IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Churn Prediction on Huge Sparse Telecom Data Using Metaheuristic

T. Sumathi

M.Phil Research Scholar, School Of Computer Science & Engg, Bharathidasan University, Tiruchirappalli, India

Abstract: Churn prediction in telecom has become a major requirement due to the increase in the number of telecom providers. However due to the hugeness, sparsity and imbalanced nature of the data, churn prediction in telecom has always been a complex task. This paper presents a Metaheuristic based churn prediction technique that performs churn prediction on huge telecom data. Particle Swarm Optimization algorithm is used as the classifier. Experiments were conducted on the Orange dataset. It was observed that PSO algorithm works best on churn data providing effective and faster results.

Keywords: Telecom churn prediction, Data Imbalance, Data Sparsity, Huge Data; PSO

I. INTRODUCTION

Increase in the telecom providers has led to a huge raise in the competitions and hence customer churns. Currently organizations have their major focus in reducing the churn by focusing on customers independently. Churn [1] can be defined as the propensity of a customer to cease business transactions with an organization. The major requirement now is identification of customers who have high probabilities of moving out. The ability of an organization to intervene at the right time could effectively reduce churn.

Churn occurs mainly due to customer dissatisfaction. Identifying customer dissatisfaction requires several parameters. A customer usually does not churn due to a single dissatisfaction scenario [2]. There usually exists several dissatisfaction cases before a customer completely ceases to do transactions with an organization. Several properties associated with the customer and their modes of operations with the organization are recorded by the organizations. This represents the customer’s behavior data. Analyzing this data would present a clear view of the customer’s current status [3]. Hence this can be used as the base data for churn prediction. The major difficulty arising from this mode of operation is that the data under discussion tends to be very huge. The hugeness can be attributed to the behavioural nature of the data, depicting all the product lines dealt with by the organization. Further, due to the requirement of structural representation of the data, all the instances are bound to contain all the properties corresponding to a generic customer in the organization [4,5]. This leads to data sparseness, since customers will be associated with only a few properties and not all the properties pertaining to the organization. The data hugeness and sparsity acts as the major difficulties in the process of churn prediction.

In-order to provide required services to the customers, large companies interacts with them to obtain their requirements [6]. One of the major differentiators for companies is the customer service. The ability to predict if a customer will leave in order to intervene at the right time can be essential for pre-empting problems and providing high level of customer service. The problem becomes more complex as customer behavior data is sequential and can be very diverse.

II. RELATED WORKS

Churn is an unavoidable process in any industry. However, though difficult, it is possible to identify the causes of churn using several approaches. This section discusses the recent approaches for churn prediction.

A risk prediction technique that identifies probable customers for churn was presented by Casement et al. in [7]. This technique utilizes Generalized Additive Models (GAM). These models relax the linearity constraints, hence allowing complex non-linear fits to the data. This technique is exhibited to improve marketing decisions by identifying the risky customers and also providing visualizations of non-linear relationships.

A neural network based customer profiling technique that can be used for churn prediction was presented by Tiwari et al. in [8]. This technique differs from the other proposed techniques from the fact that most of the techniques are only able to identify the customers who will instantaneously churn. However the neural network based churn prediction model proposes to predict customer’s future churn behavior, providing the much required buffer for the organizations to perform prevention activities. A similar neural network based model includes [23,25]. The approach in [23] is based on the 80-20 rule to identify the key attributes affecting churn, while that of [25] involves identifying the major features of the data to determine churn.

A regression based churn prediction model was presented by Awning et al. in [9]. This method identifies churn by using multiple regressions analysis. This technique utilizes the customer's feature data for analysis and proposes to provide good performance.

Class imbalance plays a major role in affecting the reliability of a classifier. The major issue existing due to class imbalance is that the minority class is not well represented and hence the classifier is undertrained on the minority classes. The technique proposed by Zhu et al. in [10] proposes to eliminate this issue by using transfer learning techniques. The approach presented in [10] operates by training the classifier using customer related behavioural data obtained from related domains. This approach has its major focus on the banking industry and the results are proposed to exhibit enhanced performance. Another technique that considers the imbalance nature of data to perform churn prediction was presented by Xiao et al. in [15]. A comparison of sampling techniques for effectively operating on churn data was presented by Amin et al. in [16]. Game theory based churn prediction techniques [17] are also on the raise.

The complex nature of churn behavior has also enabled several publications on churn prediction using multiple models. A churn prediction model based on cluster analysis and decision tree algorithm was presented by Li et al. in [11]. This technique operates on China's Telecom data. Another technique utilizing multiple prediction techniques was proposed by Le et al. in [12]. This technique utilized a combination of k-Nearest Neighbor algorithm and sequence alignment. This technique has its major focus on the temporal categorical features of the data to predict churn.

Utilizing heuristics for predictions are on the raise due to the complex nature of data. A rule generation techniques that employs heuristics for customer churn prediction in telecom services was presented by Huang et al. in [13]. A combination of Self Organizing Maps (SOM) and Genetic Programming (GP) to identify and predict churn was presented by Farris et al. in [14]. SOM is utilized to cluster the customers and then outliers are eliminated to obtain clusters depicting customer behaviours. An enhanced classification tree is built using GP.

A boosting algorithm that proposes to improve the prediction accuracy of classifier models was proposed by Lu et al. in [18]. This method boosts the learning process by using a combination of clustering and logistic regression. A similar prediction boosting technique using Genetic Algorithm was proposed by Idrus et al. in [19]. This is also an ensemble model utilizing multiple techniques for the prediction process. Other ensemble based prediction techniques include [20,21,22,24].

III. OUR APPROACH

Particle Swarm Optimization (PSO) [26,27] is a computational technique that optimizes a problem by iteratively trying to improve a candidate solution with regard to a certain measure called its fitness. PSO being a swarm based metaheuristic technique solves the problem by using a set of candidate solutions called particles. These particles are moved in a search space by using a velocity component. This process is iteratively performed to identify the best solution in the search space. This approach utilizes PSO for the process of churn prediction.

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)

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

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.

This process is continued till a stagnation behavior is encountered. The pbest value of every iteration corresponds to a probable prediction for the test data. However, there are several pbest values, hence the best of the predictions are stored in gbest. The value of gbest changes for every iteration, corresponding to the selected pbest entries. A no-change in the gbest for a defined

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

number of iterations mark the stagnation behavior. 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. Orange is a benchmark dataset that corresponds to a French Telecom company [28]. It was used as a part of KDD 2009 challenge [29].

ROC plot corresponding to actual PSO is presented in Fig. 1. It could be observed from the plot that most of the points are aggregated in the region exhibiting very high true positive rates with moderate false positive rates. The presence of false positives is attributed to the sparsity of the data. However, the indication of high levels of true positives proves the efficiency of the classifier.

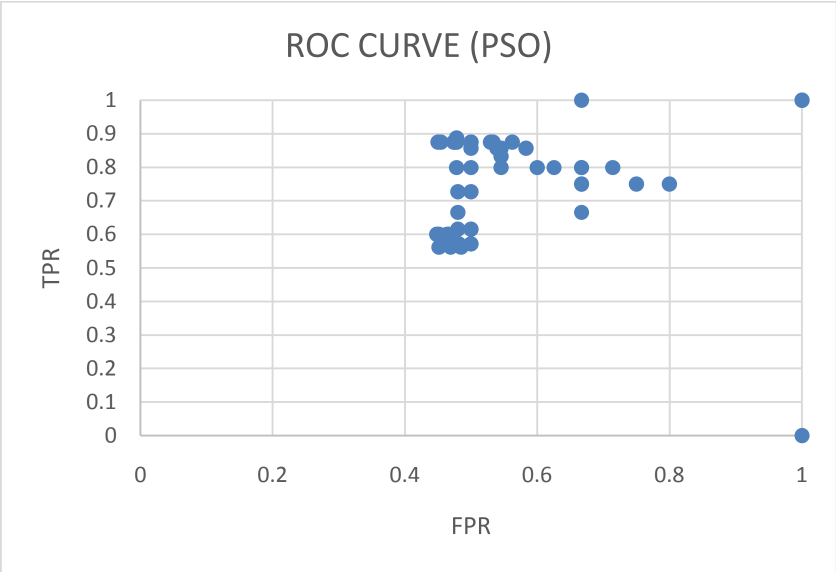

Fig.1 ROC PLOT (PSO)

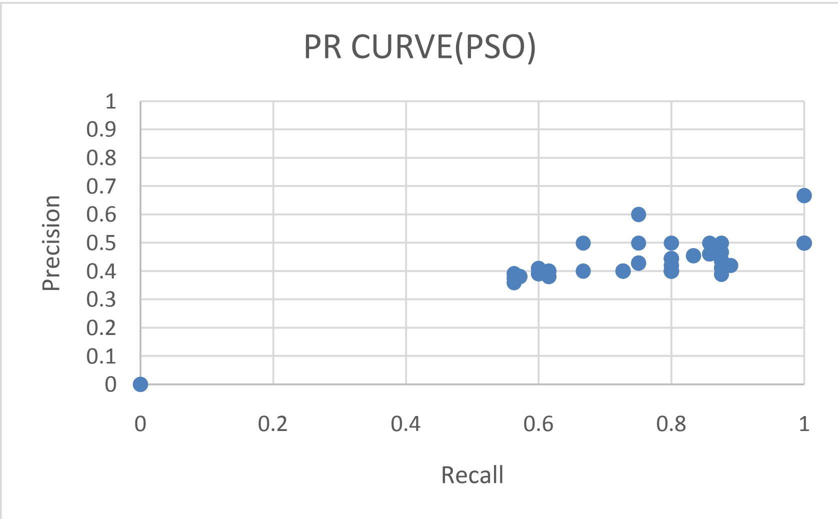

Fig. 2 PR PLOT(PSO)

Precision and Recall values exhibited by PSO is presented in the PR plots (Fig. 2). It could be observed that the PR plot exhibit moderate precision levels, but high recall levels.

The low precision levels are attributes to the selection probabilities and the absence of data (sparse nature of data) in the dataset.

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

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

modifications to reduce the False Positive rates. Further, analysis in terms of imbalance levels and data sparsity will 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] V. Effendy, and Z. A. Baizal, "Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest." In Information and Communication Technology (ICoICT), 2014 2nd International Conference on (pp. 325-330). IEEE, 2014.
[2] D. Seo, C. Ranganathan, and Y. Babad, "Two-level model of customer retention in the US mobile telecommunications service market."Telecommunications Policy, 32(3), pp.182-196, 2008.
[3] S. Y. Hung, D. C. Yen, and H. Y. Wang, "Applying data mining to telecom churn management." Expert Systems with Applications, 31(3), pp.515-524, 2006.
[4] G. Canning, "Do a value analysis of your customer base." Industrial Marketing Management, 11(2), pp.89-93, 1982.
[5] K. Eriksson, and A. L. Vaghult, "Customer retention, purchasing behavior and relationship substance in professional services." Industrial marketing management, 29(4), pp.363-372, 2000.
[6] C. B. Bhattacharya, "When customers are members: Customer retention in paid membership contexts." Journal of the academy of marketing science, 26(1), pp.31-44, 1998.
[7] K. Coussement, D. F. Benoit, and D. Van den Poel, "Preventing Customers from Running Away! Exploring Generalized Additive Models for Customer Churn Prediction." In The Sustainable Global Marketplace (pp. 238-238). Springer International Publishing.2015.
[8] A. Tiwari, J. Hadden, and C. Turner, "A new neural network based customer profiling methodology for churn prediction." In Computational Science and Its Applications–ICCSA 2010 (pp. 358-369). Springer Berlin Heidelberg, 2010.
[9] M.K. Awang, M.N.A. Rahman, and M.R. Ismail, "Data Mining for Churn Prediction: Multiple Regressions Approach." In Computer Applications for Database, Education, and Ubiquitous Computing (pp. 318-324). Springer Berlin Heidelberg, 2012.
[10] B. Zhu, J. Xiao, and C. He, " A Balanced Transfer Learning Model for Customer Churn Prediction." In Proceedings of the Eighth International Conference on Management Science and Engineering Management (pp. 97-104). Springer Berlin Heidelberg, 2014.
[11] G. Li, and X. Deng, "Customer churn prediction of china telecom based on cluster analysis and decision tree algorithm." In Emerging research in artificial intelligence and computational intelligence (pp. 319-327). Springer Berlin Heidelberg, 2012.
[12] M. Le, D. Nauck, B. Gabrys, and T. Martin, "KNNs and Sequence Alignment for Churn Prediction." In Research and Development in Intelligent Systems XXX (pp. 279-285). Springer International Publishing, 2013.
[13] Y. Huang, B. Huang, and M.T. Kechadi, "A rule-based method for customer churn prediction in telecommunication services." In Advances in Knowledge Discovery and Data Mining (pp. 411-422). Springer Berlin Heidelberg, 2011.
[14] H. Faris, B. Al-Shboul, and N. Ghatasheh, "A genetic programming based framework for churn prediction in telecommunication industry." InComputational Collective Intelligence. Technologies and Applications (pp. 353-362). Springer International Publishing, 2014.
[15] J. Xiao, G. Teng, C. He, and B. Zhu, "One-Step Classifier Ensemble Model for Customer Churn Prediction with Imbalanced Class." In Proceedings of the Eighth International Conference on Management Science and Engineering Management (pp. 843-854). Springer Berlin Heidelberg, 2014.
[16] A. Amin, F. Rahim, I. Ali, C. Khan, and S. Anwar, "A Comparison of Two Oversampling Techniques (SMOTE vs MTDF) for Handling Class Imbalance Problem: A Case Study of Customer Churn Prediction." In New Contributions in Information Systems and Technologies (pp. 215-225). Springer International Publishing, 2015.
[17] J. Kawale, A. Pal, and J. Srivastava, "Churn prediction in MMORPGs: A social influence based approach." In Computational Science and Engineering, 2009. CSE'09. International Conference on (Vol. 4, pp. 423-428). IEEE, 2009.
[18] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting." Industrial Informatics, IEEE Transactions on, 10(2), pp.1659-1665, 2014.
[19] A. Idris, A. Khan, and Y. S. Lee, "Genetic programming and adaboosting based churn prediction for telecom." In Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on (pp. 1328-1332). IEEE, 2012.
[20] L. Xie, D. Li, and J. Xia, "Feature selection based transfer ensemble model for customer churn prediction." In System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2011 International Conference on (Vol. 2, pp. 134-137). IEEE, 2011.
[21] Idris, A. and Khan, A., 2014, December. Ensemble Based Efficient Churn Prediction Model for Telecom. In Frontiers of Information Technology (FIT), 2014 12th International Conference on (pp. 238-244). IEEE.
[22] V. Effendy, and Z. A. Baizal, "Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest." In Information and Communication Technology (ICoICT), 2014 2nd International Conference on (pp. 325-330). IEEE, 2014.
[23] J. Liu, and G. Yang, "Research on customer churn prediction model based on IG_NN double attribute selection." In Information Science and Engineering (ICISE), 2010 2nd International Conference on (pp. 5306-5309). IEEE, 2010
[24] Y. Huang, B.Q. Huang, and M. T. Kechadi, "A new filter feature selection approach for customer churn prediction in telecommunications." In Industrial Engineering and Engineering Management (IEEM), 2010 IEEE International Conference on (pp. 338-342). IEEE, 2010.
[25] Q. Shen, H. Li, Q. Liao, W. Zhang, and K. Kalilou, "Improving churn prediction in telecommunications using complementary fusion of multilayer features based on factorization and construction." In Control and Decision Conference (2014 CCDC), The 26th Chinese (pp. 2250-2255). IEEE, 2014.
[26] Kennedy, J. "Eberhart. RC (1995). Particle swarm optimization." Proceedings of IEEE International Conference on Neural Networks. Vol. 1948. 1942.
[27] Shi, Yuhui, and Russell Eberhart. "A modified particle swarm optimizer." Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on. IEEE, 1998.
[28] K. Morik and H. Köpcke, "Analysing Customer Churn in Insurance Data - A Case Study", In Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, 2004.
[29] http://www.kdd.org/kdd-cup/view/kdd-cup-2009/Data