Applied Soft Computing 19 (2014) 31–40

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Churn prediction using comprehensible support vector machine: An analytical CRM application

M.A.H. Farquad[a,b,c], Vadlamani Ravi[a,*], S. Bapi Raju[b]

[a] Institute for Development and Research in Banking Technology, Castle Hills Road #1, Masab Tank, Hyderabad 500057, AP, India
[b] Department of Computer and Information Sciences, University of Hyderabad, Hyderabad 500046, AP, India
[c] School of Business, The University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 9 February 2011
Received in revised form 20 August 2013
Accepted 19 January 2014
Available online 4 February 2014

Keywords:
Churn prediction
Support vector machine
Rule extraction
Naive Bayes Tree
Machine learning and customer relationship management

ABSTRACT

Support vector machine (SVM) is currently state-of-the-art for classification tasks due to its ability to model nonlinearities. However, the main drawback of SVM is that it generates "black box" model, i.e. it does not reveal the knowledge learnt during training in human comprehensible form. The process of converting such opaque models into a transparent model is often regarded as *rule extraction*. In this paper we proposed a hybrid approach for extracting rules from SVM for customer relationship management (CRM) purposes. The proposed hybrid approach consists of three phases. (i) During first phase; SVM-RFE (SVM-recursive feature elimination) is employed to reduce the feature set. (ii) Dataset with reduced features is then used in the second phase to obtain SVM model and support vectors are extracted. (iii) Rules are then generated using Naive Bayes Tree (NBTree) in the final phase. The dataset analyzed in this research study is about Churn prediction in bank credit card customer (Business Intelligence Cup 2004) and it is highly unbalanced with 93.24% loyal and 6.76% churned customers. Further we employed various standard balancing approaches to balance the data and extracted rules. It is observed from the empirical results that the proposed hybrid outperformed all other techniques tested. As the reduced feature dataset is used, it is also observed that the proposed approach extracts smaller length rules, thereby improving the comprehensibility of the system. The generated rules act as an early warning expert system to the bank management.

1. Introduction

CRM is a process or methodology used to learn more about customers' need and behaviours in order to develop stronger relationship with them. CRM involves the continuous use of refined information about current and potential customers in order to anticipate and respond to their needs and draws on a combination of business process and Information Technology to discover the knowledge about the customers and answer questions like, "*who are the customers?*", "*what do they do?*" and "*what they like?*". Therefore the effective management of information and knowledge is central and critical to the concept of CRM for;

- Product tailoring and service innovation (web-sites tailored to customer needs, taste experience and the development of mass customisation).
- Providing a single and consolidated view of the customer.
- Calculating the lifetime value of the customer.
- Designing and developing personalized transactions.
- Multichannel based communication with the customer.
- Cross-selling/up-selling various products to customers.

Different definition of CRM put emphasis on different perspectives. CRM's technological perspective was stressed in [1], its knowledge management perspective was emphasized in [2] and its business re-engineering and continuous improvement perspective is presented in [3].

We can think about CRM at three levels, strategic, analytical and collaborative.

Strategic CRM: It is focused on development of a customer-centric business culture. Product, production and selling are the three major business orientations identified by Kotler [4].

Analytical CRM: Analytical CRM builds on the foundation of customer information. Customers' data may be found in enterprise wide repositories, sales data (purchasing history), financial data (payment history and credit score), marketing data (campaign response, loyalty scheme data) and service data. With the application of Data Mining, the industry can then interrogate this data and

* Corresponding author. Tel.: +91 40 2329 4042.
E-mail addresses: farquadonline@gmail.com (M.A.H. Farquad), rav_padma@yahoo.com, padmarav@gmail.com (V. Ravi), bapics@uohyd.ernet.in (S.B. Raju).

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2014.01.031

intelligent interrogation provides answers to questions, such as, “who are our most valuable customers?”, “which customer have the highest propensity to switch to competitors?”, “which customers would be most likely to respond to particular offer?” and so on.

Collaborative CRM: Staff members from different departments can share information collected when interacting with customers [5].

Churn prediction problem is an analytical CRM application and using the extracted rules from SVM, service providers can get transparent and efficient insight about their customers and can make better policies to retain their existing customers.

1.1. Churn prediction problem

Over the decade and half, the number of customers with banks and financial companies is increasing by the day and this has made the banks conscious of the quality of the services they offer. The phenomenon, called ‘*churn*’ i.e. shifting loyalties from one service provider to another occurs due to reasons such as availability of latest technology, customer-friendly bank staff, low interest rates, proximity of geographical location and varied services offered. Hence, there is a pressing need to develop models that can predict which existing ‘*loyal*’ customer is going to churn out or attrite in near future.

Service organizations need to be proactive in understanding the customers’ current satisfaction levels before they attrite [6]. Research indicates that the online bank customers are less price-conscious than traditional bank customers with less probability of churning out [7]. Targeting customers on the basis of their (changing) purchase behaviour could help the organizations do better business and loyalty reward programmes helps the organizations build stronger relationships with customers [8].

In the financial services industry two “critical” churn periods are identified [9], the first period is the early years after becoming a customer and the second period is after being a customer for some 20 years. A comparative study on Logistic Regression and Neural Network for subscriber data of a major wireless carrier is carried out and it is concluded that using sophisticated neural net $93 could be saved per subscriber [10].

Machine learning techniques such as; multilayer perceptron, Hopfield neural network, self-organizing neural networks [11], decision tree [12], multivariate regression analysis [13], logistic regression and random forest [14], emergent self-organizing feature maps (ESOM) [15], neural networks [10], SVM [18], genetic algorithms and rough set theory [16], ensemble with majority voting [17] and hybrid neural networks [18] are employed to solve churn prediction problems. Glady et al., proposed a churn prediction model using customer life time value (CLV), which is defined as the discounted value of future marginal earning, based on customers’ activity [19]. Hu presented a comparative study of different machine learning algorithms [20]. The trend in marketing towards building relationships with customers continues to grow and marketers have become increasingly interested in retaining customers over the long run [21]. Hyung-Su and Young-Gul suggested a performance measurement framework called CRM score card to diagnose and asses a firm’s CRM practice [22].

Churn prediction problem is one of the most important applications of analytical CRM in finance. Banks would be interested to know their *about-to-churn* customers and the proposed rule extraction approach do not only provide better predictions but also comprehensibility of the system is improved. Feature selection using SVM-RFE algorithm in the first phase reduces the dimensionality of the data by yielding the key attributes in the data. Thus, less number of rules and smaller rules are extracted resulting in the improvement of the comprehensibility of the system. During the research study in this Paper, various standard balancing techniques are employed such as, random under-sampling, random over-sampling and SMOTE.

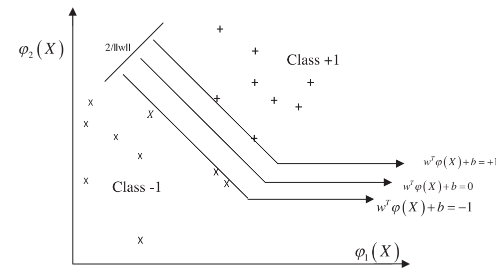

Fig. 1. Illustration of SVM optimization of the margin in the feature space.

Remaining paper is organized as follows. Section 2 provides the details about SVM and literature survey of rule extraction from SVM. Proposed rule extraction approach is then detailed in Section 3. Dataset description and experimental setup followed during this research study is presented in Section 4. Next section provides the detailed empirical analysis and observations. Section 6 concludes the paper.

2. Rule extraction from SVM

2.1. Support vector machine

The SVM is a learning procedure based on the statistical learning theory [23]. It has been used in wide variety of applications such as gene analysis [24], financial time-series forecasting [25], marketing [26], patent classification [27], face recognition [28] and predicting longitudinal dispersion coefficients in natural rivers [29].

For classification problems, the main objective of SVM is to find an optimal separating hyperplane that correctly classifies data points as much as possible and separates the points of two classes as far as possible, by minimizing the risk of misclassifying the training samples and unseen test samples [30]. The training points that are closest to the optimal separating hyperplance are called support vectors and other training examples are irrelevant for determining the binary class boundaries as shown in Fig. 1. To deal with non-linearly separable datasets problems, SVM first projects the data into a higher dimensional feature space and tries to find the linear margin in the new feature space.

Given a set of points $b \in \Re$ with $i = 1, \ldots, N$ each point x_i belongs to either of two classes with the label $y_i \in \{-1, +1\}$ [31].

The optimization problem for the SVM can be depicted as follows:

$$\begin{aligned} &\min \frac{1}{2}\langle w, w\rangle \\ &\text{Subject to } y_i(w \cdot x_i + b) \geq 1 \ \forall x_i \end{aligned} \tag{1}$$

The SVM classification function for classifying linearly separable data can be written as:

$$f(x) = \langle w, x\rangle + b = \sum_{i=1}^{l} y_i \alpha_i \langle x_i, x\rangle + b; \tag{2}$$

This is also known as hard margin, where no room is given for errors. It is observed that most of the time it is linearly non-separable. Hence slack variable ξ is introduced to allow ξ error and the optimization function takes the form of (3) as shown below:

$$\begin{aligned} &\min \frac{1}{2}\langle w, w\rangle + C\sum_{i=1}^{l} \xi_i \\ &\text{Subject to } y_i(w \cdot x_i + b) \geq 1 \ \forall x_i \end{aligned} \tag{3}$$