INITIALIZING GENETIC PROGRAMMING USING FUZZY CLUSTERING AND ITS APPLICATION IN CHURN PREDICTION IN THE TELECOM INDUSTRY

Bashar Al-Shboul [1], Hossam Faris [2], Nazeeh Ghatasheh [3]

[1, 2] The University of Jordan, Amman, Jordan.

[3] The University of Jordan, Aqaba Campus, Aqaba, Jordan.

Email: [1] b.shboul @ ju.edu.jo, [2] hossam.faris @ ju.edu.jo, [3] n.ghatasheh @ ju.edu.jo

ABSTRACT

Customer defection or "churn" rate is critically important since it leads to serious business loss. Therefore, many telecommunication companies and operators have increased their concern about churn management and investigated statistical and data mining based approaches which can help in identifying customer churn. In this paper, a churn prediction framework is proposed aiming at enhancing the predictability of churning customers. The framework is based on combining two heuristic approaches; Fast Fuzzy C-Means (FFCM) and Genetic Programming (GP). Considering the fact that GP suffers three different major problems: sensitivity towards outliers, variable results on various runs, and resource expensive training process, FFCM was first used to cluster the data set and exclude outliers, representing abnormal customers' behaviors, to reduce the GP possible sensitivity towards outliers and training resources. After that, GP is applied to develop a classification tree. For the purpose of this work, a data set was provided by a major Jordanian telecommunication mobile operator.

Keywords: Churn Prediction; Fuzzy Clustering; Genetic Programming

1.0 INTRODUCTION

Nowadays, presence of multiple service providers in the mobile telecommunication industry creates an intensive competitive environment for acquiring customers. Given a mobile subscriptions rate of 140%, Jordan poses the first place in the region in mobile penetration[1]. By shrinking the technology gap between mobile operators, customers have become the most valuable asset of each company; therefore, it became absolutely important for each operator to keep their customers; however, it is possible for a customer to have more than one subscription with single, or different, service providers concurrently. Customers may choose to have different subscriptions with the same service provider, different providers, or even switch/churn completely to another. Churn is the case when customers leave the service provider and cancel their subscriptions. Due to the importance of predicting churning customers, many researchers tackled the churn issue generally and addressed the importance of the issue in the telecommunication sector [5, 6, 14, 20, 21, 25, 30]. Basically, service providers are concerned in predicting when a churn might happen so that Customer Relationship Management (CRM) division can suggest activities intending to maximize the lifetime of a customer (e.g. acquisition, follow-up, retention, etc.) .

Detecting a churn by observation is almost impossible; therefore, service providers try to utilize artificial intelligence techniques hoping for better pre-churn prediction information feed for the CRM department [4]. Among the various techniques used for predicting customer churn [11, 18, 24, 27, 29, 30] Genetic Programming (GP) was not investigated. Given the problem described in this paper, we believe that GP features can significantly contribute to a better churn prediction. However, GP like any other evolutionary method (e.g. Genetic Algorithms, Swarm Intelligence, etc.) is sensitive to the initial steps, where encoding genes used in the

---

1 http://www.itu.int/net4/itu-d/icteye/

213

Malaysian Journal of Computer Science. Vol. 28(3), 2015

GP may contribute towards effectiveness, and therefore a method has to be used to assure good initialization (i.e. Fast Fuzzy C-Means).

Fast Fuzzy C-Means (FFCM) represents an unsupervised clustering technique that aims at shaping clusters and finding representatives for each (i.e. Centroids) efficiently, if compared to the traditional Fuzzy C-Means (FCM), and effectively [3]. The basic idea is that FFCM reduces the number of calculations within consecutive iterations by disclosing centroid calculations to the nearest neighbors appearing within a threshold area. One important feature of FFCM, differently from other statistical clustering methods, is that centroid calculations are not affected with the dataset outliers filtered by a certain membership threshold [3], and thus centroids are more representative to their neighbors.

In this paper we propose a framework for predicting customer churn in telecommunication companies. The framework is based on combining FFCM and GP as a hybrid churn prediction technique. FFCM is used to perform data refinement by eliminating outliers representing odd customer behaviors. After that, GP is applied to develop the final prediction model. The approach was later evaluated using different criterion and compared to other state of the art classification techniques.

This paper is structured as follows. In the following section, we review some machine learning and hybrid approaches applied for predicting customer churn in telecommunication companies. The proposed framework is described and illustrated in section 3. Section 4 describes the dataset used in our work. Evaluation method is described in Section 5. In section 6 experimental results and their analysis are discussed. After that we conclude in section 7.

2.0 RELATED WORK

Heuristic methods (Alternatively called approximating techniques) are optimization techniques used in problem solving, by reaching an effective solution efficiently [18]. In churn analysis, they are applied to find the churn prediction for a customer, or a set of customers. Among various approaches used in literature for predicting churners [12,24,27,29] heuristic based techniques are gaining more focus due to their ability of finding nearly optimal solutions with low time complexity [18]. In [12], for example, Neural Network (NN) based approach that estimates the possibility of churn based on customer complaints was examined. Compared to a Bayesian framework-based Neural Network (NN-BN), NN-BN was better. On the other hand, predicting non-churning customers was highly accurate by linear regression. After all, regression trees were the best among the all four methods (i.e. NN, NN-BN, Linear Regression, and Regression Trees). In [26] the use of NN based on Genetic Algorithms in predicting churn was investigated. In a nutshell, cellular service data was analyzed using a genetic algorithm based NN resulting in a better prediction than the z-score statistical model. Another attempt is [29] where authors investigated Support Vector Machine (SVM) for churn prediction. Their dataset contained different categories of input variables including: customers' demographic data, quality of service, and marketing-related options. Unfortunately, these features were unavailable in our dataset due to the company policy. Authors compared their results with other different algorithms including: NN, Decision Trees and Naive Bayes showing outstanding accuracy for SVM compared to other methods.

Fast Fuzzy C-Means (FFCM) was first proposed by [3]. The algorithm extends Fuzzy C-Means (FCM), i.e. a state of the art fuzzy clustering algorithm. Basically, FCM is an iterative clustering technique that calculates a fuzzy membership score for each data point with respect to a centroid (i.e. cluster center). At first, centroids are calculated after data points are randomly assigned a cluster. After that, membership between points and centroids is measured, and points are assigned to the nearest centroid again. The process continues until centroids stop changing significantly. As FCM centroids are calculated from the data points assigned to their clusters, memberships will be measured for all data points at every iteration. Fast Fuzzy C-Means, however, reduces the number of calculations and measures more accurate centroids by setting a membership threshold where all points with stronger membership scores will be included in centroid calculations, while others will be ignored. Membership threshold is usually set empirically. Even though FFCM was studied solely for reducing data dimensionality and applied in different other domains (i.e. [2]), to the best of our knowledge it was never been used in a hybrid approach, or applied in any churn prediction related topics.

Genetic Programming (GP) is a domain independent evolutionary algorithm, inspired by the biological evolution theories, which automatically creates computer programs [16, 17, 31]. GP has some advantages when

214

Malaysian Journal of Computer Science. Vol. 28(3), 2015

used to model complex problems including: flexibility, automatic feature selection, and interpretability [9, 15] and therefore GP has been applied and investigated in a wide range of complex problems in both business and finance domains among others. Some of these applications can be found in [8, 10, 11, 19, 23]. In this work, GP uses data provided by FFCM to build churn prediction model.

Applying hybrid approaches to the problem of Churn prediction gives some insights regarding this methodology. An approach is considered hybrid when it applies more than one approach sequentially. Theoretically, hybrid approaches can overcome some of the weaknesses or disadvantages of a single approach. Tsai and Lu, 2009, applied a hybrid approach that consists of both a Neural Network and Self-Organizing Maps (SOM) aiming to remove the outliers and perform data reduction. The results nominate a SOM-based hybrid approach in terms of better performance. Also a hybrid approach, using K-means and Neural Networks for data cleaning before applying the Genetic Programming algorithm, showed better results comparing to the standard Genetic Programing approach [24]. Moreover, [13] used K-means and Neural Network for customer's segmentation into number of clusters. Afterwards they created a decision tree model for every cluster. The results illustrate more accuracy for the segment-based model over the full database-based model. This work differs in that the combination of both FFCM then GP has never been examined to the best of our knowledge

3.0 PROPOSED CHURN MANAGEMENT FRAMEWORK

The model development method proposed in this research is based on using FFCM clustering with GP in two stages. Figure 1 represents the proposed framework. In the first stage, data reduction is performed by applying FFCM clustering algorithm on the selected training data set. This process will split the training data set into a number of smaller sets (clusters). Clusters which contain only churners or non-churners are selected to form a new training data set. While other clusters that could not separate churners and non-churners are excluded. As GP suffers different major problems (i.e. sensitivity towards outliers, variable results on various runs, and resource expensive training process, among others) FFCM was first used to cluster the data set and exclude outliers, representing abnormal customers' behaviors, to reduce sensitivity towards outliers and training resources. FFCM was selected due to its constant results on different runs, and its insensitivity towards outliers. In the second stage, GP is applied on the joined clustered resulted from stage one. Consequently, GP starts its evolutionary cycle and develops the final classification model. Finally, the developed GP model is assisted using multi-fold cross validation.
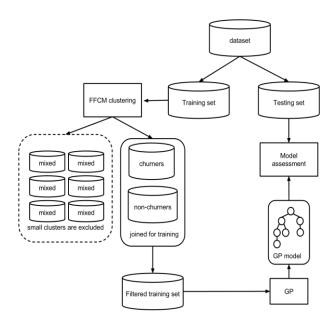

Fig. 1: Customer churn prediction framework

4.0 DATASET DESCRIPTION

Data used in this research was provided by a major cellular telecommunication company in Jordan. The dataset contains 11 attributes of randomly selected 14,573 customers subscribed to a prepaid service for a time interval of three months. The attributes cover outgoing/incoming calls statistics. Customers included in the dataset were annotated as churners or active. The total number of churners is 834 (i.e. 5.72% of total customers). The attributes in the dataset, along with their descriptions, are listed in Table 1. It is important to mention that attributes were normalized to eliminate the influence of larger values within different attribute scales.

Table 1. Dataset attributes and their descriptions

Attribute name	Description
3G	Subscriber is provided with 3G service (Yes, No)
Total Consumption	Total monthly fees(calling +SMS) in (JD)
Calling fees	Total monthly calling fees (JD)
Local SMS fees	Monthly local SMS fees(JD)
Int'l SMS fees	Monthly fees for international SMS(JD)
Int'l calling fees	Monthly fees for international calling (JD)
Local SMS count	Number of monthly local SMS
Int'l SMS count	Number of monthly international SMS
Int'l MOU	Total of international outgoing calls in minutes
Total MOU	Total minutes of use for all outgoing calls
On net MOU	Minutes of use for on-net-outgoing calls
Churn	Churning customer status (Yes, No)

5.0 EVALUATION MEASURES

In order to evaluate the developed model, we refer to the confusion matrix shown in Table 2 which is the primary source for accuracy estimation in classification problems. Based on this confusion matrix, the four different criterion used are listed in Table 3. Accuracy identifies the percentage of the total number of predictions that were correctly classified where actual churners rate shows the percentage of predicted churn in actual churn. The hit rate shows the percentage of predicted churn in actual churn and actual non-churn. Finally, lift coefficient shows the precision of model where CP represents the real churn percentage in the data set. The higher the lift is, the more accurate the model is [28].

Table 2. Confusion matrix

	Actual	
	non-churners	Churners
Predicted non-churners	A	B
Predicted churners	C	D

Table 3. Evaluation methods

Accuracy	$\dfrac{A+D}{A+B+C+D}$
Actual Churners Rate	$\dfrac{D}{B+D}$
Hit Rate	$\dfrac{D}{C+D}$
Lift Coefficient	$\dfrac{D}{(C+D).CP}$

216

Malaysian Journal of Computer Science. Vol. 28(3), 2015

6.0 EXPERIMENTS SETUP AND RESULTS ANALYSIS

Before applying FFCM to cluster the training subsets, the number of clusters has to be determined. To find the optimal number, different FFCM sizes were applied on all the dataset (i.e. 4, 9, 16 and 25). Even though FFCM approximately isolated all churning cases in a single cluster at every run, non-churning cases were mostly distributed over other clusters, especially when number of clusters is high, therefore, it was found that FFCM with the size of 4 was is the best in identifying two clusters with the highest rates of churners and non-churners. Hereafter, FFCM was configured to cluster data into 4 clusters in our experiments. In order to give a better indication of how well the developed classification model will perform when it is asked to classify new data, a cross validation with five folds is applied.

In this work, GP will develop a classification model in order to fit the given data set and minimizes the error with respect to the actual value. The final developed model as a result is a computer program created by the GP evolutionary cycle which takes the values of the independent customer variables as input and produces the values of the dependent variables as output which represents the class value (i.e. churn or not churn). The setup of the GP parameters is shown in table 4. By this process the individual programs evolve and have better fitness (reduced MSE) values by time.

Table 4. GP parameters

Parameter	Value
Mutation probability	15%
Population size	1000
Maximum generations	100
Selection mechanism	Tournament selector
Operators	+,-,*,/,AND,OR,NOT, IF THEN ELSE
Max Tree Depth	10
Max Tree Length	50
Elites	1
Fitness (Mean-Squared Error)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$

In the initialization step, GP algorithm starts by creating a predetermined number of individuals which form a population where each individual represents a computer program representing a classification model for customers. The model returns 0 or 1 if the customer is an active or a churn, respectively. Individuals can be viewed as symbolic tree structures which are graphical representations of their equivalent S-expressions in LISP programming language [1]. GP symbolic trees can be defined by a Terminal set T and a function set F. T set contains variables and constants. The set F typically contains arithmetic operators such as addition, subtraction, multiplication and division; also it could contain other non-linear functions as well as logical operator for the case of GP classification trees.

Next, fitness is evaluated for each individual. In this work, we use mean squared error (MSE) for evaluating all individuals. Then, a number of individuals are selected to reproduce new individuals and form new generation. There are different techniques to perform this selection listed in [22] from which we used tournament selection where k individuals are selected randomly, and then the fittest one is chosen for reproduction. In the reproduction step new individuals are created using reproduction operators, replacing old generations. In general, this process makes small random changes to the construction of the individuals. At first, *Crossover* produces two new individuals (children) by selecting a random sub-tree in each of the two parents and swapping the resultant sub-trees. The new individuals form a new generation or offspring. Then, *Mutation* is applied on a single GP individual where a random node is selected in the tree of the individual and then the sub-tree under this node is replaced by a new randomly generated sub-tree. Finally, *Elitism* selects one or more individuals with high fitness values and copies them to the next generation without any modification. The previously described process iterates until a termination condition is met (i.e. when it finds an individual with the required fitness value, or when the maximum number of iterations set by the user is reached).

217

Malaysian Journal of Computer Science. Vol. 28(3), 2015

In order to validate the effectiveness of the classification model, a five-fold cross validation is applied. Training and testing subsets were loaded into Heuristiclab[2] framework then a symbolic regression via GP was applied with parameters set as shown in Table 4. The best generated GP model tree was evaluated and then compared with the results obtained from basic GP without SOM, k-Nearest Neighbor (IBK), Naïve Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and GP combined with SOM for clustering as used in [27]. Comparison results for individual measures are shown in Fig. 2(a), 2(b), 2(c), and 2(d) respectively.

Figure 2 (a) shows that FFCM+GP approach is better than the basic GP and NB algorithms, while being comparable to IBK, RF, and SOM+GP. Using FFCM to eliminate the outliers enhanced the accuracy of GP by slightly over 9 points (13% accuracy enhancement). Further, it can be noticed that even though IBK and RF outperformed SOM+GP in accuracy, they absolutely failed in predicting churners with 3% and 1.5% respectively, as shown in Fig. 2(b). Additionally, FFCM+GP achieved outstanding performance in hit rate and lift coefficient, with approximately 20% improvement over the second best performing method (i.e. SOM+GP in both cases) as shown in Fig. 2(c) and 2(d) respectively. From Fig. 2(b), 2(c), and 2(d) it can be concluded that FFCM + GP is the best predictor of churners if compared to any other classification method. Therefore, the accuracy figure (i.e. Fig. 2(a)) explains that other methods (i.e. IBK and RF) achieved better accuracy because they have predicted better percentages of non-churners, otherwise they would have achieved better percentages in Fig. 2(b). Therefore, it can be safely concluded that FFCM + GP model significantly outperformed the SOM+GP model and the other approaches.
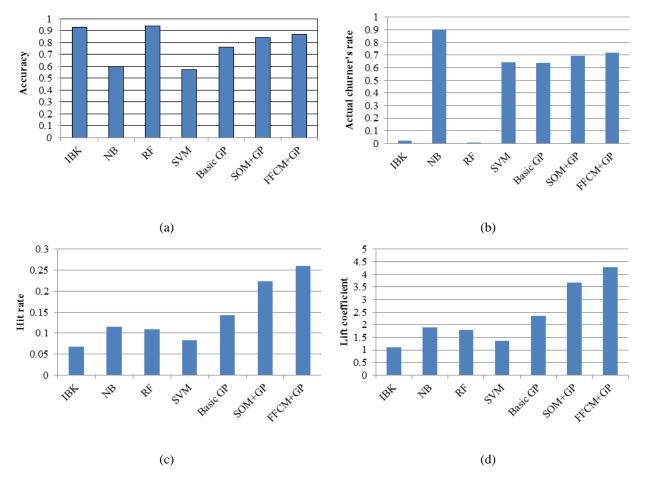

(a)

(b)

(c)

(d)

Fig 2. A comparison between kNN (IBK), Naïve Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), Basic GP, Self-Organizing Map followed by GP (SOM+GP), and the proposed method (FFCM + GP) in terms of: (a) Accuracy, (b) Actual Churner's Rate, (c) Hit Rate, and (d) Lift Coefficient

7.0 CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the application of a churn prediction approach combining Fast Fuzzy C-Means (FFCM) and Genetic programming (GP) for predicting possible churners in a Jordanian cellular telecommunication network. FFCM was used to eliminate outliers which correspond to unrepresentative customers' data, while GP was used to develop a final churn prediction model. Performance was evaluated using different criterion and compared with other state of the art classification methods. It was found that the FFCM with GP outperformed all other methods, and proved promising, yet better, capability in predicting possible churns.

REFERENCES

[1]   M. Affenzeller st al., "Genetic Algorithms and Genetic programming- Modern Concepts and Practical Applications", *CRC Press*, 2009.

[2]   B. Al-Shboul et al., "Real Time Speaker Identification System", *Proceedings of the 7th WSEAS International Conference on Applied Computer Science*, 2007, pp.424-428.

[3]   M. B. Al-Zoubi et al., "A fast fuzzy clustering algorithm", *In Proceedings of the 6th Conference on 6th WSEAS International Conference On Artificial Intelligence, Knowledge Engineering and Data Bases*, 2007, pp. 28–32.

[4]   I. Bose et al., "Hybrid models using unsupervised clustering for prediction of customer churn", *Journal of Organizational Computing and Electronic Commerce*, Vol. 19 No. 2, 2009, pp. 133–151.

[5]   R. C. Blattberg et al., "Churn management. In Database Marketing", *International Series in Quantitative Marketing*, Vol. 18, 2008, pp. 607–633.

[6]   R. C. Blattberg et al., "Statistical issues in predictive modeling. In Database Marketing", *International Series in Quantitative Marketing*, Vol. 18, 2008, pp. 291–321.

[7]   Huang L. B., Balakrishnan V. & Raj R. G. "Improving the relevancy of document search using the multi-term adjacency keyword-order model." Malaysian Journal of Computer Science 25.1, 2012, pp. 1-10.

[8]   A. E. Eiben et al., "Modelling customer retention with statistical techniques, rough data models and genetic programming", *In Rough-Fuzzy Hybridization: A New Trend in Decision Making Fuzzy Sets, Rough Sets and Decision Making Processes*, 1998, pp. 330–345.

[9]   P. G. Espejo et al., "A survey on the application of genetic programming to classification", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 40, No. 2, 2010, pp. 121–144.

[10]   C. Estébanez et al., "Gppe: a method to generate ad-hoc feature extractors for prediction in financial domains", *Applied Intelligence*, Vol. 29, No. 2, 2008, pp. 174–185.

[11]   V. Balakrishnan, F. G. Sim, R. G. Raj. "A one-mode-for-all predictor for text messaging", *Maejo International Journal of Science and Technology*, Vol. 5, No. 2, pp. 266-278, 2011.

[12]   J. Hadden et al., "Churn prediction using complaints data", *In Proceedings of World Academy of Science, Engineering and Technology*, Vol. 13, 2006, pp. 158-163.

[13]   S. Y. Hung et al., "Applying data mining to telecom churn management", *Expert Systems with Applications*, Vol. 31, No. 3, 2006, pp. 515–524.

[14]   N. Kim et al., "A new ensemble model for efficient churn prediction in mobile telecommunication", *46th Hawaii International Conference on System Sciences*, 2012, pp. 1023–1029.

[15]   M. Kotanchek et al., "Industrial strength genetic programming", *In Genetic Programming Theory and Practice*, chapter 15, 2003, pp. 239–256.

[16]   J. Koza, "Evolving a computer program to generate random numbers using the genetic programming paradigm", *In Proceedings of the Fourth International Conference on Genetic Algorithms*, 1991.

[17]   J. Koza, "Genetic Programming: On the programming of computers by means of natural selection", *MIT press*, Vol. 1, 1991.

219

Malaysian Journal of Computer Science. Vol. 28(3), 2015

[18]   Yeow, W. L., Mahmud, R., & Raj, R. G., "An application of case-based reasoning with machine learning for forensic autopsy", Expert Systems with Applications, Vol 41, No. 7, 2014, pp. 3497-3505, ISSN   0957-   4174,   http://dx.doi.org/10.1016/j.eswa.2013.10.054. (http://www.sciencedirect.com/science/article/pii/S0957 417413008713).

[19]   T. E. McKee et al., "Genetic programming and rough sets: A hybrid approach to bankruptcy classification", *European Journal of Operational Research*, Vol. 138, No. 2, 2011, pp. 436–451.

[20]   A. Qazi, R. G. Raj, M. Tahir, M. Waheed, S. U. R. Khan, and A. Abraham, "A Preliminary Investigation of User Perception and Behavioral Intention for Different Review Types: Customers and Designers Perspective," *The Scientific World Journal*, vol. 2014, Article ID 872929, 8 pages, 2014. doi:10.1155/2014/872929.

[21]   V. Miguéis et al., "Predicting partial customer churn using markov for discrimination for modeling first purchase sequences", *Advances in Data Analysis and Classification*, Vol. 6, No. 4, 2012, pp. 337–353.

[22]   W. Miller et al., "Neural Networks for Control", *MIT Press*, 1995.

[23]   Moohebat, M., Raj, R.G. , Kareem, S.B.A., Thorleuchter, D., "Identifying ISI-indexed articles by their lexical usage: A text analysis approach", *Journal of the Association for Information Science and Technology*, Vol. 66, No. 3, pp. 501–511. doi: 10.1002/asi.23194.

[24]   R. Obiedat et al., "Customer churn prediction using a hybrid genetic programming approach", *Scientific Research and Essays*, Vol. 8, No. 27, 2013, pp. 1289–1295.

[25]   M. Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", *Expert Systems with Applications*, Vol. 37, No. 6, 2010, pp. 4710 – 4712.

[26]   P. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services", *Expert Systems with Applications*, Vol. 36, No. 3, 2009, pp. 6714–6720.

[27]   C. F. Tsai et al., "Customer churn prediction by hybrid neural networks", *Expert Systems with Applications*, Vol. 36, No. 10, 2009, pp. 12547–12553.

[28]   X. Yu et al., "An extended support vector machine forecasting framework for customer churn in e-commerce", *Expert Systems with Applications*, Vol. 38, No. 3, 2011, pp. 1425 – 1430.

[29]   Y. Zhao et al., "Customer churn prediction using improved one-class support vector machine", *In Advanced Data Mining and Applications*, Vol. 3584 LNCS, 2005, pp. 300–306.

[30]   Blattberg, R., Do, K., and Scott, N. (2008). Database Marketing: Analyzing and Managing Customers, *International Series in Quantitative Marketing*, 18: 607–633.

[31]   H. Faris et al., "A Genetic Programming Based Framework for Churn Prediction in Telecommunication Industry", Vol. 8733, *LNCS*, 2014, pp. 353-362

220

Malaysian Journal of Computer Science. Vol. 28(3), 2015