Can sentiment analysis help mimic decision-making process of loan granting? 
A novel credit risk evaluation approach using GMKL model

Dailing Zhang, Wei Xu, Yingqiu Zhu, Xinwei Zhang 
School of Information  
Renmin University of China 
Beijing, 100872, China 
eileenlan@163.com, weixu@ruc.edu.cn, rozen0maiden@126.com, xinweify.zhang@gmail.com

Abstract
Credit risk assessment is a crucial process for financial institutions when granting commercial loans. However, the manual analysis of the overall condition of firms through customer due diligence reports is costly for both time and labor. This paper proposes a novel credit risk evaluation approach using GMKL model to automate the decision-making process. Sentiment indexes are generated by mining the opinions of the text content in customer due diligence reports and further used as input for model construction. The method distinguishes itself by innovatively employing sentiment analysis in credit risk assessment. A real-life loans granting dataset is utilized for verifying the performance of the method. The experiment results show that, when combining the traditional financial indicators along with the sentiment indexes, the classifiers trained by GMKL model can outperform several baseline models, successfully improving the accuracy of classification and also detecting the default loans.

1. Introduction
Business lending is a significant part of the commercial banking. When receiving loan requests, the loan officers usually assess the credit risk of the firm in order to control the quality of the loans and also to maximize the expected profit of the bank. Therefore, a good credit risk evaluation method becomes crucial to these financial institutions while credit scoring is proved to be the primary method to develop tools for credit risk assessment.

At the inception, the assessment of credit risk depended on subjective judgments according to the 5Cs rules referred to as character, capacity, capital, collateral and condition. However, the booming of the credit industry made it impossible to assess thousands of applicants completely manually but to automate the process. Hence various credit scoring models have emerged to help financial institutions enforcing efficient credit approval. Among them, statistical models are the most widely used tools including the linear discriminant analysis (LDA) model, logit model and probit model. Myers and Forgy [1] used discriminant analysis and multiple regression analysis to compare scorecard built forecasting credit risk in a financial company. Wiginton [2] was one of the first scholars to apply logistic regression method to credit scoring while West [3] employed the logit estimation along with the factor analysis to generate early warning systems for commercial banks. Probit analysis is another regression model to classify credit applicants suggested by Grablowsky and Talley [4]. In the statistical models, financial ratios were generally used as the input features to predict the default risk. Meanwhile, the operational research (OR) models also enjoy great popularity in credit scoring such as linear programming (LP) and integer programming. Hardy and Adrian [5] found the linear programming classifiers could perform as well as statistic methods while Kolesar and Showers [6] developed AT&T scorecard applying the integer programming.

In recent decades, several sophisticated models like the non-parametric statistical methods and artificial intelligence (AI) approaches emerged and they have been proved to be alternative tools for credit scoring, such as decision tree (DT) [7], genetic algorithms (GA) [8], neural networks (NNs) [9-11] and support vector machine (SVM) [12]. In general, the AI models are found to be superior to the traditional statistical methods when dealing with nonlinear patterns in credit scoring. Desai et al. [8] investigated the predictive power of neural networks and genetic algorithms in comparison to traditional techniques concluding that
NNS performed better when classifying poor loan. West [11] compared five neural networks models with traditional methods and found the neural models could achieve fractional improvement in credit scoring accuracy. Huang et al. [13] demonstrated the competitiveness of SVM-based model in terms of classification accuracy comparing to the BPNN, genetic programming and decision tree.

Furthermore, several hybrid models have been proposed and applied to credit scoring. For example, Lee and Chen [14] raised a hybrid NN-MARS model whose results outperformed others from LDA, LR, NN, and MARS. More recently, Han et al. [15] hybridized the artificial intelligence with the logistic regression, and found the new method helpful to improve credit scoring. Besides, Zhou et al. [16] employed a hybrid SVN-KNN model and proved it a promising method for credit risk evaluation. More comprehensive introductions of the credit scoring methods can be accessed referring to these surveys [17-18].

As a kind of algorithm in machine learning, the multiple kernel learning (MKL), first introduced by Lanckriet et al. [19], which is aiming at finding optimal linear representation of base kernel functions to maximize the capability of generalization, achieves a successful performance and often outperforms SVM with single kernel function. Among the MKL methods, $L_1$ norm constraint is applied and has been proved good performance. $L_p$ norm ($p>1$) is regarded as an alternative way to avoid the suboptimal problem in $L_1$-MKL but at the same time it discards its sparseness of solution [20]. Generalized multiple kernel learning (GMKL) method adopting a combination constraint was then proposed to avoid the problem in $L_1$-MKL and $L_p$-MKL [21]. The generalized multiple kernel learning has been successfully applied in several fields like fault diagnosis of local area network [22], network intrusion detection [23], but hasn’t been used in financial field.

Except the improvement of the quantitative models and algorithms, subjective information is also considered as complementary material to improve performance of financial forecasting. Thanks to the rapid development of text mining techniques, the sentiment hidden in the booming web news media now can be utilized for financial forecasting. Tetlock [24] explored the relationship between media content and stock market activity while the results revealed that high media pessimism value brought owntrend pressure on market prices. Similarly, Das and Chen [25] extracted sentiment of small investors from stock message board and proved that stock value was impacted by the investors’ sentiment and board activity.

However, partially due to the difficulty in obtaining the real-life financial risk dataset and limited access to domain expert knowledge, most researches just focus on algorithm development or improvement based on the standard financial databases but seldom consider the feasibility of the model as well as the need of the end users. The gap between the industry and academia impairs the value of data mining in financial risk detection [26].

Recent years have witnessed a huge growth in Chinese business loan. In the banks of China, the process of granting loans to corporations can be portrayed as follows. When applying for loan, the corporation first submits its recent financial reports and specifies the amount and term. The staff of the bank then conduct a thorough research and complete a document called the due diligence report describing the details of the certain corporation. Finally, the financial reports along with the due diligence report are delivered to the leadership. The decision makers, usually a group of loan managers, will scan the due diligence report along with the financial reports and make the final decision based on their own personal knowledge and experience.

According to the scenario above, it is obvious that the customer due diligence report is important for decision making since it contains overall details of a firm, like the firm’s age, level of employment, capital structure, profitability, solvency, cash flow gap and so on. However, the content in the report exists as text format and hard to be quantified while the manual analysis is a time consuming and labor intensive process. So far there are few researchers or institutions focus on the sentiment analysis of the due diligence report and automate the decision-making process. As far as we know, it maybe the first attempt to incorporate sentiment analysis in building credit scoring models.

In this paper, we propose a novel credit risk evaluation approach to mimic the decision-making process of the leadership. Besides the traditional financial conditions of the firms, we also adopt the sentiment indexes through opinion mining as the input features to construct model, which is generated from the customer due diligence report. To better meet the real need of industrial users, we consider the characteristics of the loan applicants in China and further apply a real-life loan dataset obtained from a commercial bank in China to verify the performance of the model. When building the classifiers, the GMKL approach is employed.

The remainder of this paper is organized as follow. Section 2 describes the GMKL approach used in this study, followed by Section 3 that introduces the details of our proposed method involving the framework and
In Section 4, empirical results are presented and the performance of our model is evaluated comparing to several baseline models. The conclusion and future work are summarized in the last Section 5.

2. Theoretical foundation

In this section, the basic concepts of multiple kernel learning and Generalized Multiple Kernel Learning (GMKL), proposed by Yang et al in [21], are briefly introduced.

GMKL is an extension of multiple kernel learning method. Multiple kernel learning (MKL) methods reveal the unseen similarity relationship of data by learning kernel combination weights under a specific constraint. \(L_1\)-norm constraint is adopted in most of the \(L_1\)-MKL where sparse but non-smooth solution is generated. The sparse solution brings the convenience for identifying the kernel combination format but discards useful information when two kernel functions are correlated or orthogonal. Caused by the information lose, the solutions of the \(L_1\)-MKL often turn to be suboptimal other than optimal. \(L_p\)-norm constraints \(p>1\) are considered as the alternative way and imposed into the MKL. Although all kernel information is retained, these extensions generate non-sparse solution which processes a weaker capability of noise tolerance.

GMKL is developed as an extension of the multiple kernel learning to avoid the weaknesses of both \(L_1\)-MKL and \(L_p\)-MKL, combing the \(L1\)-norm constraint and \(L2\)-norm constraint. This model adapts a linear combination constraint of \(L_1\)-norm and \(L_2\)-norm, which keep the correlation and orthogonal kernel information and produce a favorable sparse solution.

The objective of GMKL is to generate a decision function \(f\) based on the given training dataset \((x_i, y_i)_{i=1}^N\) where feature \(x_i \in \mathbb{R}^n\) and label \(y_i \in \{+1, -1\}\). In the GMKL, \(K\) kernel functions, each of which establishes a mapping from input data \(X\) space to feature space \(\mathcal{H}\), are given. The decision function \(f\) with \(Q\)-dimensional \(w\) weight vector and bias value \(b\) is represented as

\[
f_{w,b}(x) = w^T \phi_b(x) + b \quad (1)
\]

where \(\phi_b = \sqrt{\alpha_1} \phi_1 \times \cdots \times \sqrt{\alpha_K} \phi_K : X \rightarrow \mathcal{H}\) denotes the combination of the \(K\) feature mappings and \(\phi_i : X \rightarrow \mathcal{H}(i = 1, \cdots, K)\) represents the feature mapping defined by the kernel function. The optimal combination weight \(w\) and \(b\) are obtained by solving the following optimization problem:

\[
\begin{align*}
\min_{\alpha,b} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^N R(f_{\alpha,b}(x_i), y_i) \\
\text{s.t.} & \quad \Phi(\theta) \leq 1
\end{align*}
\]

where \(\Phi(\theta)\) defines a constraint of \(\theta\). \(R\) is a convex function used to ensure the optimization problem is convex. In GMKL model, the \(2\) is represents as

\[
\begin{align*}
\min_{\alpha,b} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^N R(f_{\alpha,b}(x_i), y_i) \\
\text{s.t.} & \quad \nu \|\theta\| + (1 - \nu) \|\theta\|_2 \leq 1 \\
& \quad 0 \leq \nu \leq 1
\end{align*}
\]

where \(\nu\) is the parameter which balances the \(L_1\)-norm and \(L_2\)-norm constraints. \(C\) is the constraint parameter which balances the empirical risk and regularization performance.

The dual form of \(3\) is:

\[
\begin{align*}
\min_{\alpha,b} & \quad \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \left( \sum_{k=1}^K \alpha_k K_{ij} \right) - \sum_{i=1}^N \alpha_i \\
\text{s.t.} & \quad \nu \|\theta\| + (1 - \nu) \|\theta\|_2 \leq 1 \\
& \quad \sum_{i=1}^N \alpha_i y_i = 0, \\
& \quad \alpha_i \leq C \quad (i = 1, \cdots, N)
\end{align*}
\]

3. The proposed method

In this section, a novel credit risk evaluation approach using GMKL model is proposed and the process is illustrated in details. Besides the conditional financial indicators, we simultaneously conduct opinion mining in the customer due diligence report and generate sentiment indexes. The framework of the proposed method is illustrated in Figure 1.

3.1. Research framework

There are six steps of the proposed method in the research framework shown as below.

Step 1: Data cleaning is applied to the original dataset to eliminate duplicated and invalid records.

Step 2: The dataset is then separated into two parts. One part contains the numeric data of the applicants, like the data from financial reports or data related to loan request. Another part refers to the text data of the customer due diligence reports.

Step 3: Different types of financial ratios are calculated according to the financial reports. Combining the case specific variables, we get the financial index.
Step 4: We employ an opinion mining approach to deal with the text data of the customer due diligence reports and generate sentiment indexes.

Step 5: The financial ratios variables, case specific variables along with the sentiment variables are utilized as the features input. The binary classes are used as output. Now, a modified dataset is ready for further use.

Step 6: The modified dataset is then divided into training set and testing set. Training set is put into the GMKL model to train classifiers after which the testing set is used for simulation to verify the performance of the proposed method.

Financial ratios are crucial indicators for financial statement analysis [27]. They are objective indicators that reflect the strengths and weaknesses of a company. Ratios analysis is widely used by financial institutions like the rating agencies and banks to assess the default possibility and further make rational decisions. The financial ratios can be categorized into four major types as liquidity ratios, profitability ratios, efficiency ratios and leverage ratios, each of which contains several financial ratios. Liquidity ratios involve quick ratio (F₁) and current ratio (F₂). Profitability ratios include net profit margin (F₃), return on equity (F₄) and return on assets (F₅). Efficiency ratios can be represented by the accounts receivable turnover (F₆). Finally, leverage ratios contain debt ratio (F₇), debt to tangible assets ratio (F₈) and equity ratio (F₉). The formulas of these financial ratios are listed and elaborated in the Appendix.

Case specific variables consider the scale and time span of the application, including the proposed amount (C₁) and proposed term (C₂) of loans.

Sentiment variables are novel variables introduced in our model to catch the subjective opinion and sentiment of the customer due diligence report. We use three variables as positive sentiment index (S₁), negative sentiment index (S₂) and overall sentiment index (S₃) to be the novel features for model building. The approach to generate these indicators will be illustrated in the next part.

Table 1 presents the variables and their according descriptions.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial ratio variables</td>
<td>Current ratio (C₂)</td>
</tr>
<tr>
<td>F₁</td>
<td>Quick ratio</td>
</tr>
<tr>
<td>F₂</td>
<td>Current ratio</td>
</tr>
<tr>
<td>F₃</td>
<td>Net profit margin</td>
</tr>
<tr>
<td>F₄</td>
<td>Return on equity (ROE)</td>
</tr>
<tr>
<td>F₅</td>
<td>Return on assets (ROA)</td>
</tr>
<tr>
<td>F₆</td>
<td>Accounts receivable turnover</td>
</tr>
<tr>
<td>F₇</td>
<td>Debt ratio</td>
</tr>
<tr>
<td>F₈</td>
<td>Debt to tangible assets ratio</td>
</tr>
<tr>
<td>F₉</td>
<td>Equity ratio</td>
</tr>
<tr>
<td>Case specific variables</td>
<td>Proposed amount (C₁)</td>
</tr>
<tr>
<td>Sentiment variables</td>
<td>Proposed term (C₂)</td>
</tr>
<tr>
<td>Positive sentiment index (S₁)</td>
<td></td>
</tr>
<tr>
<td>Negative sentiment index (S₂)</td>
<td></td>
</tr>
<tr>
<td>Overall sentiment index (S₃)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Framework of the proposed method

3.2. Input feature extraction

Referring to the characteristic of loan applicants in China, we utilize three categories of variables as input features in our proposed model: financial ratio variables, case specific variables and sentiment variables.

Table 1. Description of financial ratio, case specific and sentiment variables
3.3. Sentiment index development

As mentioned above, the decision makers assess the loan request depending on both financial reports and customer due diligence reports requiring subjective judgments. Comment text in the customer due diligence report of a particular firm is a combination of evaluator’s opinion, altitude and evaluation towards the firm, so it may reflect the status of operation for a firm that financial reports do not cover and may be helpful for risk assessment. Therefore, sentiment analysis is carried out on the comment text to collect the subjective information, in the due diligence reports, as a complementary material in risk assessment.

In order to extract the sentiment of the linguistic terms in the diligence reports, an opinion mining method is proposed. The following two parts will respectively introduce the preliminary work and text analysis work of our opinion mining approach.

The preliminary work of the opinion mining is expanding general dictionary. Sentiment analysis of comment in Chinese started with text segmentation. The dictionary-based method through which words are extracted and labeled simultaneously in a text is adopted to conduct the segmentation work. The original dictionary is made of common words and their corresponding labels. It is insufficient in solving the credit evaluation problem for its imprecision when segmenting the customer due diligence reports. So in order to solve the problem better, two preparative measures are taken to expand the dictionary:

First, the original dictionary is expanded with a financial lexicon to get more accurate segmentation results. The financial lexicon, used in our passage, whose content includes the words and phrases using in financial industry, is referred to some finance lexicons like Reutter’s lexicon, Goldman Sachs’ financial lexicon. After updating, the dictionary gains the ability to precisely and exactly identify the words which describe the running situation of the corporation.

Second, the sentiment part is added through the label in the dictionary. The sentiment label of every individual word is made into the dictionary, which is determined according to several financial experts’ suggestions. The orientations of the sentiment labels range from positive, neutral to negative.

After the preliminary work on dictionary, analysis work on text is carried out. Based on the dictionary with financial lexicon and sentiment labels, opinion mining is the carried out on the customer due diligence reports. The analysis work can be divided into three steps.

Step 1: Find all features of a corporation and definitions about each feature.

With the concept that feature serves as core of an opinion, it is a hitting-point in analyzing the comment text. Features of a corporation mean the attributes of a corporation which may influence the credit evaluation. With the dictionary we established in the preliminary work, the features of corporations are extracted after the segmentation based on the financial lexicon.

For each feature, there exists a sentence that contains its definition. Definitions mean the words or phrases describing the features. A definition may occur as a group of verbs, adjectives and adverbs. Verbs and adjectives are considered as orientations for sentiment identification, while adverbs are treated as sentiment level orientations. The definition for a particular feature is extracted after the segmentation as well. The definition can’t be ignored for its significance in sentiment expression.

Step 2: Identify the positive, neutral, and negative opinions for each feature group.

The whole text is divided into several groups of words according to features, each of which contains a set of definition words centered on a particular feature. The opinion for the whole group is identified by the integrated process on the sentiment labels of both the feature and definition words. The rule of the integrated process on sentiment labels are shown in the Table 2. Following the rule, the group of words would be interpreted into a particular sentiment opinion.

For example, after segmentation and filtration, the sentence “The sales volume of the corporation has increased greatly during this quarter” would turn to be a group of words, [corporation, sales volume, increase, greatly, quarter], centered on the feature [sales volume]. More precisely, the group is made of two parts: feature word — sales volume and definition words — increase and greatly, illustrated as an example in the Figure 2. As every word owns its sentiment label, the group of words owns its corresponding group of labels as well: [neutral feature (from sales volume), positive definition (from increase), and neutral definition (from greatly)]. In this case, the sentiment opinion of this sentence centered on sales volume would be identified as a positive opinion, according to the rule shown in Table 2.

The rule defines how to identify the sentiment opinion of a simple sentence. If the form of a sentence is complex, this sentence will be divided into several simple parts each of which only is centered on one feature. If a sentence contains negative words, like “not”, the sentiment opinion of the sentence will be reversed.

Step 3: Summarize the sentiment index of the whole text.

The collection of all opinions provides access to calculating the sentiment index of the comment text.
Positive index, negative index and overall index are used as the results of the sentiment analysis. A schematic graph of the sentiment analysis for a comment text is shown in Figure 2.

<table>
<thead>
<tr>
<th>Opinion</th>
<th>Definition label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>positive</td>
</tr>
<tr>
<td>Feature label</td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td>positive</td>
</tr>
<tr>
<td>negative</td>
<td>negative</td>
</tr>
<tr>
<td>neutral</td>
<td>positive</td>
</tr>
</tbody>
</table>

Table 2. Rule of the integrated process

3.4. Model construction

In this section, the detailed construction process of the credit risk evaluation model using GMKL is introduced. As the framework shown in Section 3.1, at the inception the original dataset involves the financial reports, customer due diligence reports, and the real final decisions of whether grant the loans or not. Each application is treated as one record. After data cleaning, the financial ratios listed in Table 1 are calculated based on the data from financial statement sheets for each firm. Then, opinion mining approach is carried on the diligence reports to extract the sentiment indexes. All the sentiment indexes, financial ratios along with the case specific variables are used as the features in the model. The decisions classes are binary as label ‘1’ stands for approval and label ‘-1’ for rejection.

Subsequently, the modified dataset is divided into training and testing dataset respectively. The GMKL is used to solve the binary classification problem. The features along with their according classes in training set are put into the GMKL model for learning. After training, the classifiers are further used to simulate the decision making process by applying the testing set. After the simulation, the classifiers generate a predictive label for each record in the testing set. The difference between the simulated decisions and the real decision is evaluated to judge the model performance. To assess the performance more properly and make full use of the limited dataset, k-fold cross-validation test is adopted.

3.5. Performance measurement

Accuracy and error rates are crucial indicators to measure the performance of classification algorithm in financial risk forecasting [24]. This work adopts overall accuracy, precision, true positive rate, true negative rate and tradition F-measure as measurements.

In the classification cases, the classes of samples are binary as good or bad, while the good means creditworthy and the bad is opposite. Accuracy is the percentage of the samples correctly classified, measuring the overall quality of the results of classification. Precision is the proportion of the good samples accurately classified as positive against all the positive results. The true positive rate (TP rate), also known as sensitivity measure or recall rate, detect how well a classifier can recognize the abnormal samples. A higher TP rate means less potential credit losses since the bad samples are rejected initially. True negative rate (TN rate), also called specificity measure, measures the performance of classifier recognizing the normal sample. The traditional F-measure is a
synthesized indicator as it combines precision and recall TP rate thus a higher value means better performance.

Let the number of the creditworthy samples correctly classified as good to be TP, and misclassified as bad to be FP. Similarly, let the number of default samples classified as bad to be TN and misclassified ones to be FN. The formulas of the indicators are as follow.

\[
\text{overall accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}
\]

\[
\text{precision} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{TP rate} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{TN rate} = \frac{TN}{TN + FN} \tag{7}
\]

\[
F = 2 \cdot \frac{\text{precision} \cdot \text{TP rate}}{\text{precision} + \text{TP rate}} \tag{8}
\]

4. Empirical analysis

4.1. Data description

The loan applicant dataset used in this study is obtained from a commercial bank in China. In the original dataset, there are 2132 records in total of the loan requests, each of which contains the content about a firm’s recent financial reports, customer due diligence report conducted by the staff of bank as well as the final decision result label, namely rejected or approved. After data cleaning, 513 records are eliminated and 1619 records remains. 775 of the 1619 valid applicant records are approved and the rest 844 of them are rejected. The proportion between the approvals and rejections are nearly 0.92:1.

Then the financial data and text data are separated for further processing. Financial ratios are calculated according to the financial reports by applying the formulas (shown in Appendix). Nine ratios are finally obtained covering the liquidity, profitability, efficiency and leverage of the firms, which has been mentioned in Section 3.2.

After the text data of records are processed through the opinion mining approach, the positive sentiment index and the negative sentiment index are generated along with the overall sentiment index which is the linear combination of the positive and negative sentiments.

Including the financial ratio variables, the sentiment variables and the case specific variables, 14 attributes are used as input features for our credit model.

4.2. Results of measurements

The segmentation based on the dictionary is implemented using JCSEG. We first expand the original dictionary in JCSEG with financial lexicon and sentiment labels. The positive sentiment index, negative sentiment index and overall sentiment index for each due diligence report of a record are calculated after the segmentation and sentiment analysis.

Yang’s GMKL toolbox is used to solve the GMKL’s problem. The solver implemented in the GMKL adopts MOSEK version 6. The v is set as 0.5 for simplicity. Comparison experiments are carried out among different models including artificial intelligence approaches (BPNN and SVM) and statistical model (Probit). BPNN is implemented using Matlab Neural Network toolbox and the number of neural is set as 10. LIBSVM is applied to solve the SVM classifying problem. The GRBF kernel method is adopted. The sigma and the cost are set as 1 and 18 respectively. The Probit model is implemented in the Weka 3.6. For all classifiers construction, 10-fold cross-validation is used to test the classification performance.

The overall accuracy, true positive rate, true negative rate, precision and F-measure for the 10-fold cross-validation are shown in Table 3.

<table>
<thead>
<tr>
<th>(units in %)</th>
<th>GMKL</th>
<th>BPNN</th>
<th>SVM</th>
<th>Probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS-Original (F₁-F₉, C₁-C₂)</td>
<td>Accuracy</td>
<td>60.78</td>
<td>57.39</td>
<td>67.76</td>
</tr>
<tr>
<td>Precision</td>
<td>61.36</td>
<td>55.60</td>
<td>59.77</td>
<td>59.50</td>
</tr>
<tr>
<td>TP rate</td>
<td>49.21</td>
<td>53.71</td>
<td>99.63</td>
<td>56.05</td>
</tr>
<tr>
<td>TN rate</td>
<td>71.26</td>
<td>60.76</td>
<td>100</td>
<td>63.19</td>
</tr>
<tr>
<td>F-measure</td>
<td>54.52</td>
<td>54.54</td>
<td>74.66</td>
<td>57.54</td>
</tr>
<tr>
<td>S-Mixed (F₁-F₉, C₁-C₂, S₁-S₃)</td>
<td>Accuracy</td>
<td>77.58</td>
<td>71.71</td>
<td>64.92</td>
</tr>
<tr>
<td>Precision</td>
<td>76.55</td>
<td>72.96</td>
<td>57.68</td>
<td>72.26</td>
</tr>
<tr>
<td>TP rate</td>
<td>76.58</td>
<td>64.88</td>
<td>100</td>
<td>69.53</td>
</tr>
<tr>
<td>TN rate</td>
<td>78.50</td>
<td>77.99</td>
<td>32.60</td>
<td>82.71</td>
</tr>
<tr>
<td>F-measure</td>
<td>76.45</td>
<td>68.62</td>
<td>73.11</td>
<td>74.42</td>
</tr>
</tbody>
</table>

As can be seen from Table 3, the dataset with only financial index (involving F₁-F₉, C₁-C₂) is called as

1. The access to the JCSEG: https://code.google.com/p/jcseg/
2. The access to Yang’s GMKL toolbox: http://appsrv.cse.cuhk.edu.hk/~hqyang/doku.php?id=gmkl
3. The access to MOSEK: http://www.mosek.com/
4. The access to LIBSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
NS-Original dataset for convenience. And the other dataset combining financial index with sentiment index (involving F₁-F₉, C₁-C₂, S₁-S₃) is referred as S-Mixed dataset. The consequences of the four classifiers models using NS-Original dataset are shown in the row ‘NS-Original’ and the consequences of using S-Mixed dataset are shown in the row ‘S-Mixed’.

To compare the performance of adding sentiment indexes more directly, the box plots of the F-measure and the accuracy for the 10-fold cross-validation are also shown in Figure 3 and Figure 4.

![Figure 3. Box plots of accuracy](image)

![Figure 4. Box plots of F-measure](image)

To research the effect of adding sentiment indexes, we first take a look at the performance of the model when the NS-Original dataset applied. When no sentiment indexes included, the accuracies of the four models are 60.78%, 57.39%, 67.76% and 56.68%, the mean value of which is 60.65%. The accuracy results show that the SVM get the highest accuracy and GMKL ranks second. In addition, the highest average F-measure for 10-folds cross-validation test of all the four classification models is that of SVM, 74.66%. The F-measures of GMKL, SVM and Probit are 54.52%, 54.54% and 57.54% respectively, all lower than 60%. Besides, the average F-measure value of the four models is 60.32%.

After incorporating the sentiment indexes into the input features, the classification performances of the four methods except SVM model get improved. The average accuracy values of GMKL, BPNN and Probit are 77.58%, 71.71% and 76.40% respectively, which much higher than the ones of NS-Original dataset. From the F-measure aspect, the performance of GMKL, SVM and Probit are also improved and all exceed 70%. Even the lowest F-measure performance, performed by BPNN, can also reach 68.62%. The average F-measure of the four models is improved by approximately 12.84%. It shows the robustness of the performance improvement when adding sentiment indexes.

When using S-Mixed dataset, whose features including sentiment indexes, the average F-measure of the GMKL, 76.45%, beats the other three classifiers. Both of the precision and accuracy of the GMKL rank the first among the four classification models. With the sentiment indexes, the Probit model also obtains a good performance but its robustness is lower than the GMKL. It can be seen from the box plot of accuracy that several outliers exist in the result of Probit model. Besides, the box plot also shows the distribution of classification accuracy for 10-folds cross-validation. It demonstrates that, among the four classification models, GMKL is the most robust and effective classifier when incorporating the sentiment indexes.

The false negative rate (FN rate) is also a significant indicator for measuring the performance of credit scoring. It denotes the capability of the model whether it can avoid fraud risk. The term FN represents the percentage of the high risk claims where bad loans are failed to be rejected and the applicants are misclassified as acceptable ones. The FN of SVM is 0.37% (100%-TP rate) when using NS-Original dataset, which minimizes the fraud risk to a very low level. However, when using the S-Mixed dataset, the average FN rate of SVM decreases down to 0% amazingly. It means the adding of the sentiment indexes could help avoid the fraud risk totally. Besides, the FN rate of GMKL is reduced by more than half from 50.79% to 23.42%, which demonstrates that the sentiment indexes play a great influence on reducing fraud risk.

5. Conclusions and future work

This paper presents a novel credit risk evaluation approach using GMKL model. The innovation of this proposed approach is that it applies sentiment analysis into the credit risk assessment. Compared to the traditional credit assessment method, the sentiment
analysis result of the due diligence report is considered as the supplementary reference. GMKL model is used to automate the decision-making process using the sentiment index and financial indicators as the input feature.

The empirical experiment is carried out on a loans granting dataset from real life where 1619 applicants are valid instances. The results show that, after incorporating the sentiment indexes, the classification performance successfully gets improved. Besides, the improvement on the true negative rate demonstrates the sentiment analysis result of the due diligence report enhance the capability of avoiding fraud risk. GMKL is proved to be the most suitable classification model when combing the sentiment indexes and the financial indicators.

Future research will be carried out in several aspects. First, the topic modeling and opinion mining method can be considered to be improved the forecasting performance by extracting smart topic and sentimental indicators more precisely. Second, the feature selection will be offered to eliminate redundant and irrelevant features of financial indicators. Finally, a decision support system for loan granting will be developed to support the decision-making process.

6. Appendix

The details of the financial ratios used in the proposed method.

1. Liquidity ratios
   a. Current ratio (F_1)
   \[ \text{Current ratio} = \frac{\text{Current assets}}{\text{Current liabilities}} \] (9)
   b. Quick ratio (F_2)
   \[ \text{Quick ratio} = \frac{\text{Current assets} - \text{Inventories}}{\text{Current liabilities}} \] (10)

2. Profitability ratios
   a. Net profit margin (F_3)
   \[ \text{Net profit margin} = \frac{\text{Net income}}{\text{Revenue}} \] (11)
   b. Return on equity (ROE) (F_4)
   \[ \text{ROE} = \frac{\text{Net income}}{\text{Shareholder's equity}} \] (12)
   c. Return on assets (ROA) (F_5)
   \[ \text{ROA} = \frac{\text{Net income}}{\text{Average total assets}} \] (13)

3. Efficiency ratios
   a. Accounts receivable turnover (F_6)
   \[ \text{Accounts receivable turnover} = \frac{\text{Net credit sales}}{\text{Average accounts receivable}} \] (14)

4. Leverage ratios
   a. Debt ratio (F_7)
   \[ \text{Debt ratio} = \frac{\text{Total debt}}{\text{Total assets}} \] (15)
   b. Debt to tangible assets ratio (F_8)
   \[ \text{Debt to tangible assets ratio} = \frac{\text{Total debt}}{\text{Total tangible assets}} \] (16)
   c. Equity ratio (F_9)
   \[ \text{Equity ratio} = \frac{\text{Total Shareholders's Equity}}{\text{Total assets}} \] (17)

Acknowledgments

This work was supported in part by 973 Project (Grant No. 2012CB316205), National Natural Science Foundation of China (No. 71001103, 91224008, 91324015), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075), Beijing Natural Science Foundation (No. 9122013), Beijing Nova Program (No. Z13110100413058), Program for Excellent Talents in Beijing.

References


