IJBT (2018) 02:44-63
RESEARCH ARTICLE

Sustainable co-training of mixture-of-experts for credit scoring of borrowers in social lending

Jae-Min Yu [1] · **Sung-Bae Cho**[1]

Abstract Of late, social lending has been so popular that several services for it are provided based on credit scoring with a variety of personal aspects that affect an individual's credit. The factors that are not considered in traditional banks might be more influential than the conventional scoring. Loan requisition can be registered continuously, anytime, by whoever wants loan through social lending. At the same time, large-scale unlabeled data are increasing. Labeling data is expensive. In this paper, with focus on these characteristics, we present a global-local co-training algorithm for mixture-of-experts to exploit the unlabeled data for accurate credit scoring. We conducted experiments with dataset from the Lending Club to evaluate the accuracy based on the reliability of unlabeled dataset. To show the usefulness of the proposed method, we compared the performance with other machine learning methods such as Naive Bayes, logistic regression, decision tree and SVM, and analyzed the confusion matrix. A series of repetitive experiments revealed the quantitative superiority in various characteristics.

Keywords: Social Lending · Credit Scoring · Co-training · Mixture-of-experts

1 Introduction

Social lending, also called as peer-to-peer lending, is proliferating. It is a platform enabling individuals to borrow or lend, without a bank as the financial intermediary. Recently, the credit scoring model has been extensively studied for the evaluation of loan admission with the rapidly growing financial industry [1]. It is basis on which social lending is developed with the number of new users accelerating over time to satisfy the increasing worldwide demand for financial platforms and personal loans. However, the investment of lenders in the loan is not protected generally from governments. In terms of principal guarantee, social lending differs from bank deposit.

The core problem is whether social lending can approve lending to borrowers or not [2]. This problem can be considered as binary classification (good or bad) from the point of machine learning. Most of the loan datasets consist of biased instances [3]. This can be explained by poverty of default cases in the midst of plenty of normal customer cases. Biased dataset belonging to one of the classes can be regarded as

✉ Jae-Min Yu:
yjam@yonsei.ac.kr

Sung-Bae Cho:
sbcho@yonsei.ac.kr

[1] Department of Computer Science, Yonsei University, Seoul, South Korea

imbalanced. Classification problem of imbalanced dataset occurs in both credit scoring and intrusion detection.

Another problem in social lending is the lack of systematic means/methods to check if repayments from borrowers are made duly on time. We consider this kind of data as unlabeled dataset (Loan requisition can be registered continuously anytime by whoever needs loan in social lending. Simultaneously, large-scale unlabeled data is increasing). Labeling all of this data costs high. Most of the real social lending data is unlabeled due to non-repayment.

Table 1 shows the statistics of the Lending Club data in terms of different classes. Except for fully paid and charged off, the rest of the classes mean in progress. That is, fully paid class means that borrowers repay principal in due date, whereas charged off class means that borrowers did not repay. About 74% of dataset consists of unlabeled data including late within 31-120 days, late within 16-30 days, loan in grace period and loan in current. This can support the approach of semi-supervised learning that uses both labeled and unlabeled datasets to enhance the performance of loan default classification.

Table 1 The statistics of the lending club data (2013-2014)

Classes	**No. of loans**	**Percent of loan**	**Amount ($)**	**Percent of amount**
Charged off	13,350	5.67%	200,700,075	5.73%
Late in 31-120 days	4,974	2.11%	76,178,175	2.17%
Late in 16-30 days	783	0.33%	12,539,600	0.35%
In grace period	1,928	0.81%	30,711,275	0.87%
Current	162,337	68.96%	2,452,472,400	70.06%
Fully paid	52,023	22.10%	727,834,325	20.79%

Table 2 Matured loans of lending club data (2013-2014)

Matured loans	**No. of loans**	**Percent of loan**	**Amount ($)**	**Percent of amount**
Charged off	13,350	20.42%	200,700,075	21.61%
Fully paid	52,023	79.57%	727,834,325	78.38%

The social lending data have a problem against the conventional classification method. We can see that the number of normal borrowers is larger than that of defaulted borrowers in Table 2. If the number of data with a class is abnormally high and we use a simple accuracy as metrics, machine learning algorithms tend to classify most of the data as the majority class for reducing the probability of misclassification. In this case, other minor class dataset can be considered as an important class [4]. This problem could cause to reduce the performance of machine learning classifiers. If we apply these situations to social lending, we could not find defaulted borrowers (minority) among normal borrowers (majority). Also, imbalanced dataset has a difficult problem to determine accurate classification boundary.

Several classification problems in the financial area, such as credit scoring in the social lending, lie on the complicated data spaces for growth of the data with unpredictable causal relationship between the instances. Only one model for classification which covers the space of the whole problem may not induce accurate results. For resolving these problems, Jacobs proposed a mixture-of-experts that is based on the idea of divide-and-conquer [5]. The model divides a problem into smaller sub-problems, working well for the problems in smaller and independent information.

Figure 1 shows the correlation matrix in the Lending Club dataset. As can be observed, the correlation with loan status and interest rate has the highest value of r=0.23. Most of the correlation coefficients are considerably small. The values in white boxes range in -0.05~0.05. This confirms that there is not enough relationship between each finance-related attributes. As mentioned, the mixture-of-experts has a potential to deal with this kind of problem.

Pearson's r	loan_amnt	loan_status	term	int_rate	installment	emp_length	home_ownership_MORTGAGE	home_ownership_RENT	home_ownership_OTHER	home_ownership_OWN	annual_inc
loan_amnt	-	0.058	0.424	0.208	0.962	0.130	0.184	-0.170	-0.002	-0.031	0.348
loan_status	0.058	-	0.158	0.230	0.041	-0.006	-0.049	0.046	0.005	0.007	-0.065
term	0.424	0.158	-	0.464	0.204	0.086	0.106	-0.100	-0.002	-0.013	0.081
int_rate	0.208	0.230	0.464	-	0.200	0.015	-0.092	0.087	0.002	0.011	-0.004
installment	0.962	0.041	0.204	0.200	-	0.115	0.153	-0.139	-0.001	-0.027	0.347
emp_length	0.130	-0.006	0.086	0.015	0.115	-	0.193	-0.189	-0.001	-0.011	0.098
home_ownership_MORTGAGE	0.184	-0.049	0.106	-0.092	0.153	0.193	-	-0.853	-0.016	-0.290	0.183
home_ownership_RENT	-0.170	0.046	-0.100	0.087	-0.139	-0.189	-0.853	-	-0.013	-0.251	-0.160
home_ownership_OTHER	-0.002	0.005	-0.002	0.002	-0.001	-0.001	-0.016	-0.013	-	-0.005	-0.003
home_ownership_OWN	-0.031	0.007	-0.013	0.011	-0.027	-0.011	-0.290	-0.251	-0.005	-	-0.047
annual_inc	0.348	-0.065	0.081	-0.004	0.347	0.098	0.183	-0.160	-0.003	-0.047	-
purpose_car	-0.079	-0.014	-0.019	-0.053	-0.082	-0.028	-0.017	0.010	-0.002	0.014	-0.017
purpose_credit_card	-0.023	-0.042	-0.066	-0.096	-0.016	-0.024	-0.023	0.026	0.010	-0.006	-0.017
purpose_debt_consolidation	0.121	0.023	0.074	0.050	0.115	0.045	-0.001	0.014	-0.009	-0.024	-0.025
purpose_home_improvement	-0.009	-0.009	0.010	-0.049	-0.020	0.019	0.144	-0.166	0.009	0.038	0.067
purpose_house	0.030	-0.005	0.009	-0.014	0.026	-0.012	-0.013	0.009	-0.001	0.008	0.015
purpose_major_purchase	-0.073	-0.016	-0.017	-0.054	-0.078	-0.021	-0.012	0.011	-0.002	0.003	0.009
purpose_medical	-0.061	0.004	-0.023	0.021	-0.058	-0.002	-0.009	0.004	-0.001	0.011	0.001
purpose_moving	-0.064	0.003	-0.025	0.032	-0.059	-0.037	-0.052	0.057	-0.001	-0.007	-0.006
purpose_renewable_energy	-0.010	-0.001	-0.010	0.003	-0.008	-0.001	-0.009	0.006	0.000	0.004	0.007
purpose_small_business	0.038	0.035	0.002	0.041	0.046	-0.015	-0.002	0.001	-0.002	0.003	0.036
purpose_vacation	-0.074	0.003	-0.028	0.012	-0.072	-0.005	-0.020	0.016	-0.001	0.007	-0.009
purpose_wedding	-0.033	-0.008	-0.024	0.020	-0.028	-0.031	-0.038	0.040	-0.002	-0.003	0.001
dti	0.032	0.109	0.065	0.145	0.032	0.033	-0.005	-0.005	0.005	0.018	-0.186
delinq_2yrs	0.014	0.012	0.010	0.107	0.026	0.026	0.048	-0.048	-0.001	-0.002	0.072
inq_last_6mths	0.029	0.054	0.063	0.246	0.042	-0.010	0.093	-0.091	-0.004	-0.005	0.077
total_acc	0.235	-0.028	0.115	0.003	0.218	0.129	0.248	-0.236	0.000	-0.028	0.226

Pearson's r	purpose_car	purpose_credit_card	purpose_debt_consolidation	purpose_home_improvement	purpose_house	purpose_major_purchase	purpose_medical	purpose_moving	purpose_renewable_energy	purpose_small_business	purpose_vacation	purpose_wedding	dti	delinq_2yrs	inq_last_6mths	total_acc
loan_amnt	-0.079	-0.023	0.121	-0.009	0.030	-0.073	-0.061	-0.064	-0.010	0.038	-0.074	-0.033	0.032	0.014	0.029	0.235
loan_status	-0.014	-0.042	0.023	-0.009	-0.005	-0.016	0.004	0.003	-0.001	0.035	0.003	-0.008	0.109	0.012	0.054	-0.028
term	-0.019	-0.066	0.074	0.010	0.009	-0.017	-0.023	-0.025	-0.010	0.002	-0.028	-0.024	0.065	0.010	0.063	0.115
int_rate	-0.053	-0.096	0.050	-0.049	-0.014	-0.054	0.021	0.032	0.003	0.041	0.012	0.020	0.145	0.107	0.246	0.003
installment	-0.082	-0.016	0.115	-0.020	0.026	-0.078	-0.058	-0.059	-0.008	0.046	-0.072	-0.028	0.032	0.026	0.042	0.218
emp_length	-0.028	-0.024	0.045	0.019	-0.012	-0.021	-0.002	-0.037	-0.001	-0.015	-0.005	-0.031	0.033	0.026	-0.010	0.129
home_ownership_MORTGAGE	-0.017	-0.023	-0.001	0.144	-0.013	-0.012	-0.009	-0.052	-0.009	-0.002	-0.020	-0.038	-0.005	0.048	0.093	0.248
home_ownership_RENT	0.010	0.026	0.014	-0.166	0.009	0.011	0.004	0.057	0.006	0.001	0.016	0.040	-0.005	-0.048	-0.091	-0.236
home_ownership_OTHER	-0.002	0.010	-0.009	0.009	-0.001	-0.002	-0.001	-0.001	0.000	-0.002	-0.001	-0.002	0.005	-0.001	-0.004	0.000
home_ownership_OWN	0.014	-0.006	-0.024	0.038	0.008	0.003	0.011	-0.007	0.004	0.003	0.007	-0.003	0.018	-0.002	-0.005	-0.028
annual_inc	-0.017	-0.017	-0.025	0.067	0.015	0.009	0.001	-0.006	0.007	0.036	-0.009	0.001	-0.186	0.072	0.077	0.226
purpose_car	-	0.199	-0.130	-0.026	-0.009	-0.016	-0.010	-0.009	-0.003	-0.015	-0.008	-0.010	-0.052	-0.005	0.002	-0.036
purpose_credit_card	0.199	-	-0.654	-0.129	-0.045	-0.079	-0.051	-0.044	-0.016	-0.074	-0.042	-0.052	0.009	-0.022	-0.050	-0.020
purpose_debt_consolidation	-0.130	-0.654	-	-0.287	-0.100	-0.177	-0.115	-0.098	-0.036	-0.165	-0.094	-0.117	0.104	0.001	-0.016	0.056
purpose_home_improvement	-0.026	-0.129	-0.287	-	-0.020	-0.035	-0.023	-0.019	-0.007	-0.032	-0.018	-0.023	-0.090	0.019	0.056	0.024
purpose_house	-0.009	-0.045	-0.100	-0.020	-	-0.012	-0.008	-0.007	-0.003	-0.011	-0.006	-0.008	-0.042	0.003	0.040	-0.003
purpose_major_purchase	-0.016	-0.079	-0.177	-0.035	-0.012	-	-0.014	-0.012	-0.004	-0.020	-0.011	-0.014	-0.065	-0.002	0.020	-0.028
purpose_medical	-0.010	-0.051	-0.115	-0.023	-0.008	-0.014	-	-0.008	-0.003	-0.013	-0.007	-0.009	-0.019	0.012	0.007	-0.011
purpose_moving	-0.009	-0.044	-0.098	-0.019	-0.007	-0.012	-0.008	-	-0.002	-0.011	-0.006	-0.008	-0.021	0.003	0.012	-0.024
purpose_renewable_energy	-0.003	-0.016	-0.036	-0.007	-0.003	-0.004	-0.003	-0.002	-	-0.004	-0.002	-0.003	-0.006	-0.002	0.002	-0.006
purpose_small_business	-0.015	-0.074	-0.165	-0.032	-0.011	-0.020	-0.013	-0.011	-0.004	-	-0.011	-0.013	-0.063	0.008	0.033	-0.026
purpose_vacation	-0.008	-0.042	-0.094	-0.018	-0.006	-0.011	-0.007	-0.006	-0.002	-0.011	-	-0.008	-0.005	-0.006	0.005	-0.016
purpose_wedding	-0.010	-0.052	-0.117	-0.023	-0.008	-0.014	-0.009	-0.008	-0.003	-0.013	-0.008	-	-0.026	0.006	0.008	-0.027
dti	-0.052	0.009	0.104	-0.090	-0.042	-0.065	-0.019	-0.021	-0.006	-0.063	-0.005	-0.026	-	-0.012	0.011	0.231
delinq_2yrs	-0.005	-0.022	0.001	0.019	0.003	-0.002	0.012	0.003	-0.002	0.008	-0.006	0.006	-0.012	-	0.027	0.133
inq_last_6mths	0.002	-0.050	-0.016	0.056	0.040	0.020	0.007	0.012	0.002	0.033	0.005	0.008	0.011	0.027	-	0.155
total_acc	-0.036	-0.020	0.056	0.024	-0.003	-0.028	-0.011	-0.024	-0.006	-0.026	-0.016	-0.027	0.231	0.133	0.155	-

Fig 1. Pearson correlation matrix for the lending club dataset

Figure 2 shows the nonlinear problem in social lending domain in classification. Each color means the status of borrowers: red means good, blue means bad, according to the interest rate, installment and DTI. In this paper, we propose a global-local co-training (GLCT) method for mixture-of-experts, focusing on these characteristics in social lending, which can predict the credit score by using co-training scheme [6].

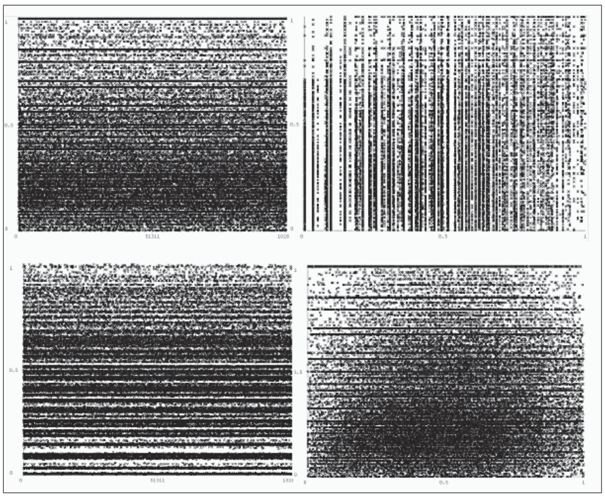


Fig 2. Distribution of the lending club dataset

2 Background

2.1 Learning issue on imbalanced data

In case of imbalanced data, the models such as artificial neural network and decision tree should consider the relative difference on classes because they tend to ignore minor class during classifying major class accurately [7]. Breiman et al. showed that minor class can be classified as major class (misclassification) when classifying imbalanced data [8].

To reduce the total misclassification, classifier learns major data more, to increase the correct classification of major class. Therefore, minor class can be classified as major class. Many studies have been performed to overcome this imbalanced data problem [9]. Table 3 shows the related works of learning imbalanced data. Oversampling extracts minor class by duplication for total dataset. Undersampling extracts less major class compared to total dataset. The mixture-of-experts is good at dealing with imbalanced data.

Table 3 Related works of learning imbalanced data

Author	Method
Chen et al. [10]	Oversampling extracts minor class duplicately for total dataset
Kubat et al. [11]	Undersampling which extracts major class, less compared to total dataset
Cardie et al. [12]	Learning method which gives a weight to minor class
Grzymala et al. [13]	Method which gives a weight to learning rule
Joshi et al. [14]	Method which boost minor class
Veropoulos et al. [15]	Gives penalty to each class
Kotsiantis et al. [16]	Utilizes mixture of experts to deal with imbalanced data

2.2 Co-training

Co-training is an approach of multi-view semi-supervised learning proposed by Blum [17]. A classifier becomes better when the features used follow the two assumptions: Each feature is appropriate for its problem space, and two features in a class are conditionally independent. Co-training is based on two learning models for different feature sets with each other. The models are independently learned, and enlarge the opponent's data instances while iteratively predict the unlabeled data instances. Originally, co-training was used for classifying web pages and has been applied to a variety of problems [18], [19].

Co-training is often compared with another conventional method of semi-supervised learning, called as Expectation-Maximization (EM) [20]. Because of a concrete model assumption problem, co-training is recently preferred to the EM algorithm [21], and it can be used for learning the mixture-of-experts at the semi-supervised scheme. Although the co-training algorithm needs two independent feature sub-spaces, these are difficult to split feature space into two sub-spaces in the real world. To overcome it, Goldman proved that two models trained with different methods become two independent feature sets [22].

2.3 Mixture-of-experts

The mixture-of-experts is a class of probabilistic models [5], [23]. It consists of several experts that manifest conditional probabilistic processes, and a gating network of combining the experts. The operational rationale in the mixture-of-experts is that each expert constructs a conditional model to cover a part of input space, while the gating network determines the weights of each expert to form an ensemble.

Due to the power of divide-and-conquer, it can transform the non-linearly separable problem into a set of the linearly separable problems. Figure 3 illustrates these characteristics with a radial decision boundary that was approximated by several linear decision boundaries. The linear decision boundaries are mapped to the localized regions in the whole space. This principle states that complex problems can be better-solved by decomposing them into smaller tasks. The mixture-of-experts assumes that separate processes are in the underlying process of generating the data. Modelling these processes is conducted by the experts, while the decision for the process is modelled by the gating network.

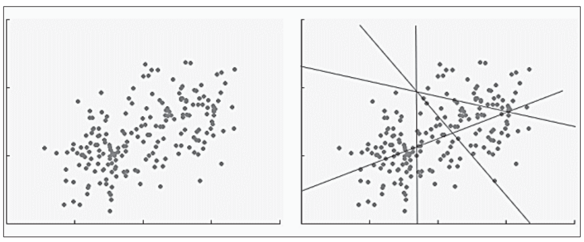


Fig 3. Non-linearly separable boundary estimated by several linearly separable boundaries

2.4 Prediction of financial data

Several researchers have studied the credit score by machine learning methods such as neural networks (NN) [24], [25], decision tree (DT) [26], support vector machine (SVM) [27], and case-based reasoning (CBR) [28]. They are different from the statistical approaches that assume a specific distribution of dataset. Table 4 shows the relevant research, each of which confirms the superiority to the statistical methods for credit scoring [29]. Similar to the credit scoring problem, many researchers have worked out the financial problems in general with machine learning methods. Min used SVM for predicting bankruptcy to verify the explanatory power and stability of the method [30]. Tsai presented a hybrid method of four different models and showed the superiority of the combination of machine learning and clustering algorithms with the dataset from a Taiwan bank [31]. In spite of the good performance, the previous works did not fully exploit the characteristics in the financial data. The co-training of mixture-of-experts are novel for the credit scoring problem.

Table 4 Relevant works of machine learning for financial data

Author	Problem	Data	Method
Min, et al [30]	Bankruptcy prediction	Corporate financial information	SVM
Li, et al [32]	Customer loan assessment	Customer financial information	SVM
Choudhry, et al [33]	Stock market forecasting	Customer credit information	GA and SVM
Tsai, et al [31]	Credit scoring	Customer credit information	EM and DT, EM and NB
Omidi, et al [34]	Forecasting stock prices	Stock price information	Neural Network
Milad, et al [1]	Risk evaluation	Social lending information	Random Forests
Zang, et al [35]	Risk evaluation	Social lending information	Neural Network
Emekter, et al [36]	Risk evaluation	Social lending information	Logistic Regression
Bitvai, et al [37]	Credit rating	Social lending information	Bayesian Non-Linear Regression
Byanjankar, et al [38]	Risk evaluation	Social lending information	Neural Network
Serrano-Cinca, et al [39]	Credit rating	Social lending information	Decision Tree
Zhang, et al [40]	Risk evaluation	Corporate financial information	Long Short-Term Memory
Huo [41]	Credit rating	Customer credit information	Logistic Regression, Neural Network

3 Proposed method

3.1 Semi-supervised learning for credit scoring

The semi-supervised approach is required to classify personal information in P2P environment. With the collection of the personal data, it is time-consuming for people to classify all of these borrowers. Personal requests to borrow money are always created as nature of social lending. Because these tasks have high labelling cost, we need to learn model through minor classified user information.

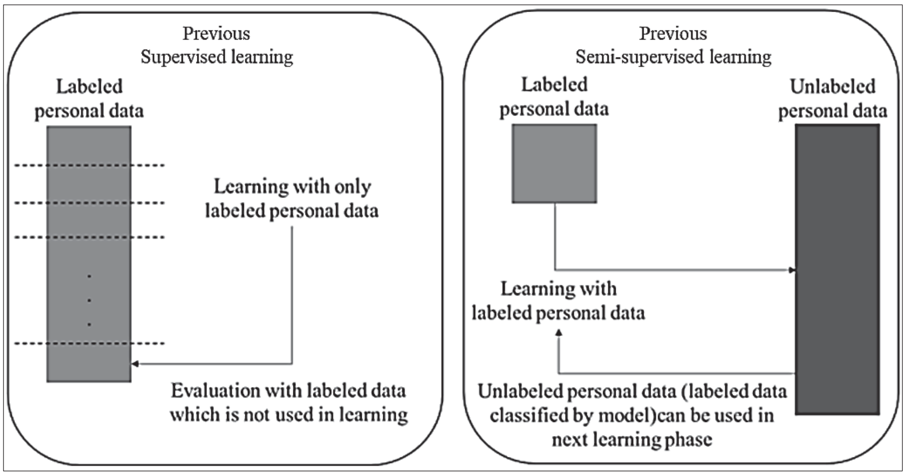


Fig 4. Semi-supervised approach for credit scoring

To predict credit scoring, previous studies proposed the methods based on supervised learning mostly. However, sample dataset is different from parent population in real social lending. In Figure 4, the classifier has a possibility to learn classification standard of other environment which is different from the real world. Therefore, we propose a semi-supervised approach to learn minor labeled personal data and major unlabeled personal data simultaneously.

3.2 Co-training with global and local perspectives

At first, local experts should be constructed in the separate space of dataset and then the mixture-of-experts model determines the credit scoring. When we divide the whole space into several sub-spaces, the data instances in each sub-space are similar with each other. The local experts become specialists to separate the similar patterns in a part of entire personal data. We use the k-means algorithm to make the clusters. Figure 5 describes the whole process of the co-training with global and local perspectives that were originally proposed at our previous study [42].

Figure 6 illustrates the algorithm based on this structure for predicting defaults. It can adopt the global and local views for the classification problem simultaneously.

Because it is not easy to divide the features of data instances according to each view, we use different learning algorithms for each model.

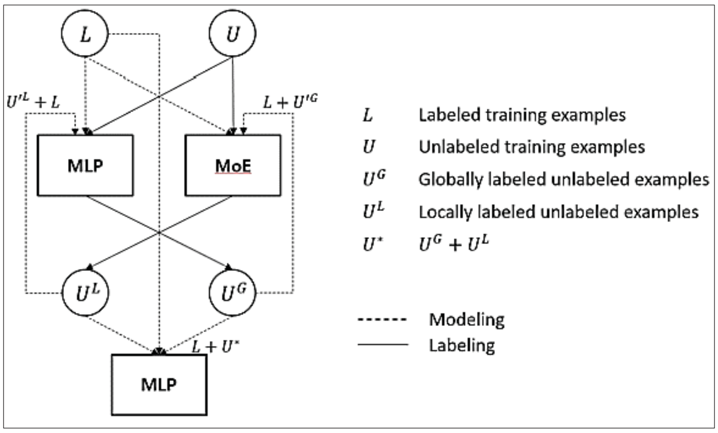


Fig 5. The overall procedure of the proposed learning method

Inputs:

Threshold e.
Labeled data L.
Unlabeled data U.
Multi-layer perceptron M_G
Mixture-of-experts $M_L = (\mu_1, \mu_2, \dots, \mu_N, g)$

Initialization:

Generate two sets of training instances $L_G \leftarrow L$ and $L_L \leftarrow L$.
Do while L_G and L_L are not the same with the previous iteration:
  Train M_G and M_L with L_G and L_L.
  Predict U with M_G andM_L.
  Get instances that the confidence degree of M_G is higher than M_L and add them to U^L.
  Get instances that the confidence degree of M_L is higher than M_G and add them to U^G.
  Sort U^L and U^G in descending order of confidence.
  For each class C
    Select N instances in U^L which have higher confidence than threshold e Add them to L_L with label C.
    Select N instances in U^G which have higher confidence than threshold e
    Add them to L_G with label C.
    Remove the selected data instances in U, U^L and U^G.

Output:

The predicted labels of unlabeled instances.
Multi-layer perceptron M_G and mixture-of-experts M_L.

Fig 6. The co-training algorithm used in this paper

The mixture-of-experts utilizes various experts focusing on their own fields. Social lending information that is able to decide credit evaluation consists of various attributes. In order to evaluate the credit of borrowers, various attributes should be recognized accurately. Fifty-four attributes are inserted to the mixture-of-experts as input value. Gating network determines the reflection degree of opinions of experts based on the input value. The output of gating network $G(x)$ can be defined as equation (1) for input value x. The final output vector $O(x)$ of combining outputs of N-experts is calculated by the following equation (2).

$$G(x) = \{g_1(x), g_2(x), g_3(x), \dots, g_N(x)\} \tag{1}$$

$$O(x) = G(x)E(x)^T = \sum_{i=1}^{N} g_i(x)e_i(x) \tag{2}$$

$$where\ E(x) = \{e_1(x), e_2(x), e_3(x), \dots e_N(x)\}$$

Among these combined opinions, the highest value is chosen as the result through the above process. Local experts and gating network were implemented by using multilayer perceptrons. In order to train gating network, the expected output about specific input value x is calculated by the accuracy of each expert. Before calculating the accuracy for each expert, the error of expert i for x can be defined as equation (3). Figure 7 shows the structure of mixture-of-experts in the proposed co-training.

$$Err_i(x) = \sum_t \left| e_i^t(x) - y^t \right| \tag{3}$$

where y is the correct vector corresponding to input value x. $e_i^t(x)$ and y^t means the values of $e_i(x)$and y at time t. After the normalization using minimum error and maximum error of each expert, the accuracy of an expert is calculated by equation (4).

$$Acc_i(x) = 1 - \frac{Err_i(x) - \min_k Err_k(x)}{\max_k Err_k(x) - \min_k Err_k(x)} \tag{4}$$

$$y_g^t = Acc_t(x)(t = 1,2,3, \dots, N) \tag{5}$$

Finally, the expected output vector y of gating network for x can be defined as equation (5). Gating network is learned based on the backpropagation algorithm using total training data instances and expected output vector.

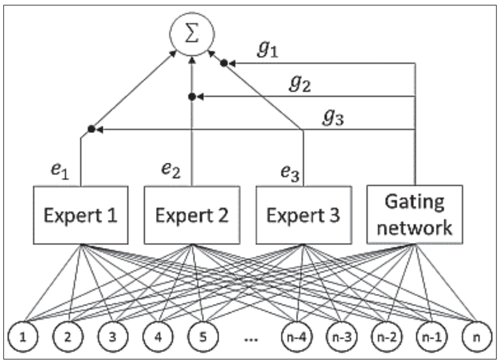


Fig 7. Structure of mixture-of-experts

Generally, it is important to carefully predict the labels of unlabeled instances when we have insufficient data for the local experts, because of the imperfect characteristics and poor performance of the models. The proposed method compares the two confidences from M_G and M_L to minimize the degradation caused from imperfection. After initial training with the data for credit scoring, the models get labeling only when the confidence of a model is higher than the other. The data samples which pass the criteria are added to the set of U^Lor U^G.

After the added samples are sorted in descending order of the confidence, N samples in each class with the highest confidence are chosen. By picking the data instances with high confidence, the local experts can be improved for the prediction performance. The training process stops when L_G and L_L are the same; in other words, the proposed method continues learning iteratively until additional data instances for $L_L$and $L_G$are not left.

In the proposed co-training method, the two models are based on the multilayer perceptron trained with backpropagation algorithm. They are trained repeatedly using gradient descent with the error function of given data instances from its opponents. Newly added data instances from the global model are used for learning the other model, local experts, at the next iteration and for the local experts are vice versa. In the following equations, E_L and E_G indicate the error functions of the additional labeled samples for the two models, respectively.

$$E_G = \|O_G - f_G(x)\| \tag{6}$$

$$E_L = \left\|O_L - \sum \pi_i f_{Li}(x)\right\| \tag{7}$$

where $O_L$is an output vector of the local experts and $f_G(x)$ is an output vector of the global model. In equation (7), an error function of a local expert, O_G, means, an output of the global model, and π_i is an output of gating network in the ith local experts, and $f_{Li}(x)$is an output of the ith local experts. In particular, O_G can be altered with an output function of the global model in the previous step, $f'_G(x)$, and an output function of the local experts in the previous step, $\sum \pi'_i f'_{Li}(x)$can be replaced with O_L as shown in the equations.

$$E_L = \left\| f'_G(x) - \sum \pi_i f_{Li}(x) \right\| \tag{8}$$

$$E_G = \left\| \sum \pi'_i f'_{Li}(x) - f_G(x) \right\| \tag{9}$$

The co-training method stops when the current and previous errors are the same, but it is not easy to satisfy the condition. In this paper, we set the maximum number of iterations for training process to that of added training instances for each model. The number of samples for training each model N(L_L) and N(L_G) are determined with the confidence degree.

3.3 Measuring confidence degrees

As a detailed method of semi-supervised learning, we classify unlabeled data as each class using the proposed models learned with minor labeled data. Among these classified data, we select the data with higher confidence and give these data to labeled dataset. Each classifier can model with updated dataset again. In the process of data selection with higher confidence, we collect the prediction score or probability of each data for each classifier. Selection method based on the probability is applied. If the probability of each classifier is the same as Table 5, this data can be classified as bad class with the highest probability.

Table 5 Data example for classification (probability-based)

Classifier	**Prob. (bad class)**	**Prob. (good class)**
MLP	0.24	0.76
Mixture-of-expert	**0.89**	0.11

For determining the confidence value, we set confidence(w, x) as an original confidence of the input x for the model w – confidence(w, x) can be changeable with the learning methods. In this paper, the models are implemented with multilayer perceptron whose outputs can be evaluated by a measure of confidence as follows:

$$confidence(w, x) = \{w(x) |\ output\ vector\ of\ w\ for\ x\} \tag{10}$$

The number of data in local experts is smaller than that in global model, because the original data was split into several sub-data for the local experts. As the data instances for some classes do not exist within sub-spaces even if the original problem space covers the whole classes, the mixture-of-experts may have a higher bias. Since the experts cannot be trained with all the classes, it can output unrelated values about missing classes at the local experts. In order to overcome these problems, the training instances in certain class should be confirmed whether they exist or not, before calculating a confidence of the model. The following equations show the modified confidence degrees for class C.

$$confidence(M_G, x, C) = \begin{cases} M_{G,C}(x), & if\ N_{M_{G,C}} > 0, \\ 0, & otherwise \end{cases} \quad (11)$$

$$confidence(M_L, x, C) = \sum_{i=1}^{N} g_i confidence(\mu_i, x, C) \quad (12)$$

$$confidence(\mu_i, x, C) = \begin{cases} \mu_{i,C}(x), & if\ N_{\mu_{i},C} > 0, \\ 0, & otherwise \end{cases} \quad (13)$$

where μ_i is the ith local experts, g specifies a gating network, and $M_{G,C}(x)$ and $\mu_{i,C}(x)$denote the outputs of M_G and μ_i for class C. g_i is an output of g for the ith local experts and $N_{M_{G,C}}$and $N_{\mu_i,C}$ are the numbers of training data samples for class C for $M_G$and μ_i.

4 Experiments and Results

4.1 Social lending dataset

The social lending dataset comes from the Lending Club, and consists of personal and financial variables, indicators of wealth, and variables specific to the loan as shown in Table 6. The data contains approximately 400K people, and consist of 56 attributes including one dependent variable. The data was collected between January 2012 and September 2015.

Table 6 Description of the lending club dataset

Section	Variable	Description
Variables specific to the loan	Loan amount	The listed amount of the loan applied for by the borrower.
	Term	The number of payments on the loan.
	Purpose	A purpose provided by the borrower for the loan request.
	Interest rate	Interest rate on the loan.
	Installment	The monthly payment owed by the borrower if the loan originates.
Indicators of wealth	Home ownership	The home ownership status provided by the borrowers.
	Annual income	The self-reported annual income provided by the borrower.
Personal and financial variables	Employment length	Employment length in years.
	Debt to income	A ratio of the borrower's total monthly debt payments on the total debt obligations.
	Delinquency	The number of delinquency in the borrower's credit file for the past two years.
	Inquiry	The number of inquiries in past six months.

	Total account	The total number of credit lines currently in the borrower's credit file.

Credit scoring is performed by the proposed method that classifies the criteria such as good or bad credit of the client. The proposed method uses a lot of variables which can be acquired from information of client including both unlabeled and labeled ones. After learning with events (loan) that have already occurred and the results (payment or delinquent), we evaluate the credit scoring in a way asking whether this client can repay or not. The purpose of analysis of data is to evaluate credit of clients.

4.2 Reliability test

First, we conducted the experiments for evaluating the accuracy according to the reliability of unlabeled dataset. We divide the labeled dataset into unlabeled data (U) and labeled data (L) intentionally to check the accuracy of unlabeled dataset. The number of instances of labeled data is 45,304. The number of instances of unlabeled data is 15,102. The dataset ratio is 75 (labeled): 25 (unlabeled).

Table 7 shows the confusion matrix of unlabeled dataset with respect to the reliability threshold. We conducted experiments with two semi-supervised classifiers (the proposed method and self-training) and confirmed that higher reliability led to higher accuracy of dataset. Instances of data used in this experiment are reduced in case of higher reliability. However, the accuracy of used dataset can increase as shown in Figure 8.

Table 7 Confusion matrix of unlabeled dataset based on reliability

Reliability	Dataset	TP	TN	FP	FN
The proposed method					
0.99	8,413	7,556	423	362	72
0.95	11,335	9,543	912	714	166
0.90	12,229	9,971	1,178	840	240
0.85	13,137	10,643	1,180	1,083	231
0.80	13,629	10,697	1,468	1,091	373
Self-training					
0.99	8,648	7,791	404	58	395
0.95	12,965	10,621	1,017	1,093	234
0.90	13,116	10,626	1,205	1,058	227
0.85	14,168	10,912	1,592	1,212	452
0.80	14,131	10,775	1,685	1,125	5,476

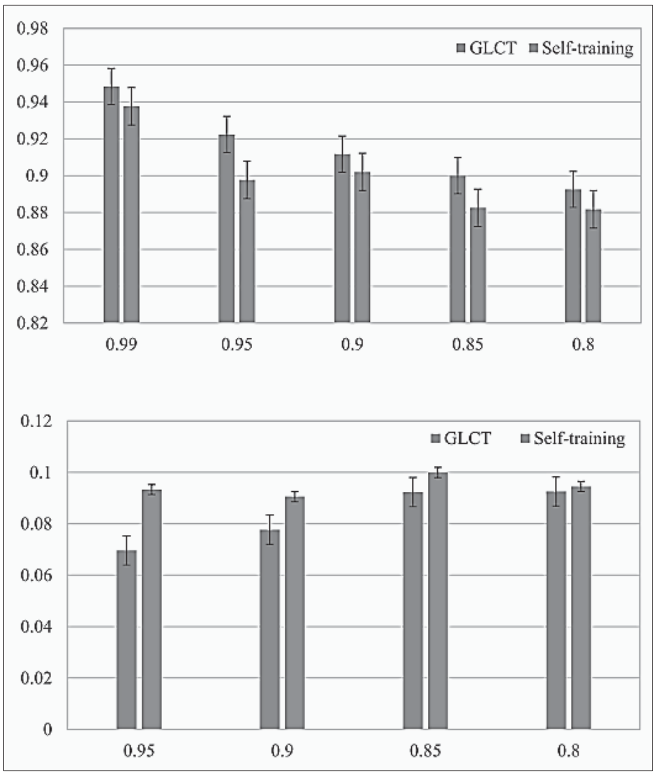


Fig. 8 Accuracy (top) and false positive rate (bottom) for U*

4.3 Quantitative analysis

Next, we present the experiments to evaluate the usefulness of predicting the default in social lending. We chose k-means clustering algorithm for obtaining training data for the mixture-of-experts with three experts. The parameters used are as follows: the number of hidden nodes in the multilayer perceptron is 10, the learning rate is 0.3, and the number of cluster is 3. In the initial step, the labeled data samples are used for training, and then the unlabeled data samples are labeled based on the previous model.

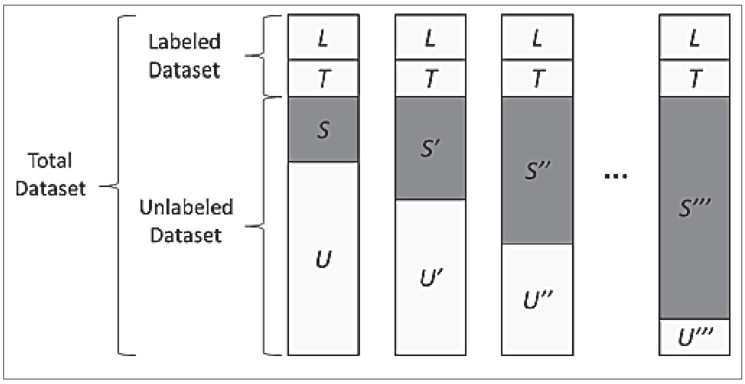


Fig. 9 Distribution of dataset

Labeled data used for experiment is the Lending Club 2014 dataset. The number of training samples is 52,857, and that of test samples is 22,651. The unlabeled data (U) used is also taken from Lending Club 2014 dataset. The number of samples is 161,121. Ratio of the selected unlabeled data is decile. It ranges from 10% to 30% as shown in Figure 9. Additionally, we conducted experiment with the whole unlabeled dataset. Each sample according to the dataset ratio is below 10% (16,011), 20% (32,022), 30% (48,033), and 100% (160,121).

Figure 10 shows the result with the unlabeled dataset ratios. We can confirm that utilizing unlabeled dataset leads to higher accuracy. However, when utilizing all of unlabeled dataset, it does not guarantee improvement in performance in semi-supervised learning. We compared with another conventional semi-supervised learning to show the usefulness of the proposed method. Figure 11 shows the result compared with self-training. We can confirm that the proposed method had better performance than the other semi-supervised learning.

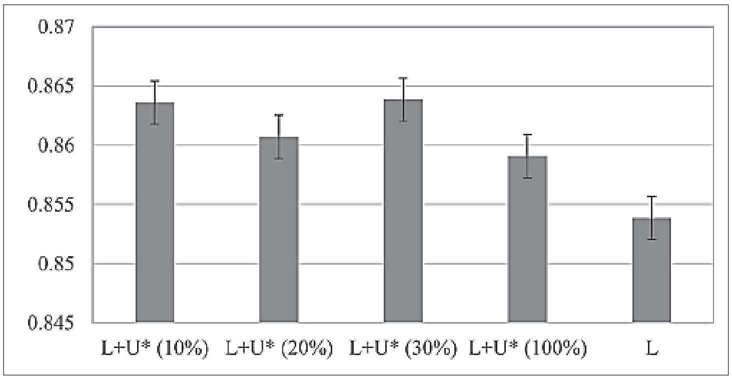


Fig. 10 Comparison of accuracies according to the unlabeled dataset ratios

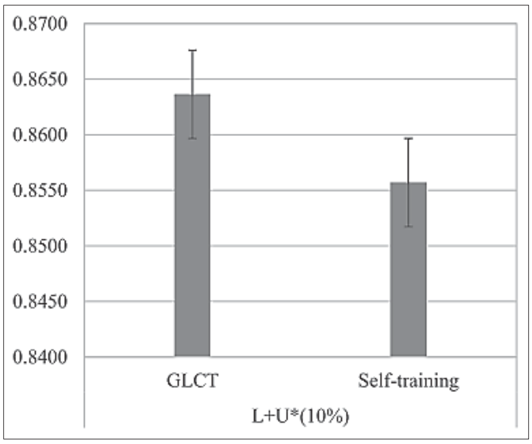
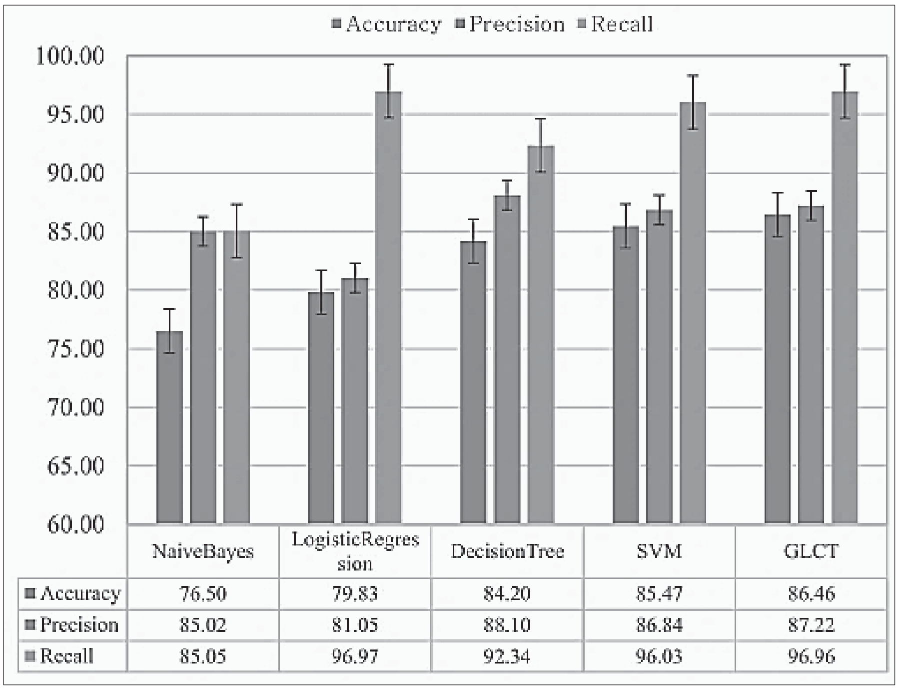


Fig. 11 Comparison of accuracies with self-training

4.4 Comparing with other machine learning methods

We compared the performance against other machine learning methods such as Naive Bayes, logistic regression, decision tree, and SVM. The number of training data used in other machine learning methods is 52,857, and that of test data is 22,651. In the experiment, we used unlabeled dataset additionally, and the number of unlabeled data becomes 16,011. Figure 12 shows higher accuracy of the proposed method compared with other machine learning methods. To confirm the qualitative results, we plotted ROC curve for all the methods. The proposed method was the best with the highest AUC value as shown in Figure 13.



	NaiveBayes	LogisticRegression	DecisionTree	SVM	GLCT
Accuracy	76.50	79.83	84.20	85.47	86.46
Precision	85.02	81.05	88.10	86.84	87.22
Recall	85.05	96.97	92.34	96.03	96.96

Fig. 12 Comparison with other machine learning methods

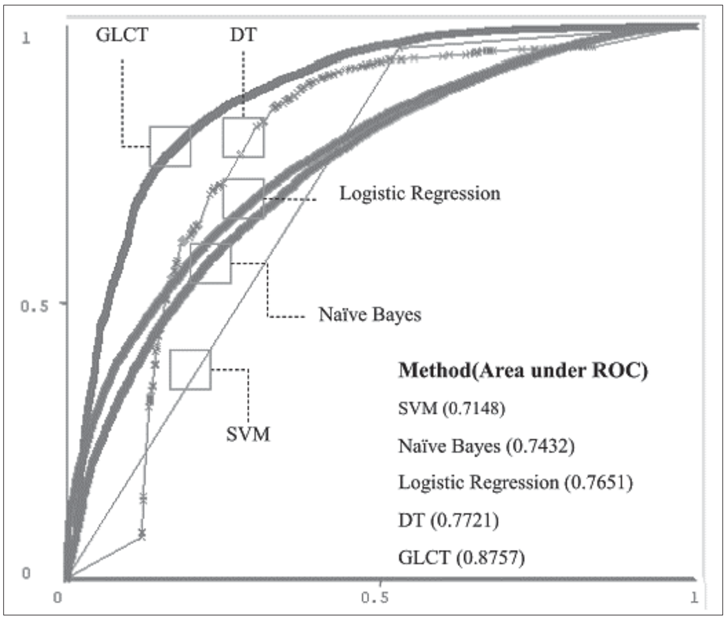


Fig. 13 ROC curve with other machine learning methods

5 Conclusion and future work

Identifying potential default borrower is important for maintaining the sustainability of social lending market. To predict the default of borrowers, financial features such as the number of delinquent accounts, debt-to-income ratio and so on are used. A large scale of unlabeled data from people in the world can be one of the main characteristics in financial data. In this paper, we incorporate a co-training algorithm based on mixture-of-experts for distinguishing default borrowers in social lending, and use the world's largest social lending platform, Lending Club's, publicly available dataset. The experiment was conducted with the ratio of labeled data. We obtained the quantitative results evaluated on various aspects through a series of iterative experiments, confirming the usefulness of the proposed method.

In the future, we will conduct additional experiment to validate reliability. Another research direction can be the integration with the social media dataset to get better default prediction. There exists a large dataset that borrower characteristics can be identified by, such as Twitter and Facebook. The data from social media have a great potential to improve the method with higher accuracy for predicting borrower's default by integrating with the social lending data.

Acknowledgement

This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2015-0-00369) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

1. M. Malekipirbazari and V. Aksakalli: Risk assessment in social lending via random forests. Expert Systems with Applications. **42**(10), 4621–4631 (2015)
2. A. I. Marqués, V. García, and J. S. Sánchez: Exploring the behaviour of base classifiers in credit scoring ensembles. Expert Systems with Applications. **39**(11), 10244–10250 (2012)
3. L. Zhou and H. Wang: Loan default prediction on large imbalanced data using random forests. TELKOMNIKA Indonesian Journal of Electrical Engineering and Computer Science, **10**(6) (2012)
4. G. Weiss and F. Provost. The effect of class distribution on classifier learning: an empirical study. Rutgers University (2001)
5. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton: Adaptive Mixtures of Local Experts. Neural Computation. **3**(1), 79–87 (1991)
6. W. Wang and Z.-H. Zhou: Analyzing co-training style algorithms. European Conference on Machine Learning, 454–465 (2007)
7. T. Jo and N. Japkowicz: Class imbalances versus small disjuncts. ACM Special Interest Group for Knowledge Discovery and Data Mining Newsletter, **6**(1), 40 (2004)
8. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone: Classification and Regression Trees. 19, 1984
9. L. A. Zadeh. A Simple View of the Dempster-Shafer Theory of Evidence and Its Implication for the Rule of Combination. AI Magazine, **7**(2), 85, June 1986
10. X. W. Chen, B. Gerlach, and D. Casasent: Pruning support vectors for imbalanced data classification. International Joint Conference on Neural Networks, 3, 1883–1888 (2005)

11. M. Kubat and S. Matwin: Addressing the curse of imbalanced training sets: one sided selection. International Conference on Machine Learning, 97, 179–186 (1997)
12. C. Cardie: Improving minority class prediction using case-specific feature weights. International Conference on Machine Learning, 57–65 (1997)
13. J. W. Grzymala-Busse, L. K. Goodwin, W. J. Grzymala-Busse, and X. Zheng: An approach to imbalanced data sets based on changing rule strength. Rough-Neural Computing, 543–553, 2004
14. M. V. Joshi, V. Kumar, and R. C. Agarwal: Evaluating boosting algorithms to classify rare classes: comparison and improvements. IEEE International Conference on Data Mining, 257–264
15. K. Veropoulos, C. Campbell, N. Cristianini, and Others: Controlling the sensitivity of support vector machines. International Joint Conference Artificial Intelligence, 55–60 (1999)
16. S. B. Kotsiantis and P. E. Pintelas: Mixture of expert agents for handling imbalanced data sets. Annals of Mathematics, Computing & Teleinformatics, **1**(1), 46–55 (2003)s
17. A. Blum and T. Mitchell: Combining labeled and unlabeled data with co-training. Annual Conference on Computational Learning Theory, 92–100 (1998)
18. S. Kiritchenko and S. Matwin: Email classification with co-training. Conf. Cent. Adv. Stud. Collab. Res., 301–312 (2001)
19. Y. Fang, J. Cheng, J. Wang, K. Wang, J. Liu, and H. Lu: Hand posture recognition with co-training. International Conference on Pattern Recognition, 1–4 (2008)
20. K. Nigam, A. K. Mccallum, S. Thrun, and T. Mitchell: Text classification from labeled and unlabeled documents using EM. Machine Learning, 39, 103–134 (2000)
21. K. Nigam and R. Ghani: Analyzing the effectiveness and applicability of co-training. International Conference on Information and Knowledge Management, 86–93 (2000)
22. S. a. Goldman and Y. Zhou: Enhancing supervised learning with unlabeled data. International Conference on Machine Learning, 3, 327–334 (2000)
23. M. I. Jordan and R. A. Jacobs: Hierarchical mixtures of experts and the em algorithm. Neural Computation, **6**(2), 181–214 (1994)
24. V. S. Desai, J. N. Crook, and G. A. Overstreet: A comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, **95**(1), 24–37 (1996)
25. D. West: Neural network credit scoring models. Computers & Operations Research, **27**(11–12), 1131–1152 (2000)
26. C. Hung and J. H. Chen: A selective ensemble based on expected probabilities for bankruptcy prediction. Expert Systems with Applications. **36**(3-1), 5297–5303 (2009)
27. K. B. Schebesch and R. Sleeking: Support vector machines for classifying and describing credit applicants: Detecting typical and critical regions. Journal of the Operational Research Society, **56**(9), 1082–1088 (2005)
28. K. Shin and I. Han: A case-based approach using inductive indexing for corporate bond rating. Decision Support Systems, **32**(1), 41–52, (2001)
29. Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu: Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems, **37**(4), 543–558 (2004)
30. J. H. Min and Y.-C. Lee: Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications, **28**(4), 603–614 (2005)
31. C.-F. Tsai and M.-L. Chen: Credit rating by hybrid machine learning techniques. Applied Soft Computing, **10**(2), 374–380 (2010)
32. S. T. Li, W. Shiue, and M. H. Huang: The evaluation of consumer loans using support vector machines. Expert Systems with Applications, **30**(4), 772–782 (2006)
33. R. Choudhry and K. Garg: A hybrid machine learning system for stock market forecasting. World Academy of Science, Engineering and Technology, **2**(15), 315–318 (2008)
34. A. Omidi, E. Nourani, and M. Jalili: Forecasting stock prices using financial data mining and Neural Network. International Conference on Computer Research and Development, **3**(30), 242–246 (2011)
35. D. Zang, M. Qi, and Y. Fu: The credit risk assessment of P2P lending based on BP neural network. International Conference on Industrial Engineering and Management Science, 2, 91 (2014)
36. R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu: Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Applied Economics, **47**(1), 54–70, January 2015

37. Z. Bitvai and T. Cohn: Predicting peer-to-peer loan rates using bayesian non-linear regression. AAAI Conference on Artificial Intelligence, 2203–2209 (2015)
38. A. Byanjankar, M. Heikkila, and J. Mezei: Predicting credit risk in peer-to-peer lending: a neural network approach, in IEEE Symposium Series on Computational Intelligence, 2015, 719–725.
39. C. Serrano-Cinca and B. Gutiérrez-Nieto: The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. Decision Support Systems, 89, 113–122, September 2016
40. Y. Zhang, D. Wang, Y. Chen, H. Shang, and Q. Tian: Credit risk assessment based on long short-term memory model. International Conference on Intelligent Computing, 700–712 (2017)
41. Y. Huo, H. Chen, and J. Chen: Research on personal credit assessment based on neural network-logistic regression combination model. Open Journal on Business Management, **5**(2), 244–252, April 2017
42. Y. S. Lee and S. B. Cho: Activity recognition with android phone using mixture-of-experts co-trained with labeled and unlabeled data. Neurocomputing. 126, 106–115 (2014).