1] \; ve \; \beta = 1-\alpha \quad (1)$$

- Fitness function is divided into two parts as classification accuracy and the reduction ratio [24].

$$Fitness(J) = \alpha * AccRate(J) + (1-\alpha) * RedRate(J) \quad (2)$$

$$AccRate(J) = K - NNAccuracy(J) \quad (3)$$

$$RedRate_{Features} = \frac{\#Features\ Selected}{M} \quad (4)$$

- Processing time and classification accuracy of the selected attributes are used as a fitness function [25]
- In DTRS models the significance level of the reduct and the expected value of the overall risk are considered together [26].

$$f(A) = \frac{|C|-|A|}{|C|} + \sum_{[x_i]A\in\pi_A} R(a_P|[x_i]_A) + R(a_B|[x_i]_A) + R(a_N|[x_i]_A) \quad (5)$$

- Min et al. proposed an algorithm, which minimizes cardinality, test cost of the reduct, and maximizes elements number of the reduct in the positive region [27].

$$c(B) = \sum_{a\in B} c(a) \;\; \forall\, B \subseteq C \quad (6)$$

$$Min\; c(B) \quad (7)$$

The cost of the test is calculated as the cost of testing each criterion given as input to the decision table. For example, the cost of doing a blood test to diagnose the disease is low; the cost of drawing tomography is high. Therefore, this model tries as much as possible to eliminate the drawing tomography criteria.

III. Decision-Theoretic Rough Set (DTRS)

In this part, firstly some basic notions and feature selection in DTRS models are reviewed. It is known that the DTRS model is a special case of probabilistic rough set models. In these models, a decision system (DS) can be defined as a 4-tuple, $DS = (U, A = C \cup D, V = \cup_{a\in A} V_a, \{f_a\})_{a\in A}$, where U is a finite non-empty set of objects, both $C\ and\ D$ are finite non-empty sets of attributes, condition and decision attribute set, respectively, where $C \cap D = \emptyset, V_a$ is a non-empty set of values of attribute $a \in A$, called the value domain of attribute a, and the number of elements in V_a is called the cardinality of V_a, denoted as $|V_a|; f_a: U \rightarrow V$ is an information function from $U\ to\ V$, which maps an object in U to a value in V_a.

Lower and upper approximations are important concepts in rough set theory.

$$\underline{BX} = \{x \in U | P(X|[x]_B) = 1\} = \cup\, \{[x]_B | P(X|[x]_B) = 1, x \in U\} \quad (8)$$

$$\overline{BX} = \{x \in U | P(X|[x]_B) > 0\} = \cup\, \{[x]_B | P(X|[x]_B) > 0, x \in U\} \quad (9)$$

$P(X|[x]_B = |X \cap [x]_B|/|[x]_B|$, denotes the degree to which condition class $[x]_B$ belongs to subset X.

There are two important parameters, α and β, where $\beta < \alpha$, are used to define the concepts of lower and upper approximations:

$$\underline{BX} = \cup \{[x]_B | P(X|[x]_B) \geq \alpha, x \in U\} \tag{10}$$

$$\overline{BX} = \cup \{[x]_B | P(X|[x]_B) > \beta, x \in U\} \tag{11}$$

The most significant difference between general probabilistic rough set models and DTRS models is that it is based on Bayesian risk theory. Through α and β parameters, DTRS models ensure that avoiding using subjective experience by theoretical support. According to the lower and upper approximations, (α,β)-positive region, (α,β) –boundary region, and (α,β)-negative region of a subset X in DTRS models are defined by:

$$POS_B^{(\alpha,\beta)}(X) = \underline{BX}, \tag{12}$$

$$BND_B^{(\alpha,\beta)}(X) = \overline{BX} - \underline{BX}, \tag{13}$$

$$NEG_B^{(\alpha,\beta)}(X) = U - \overline{BX} \tag{14}$$

IV. Minimum Cost Genetic Approach TO Feature Selection in DTRS

There are many applications of feature selection models with optimization algorithms. The minimum cost feature selection optimization problem is modelled as several heuristic algorithms such as genetic algorithm [28]. Jia et al. defined a genetic approach to minimum cost feature selection algorithm for the DTRS models [21]. This method applied genetic algorithm according to minimum cost approach. Also its objective function is to reduce the decision cost.

A decision table is the following tuple:

$$S = (U, At = C \cup \{D\}, \{V_a\}, \{I_a\}) R \subseteq C \tag{15}$$

Let $p(wj|x)$ be the conditional probability of an object x being in state wj.

The cost functions values relative to each other are as follows: $\lambda_{PP} = 0 \leq \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} = 0 \leq \lambda_{BN} < \lambda_{PN}$ that is, to classify a positive/negative region object x to positive/ negative region has not cost [21].

Jia et al. [21] presented the cost function as where A $\subseteq C$:

$$COST_A = \sum_{P_i \geq \alpha}(1-p_i).\lambda_{PN} + \sum_{\beta < P_j < \alpha}(p_j.\lambda_{BP} + (1-p_j).\lambda_{BN}) + \sum_{P_k \leq \beta} p_k.\lambda_{NP} \tag{16}$$

The reduct definition is:

$$(1) COST_R \leq COST_C \tag{17}$$

$$(2) \forall R' \subset R, COST_{R'} > COST_R \tag{18}$$

In this definition, condition (1) is the jointly sufficient condition and ensures that reduct cost minimum. Condition (2) is the individual necessary condition and ensures that reduct element number is minimum.

The chromosome structured as a binary string and each of them length is number of feature. The value "1"shows that the correponding attribute is selected to subset feature and "0" means that it is not selected.

The fitness function of algorithm:

$$f = COST_R + (\tfrac{|R|}{|C|})^\theta \tag{19}$$

The fitness function obtains both minimum cost and fewer reduct size. The parameter θ is set to 1 in our applications.

The genetic approach to minimum cost is described:

Input: A decision table
Output: A reduct R
1: BEGIN
2: create an initial random population
3: evaluate the population
4: **WHILE** the result is not convergent AND number of generations < maximum number of generations
5: select the fittest individuals in the population
6: perform crossover on the selected individuals to create offspring
7: create the new population evaluate the new population
8: END WHILE
9: selected fittest individual from current population and output is as R
10: END BEGIN

V. The Case Study

The case company is one of the telecommunication companies "TURKNET" that provide internet access, landline phone and data centre services in Turkey. In order to evaluate the proposed minimum cost genetic approach, 499 customers are randomly selected from TURKNET telecom. There are 167 churners, 332 non churners, and a total of 499 customers in the dataset. In addition, there are 21different attributes in the dataset. The dataset attributes are given in the literature [29]-[35]. The main topic of features and hierarchical structure is given in Fig. 1.

For TURKNET data set genetic approach is implemented. The cost functions are randomly generated according to constraint condition: $\lambda_{PP} = 0 \leq \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} = 0 \leq \lambda_{BN} < \lambda_{PN}$. In the experiments, 10-fold cross validation is employed, and the population size: 100, maximum evaluations are 200 generations. According to experiments the best result we obtain is:

$$\lambda_{PP} = 0, \lambda_{BP} = 1, \lambda_{NP} = 3, \lambda_{NN} = 0, \lambda_{BN} = 3 \text{ and } \lambda_{PN} = 6$$

According to cost functions the parameters obtained: $\alpha = 0{,}75$ and $\beta = 0{,}6$. The reduct size is 7 and the cost value is 59,408. Moreover, the reduct set is:

1. Using few months, the service first get
2. Is there commitment
3. Type of service
4. The period starting from the date of contract
5. Elapsed time from the contract completion date

6. The period of contract
7. The first date of the customer's call in a certain time period.

VI. Conclusion

This paper, we presented the application of a cost based feature reduction approach for churn management in telecommunication sector in Turkey. The experiments applied with minimum cost approach for DTRS models. The experiments results are considered with company and the obtained reduct is confirmed to be consistent. In the future researches, we will study feature reduction algorithm with ensemble learning algorithms.

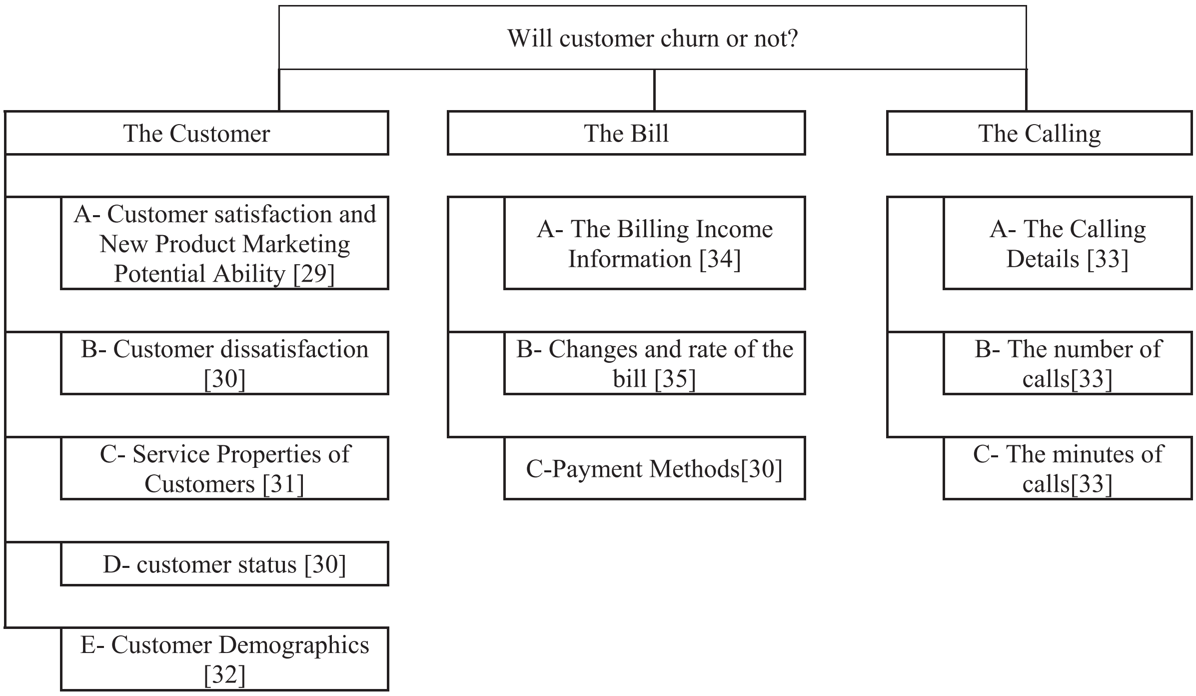

Fig. 1 Hierarchical Structure of Features

Acknowledgments

Authors would like to thank the Ministry of Science, Industry and Technology (Republic of Turkey; Project No: 0777.STZ.2014) for their contributions to the study.

References

[1] Berson, A., & Smith, S. J. (2002). Building data mining applications for CRM. McGraw-Hill, Inc.
[2] Mattersion, R. (2001). Telecom churn management. Fuquay-Varina, NC: APDG Publishing.
[3] Au, W., Chan, C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation, 7, 532–545.
[4] Coussement, K., & den Poe, D. V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter selection techniques. Expert Systems with Applications, 34, 313–327.
[5] Lu, J. (2002). Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS. SAS User Group International (SUGI27) Online Proceedings, 114-27.
[6] John, H., Ashutosh, T., Rajkumar, R., Dymitr, R. (2007). Computer assisted customer churn management: State-of-the-art and future trends.
[7] Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: A data mining approach. Expert Systems with Applications, 23, 103–112.
[8] Kim, M. K., Park, M. C., & Jeong, D. H. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. Telecommunications policy, 28(2), 145-159.
[9] Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), 1414-1425.
[10] Huang, B., Buckley, B., & Kechadi, T. M. (2010). Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. Expert Systems with Applications, 37(5), 3638-3646.
[11] Q.H. Hu, H. Zhao, Z.X. Xie, D.R. Yu, Consistency based attribute reduction, in: Proceedings of PAKDD2007, LNAI, vol. 4426, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 96–107.
[12] X.Y. Jia, K. Zheng, W.W. Li, T.T. Liu, L. Shang, Three-way decisions solution to filter spam email: an empirical study, in: Proceedings of RSCTC2012, LNAI, vol. 7413, 2012, pp. 287–296.
[13] H.X. Li, X.Z. Zhou, Risk decision making based on decision-theoretic rough set: a three-way view decision model, International Journal of Computational Intelligence Systems 4 (1) (2011) 1–11.
[14] D. Liu, H.X. Li, X.Z. Zhou, Two decades' research on decision-theoretic rough sets, in: Proceedings of ICCI, 2010, pp. 968–973.
[15] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, MA, 1991.
[16] Y.Y. Yao, Probabilistic approach to rough sets, Expert Systems 20 (2003) 287–297.
[17] Y.Y. Yao, Probabilistic rough set approximations, International Journal of Approximate Reasoning 49 (2008) 255–271.
[18] Y.Y. Yao, S.K.M. Wong, A decision theoretic framework for approximating concepts, International Journal of Man–Machine Studies 37 (6) (1992) 793–809.
[19] Y.Y. Yao, S.K.M. Wong, P. Lingras, A decision-theoretic rough set model, in: Proceedings of the 5th International Symposium on Methodologies for Intelligent Systems, 1990, pp. 17–25.
[20] Y.Y. Yao, Y. Zhao, Attribute reductions in decision-theoretic rough set models, Information Sciences 178 (2008) 3356–3373.

[21] Jia, X., Liao, W., Tang, Z., & Shang, L. (2013). Minimum cost attribute reduction in decision-theoretic rough set models. Information Sciences, 219, 151-167.
[22] Wang, X., Yang, J., Teng, X., Xia, W., & Jensen, R. (2007). Feature selection based on rough sets and particle swarm optimization. Pattern Recognition Letters, 28(4), 459-471.
[23] Wang, X., Yang, J., Jensen, R., & Liu, X. (2006). Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma. Computer methods and programs in biomedicine, 83(2), 147-156.
[24] Derrac, J., Cornelis, C., García, S., & Herrera, F. (2012). Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. Information Sciences, 186(1), 73-92.
[25] Sun, L., Xu, J., & Tian, Y. (2012). Feature selection using rough entropy-based uncertainty measures in incomplete decision systems. Knowledge-Based Systems, 36, 206-216.
[26] Chebrolu, S., & Sanjeevi, S. G. (2015). Attribute Reduction in Decision-Theoretic Rough Set Model using Particle Swarm Optimization with the Threshold Parameters Determined using LMS Training Rule. Procedia Computer Science, 57, 527-536.
[27] Min, F., Hu, Q., & Zhu, W. (2014). Feature selection with test cost constraint. International Journal of Approximate Reasoning, 55(1), 167-179.
[28] Aydogan E. K.et.al, (2012) hGA: Hybrid genetic algorithm in fuzzy rule-based classification systems for high-dimensional problems, Appl. Soft Comput., vol. 12, no. 2, pp. 800806.
[29] Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. Applied Soft Computing, 24, 994-1012.
[30] Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommunications policy, 30 (10), 552-568.
[31] Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. Expert Systems with Applications, 31(3), 515-524.
[32] Kim, N., Jung, K. H., Kim, Y. S., & Lee, J. (2012). Uniformly subsampled ensemble (USE) for churn management: Theory and implementation. Expert Systems with Applications, 39(15), 11839-11845.
[33] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1-9.
[34] Huang, B., Buckley, B., & Kechadi, T. M. (2010). Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. Expert Systems with Applications, 37(5), 3638-3646.
[35] Zhang, Y., Qi, J., Shu, H., & Li, Y. (2006, October). Case study on crm: Detecting likely churners with limited information of fixed-line subscriber. In Service Systems and Service Management, 2006 International Conference on (Vol. 2, pp. 1495-1500). IEEE.