Efficient Customer Churn Prediction Model Using Support Vector Machine with Particle Swarm Optimization

T.Kamalakannan
AP BCA DEPT
VELS UNIVERSITY,
CHENNAI.
kannan722003@yahoo.co.in

Dr.P.MAYILVAHANAN,
PROFESSOR,
DEPARTMENT OF MCA, SCHOOL OF COMPUTING SCIENCES, VELS UNIVERSITY,
PALLAVARAM, CHENNAI - 600 117
mayilkadir@yahoo.com

Abstract—Today, telecommunication market all over the world is facing a severe loss of revenue due to fierce competition and loss of potential customers. To keep the competitive advantages and acquire as many customers as possible, most operators invest a huge amount of revenue to expand their business in the very beginning. Customer Relationship Management (CRM) strategy helps an organization to improve the business processes and technology solutions around selling, marketing, and servicing functions across all customer touch-points. One of the important issues in customer relationship management is churn prediction. It aims at identifying potential churning customers based on past information and prior behaviors. In this proposed work the data mining techniques were utilized for efficient churn prediction. Here the normalized k means algorithm is utilized for dataset preprocessing. Then the attributes are selected from preprocessed image by utilizing minimum Redundancy and Maximum Relevance (mRMR) approach. It tends to select attributes with a high correlation with the class (output) and a low correlation between themselves. Based on the selected attributes the customer churn separation or prediction is examined with the help of Support Vector Machine with Particle Swarm Optimization (SVM with PSO). In order to optimize the hyper parameters of SVM the PSO is used. And also it overcome the local optimal solution problem and obtains higher classification accuracy. The experimental results show that the proposed system achieves better performance compared with the existing system in terms of accuracy, true positive rate, false positive rate and processing time.

Keywords—Customer Relationship Management (CRM), Particle Swarm Optimization (PSO), Churn Prediction and Support Vector Machine (SVM).

I. INTRODUCTION

The emergence of service organizations in the corporate sector, the growing competition due to liberalization, and the growing expectations of customers propelled by globalization and facilitated by IT revolution - are defining new rules of game for existing private and public enterprises in India [1]-[4]. Telecom is one of the fast growing sectors among the services. The mobile revolution has created a new wave of interest among people to utilize telephone services. The firms offering the services are vying with each other to capture this sudden spurt in demand [5]. They are using their technology and marketing prowess to attract new customers and simultaneously retain their existing customers, make inroads into competition and at the same time build entry barriers to competition to defend their position.

Customer-driven initiatives to attract, retain and build intimate long term relationship with profitable customers, innovation and delivery of Quality Service have become the key elements in the marketing strategies [6]. Relationship Management with the three focal points – Customer Perceived Value, Customer Satisfaction and Customer Loyalty – has become key success factor in achieving sustained customer patronage and profitability to the firm. Customer Relationship Management (CRM) is an effective tool to achieve this goal. The philosophy and practices of CRM in telecom has caught the attention of policy makers, academicians and researchers.

Customer Relationship Management (CRM) is a process focused on using customers’ information to create, develop and maintain long-term, profitable relationships through customers’ value perception increment that will reflect on maximization of return for shareholders [7]. Based on this concept, it would be inadequate to consider CRM as an Information Technology system, as this is not enough to understand and nurture the relationship between customer and company. Instead, CRM connects Information and Communication Technologies (ICTs) with the strategies of Relationship Marketing, through deliverance of maximum value to customers. CRM has become a relevant strategy for organizations, since its application in business may successfully improve focus on customer needs [8].

Identification of churners is a challenging task and as markets have become increasingly saturated, global telecommunications service companies have acknowledged that managing customer churn is of great concern. It is becoming a more serious problem as the market matures. In [9] focused on building an accurate and succinct predictive model with the purpose of churn prediction by using a Partial Least Squares (PLS) based method on highly correlated data sets among variables. They not only present a prediction model to accurately predict customers churning behaviour, but also a simple but implementable churn marketing program was employed. The proposed methodology allows the marketing
managers to maintain an optimal (atleast a near optimal) level of churners effectively and efficiently through the marketing programs. Here, PLS is employed as the prediction modeling method.

II. LITERATURE REVIEW

Rajini&Sangamaheswary argued that telecom companies must enhance their trust worthiness by holding customers at heart, offering customized services and listening to their customers to make every interaction of their customers a lifelong experience. Aggressive customer centric strategies can only help the telecom companies to survive and retain their market share. In the highly dynamic business environment, companies with customer focused strategies can only win the battle for customers. Knowledge about the choices and preferences of the customers help telecom companies to improve quality of services, better network coverage, relationship development, price perception, brand image, trust and customer expectations. This not only helps in maximizing customer satisfaction but also leads to increased loyalty [10].

Ning Lu [11] designed the use of boosting algorithms to enhance a customer churn prediction model in which customers are separated into two clusters based on the weight assigned by the boosting algorithm. As a result, a highly risky customer cluster has been found. Logistic regression is used as a basis learner, and a churn prediction model is built on each cluster, respectively. The experimental results showed that boosting algorithm provides a good separation of churn data when compared with a single logistic regression model.

Lee et al. [12] focused on building an accurate and succinct predictive model with the purpose of churn prediction by using a Partial Least Squares (PLS) based method on highly correlated data sets among variables. They not only present a prediction model to accurately predict customers churning behaviour, but also a simple but implementable churn marketing program was employed. The proposed methodology allows the marketing managers to maintain an optimal (atleast a near optimal) level of churners effectively and efficiently through the marketing programs. Here, PLS is employed as the prediction modelling method.

Bose et al designed a two-stage hybrid models to combine unsupervised learning technique with supervised learning technique. It developed a model for the prediction of customer churn. The important decision is the separation of churners from non-churners in customer churn management. Decision tree model are very popular in prediction of churn. It used multiple variables for clustering and examines different hybrid approaches for utilizing the results of clustering in order to build supervised learning models for prediction of churn. In the hybrid method, clustering used as a first stage and decision tree used as a second stage. C5.0 decision tree models with boosting improved the performance of models in term of top decide lift [13].

WouterVerbeke designed the application of Ant-Miner+ and ALBA algorithms on a publicly available churn prediction dataset in order to build accurate as well as comprehensible classification rule-sets churn prediction models [14]. Ant-Miner+ is a high performing data mining method based on the principles of Ant Colony Optimization which allows to include domain knowledge by imposing monotonicity constraints on the final rule-set. The advantages of Ant-Miner+ are high accuracy, comprehensibility of the generated models and the possibility to demand intuitive predictive models. Active Learning Based Approach (ALBA) for SVM rule extraction is a rule extraction algorithm, which combines the high predictive accuracy of a non-linear support vector machine model with the comprehensibility of the rule set format.

III. PROPOSED METHODOLOGY

Today is the competitive world of communication technologies. Customer Churn is the major issue that almost all the Telecommunication Industries in the world faces now. In telecommunication paradigm, Churn is defined to be the activity of customers leaving the company and discarding the services offered by it due to dissatisfaction of the services and/or due to better offering from other network providers within the affordable price tag of the customer. This leads to a potential loss of revenue/profit to the company. Also, it has become a challenging task to retain the customers. Therefore, companies are going behind introducing new state of the art applications and technologies to offer their customers as much better services as possible so as to retain them intact. Before doing so, it is necessary to identify those customers who are likely to leave the company in the near future in advance because losing them would results in significant loss of profit for the company. This process is called Churn Prediction. Data mining techniques are found to be more effective in predicting customer churn from the researches carried out during the past few years.

3.1. System overview

The overall performance of proposed system is illustrated in fig1. It is shows that the initial phase is data preparation, it is collected the data, integrated and cleaned. The data preprocessing is done using Normalization based K means algorithm. Then the attributes are selected using mRMR algorithm to search for relevant attributes, eliminate irrelevant or redundant ones. Finally the SVM with PSO algorithm is utilized for decision making.
Clustering is defined as grouping set of objects given in Eqs. (1) and (2).

$$V = \frac{(v - \min(e)) \cdot (\max(e) - \min(e))}{(v - \min(e)) \cdot (\max(e) - \min(e))}$$

where, min(e) and max(e) are the minimum and maximum values for attribute E.

4. Pass the number of clusters and generate initial centroids using algorithm 2.
5. Generate clusters.

Algorithm 2: Initialization of centroids
1. Calculate the average score of each data point.
   a. di = x1, x2, x3, x4, ..., xn
   b. di(avg) = (w1 * x1 + w2 * x2 + w3 * x3 + ..., wm * xm)/m

2. Sort the data based on average score.
3. Divide the data based on k subsets.
4. Calculate the mean value of each subset.
5. Take the nearest possible data point of the mean as the initial centroid for each data subset.

3.3. MRMR based feature reduction

The minimum Redundancy and Maximum Relevance (mRMR) is a feature selection approach that tends to select features with a high correlation with the class (output) and a low correlation between themselves. The mRMR works by selecting those features which show strong correlation with class labels while not being dependent on each other. The mRMR extracts the subset of features from the preprocessed dataset, which are not redundant with class labels thus showing maximum relevance. While maintaining high correlation with class labels, the instances can be selected to be mutually far away from each other, resulting into minimum redundancy. The maximum relevance criterion is implemented with the help of expressions given in Eqs. (1) and (2). In this section let us consider that a feature is represented by a random variable v, and the target classification as c.

$$\max D(S, c), D(|\{v_i, i = 1, ..., m\}; c)$$

$$\max D(S, c), D = \frac{1}{|S|} \sum_{i} e(I(v_i; c))$$

Both the above criteria of minimizing redundancy and maximizing relevance are combined and U operator is defined as given in Eq. (4). This simplest form is used to optimize both D and R.

$$\min R(S), R = \frac{1}{|S|} \sum_{v_i, v_j} I(v_i; v_j)$$

Both the above criteria of minimizing redundancy and maximizing relevance are combined and U operator is defined
as given in Eq. (4). This simplest form is used to optimize both D and R.

\[
\max \phi(D,R), \phi = D - R \quad (4)
\]

The mRMR selects a feature subset from preprocessed dataset, which has strong relevance with class targets and at the same time having maximum unique values, which consequently improves the classification performance.

### 3.4 Decision making using SVM with PSO

SVM is a kind of learning procedure in accordance with the statistical learning theory is illustrated in Figure 2.

![Figure 2. Illustration of SVM optimization of the margin in the feature space](image)

In case of classification problems, the major intention of SVM is to discover an Optimal Separating Hyper-plane (OSH) that appropriately classifies data points to the extent that is possible and separates the points of two classes to the extent that is possible, by minimizing the risk of misclassifying the training samples and hidden test samples.

Provided a set of points \( b \in R \) with \( i = 1, \ldots, N \) each point \( x_i \) belongs to either of two classes with the label \( y_i \in \{-1, +1\} \). The optimization problem for the SVM can be given as below:

\[
\min \phi(w, b) = \frac{1}{2}(w, w)
\]

subject\( \sum y_i (w, x_i + b) \geq 1 \quad (5)
\]

This is also regarded as hard margin, where no space is provided for errors. It is to be noted that most of the time it is linearly non separable. As a result, slack variable is introduced for the purpose of permitting error and the optimization function takes the form of (6) as given below:

\[
\min \phi(w, b) = \frac{1}{2}(w, w) + c \sum_{i=1}^{l} \xi_i
\]

subject\( \sum y_i (w, x_i + b) \geq 1 \quad (6)
\]

The OSH binary decision classes is given as,

\[
f(x) = \sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \quad (7)
\]

\( \alpha_i, K \) represents constants.

Nonlinear separable problems can be solved as given below:

\[
\min \phi(w, b) = \frac{1}{2}(w, w) + c \sum_{i=1}^{l} \xi_i
\]

subject\( \sum y_i (w, \phi(x) + b) \geq 1 \quad (8)
\]

Where \( K(x_i, x) = \phi(x_i)^T + \phi(x) \) is taken with a semi positive kernel meeting the mercer theorem.

For kernel function \( K(x_i, x) = (x_i^T, x) \)

**Particle swarm optimization**

In the scenario of obtaining optimized hyper parameters of SVM, an effective Swarm Intelligent technique, PSO is incorporated into SVM. PSO algorithm was formulated for real parameter optimization. Consider, in PSO algorithm, the solution space of the problem is D-dimensional, where D indicates the amount of parameters to be optimized. Here, the parameters involved are \( w, b \) and \( \beta \). The fitness value of the randomly selected site is given in the following equation (9):

\[
\text{fit}_i = \frac{1}{(1 + \text{obj. Fun}_i)} \quad (9)
\]

In PSO, a group of particles explores the solution space to find the best solution. In particle swarm optimization, each solution is treated as an individual bird in the swarm, i.e., a particle in the search space. Position of \( i \) th particle is denoted by \( X_i = (x_1, x_2, \ldots, x_D) \) and velocity \( V_i \) be \( (v_1, v_2, \ldots, v_D) \) in the D-dimensional search space. Each particle uses its individual experience and its neighboring particles experience to move in the search space, that is, every particle has knowledge about its best position it has found so far and the best position found by swarm. We denote Pbi as the best position (pbest) of \( i \) particle, and gbest as the best position found by the swarm so far. Every particle is given random initial position and random initial velocity. Fitness function evaluates every particle in the search space. It determines the quality of the particle. Particles are attracted towards best particles in their search space which have best fitness values. The velocity and position of particles are updated according the equation

\[
V_{i+1} = \omega * V_i + c_1 * \text{rand } 1 * (P_{bi} - X_i) + c_2 * \text{rand } 2 * (\text{gbest} - X_i) \quad (10)
\]

where, \( i = 1, N \) represents the swarm size, rand1 and rand2 are two random numbers uniformly distributed in the range 0 and 1, and \( c_1 \) and \( c_2 \) are constants. Velocity is updated based on the inertia of previous velocity, experience of the particle itself and the experience of the neighboring particles. Particle position is updated by:

\[
X_{i+1} = X_i + V_{i+1} \quad (11)
\]

In every iteration, fitness of every particle is evaluated. Velocity is updated based on pbest and gbest values and parameters like \( c_1 \), \( c_2 \) and \( w \). Position is updated based on velocity. This process is repeated until maximum iterations or error criterion is reached.

With the intention of overcoming the fault of falling into local optimal solution which the entire common SVM parameters optimization schemes had in different degree, a new SVM parameters optimization scheme in accordance with PSO algorithm is proposed and applied to churn prediction. Kernel function constraints of SVM were considered as the optimization object, and classification accuracy of SVM was taken as fitness value. PSO is implemented here, in order to
realize the global optimal solution of parameter \( w \) and \( b \). It is observed that this scheme overcame the local optimal solution problem and obtain higher classification accuracy. The cost time of searching optimized parameters of small number classification problem was also considerably diminished. Hence, this scheme was applied to customer churn prediction for effective analysis.

**Figure 3. Flow chart of hybrid SVM with PSO**

**Pseudo code of PSO algorithm**

Step 1: Initialize the population of solution
Step 2: Parameters \( w, b \) initialization
Step 3: For every particle \( i \) (sample)
Step 4: Evaluate fitness value \( f_i \) with SVM in accordance with the training set
Step 5: If the fitness value is better than the best fitness value (pBest) in history
Step 6: set current value as the new pBest
Step 7: end
Step 8: Choose the particle with the best fitness value of all the particles as the gBest
Step 9: For each particle
Step 10: Calculate particle velocity according equation (10)
Step 11: Update particle position according equation (11)
Step 12: End
Step 13: While maximum iterations or minimum error criteria is not attained
Step 14: When check cycle value is maximum, show optimal w,b. Else, generate new solution again replicate all the steps.

Finally the hyperplane separates the churn customers. It is the effective means of implementing a customer relationship management strategy and helping telecommunications companies to keep their customers happy.

**IV. EXPERIMENTAL RESULTS AND DISCUSSION**

In this section, in order to assess the performance of SVM with PSO over a six-month period. Every customer is categorized into one of two predetermined groups and his/her churn propensity is monitored and revised in accordance with his/her most recent three-month information.

In this manner, it is simple to simulate the real-world setting of churn prediction. A churn prediction system is supposed to be measured through its capability to recognize churners for marketing purpose, and in this work use the Receiver Operating Characteristic (ROC) curve and top-quantile.lift values to provide a complete evaluation of SVM with PSO scheme.

**4.1. Data collection**

In this research work, a data set is selected from a telecommunication organization which comprises a segment of mobile customers.


This data set provides information of behavior to retain customers. A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors.

They need to understand who is leaving, and they have to analyst at this company and find out who is leaving and why. The data set includes information about: Customers who left within the last month -- the column is called Churn. Services that each customer has signed up for -- phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information -- how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges.

**4.2. Evaluation criteria**

In this study, the Area Under Receiver Curve (AUC), sensitivity, and specificity are used to quantify the accuracy of the predictive models. If True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are the TP, FP, TN and FN in the confusion matrix, then the sensitivity is \((TP / (TP + FN))\): The proportion of positive cases which are predicted to be positive.

The specificity is \((TN / (TN + FP))\): The proportion of negative cases which are predicted to be negative [19]. To assess the accuracy of a classifier independent of any threshold, ROC analysis can be used. The horizontal axis and the vertical axis of an ROC curve are defined by Equations 12 and 13 respectively.

\[
x = 1 - \text{specificity}(t) \quad (12)
\]

\[
y = \text{sensitivity}(t) \quad (13)
\]

To measure the accuracy of a model, the AUC can be measured.
The proposed SVM with PSO and existing SVM based churn prediction methods are compared in terms of Processing time. In x-axis number of prediction is taken and processing time is taken as y-axis. In this proposed work, the hyperplane parameters such as $w, b$, and $\theta$ are optimized by using PSO algorithm, it reduced the processing time of the proposed system.

**Accuracy comparison**

Accuracy of the given prediction model is defined as below

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(14)

Figure 6. Accuracy comparison

The SVM with PSO prediction model discovers frequent churn prediction and much greater accuracy results than existing prediction model which is shown in Fig 5. In this proposed research, mRMR extracts optimal attributes from the preprocessed dataset, which have strong correlations with class labels thus showing maximum relevance. When the number of prediction is increases the accuracy of the result is increases.

The proposed HGAPSO produces high accuracy rate when compared to existing system.

**V. CONCLUSION**

Predicting churn customers using CRM strategies and data mining techniques is an important issue for telecom companies in competitive markets of today. In this work, dataset preprocessing is performed by using normalized k-means algorithm. The attributes are selected by using mRMR approach. Based on the selected attributes the decision making about churn is performed using SVM with PSO approach. In this work, the PSO algorithm is utilized for optimal selection of hyperplane parameters. The experimental results show that the proposed system achieves better performance compared with the existing system in terms of accuracy, true positive rate,
false positive rate and processing time. However, usage and billing features were still of primary importance, while demographics may also affect the churn prediction. In the future, research is planned for mining further churning behaviours and developing retention strategies.

REFERENCES


