VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2016 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com

PREDICTIVE MODELING FOR TELCO CUSTOMER CHURN USING ROUGH SET THEORY

N. S. M. Nafis[1], M. Makhtar[1], M. K. Awang[1], M. N. A. Rahman[1] and M. M. Deris[2]
[1]Faculty of Informatics and Computing, Universiti Sultan, Kuala Terengganu, Malaysia
[2]Faculties of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Malaysia
E-Mail: nsyafiqahmnafis@gmail.com

ABSTRACT

A rough set is a mathematical tool to handle imprecise and imperfect information. It has been increasing in popularity recently in Knowledge Discovery in Database (KDD) and Machine Learning application. Rough set is one of the techniques used in KDD data mining. Data mining is an approach to extract useful information from a massive database for business purposes, for example, classifying customer churn. Churn is customer behaviour to terminate a service in favour of a competitor. Identifying customers who are likely to churn in the early stage will help firms to increase profitability since acquiring new customers is costly compared to retaining existing one. Limited research in investigating customer churn using machine learning techniques had led this research to discover the potential of rough set theory to enhance customer churn classification. This paper proposes a rough set predictive classification framework for customer churn in Telecommunication Companies. Experimental results show that the classification model is able to classify up to 83% to 98% accuracy for customer churn dataset. Overall, this indicates that the rough set theory is effective to classify customer churn compared to traditional statistical predictive approaches.

Keywords: customer churn, rough set, classification model, telecommunication.

INTRODUCTION

Customer churn or defection is a global phenomenon that threatens firms in a variety of industries. Churn can be defined as customer behaviour to leave or terminate a service for a competitor. In the telecommunication industry, they face the same problem in customer churn management since the industry has been growing rapidly in recent years. However, firms are not aware that information stored in their database can be useful for business strategies. However, information regarding customer's personal records, daily transaction or even monthly usage is normally incomplete, imprecise and vague, making it difficult to predict future churners. In order to handle this problem, firms should have an efficient churn predictive model since companies already spend significant sums to acquire new customers. In short, acquiring new customers is expensive compared to retaining existing ones [1].

Sunil Gupta, the Edward W. Carter Professor of Business Administration at Harvard Business School, argues that companies often fail to take into account the complete value of the customers that they are trying to retain. Churn can be distinguished into two categories, involuntary and voluntary. ¬Involuntary churn occurs when the service provider terminates customer account because of fraud and non-payment. Meanwhile, voluntary churn can be varied and is more complex to handle since it is an unexpected customer behavioural action such as the customer decides to churn due to terrible service. The focus in this paper is voluntary churn since it is an unexpected occurrence.

Data mining in Knowledge Discovery in Database (KDD) is one of the famous approaches for classification. Data mining can be divided into two types, predictive data mining and descriptive data mining [2]. Predictive data mining aims to predict the performance of various variables while descriptive data mining generates new data and performance that describe the behavioural patterns of variables based on the available data set. Classification, regression and clustering are the most common tasks in data mining. Prediction through classification is one of the common techniques used in data mining [3]. Thus, the terms predict and classify are interchangeably used in this paper. This paper mainly proposes how to apply rough set theory KDD method to classify and analyse customer churn in Telecommunication Company.

This paper is organised as follows: Section 2 describes all the related works while section 3 recalls Rough Set Theory methods and techniques used in the experiment. Section 4 provides detailed explanation on the result and discussion. Finally, the conclusion is discussed in section 5.

RELATED WORK

Past research in customer churn management showed that there are many approaches involved in handling customer churn problems. Some researchers found that decision tree is sufficient to tackle customer churn predictive modeling [4] but decision tree classification is not suitable for continuous attribute value. Meanwhile, Zhang *et al.* [5] applied Bayesian neural network to design a behaviour-based telecom customer churn prediction system. Naïve Bayes classifier had also been applied to solve customer churn prediction [6] but this simple probabilistic classifiers did not offer a readable

©2006-2016 Asian Research Publishing Network (ARPN). All rights reserved.

if-else rule. Furthermore, in recent years, combined sampling and weighted random forest are proposed to handle imbalanced data in customer churn prediction [7]. The mixed methods also offer complex model and algorithms.

Nevertheless, soft computing technique is becoming famous nowadays and is relevant in predicting customer churn in various fields [8]. For example, [9] implement multilayer perceptron to predict customer churn. However, Rough Set is one of the most common approaches used because Rough set theory is a powerful mathematical framework to acquire information about the data that cannot be categorised with traditional set theory [10]. Moreover, rough set can tackle some basic problems in data analysis such as discovery of dependencies between attributes, superfluous attributes reduction as well as seeking the most significant attributes and decision rules generation from the reduction set. In addition, it offers readable minimal sets of decision rules from the data (if-then rules). Thus, it is easy to understand and offers straightforward interpretation of result analysis.

Rough set has been used worldwide in numerous applications, including feature selection [11], stock market prediction [12] and hybridisation with fuzzy set. Rough set has already been used for medical image segmentation [13] and designing diabetic diagnose system in India [14]. Rough set also has a good research base in multimedia data management [15] and accident chain exploration [16]. The approximation concept in Rough Set can be implemented for classifying customer churn. In short, employing rough set to classify customer churn is currently relevant.

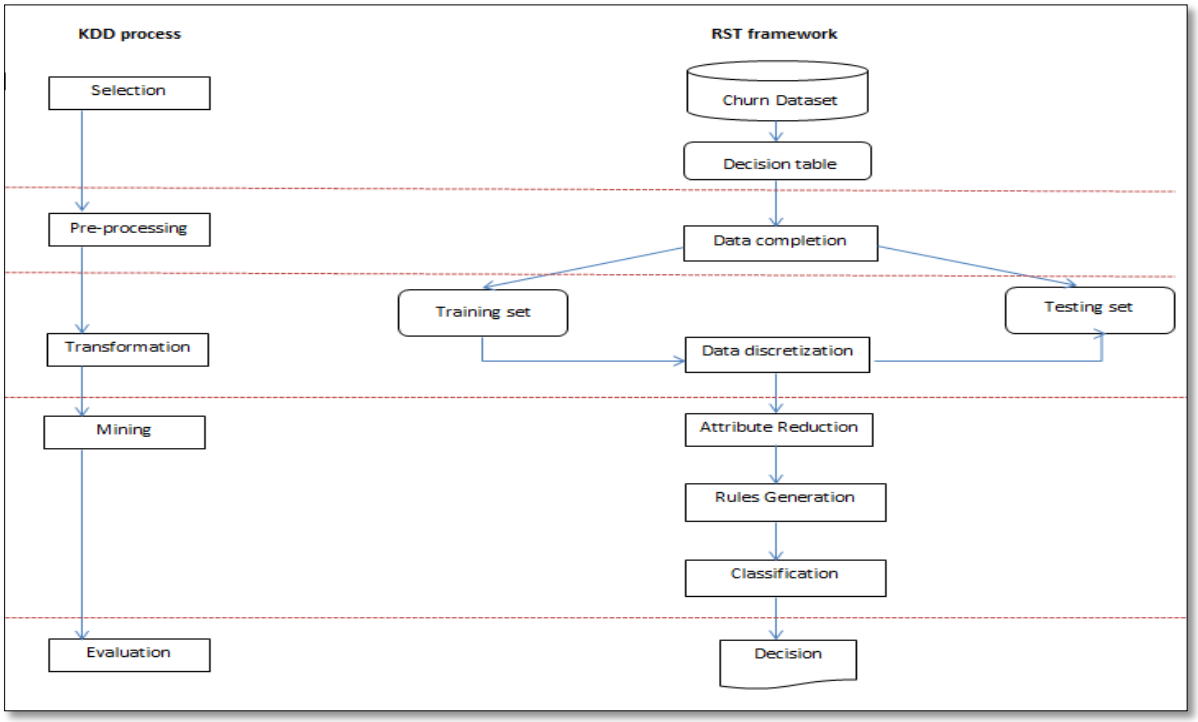

Figure-1. KDD Rough set predictive modelling framework.

METHODS AND TECHNIQUES

The proposed KDD rough set predictive modelling framework mainly comprises of five phases, which are data selection, pre-processing, transformation, mining and evaluation as shown in Figure-1. Firstly, specific data had to be selected. In this case, TELEKOM Malaysia customer churn dataset is retrieved. The dataset contains 312 objects with 20 attributes excluding one decision attribute. From the 20 attributes, only 7 are selected using WEKA data mining. Table-1 shows decision table for customer churn dataset.

Table-1. Customer churn dataset sample.

Object	C					D
	A1	A2	A3	A4	A5	Churn
O_1	1	1	1	1	1	1
O_2	1	2	2	1	4	1
O_3	1	3	3	2	2	0
O_4	1	4	4	2	2	0
O_5	0	4	4	2	2	1
O_6	0	5	5	2	2	1

VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2016 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com

Rough set theory

In rough set theory, the relation between two objects or more is called indiscernibility where all the values are identical in relation to a subset of a considered attribute. Given a subset of attributes, $\alpha \in A$ and $B \subseteq A$, each such subset defines an equivalence relation $IND_A(B)$ called an indiscernibility relation. This indiscernibility relationship can be defined as follows:

$$INDA\ (B) = \{(x,x) \in U^2 \mid \forall \alpha \in B, \alpha(x) = \alpha(x)\} \quad (1)$$

From equation (1), the subset of attributes will be defined as a classification process of the universe into sets such that each object in a set cannot be distinguished from other objects in the set using only the attributes in *B*. The sets into which the objects are divided are known as equivalence class. From Table-1, objects O_4 and O_5 can be classified as indiscernibility relation. The examples of indiscernibility relations available in Table-1 are:

$$IND\ (A1) = \{\{O_1, O_2, O_3\}, \{O_4, O_5, O_6\}\}$$
$$IND\ (A4, A5) = \{\{O_1\}, \{O_2\}, \{O_3, O_4, O_5, O_6\}\}$$
$$IND\ (A1, A2, A3) = \{\{O_1\}, \{O_2\}, \{O_3\}, \{O_4, O_5\}, \{O_6\}\} \quad (2)$$

Then, the concept of approximation is required to class the objects based on the equivalence class. Now, to define two approximations, called the P-lower and the P-upper approximation of *X* respectively where,

$$\underline{P(X)} = \{x \in U : [x]P \subseteq X\}_{(2)}$$
$$\overline{P(X)} = \{x \in U : [x]P \cap X \neq \Phi\} \quad (3)$$

Lower approximation is a set that contains all objects for which the equivalent class corresponding to the object is the subset of the set. Besides, this set contains all objects that certainly belong to set *X*. Meanwhile, upper approximation is a set that contains objects for which the intersection of the object has an equivalent class and the set is not an empty set. In fact, this set contains all objects that might possibly belong to set *X*. Moreover, the boundary of *X* for given $B \subseteq A$ and $X \subseteq U$ in IS can be defined as:

$$PN_P = \overline{P(X)} - \underline{P(X)} \quad (4)$$

PN_P Consists of objects that do not certainly belong to X on the basis of A. Based on Table *, the lower and upper approximation of X can be classified as in the following:

$$\underline{P(X)} = \{O_1, O_2, O_3, O_6\}.$$

$$\overline{P(X)} = \{O_1, O_2 O_3, O_4, O_5, O_6\}.$$

$$Then,\ PN_P = \{O_4, O_5\}.$$

A reduction set or so called 'reduct' is a minimal set of attributes that is obtained after removing the redundant and insignificant attributes, but preserving the original classification. Nevertheless, in some cases, not all attributes are required to classify an object. A reduct of A is defined as a minimal set of attributes in the original classification defined by $B \subseteq$ A such that $IND_A(B) = IND_A(A)$. Therefore, in this designed model example, in order to discern between the different equivalent classes, only if attributes for Call Plan and Service Rental had been necessary and the example of reduct is:

$$IND_A(\{Packages, Call\ Plan\}) = IND_{(A)}$$

In addition, Table-2 depicts an example of decision table after the reduction process, whereby attributes Gender, Monthly Commitment, and service Rental had been dropped. As a result, the decision rules in the 4th and the 5th rows in Table displayed similar conditional attributes, but varying decisions. Thus, the rules had been marked as inconsistent. Meanwhile, rules in the 1st and the 2nd rows exhibited consistency. Hence, in order to handle inconsistency in such decision table, rough set approximation concept is required [17].

Table-2. Example of decision table after reduction process.

Object	Decision Attribute		Condition Attribute
	Packages	Call Plan	Churn?
O_1	1	1	1
O_2	2	1	1
O_3	3	2	0
O_4	4	2	0
O_5	4	2	1
O_6	5	2	1

Furthermore, the approximation of the decision, D, can be defined by constructing the decision rules from Table-2 depicted above. From these generated approximate decision rules, the objects can be successfully classified. From Table-2, the rules cannot be exactly classified, but only be determined approximately. Basically, the rules were presented as implication *"if....then..." rules*. The rules that had been constructed are listed in the following:

1. Rule 1, if (packages, 1) and (call plan, 1) then (churn, 1)

2. Rule 2, if (packages, 2) and (call plan, 1) and then (churn, 1)

©2006-2016 Asian Research Publishing Network (ARPN). All rights reserved.

3. Rule 3, if (packages, 5) and (call plan, 2) and then (churn, 1)

4. Rule 4, if (packages, 3) and (call plan, 2) and then (churn, 0)

5. Rule 5, if (packages, 5) and (call plan, 2) and then (churn, 1)

6. Rule 6, if (packages, 5) and (call plan, 2) and then (churn, 0)

In conclusion, Rules 1, 2, and 3 can be certainly classified as causing churn, while Rule 4 can be certainly classified as not causing churn. Lastly, Rules 5 and 6 can be possibly classified as causing churn and not causing churn.

EXPERIMENTAL RESULTS

This section presents the evaluation of Rough Set based classifier and Non-Rough Set Classifier using different split factors. In the experiment using Rough Set classifier, attribute reduction is involved prior to classification. Meanwhile, for Non-Rough Set classifier experiment, no attribute reduction is employed because Naïve Bayes classifier is based on probability calculation. Thus, attribute reduction is not required [18].

Table-3 and Table-4 summarise results obtained from both classifiers. From Table-2 below, Rough Set classifier performed the best at split factor 0.9. At 0.9, 90% of dataset is used for learning purpose while 10% for testing set. So, rules from learning set are adequately capable to classify churners and non-churners. However, at split factor 0.4 to 0.1, accuracies are decreasing. This may be caused by under-fitting problem. Under-fitting occurs when rules or patterns retrieved from learning set are not enough to classify the whole testing set.

Table-3. Standard voting/tuned (RSES) classification accuracy.

Split Factor	Accuracy (%)
0.9	90.32
0.8	75.81
0.7	75.53
0.6	76.80
0.5	76.92
0.4	69.52
0.3	71.56
0.2	72.40
0.1	60.85

Table-4. Naïve Bayes classification accuracy.

Split Factor	Accuracy (%)
0.9	77.40
0.8	80.60
0.7	87.20
0.6	82.40
0.5	78.20
0.4	75.40
0.3	72.90
0.2	59.60
0.1	64.80

Referring to Table-3 above, Naïve Bayes classifier achieved the best accuracy at 0.7 split factors. It means that 70% of the dataset is enough to generate rules which are capable of classifying customer churn at optimum rate. Meanwhile, accuracies began to decrease when split factor was changed from 0.4 to 0.1 respectively, cause by under-fitting in testing set.

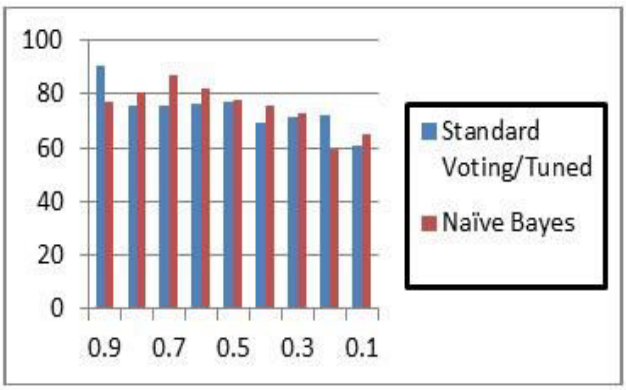

Figure-2. Comparisons of accuracy.

Based on Figure-2 above, it can be concluded that Rough Set based classifier performed better at all split factors.

CONCLUSIONS

This paper presents a comparative study of proposed rough set theory predictive modelling for customer churn with respect to non-rough set classifier to rough set classifier. This paper analyses rough set predictive model using Rough Set Technical Analysis Software. The experimental result shows that the proposed predictive modelling leads to significant improvement compared to the existing classifier. Thus, it can be concluded that rough set theory offers a powerful mathematical framework to classify customer churn for business purposes. In future, this research will investigate more on predictive modelling for churn dataset using different rough set toolkits.

©2006-2016 Asian Research Publishing Network (ARPN). All rights reserved.

ACKNOWLEDGEMENTS

This work is partially supported by Universiti Sultan ZainalAbidin (UniSZA) and The Ministry of Education, Malaysia (Grant No.FRGS/2/2013/ICT07/UniSZA/02/2).

REFERENCES

[1] A. Sharma. 2011. Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications. 27(11): 26–31.

[2] R. A. Soeini, K. V. Rodpysh. 2012. Applying Data Mining to Insurance Customer Churn Management. IACSIT Hong Kong Conference. 30: 82–92.

[3] J. Hadden. 2008. A Customer Profiling Methodology for Churn Prediction.Cranfield University.

[4] L. Bin, S. Peiji, L. Juan. 2007. Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service. International Conference Service Systems and Service Management.

[5] Y. Zhang, M. Berry. 2011. Behavior-Based Telecommunication Churn Prediction with Neural Network Approach. International Symposium on Computer Sciance and Society. pp. 307–310.

[6] C. Kirui, L. Hong, W. Cheruiyot, H. Kirui. 2013. Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. International Journal of Computer Science. 10(2): 165–172.

[7] V. Effendy. 2014. Handling Imbalanced Data in Customer Churn Prediction Using Combined Sampling and Weighted Random Forest. Information and Communication Technology (ICoICT), 2014 2nd International Conference. pp. 325–330.

[8] S. Sarkar. 2014. An Improved Rough Set Data Model for Stock Market Prediction. Business and Information Management (ICBIM), 2014 2nd International Conference. pp. 96–100.

[9] M. R. Ismail, M. K. Awang, M. N. A. Rahman, M. Makhtar. 2015 Multi-Layer Perceptron Approach for Customer Churn Prediction. International Journal of Multimedia and Ubiquitous Engineering. 10(7): 213–222.

[10] Z. Pawlak. 1996. Rough Sets and Data Analysis. Fuzzy Systems Symposuim. pp. 9–14.

[11] R. W. Swiniarski, A. Skowron. 2003. Rough set methods in feature selection and recognition. Pattern Recognition Letters. 24(6): 833–849.

[12] M. Khoza, T. Marwala. 2011. A rough set theory based predictive model for stock prices. Computational Intelligence and Informatics (CINTI), 2011 IEEE 12th International Symposium. pp. 57–62.

[13] N. Senthilkumaran, R. Rajesh. 2009. A Study on Rough Set Theory for Medical Image Segmentation. International Journal of Recent Trends in Engineering. 2(2): 2–4.

[14] M. Anouncia R. T. Nandhini. 2013. Design of a Diabetic Diagnosis System Using Rough Sets P. Cybernetics and Information Technologies. 13(3): 124–139.

[15] M. N. A. Rahman, Y. M. Lazim, F. Mohamed. 2011. Applying Rough Set Theory in Multimedia Data Classification. International Journal on New Computer Architectures and Their Applications (IJNCAA). 1(3): 683–693.

[16] J.-T. Wong,Y.-S. Chung. 2007. Rough set approach for accident chains exploration. Accident analysis and prevention. 39(3): 629.

[17] Z. Pawlak. 1996. Rough sets and data analysis. Soft Computing in Intelligent Systems and Information Processing. Proceedings of the 1996 Asian Fuzzy Systems Symposium. 147: 1–12.

[18] N. Syafiqah, M. Mokhairi, M. N. A.Rahman, M. D. Mustafa, A. Mohd Khalid. 2015. Feature Selections and Classification Model for Customer Churn. Journal of Theoretical and Applied Information Technology. 75(3): 356–365.