ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

Enhanced Churn Prediction on Huge Telecom Data using Particle Swarm Optimization

R.Suganji, Dr. G. Ravi

M.Phil. Student, Department of Computer Science,Jamal MohamedCollege (Autonomous), Tiruchirappalli, Tamil Nadu, India.

Associate Professor&Head, Department of Computer Science,Jamal Mohamed College(Autonomous),Tiruchirappalli,Tamil Nadu, India.

ABSTRACT: Churn prediction in telecom has gained a huge prominence in the recent times due to the extensive interests exhibited by the stakeholders, large number of competitors and huge revenue losses incurred due to churn. Predicting telecom churn is challenging due to the voluminous and sparse nature of the data. This paper presents a technique for telecom churn prediction that employs PSO for churn prediction. Experiments reveal that the performance of metaheuristics was more efficient and they also exhibited better predictability levels.

KEYWORDS: Churn prediction; Classifier; Feature Selection; Imbalance; Metaheuristics; Particle Swarm Optimization

I. Introduction

Customer Relationship Management is the major component of an organization. The major issue in any customer related organization is maintaining customers. This is due to high competitive market and a level of dissatisfaction experienced by customers. Customer retention is a major necessity for successful operation of an organization. Obtaining a new customers is considered to be five to six times more costly [1]. Loss of customers will lead to revenue loss and loss of brand loyalty and the company's morale.

Customer churn/ abrasion is the tendency of a customer to stop doing business transactions with an organization [2]. The major issue in churn prediction is that there are several reasons for a customer to churn. Identifying these reasons is difficult, as the reasons are not direct. They depend on customer's personal views and the products or services that the customers utilize from the company. Organizations need to predict churn before it happens. Huge customer data is available and predictions can be performed, however the nature of the data is a huge drawback to the prediction mechanism.

II. Related Works

This section discusses some of the recent contributions in the domain of churn prediction.

Xiao et al. presented a dynamic transfer ensemble based churn prediction model [3]. It uses feature selection as a part of pre-processing mechanism in customer churn prediction. Pre-processing is used to eliminate unnecessary entries. Feature selection is actually carried out in two phases. The first phase performs feature selection using GMDH-type neural network and the next phase analyses the source domain and adds entries to the feature list.

Chen et al. proposed a study on predicting churn in logistics industry [4]. This model considers length, recency, frequency, monetary and profit (LRFMP) to determine churn. This paper also discusses the managerial implication with additional insights on the predictor variables.

Idris et al. proposed a telecom based churn prediction technique [5]. It uses minimum redundancy maximum relevance (mRMR) for the prediction process. This technique also employs feature selection and also uses Random Forest, Rotation Forest, RotBoost and DECORATE for churn prediction. Statistical classifier based churn prediction techniques include [6,7,8,9]. They use KNN to identify churn. Hybridized KNN based churn prediction is presented in [10]. Alessandro et al. presented an agent based technique to predict unsatisfactory 3G to 4G in cellular networks [11]. It is an extended recommendation based method to avoid churn.

Droftina et al. proposed a real time churn prediction model based on sociometric clique and social status theory [12]. Xiao et al. proposed an ensemble based technique for churn prediction [13]. Other similar churn prediction techniques

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

include [17, 18]. From the above discussion it was observed that ensemble techniques are the most preferred for churn prediction. However, it is costly and time consuming.

III. OUR APPROACH

Churn data is usually composed of customer details and their associations with products provided by the organizations. This paper deals with telecom churn prediction. From the telecom data, it could be observed that the attributes in the dataset represents all the services presented by the organization. Hence the data is huge. Besides on including every service, the dataset keeps growing horizontally. Hence the dataset has very high attribute levels. This in turn leads to sparsity in the dataset due to large number of empty null valued entries.

Customers are usually associated with a few products in the organization. Hence the attributes corresponding to those entries are filled, while the others are left empty. leading to the highly sparse nature of the data.

Metaheuristic is a higher-level heuristic used to identify or select a heuristic that provides an optimal solution. The advantage of metaheuristics is that it can work with incomplete or imperfect information [14]. Hence the current problem of churn prediction is applied on a metaheuristic technique to obtain optimal solutions.

The properties of churn clearly expose the complexity involved in the process of mining churn data. Applying data mining algorithms on such data tends to be disastrous, due to its inefficiency in handling such data. Not only that data mining techniques require their input data to be in specific formats, they also require the entire data to be present in-order to perform the prediction process. The complexity of such algorithms shoot up as the amount of data increases. Hence in terms of scaling, most of the statistical prediction techniques fail. Hence opting for meta-heuristic techniques in such scenarios is the best decision. This paper uses PSO [15,16] as a classifier for predicting churn.

The search space is created using the Orange Data. The dimensions of the search space correspond to the number of attributes contained in the dataset. Particle initialization is carried out by identifying the placement location of the particles using a uniform distribution function. Particle distributions are confined to the search space boundaries. The number of particles that are to be used for a given problem is itself an optimization problem. The number of particles used for the current implementation scenario is obtained by trial and error. Initial particle velocities are then identified in random using eq. (1)

$$V_i \sim U(-|b_{up} - b_{lo}|, |b_{up} - b_{lo}|) \qquad (1)$$

where b_{up} and b_{lo} are the upper and lower bounds of the search space.

The velocity identification marks the beginning of particle movement. Every particle moves according to its velocity component. Due to the random identification of velocities, the particles are dispersed in the search space in a random fashion. The particle best (*pbest*) and the global best (*gbest*) values are identified.

Particle best (*pbest*) corresponds to the best solutions identified by a particle since the beginning of the iteration and global best (*gbest*) is the best solution obtained from the swarm, which is the best of all the *pbest*values. In-order to identify the *pbest*of each particle, the fitness of the particle is identified. The identified current fitness is compared with the previous *pbest.* If the current solution that has been identified has better fitness compared to the already existing fitness, the existing *pbest*is replaced with the current fitness. The new fitness value is then compared against the current *gbest*and the best of the values is retained as the current *gbest.*

After the initial identification of *pbest*and*gbest*, the particles are again triggered for movement towards the best solution that has been identified so-far. The new velocity is identified using eq. (2)

$$V_{i,d} \leftarrow \omega V_{i,d} + \varphi_p r_p (P_{i,d} - X_{i,d}) + \varphi_g r_g (g_d - X_{i,d}) \quad (2)$$

where $P_{i,d}$ and g_d are the parameter best and the global best values, r_p and r_g are user generated random numbers, $x_{i,d}$ is the value current particle position, and the parameters ω, φ_p, and φ_g are selected by the practitioner.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

The process of churn prediction is internally a classification system that identifies the current state of a customer using the properties of the customer. This in general can be expressed as follows.

Given a database with two classes (representing churn and no churn) and N attributes, the process of classification is to identify the optimal centroid C in an N dimensional space for each class. Hence a particle i is represented as

$$P_i = \{P_{i,1}, P_{i,2}, P_{i,3} - - - P_{i,N}\} \qquad (3)$$

with a velocity component

$$V_i = \{V_{i,1}, V_{i,2}, V_{i,3} - - - V_{i,N}\} \qquad (4)$$

The fitness of particles are computed by identifying the Euclidean distance between the training data and the current position of the particle eq. (5). Particle exhibiting the closest distance with the centroid in the N dimensional region is considered as the best solution for the current training data and hence the churn value of the best solution is identified to be the prediction for the training data.

$$\psi_i = \sum_{k=1}^{N} \sqrt{\left(P_{i,k} - \tau_k\right)^2} \qquad (5)$$

where$P_{i,k}$refers to the current position of particle I in the k^{th}dimension and τ_k represents the value of the training data in its k^{th}dimension.

This process is continued till a stagnation behavior is encountered. After stagnation, the result contained in *gbest*is considered to be the best solution for the current churn record.

IV. RESULTS AND DISCUSSION

PSO was implemented using C#.Net on Visual Studio 2012. Experiments were conducted with the Orange Dataset on PSO.

Table 1: Dataset Analysis

Property	**Orange Dataset**
Attribute density	230
Instances	50000
Missing Value Levels	60%
# Numerical Attributes	190
# Categorical Attributes	40

Comparison is carried out using PSO and J48. J48 is a tree based algorithm based on C4.5. Results for J48 were calculated using Weka. Accuracy and F Measure were used as performance metrics.

Accuracy comparison of PSO with J48 is presented in Fig.1. It could be observed that PSO exhibits accuracy similar to J48, exhibiting an accuracy level of ~85%. Even though metaheuristics are usually expected to provide near optimal results, it could be noted that PSO exhibits accuracy comparable to J48, a tree based algorithm.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

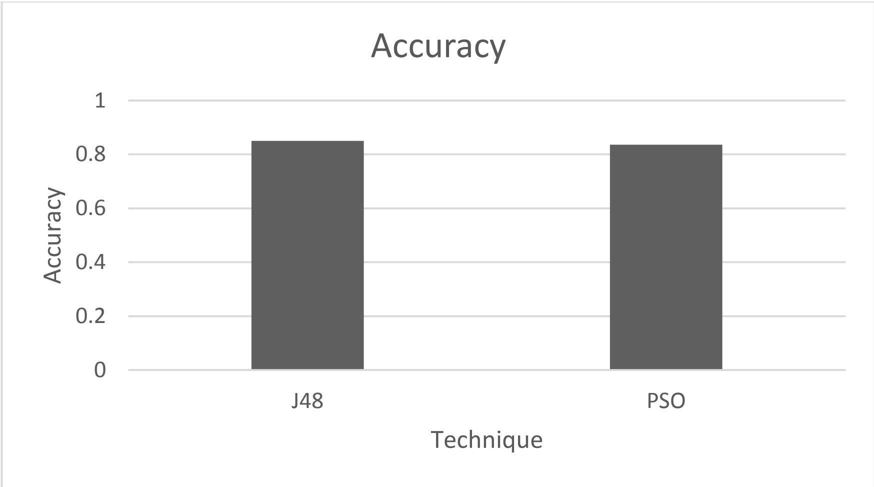

Fig.1 Accuracy

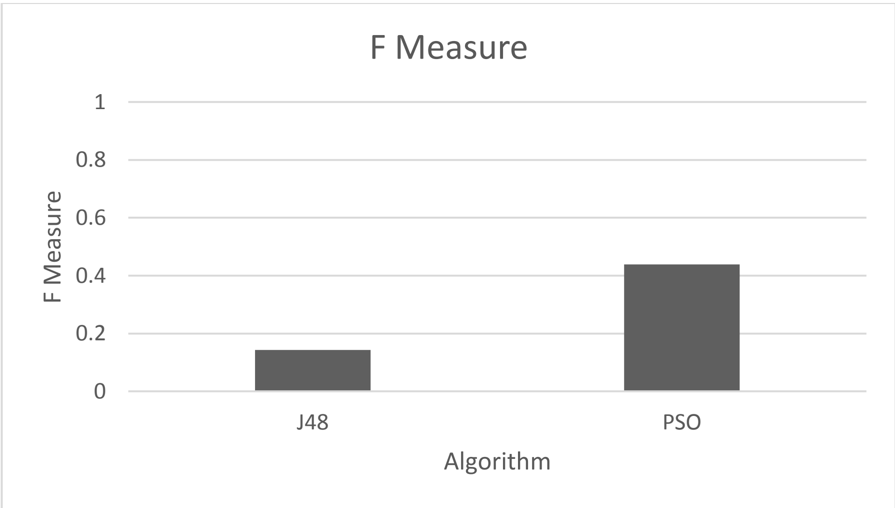

Fig. 2. F Measure

F-Measure obtained from the results of PSO and J48 are presented in Figure. F-measure is an analysis used to measure the accuracy of the binary classifier. It is calculated using the eq (6)

$$FMeasure = \frac{2.Precision.Recall}{Precision+Recall} \qquad (6)$$

It could be observed that PSO exhibits a better F-Measure compared to J48. F-Measure reflects the retrieval rates of a classifier. Hence it could be concluded that PSO exhibits better positive retrieval levels compared to J48.

V. CONCLUSION

Churn prediction is one of the major requirements of the current competitive environment. This paper deals with identifying and predicting churn in telecom data using Particle Swarm Optimization. Analysis of the algorithms was carried out on the basis of ROC and PR Curves. Future directions will include incorporation of schemes or modifications to reduce the False Positive rates. Further, analysis in terms of imbalance levels and data sparsity will

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 9 , September 2016

also be carried out. Incorporation of Game theory in the decision making process will also help improve the accuracy levels and in the identification of churn.

REFERENCES

[1] C.B. Bhattacharya, "When customers are members: Customer retention in paid membership contexts." Journal of the academy of marketing science 26.1 (1998): 31-44.
[2] D. Jill. “The CRM handbook: a business guide to customer relationship management.” Addison-Wesley Professional, 2002.
[3] X, Jin, et al. "Feature-selection-based dynamic transfer ensemble model for customer churn prediction." Knowledge and Information Systems 43.1 : 29-51,2015.
[4] C. Kuanchin, H. Ya-Han, and H. Yi-Cheng. "Predicting customer churn from valuable B2B customers in the logistics industry: a case study." Information Systems and e-Business Management 13.3 (2015): 475-494.
[5] I. Adnan, A. Khan, and Y. S. Lee. "Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification." Applied intelligence 39.3: 659-672,2013.
[6] R. Dymitr, D. Nauck, and B. Azvine. "K nearest sequence method and its application to churn prediction." Intelligent Data Engineering and Automated Learning–IDEAL 2006. Springer Berlin Heidelberg, 2006. 207-215, 2006.
[7] K. Asifullah, M. F. Khan, and C.Tae-Sun. "Proximity based GPCRs prediction in transform domain." Biochemical and biophysical research communications 371.3 : 411-415,2008.
[8] T. Songbo. "An effective refinement strategy for KNN text classifier." Expert Systems with Applications 30.2 (2006): 290-298, 2006.
[9] Z. Li, L. Wang, and Q. Xu. "Data stream classification with artificial endocrine system." Applied Intelligence 37.3 : 390-404, 2012.
[10] Z. Yangming, et al. "A hybrid KNN-LR classifier and its application in customer churn prediction." Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on. IEEE, 2007.
[11] S. D’Alessandro, et al. "Consumer satisfaction versus churn in the case of upgrades of 3G to 4G cell networks." Marketing Letters 26.4 (2015): 489-500.
[12] D. Uroš, M. Štular, and A. Košir. "A diffusion model for churn prediction based on sociometric theory." Advances in Data Analysis and Classification 9.3 : 341-365, 2015.
[13] X. Jin, et al. "One-Step Classifier Ensemble Model for Customer Churn Prediction with Imbalanced Class." Proceedings of the Eighth International Conference on Management Science and Engineering Management. Springer Berlin Heidelberg, 2014.
[14] B. Leonora, et al. "A survey on metaheuristics for stochastic combinatorial optimization." Natural Computing: an international journal 8.2 : 239-287,2008.
[15] J.Kennedy, R.C. Eberhart. “Particle swarm optimization." Proceedings of IEEE International Conference on Neural Networks. Vol. 1948. 1942, 1995.
[16] S. Yuhui, and R. Eberhart. "A modified particle swarm optimizer." Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on. IEEE, 1998.
[17] J. Qi, L. Zhang , Y. Liu, L. Li, Y. Zhou, Y. Shen, L. Liang and H. Li,"ADTreesLogit model for customer churn prediction", Annals of Operations Research, 168:247, 2009.
[18] W. Au, KCC. Chan and X. Yao, “A novel evolutionary data mining algorithm with applications to churn prediction”. IEEE T EvolComput 7(6):532–545, 2004.