Applications of Business Intelligence and Decision Making for the Customer Behaviour Analysis in Telecom Industry

AttiliVenkataRamana, AnnaluriSreenivasaRao, E. Kesavulu Reddy

Abstract: ***The decision-making (DM) systems play a vital role in the telecommunication industry to locate the customers at a particular time. The business data analysis techniques allows to find the customers data and produce analysis reports helps to implement best policies in a business environment. On the other hand, business intelligence (BI) tools ensure to predict and analyse the historical data with the current data, to predict the events of a business operation. Thismethod seems to be more complex due to the involvement of large data from the millions of the customers. Using the data mining techniques along with DM and BI, the data can be processed effectively so as to deliver accurate decisions. In this work, application of customer leaving in the telecom industry is tested for prediction of customer behaviour using data mining techniques. Large number of data sets will be process from the telecom industry by introducing a unique method by combining hybrid genetic algorithm (GA) and particle swarm optimization (PSO) as HGAPSO, which allows extracting specific and useful data. By using the advantages of both the methods, one can accurately determine the decisions from the customer churn predictions using auxiliary vector classification method. The performance metrics in this work considered accuracy, true positive response rate (TPR), false positive rate (FPR), time complexity and receiver operating characteristics (ROC). The results from this work delivered high scalability using HGAPSO and well suited for predictions in the area of business analysis.***

Index Terms: CRM, BI, DM systems, Attribute Selection, customer churn prediction

INTRODUCTION

In the current scenario of telecommunications, customers have a liberty to select their service providers for voice and wireless internet connectivity [1]. Various service providing network companies from different countries such as UK, USA, Australia, New Zeland, etc. are competing to provide the internet service providers (ISPs). In the recent times, deregulation / privatization of communication markets in Europe, Latin America and Asia seem to create a serious competition among the service providers and this helped customers to a larger extent [2]. The customers are having a leverage of changing their service premiums, networks, and can influence the overall construction of the organizations.

With the technological transformations, most of the telecom companies are in a position to identify the important demands / needs of the customers for facilitating customized products [3].

Revised Manuscript Received on March 26, 2019.

Dr AttiliVenkataRamana, Associate Professor, Dept.ECM,Sreenidhi Institute of Science and Technology, Hyderabad.
Dr AnnaluriSreenivasaRao, AsstProf,Dept.IT, VNRVJIET,Hyderabad.
Dr E. Kesavulu Reddy, Asst Prof, Dept.CS, SVUCMCS, SV University, Tirupathi,

Such transformations are not only helping the service providers but also increasing the risk of maintaining the existing customers from the attractive offers given by the competitors. There is a massive shift in the number of customers from outdated product operations to consumer operations due to deregulation and private parties [4]. The attention of customers in such circumstance directed towards new services at the lower costs and with a significant improvement in the efficiency [5].

The business processes of most of the telecommunication providers are under the scrutiny to modify as per the requirements of the customer needs. This in turn demands a careful attention towards converting larger databases of the customers into reliable evidences for future predictions and decisions making. Using the BI systems most of the service providers solved different challenges of the value chain to achieve the goals of a telecom industry.

The globalization allowed the telecom industries to expand and support their services in different international markets, regions, cities and rural areas with different types of dynamic models that suit the necessaries of different customers [6]. In order to meet the requirements of the local customers at different places, the right product development has been considered by using the customer relationship management (CRM) and BI solutions. At each stage of CRM process the large amounts of data will be processed by the telecom companies to influence customers at the time of designing new products. The three phases of CRM process used by telecom companies are listed below:

A. Identifying potential customers with the help of customer interactions;

B. Understanding the customer needs and buying patterns; and

C. To work with customers for ensuring their needs are fulfilled.

In the proposed method, the above three phases are achieved by providing the HGAPSO algorithm, selecting most appropriate and useful information from attribute set.This way helps to achieve best DM process for most of the customer departures. To implement the proposed method Java simulation environment was used effectively and also helped to evaluate for obtaining the best and accurate information from the recognition results.

Apart from all the above scenarios, most of the Indian companies supply chain managements (SCM) are facing a biggest challenge towards organizing the rewards and to expand the order fulfillment processes for a long time. SCM helps to examine the impact of different types of return policy for various goods which are unsold. Therefore by it is important to adopt a common model or structure that defines the return

policies based on the different markets conditions and retailers interests. The role of SCM and related issues can be solved by the proposed economic intelligence and by using the BI with the framework for classification and analysis from this work. The organization of the paper is given here with the following sections included with a detailed discussion on CRM and the roles of BI are given in section 2. A detailed discussion on the proposed testing method is discussed with appropriate examples and description in section 3. Experiment results and evaluation of the existing and proposed test methods are given in section 4 and finally a summary of the complete work from the proposed research work and results are given in conclusions of the section 5.

OPTIMAL CRM IN TELECALLING INDUSTRY

The role of supply chain plays a vital role in the telecalling industry which establishes a crucial relationship between different suppliers, distributors, customers and related companies which provides the designing, construction and business associations. Adding the role of BI to SCM helps to include key information of business events and helps to detect various events and prospective business environments and expectations. BI helps to identify the business trends, changes and allows making business decisions with the uncertainties and inconsistent information of various business entities. Such applications of BI are found and followed in many government, retail, financial and service sectors. Therefore with the positive experience of BI in all those sectors makes this to implement in telecommunications as well due to involvement of wide range of people and their needs. In the recent times, various telecom industries using BI reported with their success stories towards improvements, profits for the managements, development strategies, opportunities, quality of decision-making processes, etc.

A. System Overview

In the proposed architecture, the data was processed by using the K-means algorithm from the data mining techniques to improve the overall performance of the proposed system as shown in Figure 1. Ahead of this the data to be processed is prepared in a scientific way using proper data collection method, then followed by integration and data clean up processes. The HGAPSO algorithm is used in this work to select different attributes and also for the elimination of unnecessary attributes too.

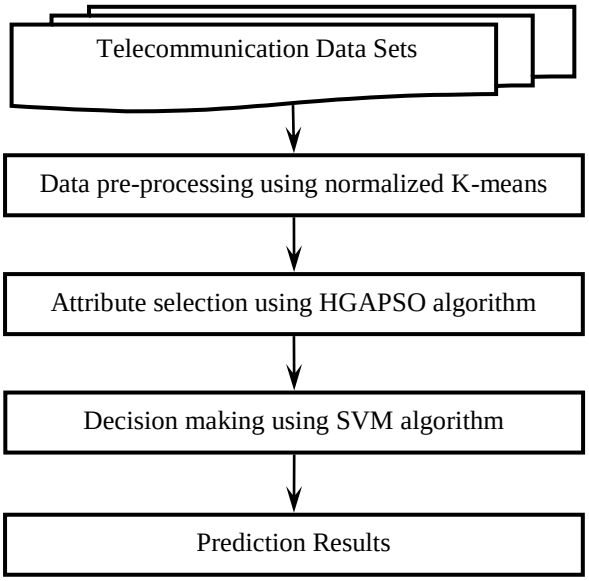

Figure 1. Overall Architecture of Proposed System

B. Data Pre-processing using Normalized K-means Algorithm

There are many definitions for cluster as "… a set of point groups into the cluster according to the distance metric". The processing of the clustering is a result of different clusters with individual set of points which are formed with small distances to each other and are expected to be very much far from the other clusters. The similar points in a cluster from the remote values are detected by this method. Some of the points will be dropped from the clustering process and are known by outliers. Valid results can be produced using the data mining applying the normalization process effectively on the data sets. It can be used as a tool to adjust the properties of datasets and to assign them equal weights, so as to filter the unnecessary objects and to produce the consistent and data with high accuracy. Here K-means use Euclidean distance due to its high susceptibility for different types of feature sizes.

In the market we have different types of normalization schemes and the best scheme will be based totally on the data to be normalized. At this juncture, in the algorithm a flow chart minimum and / or maximum is used, since data set is finite and due to inconsistency between both values (i.e. maximum and minimum). This Min-Max scheme helps to perform the linear data transformation and in the present scenario the data is matched with the predetermined limit and it within the range defined.

To obtain the results using uniform scheme for each attribute the data points are considered and will produce the initial centroids which follow the average of different data distributions. Later a different sorting scheme will be implemented on the obtained results and classify them into subsets, which in turn represent different number of clusters within the different data sets. In the final stage, a closest average value from the subsets is considered as initial center of gravity. Using this scheme each attribute is defined with a new weight based on the related attribute and the algorithm for the same is shown below.

Algorithm 1: Step involved in N-K means algorithm

INPUT: A dataset with dimensions d

OUTPUT: Clusters

1. Providing the initial datasets
2. Discovering both Max and Min values of each feature from datasets.
3. Applying the normalization process to the real scalar values of datasets using extreme values by using the equation: '=(- ( ) / (( ) - ( )))(1)

Here, ( ) and ( ) represent minimum and maximum values of the attributes.

4. Using the algorithm 2, pass all the number of clusters and then generate initial centroids.
5. Finally, produce the clusters.

Algorithm 2: Initialization of Centroids

1. First compute the average score for individual data points.

a= 1, 2, 3, 4, …

a.( ) = (1 * 1+2 * 2+3 * 3+4 …….*)/, in which ___ represents the attributes values; denotes the number of attributes and indicates the weight to be multiplied for confirming the distribution of the cluster.

2. Organize data based on the average scores.
3. Segment the data based on the subsets
4. Computing mean value of individual subsets.

Consider the nearest possible data point from the mean as initial centroids for individual data subsets.

C. HGAPSO ForAttribute Selection

To increase the accuracy of the predictions and reducing the cursing problems, the choice of attributes plays a vital role. Such choices are in general designed to identify the important and key information from specific set of the attributes. This process can be considered to be an optimization complexity, which is combined with variety of developments from the possible solutions that may not be suitable for a comprehensive research. In this research, twenty (20) attributes are considered as listed in below. The attributes listed below have all the information about the customers who recently passed away, which is called as the Churn. The detailed information of the customer on a phone with different applications usage is considered in these attributes and also contains the customer account information such as gender, age, marital status, etc. with overall payment details as well.

gender, seniors, finding partners, long-term relationships, telephone support,	multiple lines, Internet services, security, online, online backup,	protection devices, technical support, TV streaming, streaming movies, contracts,	paperless settlement, payment methods, monthly fees, total payments, departures

The above attributes are in general classified into different categories with a 'yes' or 'no' and / or '0' or '1' values. These values of different attributes are selected based on HGAPSO algorithm. The overall approach of evolutionary learning depends on the hybrid GA and PSO[12,13,14,15]. Using HGAPSO an elite strategy is considered, where upper half of best-performing person is considered to be an elite and these elites are improved at the initial stages, but not supposed to be copied for the next generation. Using the PSOs the elites are strengthened to resemble the maturity of the nature. Such intensive elite accounts for helping the new generations and covers half of the population and other half will be created by cross and mutating the elites.

The PSO systems will generate the initial population with a closest value to global optimality to solve the optimization problems in the first phase. Using the algorithm GA acquires initial population and continuously solves the optimization problems. The HGAPS based algorithm used for attribute selection is given below step-by-step.

Step 1: executing the declared attributes and total number of decision process

Step 2: generating the initial unit quantities that are based on the CRM limits and then to check only one attribute for each line

Step 3: after running the attributes, evaluate the fitness function as defined in [7]. This should be carried out for each person in that group

Step 4: updating the speed and population using [9,10,11]

Step 5: go to next step and then go to Step (3) when maximum number of iterations occurred

Step 6: consider the last population as the initial population and then update the same with GA

Step 7: evaluate the fitness function as defined by [8] after activating the attributes in the group for each person

Step 8: go to Step 9 and then to Step 6 when the Stop criteria are met

Step 9: display the output results

D. Decision Making Using SVM

The vector machines are used to check the separations or the predictions of customer departures. Auxiliary vector machines (AVM) and supervisory algorithms will be used for the purpose of predictions in the data sets. SVMs are used for customer churn predictions as it is known to be the best for separation of hyperplanes among data classes. The same can be used to predict the data which is collected for more than six months. The customers are divided into one of the two predefined groups and the observation of each resignation will be made based on the information from last three months. In this manner actual setting of simulated weight loss prediction will not be complicated. Based on the capabilities of an individual the departure prediction system (DPS) will measure the prospects of market. In this work the receiver operating characteristics (ROC) curve and maximum quantile evaluation (MQE) values are used for providing the HGAPSO scores and finally compare the results obtained from GA and PSO by not separating the customers.

DATA COLLECTION

The data set was collected from a telecom organization (IBM Watson Analytics Telco) with complete set of mobile segment for this research work. The dataset contains information related to the behavior of the customers and the customer events that are taken place before they left the operator. Also the dataset contains the information about overall customers in the past one month. Some of the attributes available in the data set are listed here: a) customer signs on the phone, b) multiple lines, c) internet,

d) online security, e) online backup, f) device protection, g) technical support, and h) TV and video streaming.

Formula

Training dataset (D)

D = { , } , ∈, ∈ {-1, 1}

= 1 (1)

Here, D is the training dataset;

x and y are the input variables

[| + |] ≥ 1 ; i = 1 to N (2)

Separated variables

To reduce the error minimization the following formula can be used

Φ( ) = ½ || ||² (3)

Estimating Function

F(x) = Σ-1 (,) (,) + (4)

A. Evaluation Criteria

The data is evaluated based on different formulas and algorithms with different procedures as explained below. Based on which the results and summary for this work was concluded.

To quantify the accuracy of predicting the model, the area under receiver curve (AUC), sensitivity and specificity are used in this work.

If True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) are available in the confusion matrix.

Then the sensitivity is given by

(/(+)):

The proportion of positive cases is predicted to be positive.

The specificity is (/((+)):

The proportion of negative cases is predicted to be negative.

To assess the accuracy of a classifier independent of any threshold, ROC analysis can be used.

The horizontal axis and the vertical axis of ROC curve are given by using Eq. 5 and Eq. 6 respectively.

B. SVM Algorithm Procedure

The obtained dataset is $D = (x_1, y_1), \ldots\ldots\ldots (x_n, y_n)$, C

x and y are labeled samples; C is Class; Initialize vector v = 0, b = 0

v – vector and b – bias train an initial SVM and learn the model

For each ∈ dp // x_i is a vector containing feature describing example i

Classify using f( )

If f( ) < 1 // prediction class label

Find ' , ' for known data // ' , ' for new features add to known data

Minimize the error functioning using (3) and now estimate using (4).

If the prediction is found to be wrong then retain **Repeat**

End

Classify attributes as normal or abnormal

EXPERIMENT RESULTS

To evaluate the performance of HGAPSO, attribute selection used the training set of the customer information is given by

$x = 1 - specificity(t)$ (5)

$y = specificity(t)$ (6)

For measuring the accuracy of the model, AUC can be measured.For the best sensitivity and specificity the performance of the separation is shown in Fig. 2 with cutoff points too. The proposed HGAPSO model will be applied to training data for generating the customer samples. For the proposed HGAPSO schema is 96% with a specificity value of 90% of training samples. The customer sample data that is used for training samples, the ROC curve, that is proposed churn prediction model found to be located above the models with an increased curve between 80.13% - 83 %. The results obtained using HGAPSO model are far better than the results obtained using GA and PSO methods.

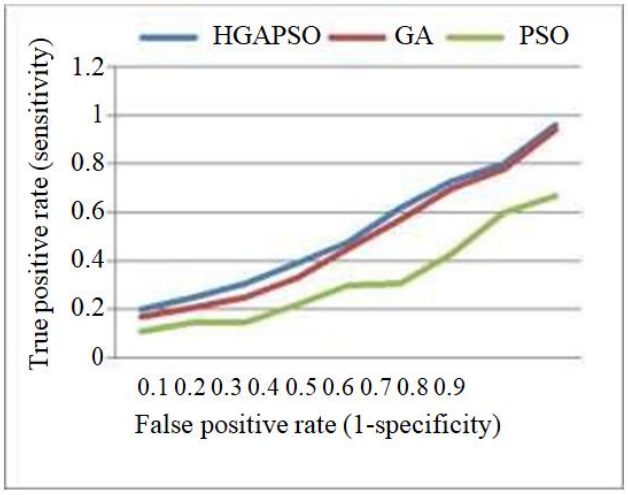

Figure 2: ROC Curves of the Separation

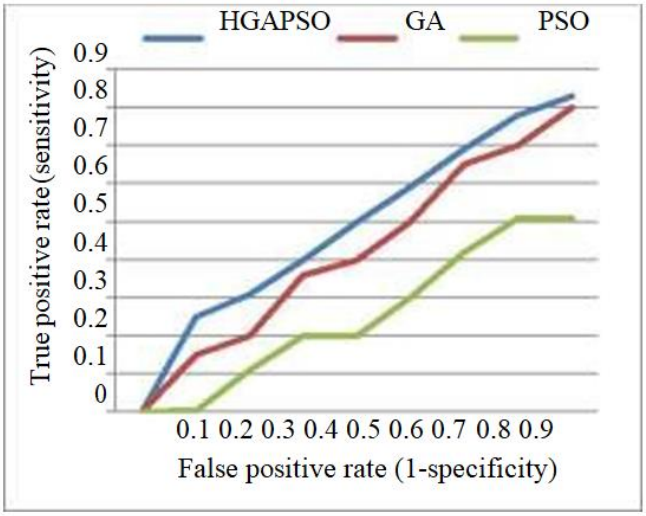

Figure 3: ROC Curves of the Predictions on Samples

A. Processing Time Comparison

From the above results it is clearly seen that greater efficiency and frequent churn prediction was possible by using the proposed HGAPSO model as compared to GA and PSO, is shown in Fig. 4. This model takes much less computation time for performing the prediction operation on the customer churn as compared to the existing models. Also the proposed model proved to be working efficiently and reported to be much faster as compared to GA and PSO.

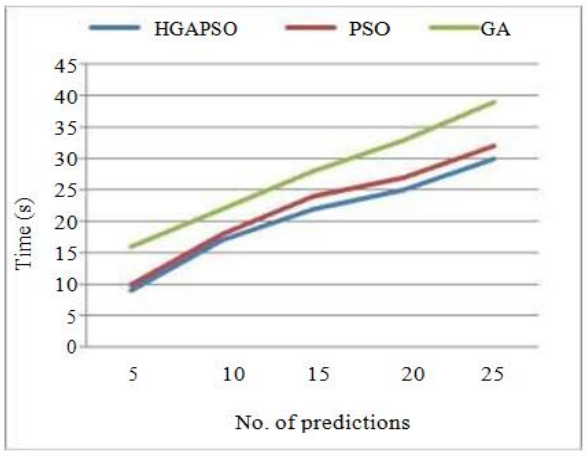

Figure 4: Processing the Predictions time Comparisons

B. Accuracy Comparison

From the Fig. 5, it can be seen that the proposed HGAPSO prediction model determines the frequent churn prediction at a greater level with good accuracy. The accuracy results are based on the number of predictions and both are proportional in relation between them. High accuracy rate was obtained using the proposed HGAPSO model and the area under receiver (AUR) curve increased from 98.13 - 98.3%.

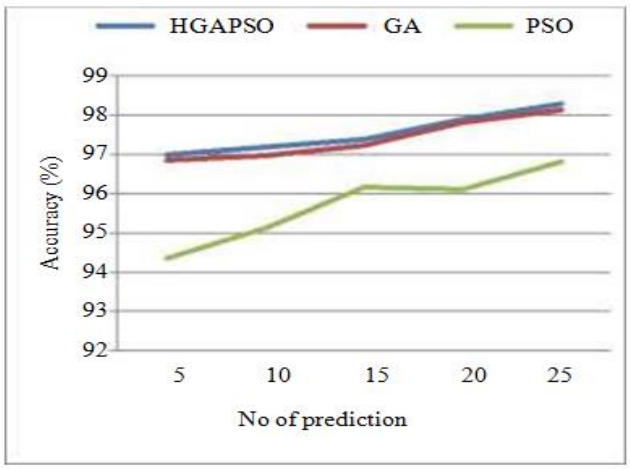

Figure 5: Processing the Accuracy Comparisons

CONCLUSIONS

In this paper the large number of customer's data from the telecom industry is considered for the customer churn prediction and the objective of this work to introduce a noble method to prediction process using the data mining techniques. This method is more suitable for the telecom industry for finding the opinion of people visiting the telecom outlets and their expectations from the company to obtain good services. The data sets obtained from the industry are used effectively to process the data and predict the customer behavior with respect to various aspects and their opinions. This is carried out using hybrid genetic algorithm with particle swarm optimization (HGAPSO) which provides an additional provision of using large quantity of telecom data. The decision making on the customer churn prediction is carried out successfully using this method with more accuracy. The performance metrics considered in this paper includes precision recall, f-measure, accuracy, TPR, FPR, time complexity and ROC. The proposed method is highly scalable and is very much useful for prediction and to incorporate the business intelligence in the telecom industry.

REFERENCES

1. Rashidirad M, Salimian H, Soltani E &Fazeli Z, "Competitive strategy, dynamic capability, and value creation: Some empirical evidence from UK telecommunications firms", Strategic Change, Vol.26, No.4, (2017),pp.333-342.
2. Zhao ZY, Tang C, Zhang X &Skitmor M, "Agglomeration and competitive position of contractors in the international construction sector", Journal of Construction Engineering and Management, Vol.143, No.6,(2017).
3. Purdy C, Evelyne, ROCH, Schubert D & Strong P, "U.S. Patent Application No. 15/046,039",(2016).
4. Pucciarelli F & Kaplan A, "Competition and strategy in higher education: Managing complexity and uncertainty", Business Horizons, Vol.59, No.3, (2016),pp.311-320.
5. Brockhoff K, Customer Integration into Continuous Development of IT-based Services, The Palgrave Handbook of Managing Continuous Business Transformation, (2017),pp.315-334.
6. Walo MT, "Net chain: An analytical approach to local economic development", The Marketing Review, Vol.16,No.1,(2016),pp.46-61.
7. Wynn M, Turner P, BanikA& Duckworth G, "The impact of customer relationship management systems in small business enterprises", Strategic Change, Vol.25, No.6, (2016),pp.659-674.
8. Pizam A, Pizam A, Shapoval V, Shapoval V, Ellis T & Ellis T,

"Customer satisfaction and its measurement in hospitality enterprises: a revisit and update", International Journal of Contemporary Hospitality Management, Vol.28, No.1, (2016), pp.2-35.
9. Guo-en, Wei ," Model of Customer Churn Prediction on Support Vector Machine", Systems Engineering Theory &Practice,Vol. 28,2008,pp.71-77
10. Hadden ,Tiwari , R.Roy ,D. Ruta.(2005). ," Computer assisted customer churn management: Stateof-the-art and future trends", Computers &Operations Research, Vol.34,2005,pp.2902-2917
11. Hwang, Jung ,Suh (2004) ,"An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry",Expert Systems with Applications, Vol.26,2004, pp.181–188 [14] Chiang , Wang , Lee , Lin, "Goal-oriented sequential pattern for network banking churn analysis". Expert Systems with Applications, Vol 25,2003,pp.293–302
12. Vaishali, R., Sasikala, R., Ramasubbareddy, S., Remya, S., &Nalluri, S. (2017, October). Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. In 2017 International Conference on Computing Networking and Informatics (ICCNI) (pp. 1-5). IEEE.
13. Reddy, S. R. S., Nalluri, S., Kunisetti, S., Ashok, S., &Venkatesh, B. (2019). Content-Based Movie Recommendation System Using Genre Correlation. In Smart Intelligent Computing and Applications (pp. 391-397). Springer, Singapore.
14. Rajasekaran, R., Kanumuri, U., Kumar, M. S., Ramasubbareddy, S., & Ashok, S. (2019). Sentiment Analysis of Restaurant Reviews. In Smart Intelligent Computing and Applications (pp. 383-390). Springer, Singapore.
15. Chugani, S., Govinda, K., &Ramasubbareddy, S. (2018, February). Data Analysis of Consumer Complaints in Banking Industry using Hybrid Clustering. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) (pp. 74-78). IEEE.

AUTHORS PROFILE

Dr AttiliVenkataRamana,
Associate Professor,
Dept.ECM,
Sreenidhi Institute of Science and Technology,
Hyderabad.
avrrdg@gmail.com

Dr AnnaluriSreenivasaRao,
Asst Prof,
Dept.IT,
VNRVJIET,
Hyderabad.
annaluri.rao@gmail.com

Dr E. Kesavulu Reddy,
Asst Prof,
Dept.CS,
SVUCMCS,
SV University,
Tirupathi,
ekreddysvu2008@gmail.com