



## Combination of feature selection approaches with SVM in credit scoring

Fei-Long Chen<sup>a</sup>, Feng-Chia Li<sup>a,b,\*</sup>

<sup>a</sup> Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, ROC

<sup>b</sup> Department of Informational Management, Jen Teh Junior College, Taiwan, ROC

### ARTICLE INFO

#### Keywords:

Support vector machine  
Linear discriminate analysis  
Decision tree  
Rough sets theory  
F-score

### ABSTRACT

The credit scoring has been regarded as a critical topic and its related departments make efforts to collect huge amount of data to avoid wrong decision. An effective classificatory model will objectively help managers instead of intuitive experience. This study proposes four approaches combining with the SVM (support vector machine) classifier for features selection that retains sufficient information for classification purpose. Different credit scoring models are constructed by selecting attributes with four approaches. Two UCI (University of California, Irvine) data sets are chosen to evaluate the accuracy of various hybrid-SVM models. SVM classifier combines with conventional statistical LDA, Decision tree, Rough sets and F-score approaches as features pre-processing step to optimize feature space by removing both irrelevant and redundant features. In this paper, the procedure of the proposed approaches will be described and then evaluated by their performances. The results are compared in combination with SVM classifier and nonparametric Wilcoxon signed rank test will be held to show if there is any significant difference between these models. The result in this study suggests that hybrid credit scoring approach is mostly robust and effective in finding optimal subsets and is a promising method to the fields of data mining.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Consumer credit prediction is a very important issue in the credit industry. With the rapid growth in this field, credit scoring models have been widely used for the credit admission evaluation. The credit scoring models are developed to distinguish which customers belong to good class (accepted) or bad class (rejected) with their related attributes such as income, marital status and age, or based on the past records. Most credit scoring models have been widely developed by reducing redundant features to improve the accuracy of credit scoring models during the past few years. Dash and Liu (1997) provided a detailed survey and overview of the existing methods for feature selection and suggested a feature selection process that consists of four parts including feature generation, feature evaluation, stopping criteria and testing. The classic evaluation measures such as accuracy, information, distance and dependence were used for removing irrelevant features. However, artificial intelligence and machine learning techniques have been used to solve these decision-making problems. The modern data mining techniques have been adopted to build the credit scoring models (Huang, Chen, & Wang, 2007). Researchers have developed a variety of conventional statistics approaches which involve

linear discriminate approach (Bellotti & Crook, 2008; Lee & Chen, 2005; Thomas, 2000), decision tree approach (Huang, Tzeng, & Ong, 2006), Rough sets theory approach (Caballero, Alvarez, Bel, & Garcia, 2007), F-score approach (Chen & Lin, 2005) and genetic programming approach (Ong, Huang, & Tzeng, 2005). Recently, researchers have proposed the hybrid data mining approach in the design of an effective credit scoring model. Lee and his colleagues (Lee, Chiu, Lu, & Chen, 2002) integrated neural network with traditional discriminate analysis approach and Chou (Chou, Lin, Liu, & Chang, 2006) applied machine learning techniques such as ANN, DT and SVM to solve decision-making problems. Generally, credit scoring can be regarded as the binary classification problem of classifying an observation into pre-defined groups. Previous studies focused on increasing the accuracy rate of credit scoring model. However, even a little bit improvement will cause noteworthy cost savings. According to previous studies, machine learning techniques are superior to that of traditional approaches in dealing with credit scoring problems, especially in non-linear pattern classification. For conventional statistical classification, an underlying probability model should be assumed. The more recently developed data mining techniques can perform the classification task without this limitation and achieve better performances than traditional statistical approaches (Huang et al., 2007).

Feature subset selection algorithms can be classified into two categories: the filter approach and the wrapper approach (Liu, 1998). The filter approach first selects important features subsets. It separates features from classifier that are independent of any

\* Corresponding author. Address: Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan, ROC. Tel.: +886 3 5715131; fax: +886 3 5730902.

E-mail addresses: [lear999@gmail.com](mailto:lear999@gmail.com), [d937818@oz.nthu.edu.tw](mailto:d937818@oz.nthu.edu.tw) (F.-C. Li).

learning algorithm. The filter approach relies on various measures of the general characteristics of the training data such as distance, information, dependency and consistency. The wrapper model usually uses the predictive accuracy of a pre-determined learning algorithm to determine the accomplishment of the selected subsets. These methods of learning algorithm are computationally expensive for data with a large number of features (John, Kohavi, & Pfleger., 1994). Generally, filters are faster and can be used as a pre-processing step to reduce space dimensionality and over-fitting. On the other hand, wrapper approach may perform better in finding useful subsets of relevant variables (Guyon & Elisseeff, 2003). However, the problem is known to be NP-hard (Amaldi & Kann, 1998) and the search becomes quickly computationally intractable. In this study, both filter and wrapper approaches will be used. SVM classifier is combined with feature selection approaches to perform better classification.

Before designing the hybrid-SVM classifier, a kernel function must be chosen. The grid search algorithm is a proper procedure to finding the best penalty parameter  $C$  and the kernel function parameter  $\gamma$ . To compare various credit scoring models completely, this study tries the whole features into SVM classifier and four SVM-based feature selection approaches, i.e., (1) using grid search to optimize model parameters without choosing features, and (2) using grid search to optimize model parameters after selecting features by LDA, RST, DT and F-score approaches.

This paper is organized as follows. Sections 2 and 3 describe the concepts of the four strategies combined with SVM and basic SVM in this research respectively. Section 4 presents the experimental results from the proposed approaches to classify two real-world data sets. Section 5 give remarks and provide a conclusion.

## 2. Basic concepts of feature selection approaches

### 2.1. Linear discriminate analysis approach

Linear discriminate analysis (LDA) is a well-known technique which was first proposed by Fisher as a classification technique (Fisher, 1936). LDA has been regarded as a data mining technique in handling classification problems which reduces the observed variables into a smaller number of dimensions that would result in decreasing the number of features to be considered by the classifiers. Rather than directly eliminating irrelevant or redundant variables from the original feature space, LDA merely transforms the original variables through linear combination into a new subset of variables. Thus, the linear methods provide a way of understanding the data, but they are not able to reduce the number of original features (Li, 2006). The LDA can be expressed as

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \tag{1}$$

where  $y$  represents the discriminate score,  $\beta_0$  is the intercept term.  $\beta_i$  ( $i = 1, \dots, n$ ) represents the  $\beta$  coefficient associated with the explanatory variable  $x_i$  ( $i = 1, \dots, n$ ). LDA is a traditional statistical method and the credit scoring classification accuracy of LDA has been treated as the benchmark to other modern classification approaches. The simple parametric model was the first model employed for credit scoring. However, the covariance matrices of the good and bad credit classes may be unequal for the nature of the credit data. Researchers are investigating hybrid models to overcome the deficiencies of the LDA model. One of the efforts is combined with SVM for credit scoring applications.

### 2.2. Rough sets theory approach

Rough sets theory (RST) is a mathematical tool that had been used successfully to discover data dependencies and reduce the

number of attributes contained in a data set by purely structural methods. RST was first proposed by Pawlak (1984) to deal with vagueness or uncertainty. Rough sets do not need any pre-assumptions or preliminary information about the data. One attribute is chosen as the decision variable and the rest of them are the condition attributes. Two partitions are formed in the mining process. The approach is based on the refusing certain set boundaries, implying that every set will be defined using a lower and an upper approximation. As can be observed from Fig. 1, the object that belongs to a set with certainty is called lower approximation while upper approximation contains all objects that may possibly belong to the set. Decision rules derived from lower approximation represents certain rules as well as extracted from upper approximation corresponds to possible rules. An important issue in the RST is about feature reduction based on reduct concept. A reduct is a minimal set of attributes  $B \subseteq A$  such that  $IND(B) = IND(A)$ , where  $IND(X)$  is called the  $X$ -indiscernibility relation. In other words, a reduct is a minimal set of attributes from  $A$  that preserves the partitioning of universe and hence the ability to perform classifications. RST has been successfully applied to real-world classification problems in a variety of areas, such as pattern recognition. Wang and his colleagues (Wang, Yang, Teng, Xia, & Jensen, 2007) proposed a new feature selection strategy based on rough sets and particle swarm optimization. Zhao and his colleagues (Zhao, Yao, & Luo, 2007) also made an empirical experiment for letter recognition for demonstrating the usefulness of the discussed relations and reducts. There are many other rough sets algorithms for feature selection. The basic solution to finding minimal reducts is to generate all possible reducts and choose any with minimal cardinality, which can be done by constructing a kind of discernibility function from the dataset and simplifying it. However, this is time-consuming and therefore is only practical for simple datasets. Finding minimal reducts or all reducts has been shown as NP-hard problems (Skowron & Rauszer, 1992).

### 2.3. Decision tree approach

Decision tree models are able to represent knowledge in a flexible and easy form. Their popularity is as a result of interpretability and implementation easily. The first decision tree generating algorithm is introduced by Quinlan (1979). Selecting an attribute to place at the root node is the first step to construct a decision tree, and then make one branch based on an attribute value test. This process is repeated recursively on each branch and only those instances that actually reach the branch. Once all cases at a node have satisfied a certain criterion, stop developing the part of the tree. Calculate the information gain for each attribute and choose the one that gains the most information to split on. The first

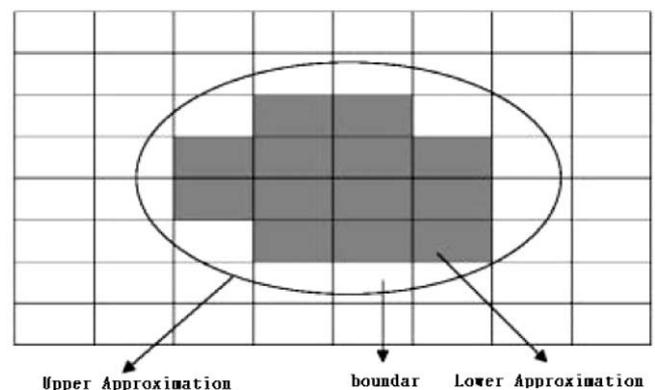


Fig. 1. Rough sets approximation.

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات