Expert Systems with Applications 39 (2012) 11709–11717

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Probabilistic and discriminative group-wise feature selection methods for credit risk analysis

Gülefşan Bozkurt Gönen [a,*], Mehmet Gönen [b], Fikret Gürgen [a]

[a] Department of Computer Engineering, Boğaziçi University, TR-34342 Bebek, İstanbul, Turkey
[b] Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University School of Science, FI-00076 Aalto, Espoo, Finland

ARTICLE INFO

Keywords:
Credit risk analysis
Feature selection
Probit classifier
Multiple kernel learning
Sparsity

ABSTRACT

Many financial organizations such as banks and retailers use computational credit risk analysis (CRA) tools heavily due to recent financial crises and more strict regulations. This strategy enables them to manage their financial and operational risks within the pool of financial institutes. Machine learning algorithms especially binary classifiers are very popular for that purpose. In real-life applications such as CRA, feature selection algorithms are used to decrease data acquisition cost and to increase interpretability of the decision process. Using feature selection methods directly on CRA data sets may not help due to categorical variables such as marital status. Such features are usually are converted into binary features using 1-of-*k* encoding and eliminating a subset of features from a group does not help in terms of data collection cost or interpretability. In this study, we propose to use the probit classifier with a proper prior structure and multiple kernel learning with a proper kernel construction procedure to perform group-wise feature selection (i.e., eliminating a group of features together if they are not helpful). Experiments on two standard CRA data sets show the validity and effectiveness of the proposed binary classification algorithm variants.

1. Introduction

Credit risk is the loss of capital in case of the credit borrower's failure for refunding the total amount of debt to recover the liability. Credit risk analysis (CRA) is an important topic in the financial management field and used by many financial organizations such as banks and retailers. Due to recent financial crises and more strict regulations, CRA becomes the major focus point of financial and banking industry since accurate estimation of credit risks enables a more efficient funding for world economy (Basel, 1988).

Banks are required to manage financial and operational risks for providing a safer environment for them within the pool of all international banks (Basel, 2004). A safer financial environment facilitates the transmission of money for convenient use in the economy. If a bank or financial organization wants to accomplish a long term success, it should follow an exhaustive and powerful strategy for CRA (Basel, 2011). Measuring the credit risk accurately also allows banks to organize upcoming lending transactions to achieve targeted return/risk characteristics. Nowadays, financial organizations are building their own software solutions to analyze their gathered data. Another benefit of CRA is for accounting companies. If an accounting company monitors a potentially troubled company and forgets to notify the credit borrower with a warning signal then the company can face with a costly lawsuit. Therefore, as the credit industry expands, CRA methods become comprehensively used to evaluate the credit applications (Thomas, 2000).

Up to now, many computational methods are proposed for CRA (Atiya, 2001). CRA methods generally aims to classify credit applicants into two groups, namely, *approved* and *disapproved*, according to properties of the applicants such as income, profession, possession, marital status, the number of people liable to look after, previous credit history, and age. Credit suppliers want to increase the volume of credit supply without increasing the failure ratio extremely (Huang, Chen, & Wang, 2007) and developing reliable computational models is the key to successful credit operations. CRA methods also ensure better examination of existing accounts, faster results in decision processing, and better precedence assignments for credit collections (Brill, 1998).

The main motivation of computational CRA methods is to have a robust binary classification algorithm for classifying credit applicants or a robust clustering algorithm for assigning them into predefined applicant categories. Most of the existing solutions in the literature formulates the problem as a binary classification problem and applies the standard classification algorithms such as decision trees (DTs), neural networks (NNs), and support vector

* Corresponding author.
E-mail addresses: gulefsanbozkurt@gmail.com (G. Bozkurt Gönen), mehmet.gonen@aalto.fi (M. Gönen), gurgen@boun.edu.tr (F. Gürgen).

0957-4174/$ - see front matter
http://dx.doi.org/10.1016/j.eswa.2012.04.050

machines (SVMs). The complexity of such learning algorithms mostly depends on the number of input features. Feature selection algorithms are proposed to reduce the number of features used for prediction. It is not necessary to use all of the input features since redundant features do not provide any useful information for classification. When we are able to explain the data with fewer features, we acquire better knowledge about the process that generates this data (Alpaydın, 2010). In this study, we also follow these lines of research using probabilistic and discriminative classification algorithms, namely, the probit classifier and multiple kernel learning (MKL), coupled with feature selection capability for lower data acquisition cost, better interpretability of the decision process, and lower test time complexity.

CRA problems usually contain categorical variables (e.g., marital status) and these variables are converted into binary features using 1-of-k encoding. When we have grouped features such as these, performing feature selection at the feature level does not produce sparse results because not eliminating a single feature from a group requires to collect data about the corresponding variable for test data points (i.e., new credit applicants in our case). We propose to use the probit classifier with a proper prior structure and MKL with a proper kernel construction procedure to perform feature selection in a group-wise manner. By doing these, we can decide whether we need to include a categorical variable into the final decision function or not.

In Section 2, we give an overview of the related work by considering existing machine learning solutions for CRA. Section 3 introduces the probit classifier and its inference mechanism with a deterministic variational approximation, and explains the details of our proposed extension towards group-wise feature selection. Section 4 introduces MKL and shows how we can perform group-wise feature selection with suitable kernel calculations and sparsity on the kernel-level. In Section 5, we evaluate the performances of our proposed group-wise feature selection methods on two well-known CRA data sets. In Section 6, we summarize our contributions and conclude the paper.

2. Related work

There are two main categories for CRA methods: (a) structural approaches and (b) statistical approaches. Structural approaches directly depend on some financial measures of the credit applicant such as total asses, yearly profit/loss rate, and growth rate. The credit interest rate is decided by looking at these measures (Kotsiantis, Tzelepis, Koumanakos, & Tampakas, 2005). The most obvious problem of such approaches is the lack of proper mathematical or statistical tools. Statistical approaches depend on empirical tools that use the credit history to build a predictor used for new credit applicants. Different machine learning algorithms such as NNs and SVMs are applied to CRA problems. In this study, we are focusing on the second approach by considering two different classification schemes for CRA. We first give a structured review of the recent machine learning studies by considering commonly used methods.

NNs are frequently used for credit risk estimation due to the fact that most of the existing statistical softwares include them as standard computational tools (Atiya, 2001). For example, Angelini, Di Tollo, and Roli (2008) give a successful application scenario for Italian small businesses using two different NN strategies. Abdou, Pointon, and El-Masry (2008) compare NNs with other standard methods for CRA of Egyptian banks and obtain the best results using NNs. Yao, Wu, and Yang (2009) propose a CRA method using fuzzy NNs for Chinese commercial banks. Khashman (2010) compare different learning algorithms for supervised NNs on German credit data set. Derelioğlu and Gürgen (2011) propose a rule extraction system using a NN-based approach for CRA of small medium enterprises in Turkey.

SVMs are also excessively used for CRA applications due to their good empirical performance. Huang, Chen, Hsu, Chen, and Wu (2004) show that SVMs slightly outperform NNs for CRA on two data sets from Taiwanese financial institutes and USA commercial banks. Chen and Shih (2006) also report a very similar result on a Taiwan banking data set. Yongqiao, Shouyang, and Lai (2005) propose a new fuzzy SVM algorithm for classifying credit applicants. Van Gestel et al. (2006) develop a Bayesian least squares SVM classifier and test the classifier on a commercial credit data set based on Belgian and Dutch firms. Li, Chen, and Xu (2006) formulates a least squares SVM classifier for joint classification and feature selection to provide interpretability using MKL framework. Huang et al. (2007) show that SVMs outperform NNs and DTs on two standard credit risk data sets and propose a hybrid method that performs feature selection for SVMs using genetic algorithms (GAs). Yoon and Kwon (2010) propose a CRA method built on credit card sales information using SVMs to solve the missing financial data problem. Kim and Sohn (2010) build a credit risk estimation method for Korean small-and-medium enterprises using SVMs.

Inspired from SVMs, multiple criteria programing framework is proposed for classification and applied to CRA problems (Shi, 2010). Peng, Kou, Shi, and Chen (2008) introduce a multiple criteria convex quadratic programing model that tries to maximize the intra-class distance between classes and to minimize the within-class distance, and test the proposed algorithm on four different CRA data sets. Li, Wei, Li, and Xu (2011) extend the same idea with a combination of GAs and MKL towards coupled classification and feature selection.

No single classification algorithm can produce the best results for all classification problems. There are two standard approaches to solve this issue: (a) classifier selection and (b) classifier combination. In classifier selection, different classifiers are trained and evaluated using a cross-validation approach. At the end, the best performing classifier is used for testing the system. Instead of relying on a single classifier, we can construct a meta-classifier that combines the predictions of multiple classifiers, known as classifier combination or classifier ensemble. This combination strategy is also applied to CRA extensively. Lai, Yu, Wang, and Zhou (2006a) propose to use an NN ensemble by training a diverse set of networks and combining uncorrelated ones to obtain a reliable prediction scheme. Lai, Yu, Wang, and Zhou (2006b) also formulate an NN ensemble method by training different NNs on different subsets of the training data (known as bagging) and combining the predictions of these networks with another NN to get the final prediction. Huang, Hung, and Jiau (2006) examine various classification algorithms and construct classifier ensembles using random committee and voted perceptron techniques to evaluate the customers of a Chinese bank. Hsieh and Hung (2010) present a CRA method that uses NNs, SVMs, and Bayesian networks as the base classifiers of the ensemble. Zhou, Lai, and Yu (2010) offer an ensemble method that uses least squares SVMs with different kernels for the combination. Twala (2010) compares different ensemble strategies and shows that using a classifier ensemble outperforms single classifiers on four different CRA data sets. Peng, Wang, Kou, and Shi (2011) propose three different methods to compare and to combine base classifiers using their predictions as inputs.

In addition to such machine learning methods, there are also evolutionary algorithms such as GAs proposed for CRA. Finlay (2009) applies a GA approach to optimize business measures instead of a statistical model objective as in machine learning algorithms. Min and Jeong (2009) propose a binary classifier using a GA-based formulation and obtain comparable results to statistical approaches.