Borrower-lender Information Fusion for P2P Lending: A Nonparametric Approach

Yanhong Guo*, Shuai Jiang, Feiting Chen, Yaocong Li, Chunyu Luo

School of Economics and Management, Dalian University of Technology, Dalian 116024, China

Corresponding Author Email: guoyh@dlut.edu.cn

ABSTRACT

As known to us, tremendous efforts have been made to exploit information from borrower's credit for loan evaluation in P2P lending, but seldom researches have explored information from investors and borrowers. To this end, we propose an integrated loan evaluation model that exploits and fuses multi-source information from both the borrower and the investor for improving investment decisions in P2P lending. First, based on the borrower's credit, we build a kernel-based credit risk model to quantitatively evaluate each loan. Second, we build an investor composition model that exploits information from the investor's investment behavior for loan evaluation. Then, based on the above two quantitative models and correlation, we define a multi-kernel weight and develop an integrated loan assessment model that can evaluate a loan with both the return and risk. Furthermore, based on the integrated information loan evaluation model, we formalize the investment decisions in P2P lending as a portfolio optimization problem with boundary constraints to help investor make better investment decisions. To validate the proposed model, we perform extensive experiments on the real-world data from the world's largest P2P lending marketplace. Experimental results reveal that the integrated loan evaluation model can significantly enhance investment performance beyond the existing model in P2P market and other baseline models.

1. INTRODUCTION

Peer-to-peer (P2P) lending platform is a novel online market for consumers' debt where individuals can directly lend to and borrow from each other in absence of the traditional intermediation. Borrowers describe the total amount of money needed and provide information about their current financial situation to generate listings. Then lenders offer the amount they would like to provide with an interest rate based upon these listings. Once the amount requested by one listing is collected from lenders within the specified time, this listing will become a loan. Different from traditional banks, P2P lending has several new properties in the following aspects. First, the investment relation in P2P lending is a many-to-many investment relation. To reduce risk, a lender will diversify his investment by allocating money on many different loans, and a loan can gain investments from different lenders. Moreover, P2P lending markets have provided large amounts of public real-world P2P lending transaction data. It is possible to track all activities of each participant and make good use of such data. All of these have enabled new research opportunities, especially on how to accurately evaluate a loan for making effective investment decisions.

In P2P lending marketplace, investors not only need to decide which loans to fund, but also how much money to allocate to each of them, which minimizes risk for a given expected return. While this feature presents as a typical portfolio optimization problem, it is very challenging to accurately assess the expected return and credit risk of each individual loan, which will serve as necessary input for portfolio optimization. To this end, tremendous efforts have been made to exploit information from borrowers' credit for

loan evaluation and investors behavior information separately. Few attempts have been made to integrate the two sources information. In this study, we study how to integrate two information sources from the two aspects of investor composition and borrower credit to establish a comprehensive loan evaluation model for P2P lending quantitative investment decisions.

We introduce a multi-kernel loan evaluation model that can effectively integrate two information sources through a kernel-based information fusion method. Specifically, we first use the mathematical framework of kernel regression to convert two information sources into two kernel weights, in which the investor kernel weight is calculated according to the investor composition score distance in the investor portfolio model, and the borrower kernel weight is calculated according to the loan default distance in the instance-oriented credit risk model. Secondly, we integrate the two kernel weights into a unified multi-kernel weight based on the correlation coefficient between the historical loan performance and the predicted performance using two information sources. In addition, we establish a multi-kernel loan evaluation model with multi-kernel weights, which predicts the return of each loan as the weighted average of similar loans, and the risk as the weighted variance. Finally, based on the multi-kernel loan evaluation model with expected returns and risks, we propose a quantitative investment decision algorithm based on modern portfolio theory. To validate our model, we conduct extensive experiments on real-world data from the world's largest P2P lending market. The experimental results show that compared with the existing model and other baseline models in P2P market, the proposed loan evaluation model integrating two information sources can significantly improve the investment

performance.

The rest of this paper is organized as follows. Section 2 outlines the related work. Section 3 describes the operation of two information sources in P2P lending. In section 4, we construct a multi-kernel loan evaluation model to enhance investment decisions. Section 5 presents experiments on actual data to verify the validity of the proposed model. Finally, section 6 summarizes this work.

2. RELATED WORK

2.1 P2P lending

In recent years, P2P lending has been introduced as a new e-commerce phenomenon in the financial field [1-4]. There exist substantial literatures studied the social aspects of P2P lending. Lin et al. [5] found that the online friendships of borrowers' act as signals of credit quality. Friendships increase the probability of successful funding, lower interest rates on funded loans. Liu et al. [6]; Faia and Paiella [7] also encountered in the literature. Duarte et al. [8] studied the role of appearance in peer-to-peer lending and argued that borrowers who appear more trustworthy have higher probabilities of having their loans funded. Lin and Viswanathan [9] found evidence that home bias exists in P2P lending market and showed that rationality-based explanations cannot fully explain such behavior Chen et al. [10] examined higher education level will lead to lower interest rates and lower risk of default in P2P lending platform. Dorfleitner et al. [11] concluded that the soft information (spelling errors, text length etc.) influence the funding probability and probability of default. On the other hand, researchers also studied the economic aspects of P2P lending extensively. Wei and Lin [12] studied the market mechanisms in online P2P lending and found that under platform-mandated posted prices, loans are funded with higher probability, while loans funded under posted prices are more likely to default. Zhao et al. [13] proposed a focused study on market state modeling to discern the hidden market states of the listings in online P2P lending. Redmond and Cunningham [14] revealed the unprofitability of arbitrage in P2P lending market. Rigbi [15] found that higher interest rate caps increase the probability that a loan will be funded. These studies provide a foundation for the successful development of this emerging marketplace.

Among all the information provided by the participants, researchers devoted their efforts to the study of effective methods which can help the participants achieve their goals. From borrower's perspective, researchers aimed to find determinants of the success of a loan and develop decision support systems to help borrowers optimize their decisions, such as requested amount and interest rates [16-18]. When considering the lenders' perspective, researchers investigated lender's investment decision making and bidding behavior. For example, Herzenstein et al. [19] evidenced that lenders have a greater likelihood of bidding on an auction with more bids. Puro et al. [20] defined and identified bidding strategies for lenders and Herzenstein et al. [21] found that unverifiable information affects lending decisions above and beyond the influence of verifiable information. Researchers proposed some methods on how lenders screen the loans, such as profit scoring system [22], decision support tool [23], random forest based classification method etc [24]. Existing studies provided valuable insights into how lenders screen and select loans, and

succeed in bidding.

2.2 Loan evaluation and decision making

Risk assessment and decision making are the main tasks for participants in the P2P lending marketplace, where traditional loan evaluation model can be used. Many analytical techniques have been proposed to distinguish good loan applications from bad applications. For instance, logistic regression [25], linear discriminate analysis [26], and k-nearest neighbor (KNN) classifiers [27], classification tree [28-29], markov chain [30-31], survival analysis [32], linear and nonlinear programming [33], neural networks [34-35], Support Vector Machines (SVMs) [36-38], genetic methods [39-41] and so on. Hybrid approaches include fuzzy systems and neural networks [42], fuzzy systems and support vector machines and neural networks and multivariate adaptive regression spines [43]. For ensemble models, the neural network ensemble is a typical example. Hill and Ready-Campbell [44] found that user-generated content is an acceptable indicator in which crowd wisdom can be used to identify good stocks and provide a genetic algorithm to learn the appropriate contributions of independent users through the use of observed past individual performance.

All proposed models mentioned above has been proved that they can effectively enhance investors' decision in the appropriate context. Most of them are classification methods which aim to help investors distinguishing good loans and bad loans by dividing the loans into different rated groups. Since investors aim to assess the expected return and credit risk of each individual loan accurately, this feature may be too coarse to meet the needs of personal investors in P2P lending. Hence, there remains the need of effective decision support for personal investors to select a subset of investments and determination of optimal amounts to put forward in each of them.

2.3 Information fusion

Information fusion is a formal framework which used as a tool for the alliance of data originating from different sources. It aims at obtaining information of greater quality [45]. This technique originated in the field of military at the beginning of the eighties and is widely used in different fields with different goals, such as dimensionality reduction, precision and certainty etc [46]. Mathematical tools play an important role in data fusion and many kinds of theories are used by scientists in the information fusion field, such as probability theory [47], neural networks [48-49] and fuzzy subset theory [50-52] and evidence theory [53-54].

In this study, we apply the mathematical framework of the multi-kernel regression to extract the regression coefficient as the optimal weighting for credit risk assessment. Kernel regression [55-56] is a non-parametric technique in statistics to estimate the conditional expectation of a random variable. Multiple kernel learning (MKL) method was originally proposed in the field of bioinformatics which is used to synthesize information from disparate types of genomic data [57-58]. Gnen and Alpaydin [59] gave a review several multiple kernel learning algorithm and perform experiments on real datasets for better illustration and comparison of existing algorithms. Noble [60] reviewed the state of the art with respect to SVM applications in computational biology and groups the method of data fusion into the intermediate

combination, the early combination, and the late combination methods.

The lending of money is traditionally handled by financial institutions. However, individual lenders bid on unsecured microloans sought by other individual borrowers and P2P lending is characterized by unique uncertainty and risk accordingly. In order to reduce the uncertainty arising from P2P lending, lenders should make fully use of all the market information. On the one hand, as is widely recognized, borrower credit information plays an important role in loan evaluation. On the other hand, lender information is shown to be useful for indicating loan values. Specifically, Luo et al. [61] found that investor composition is a powerful screening tool of valuable loans. However, it remains unclear how to put lender and borrower/loan information together to guide personal investors' selection of investments more effectively.

To the best of our knowledge, there is no existing work to build multi-kernel information fusion loan evaluation models, which combine both borrower credit information and investor composition information to quantitatively evaluate a loan for making better investment decisions in P2P lending. In this study, we use multiple kernels instead of selecting one specific kernel function and its corresponding parameters to propose a multi-kernel loan evaluation model for investment decisions in P2P lending.

3. OPERATIONALIZATION OF TWO INFORMATION SOURCES IN P2P LENDING

In this section, we explain the operationalization of two information sources in P2P lending. We first introduce two information sources in P2P lending which are fused to help investors enhance their investment decisions. To be specific, we take borrower credit information and investor composition information into consideration and then we describe the operationalization of the two sources information respectively before information fusion

3.1 Two information sources in P2P network

Different from traditional banks, individuals are allowed to lend to and borrow from each other directly through P2P lending online platform. In a P2P lending marketplace, the borrower applies for a loan, called listing, and other people are allowed to bid on it. A listing automatically becomes a loan when certain criteria are satisfied. Typically, a lender would like to spread his money across many different loans to reduce risk through diversification, and as a result, a loan is usually funded by many lenders.
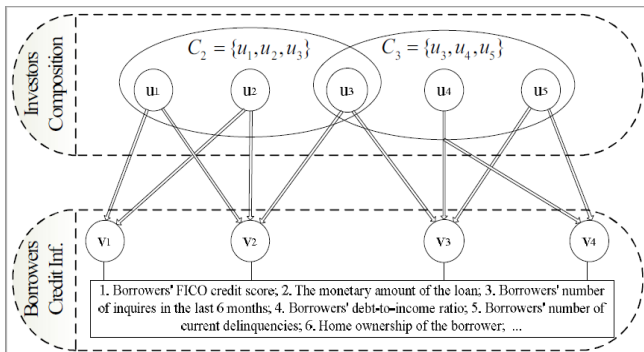

Figure 1. Information sources in P2P network

Considering investors, loans, and their relation (i.e. the actual amount invested), the P2P lending marketplace can be modeled as a bipartite investment network that represents a typical many-to-many investment relationship. The information sources in P2P investment networks can be grouped into two categories as shown in Figure 1: the borrower's credit information of loans and the investment information of investor.

The borrower's credit information of loans has been considered as the most important information source to assess the risk of loan. Tremendous efforts have been made to establish credit records for loan evaluation in traditional bank industries. To help lenders reduce risk, Prosper, the most advanced P2P lending platform, provides the rating of each loan based on the borrower's credit information such as loan amount, borrower's credit and so on. This process is similar to traditional commercial banks evaluating the risk levels of businesses. Therefore, loans are divided into seven groups, each group with the same risk. We call such a model rating-based model (R-model).

However, the information from investors remains under-explored for improving investment decisions in P2P lending. For many-to-many investment relationships in the P2P lending market, different investors can have different investment preferences, and different loans could be funded by different types of investors. We define the group of investors who invested on the same loan as the Investor Composition of the loan. In Figure 1, the investor compositions $C_2 = \{u_1, u_2, u_3\}$ and $C_3 = \{u_3, u_4, u_5\}$ correspond to two different loans v_2 and v_3. It is obvious that each investor composition contains a different combination of investors, who may have different investment histories. The group of investors who invest in this loan may exhibit certain characteristics, some of which can be quantified by various statistics. Therefore, it is promising to exploit investor composition for improving investment decisions in P2P lending.

3.2 Operationalization of lender information

Actually, composition analysis has been widely used in finance and economics, Luo et al. developed an investor composition analysis model based on the past investment behavior of an investor. For a P2P lending platform which has m investors $U = \{u_1, u_2, u_3, \dots, u_m\}$, n investees and $V = \{v_1, v_2, v_3, \dots, v_m\}$, $E = \{e_{11}, e_{12}, \dots, e_{ij}, \dots, e_{mn}\}$ are the edges connecting them, the composition of investor performance CR_j is calculated as follows:

$$CR_j = \sum_{i=1}^{m} \lambda_{ij} \overline{R}_i \tag{1}$$

where, $\lambda_{ij} = \frac{e_{ij}}{\sum_{i=1}^{m} e_{ij}}$ is the investees' investment weight, $\overline{R}_i = \sum_{j=1}^{n} w_{ij} R_j$ is the past investment performance of an investor and $w_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}}$ is the investors' investment weight.

The composition of investor risk preference CP_j is given as follows:

$$CP_j = \sqrt{\sum_{i=1}^{m} (\lambda_{ij}^2 P_i^2 + 2\sum_{k=1}^{m-i} (J_{ik} \lambda_{ik} \lambda_{i,i+k} P_i P_{i+k}))} \tag{2}$$

where, $P_i = \sqrt{\sum_{j=1}^{n} w_{ij}\left(R_j - \overline{R}_i\right)^2}$ is the investment risk preferences, $J_{ik} = \frac{D_{ik}}{D_i + D_k - D_{ik}}$ is the correlation between investors u_i and u_k.

Therefore, the investor composition score CS_j is defined as the ratio of CR_j and CP_j:

$$CS_j = \frac{CR_j}{CP_j} \qquad (3)$$

The score of investor composition CS_j is a comprehensive investor composition index that describes the composition of investor performance CR_j and risk preferences CP_j.

3.3 Operationalization of borrower information

The investors make their choices by capturing the risk of default in P2P lending. Individual investors bear the credit risk and suffer a severe problem of information asymmetry. On the one hand, borrowers are better informed than lenders of their ability and willingness to repay. On the other hand, lenders, unlike financial institutions, are not expert in dealing with risk assessment and have inadequate historical information to predict borrowers' behavior.

To reduce the information asymmetry suffered by lenders, transaction platform provides information on borrower's characteristics and loan's details. Ravina [62] has studied the effect of personal characteristics in P2P lending sites, finding that beauty, race, age, and other personal characteristics are taken into account by lenders. Pope and sydnor [63] found evidence of significant racial disparities and Gonzalez and Loureiro [64] studied the effect of photographs in lending. Serrano et al. [65] concluded that loan characteristics (such as loan purpose and loan amount), borrower characteristics (such as current housing situation, annual income), Credit history (a record of a consumer's ability to repay debts), and personal indebtedness is related to the probability of default in P2P lending. Some influential indices are listed in Table 1.

Table 1. Index and description

Index	Description
Amount Funded	The sum of bid amounts or requested amount if fully funded
Amount Requested	The amount requested in the listing
Bid Count	The number of bids on this listing
Borrower City	The home city of the borrower
Borrower Starting Rate	The starting rate of the listing
Credit Grade	The credit grade of the borrower AA-HR
Income	The annual income range of the borrower
Debt-to-Income Ratio	The debt-to-income ratio of the borrower
Is Borrower Homeowner	Specifies if the borrower is a verified homeowner
Amount Delinquent	The amount delinquent at the time the listing was created
Inquiries Last 6 Months	The number of inquiries in the last 6 months
Delinquencies Last 7 Years	The number of delinquencies in the last 7 years
Employment Status	The employment status of the borrower

There are several statistical techniques for default prediction, such as discriminant analysis, logistic regression, neural networks or classification trees, among others. These methods can be divided into two categories: statistical and judgmental [66]. The judgmental approach assigns 0 to defaulted loans and 1 to non-defaulted loans. While statistical approach provides the probability of default. The logistic regression is a well-established technique employed in evaluating the probability of occurrence of a default [67]. A default likelihood prediction model P transforms attributes describing a loan into a default likelihood score:

$$p_i = P\left(X_i\right) \qquad (4)$$

where, X_i is the observation of loan i's attributes, which is another source of information in P2P lending described in Figure 1. The default likelihood prediction model P can be any prediction model. Consider a logistic regression model as follows:

$$p_i = \frac{1}{1 + e^{-\left(\beta'X_i + \beta_0\right)}} \qquad (5)$$

where, $\beta = \{\beta_1, \beta_2, \ldots, \beta_d\}$ and β_0 are the coefficients to be estimated in the model. For support vector machines, P can be the distance to the margin.

Overall, we could predict a loan default likelihood through logistic regression, which is simple but too coarse to meet the accuracy needs of loan evaluation in P2P lending. Therefore, we propose a novel idea to combine borrower and investor information together to enhance the investment decision making in P2P lending.

4. MULTI-KERNEL BASED LOAN EVALUATION MODEL FOR INVESTMENT DECISION

In this section, we describe the multi-kernel-based loan evaluation model which fuses the borrower credit information and investor composition information. First, we construct the framework of two sources information fusion. And then we explain our model in detail, which could accurately assess the credit risk and expected return of each individual loan for P2P lending. Finally, the risk and return of each loan are served as input for portfolio optimization.

4.1 Framework of information fusion in P2P lending

As shown in Figure 1, both borrower's credit information and investor composition information are the important information sources in a P2P network for loan evaluation. However, a key problem is how to integrate the two types of information to develop a unified loan evaluation model for

decision making in P2P lending.

In this subsection, we introduce the statistical framework of multi-information fusion in Figure 2: a multi-kernel information fusion model (MKL). This method is applied to kernel method which is a non-parametric estimation method. Kernel function could make the raw data to get a better expression by mapping the raw features into new feature space and, consequently, improve the classification accuracy or prediction accuracy significantly. More worth mentioning is that multi-kernel information fusion model is referred to as intermediate integration where a heterogeneous kernel is used in which the kernel values for each type of data are pre-computed separately, and the resulting values are added together.
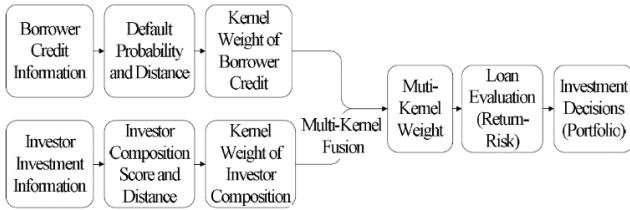

Figure 2. Framework of multi-information fusion

There can be two advantages of MKL in our study: firstly, the two types of information come from heterogeneous sources with different characteristics. Using one specific kernel can easily become a source of bias and it is better to use a combination of different kernels which correspond to different notions of similarity. Secondly, MKL in our research is a kind of intermediate combination method which trades off making too many independence assumptions versus allowing too many dependencies. Bayesian method, as a mathematical tool, is also widely used in information fusion technique. However, in this method, sources are considered as independent and it is difficult to identify the probability distribution.

The schematic diagram of multi-kernel information fusion model is shown in Figure 2. In this model, the borrower's information and the investor composition information are processed separately and converted to two kernel weights based on two information sources. Then, two correlation coefficients between the two information sources and the historical performance of the loans are computed. Furthermore, the model integrates two kernel weights and two correlation coefficients into a multi-kernel weight to build a unified loan evaluation model.

4.2 Introduce kernel regression into P2P lending

Kernel regression is a statistical technique used to estimate the conditional expectation of a random variable. In our research, kernel regression can extract weights of past loans and thus make a joint assessment of return and risk of each loan, which can meet the characteristics of P2P lending: crowdfunding, small amount and high risk.

There exist large amounts of loans which are completed and kernel weight can evaluate the similarity between the objective loan and these historical loans in defined feature space. Kernel weight usually takes a general form: $\frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right)}$, the x_i and x represent the feature vector of historical loan and

objective loan respectively, we will describe how to identify the kernel weight in detail in 4.3.

In general, suppose that each loan is evaluated on two dimensions, X and Y, where X is the independent variable and Y is the dependent variable. The observations of n neighboring instances are $\{(x_i, y_i)|i = 1,2,...,n\}$. Then the estimation of the outcome y for the loan under study is:

$$y = \hat{f}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h})y_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})} \qquad (6)$$

here, $K(\cdot)$ is a kernel function that is usually considered as a Gaussian function, thus, assigning more weights to observations closer to x and less weights to those farther away. The h>0 is a bandwidth parameter that determines the proportion of local and remote information used in the sum. And the optimization method of h is consistent [68].

4.3 Weight identification in P2P lending loan evaluation

In this subsection, we introduce a multi-kernel loan evaluation model that combines both the borrower's credit information and investor composition score using the multi-kernel method.

First, in order to quantify the distances among loan instances, the default likelihood distance between loans i and j is defined as:

$$db_{ij} = |p_i - p_j| \qquad (7)$$

where, p_i and p_j are the probabilities of default for loans i and j, computed by Equation 5 respectively. Then, we convert the distance into wb_{ij} computed by Equation 8 as the borrower weight:

$$wb_{ij} = \frac{K(\frac{db_{ij}}{h})}{\sum_{i=1}^{n} K(\frac{db_{ij}}{h})} \qquad (8)$$

similarly, for given loans' investor composition scores CS_i and CS_j, we define the *investor composition distance dc_{ij}* as:

$$dc_{ij} = |CS_i - CS_j| \qquad (9)$$

where, CS_i and CS_j are defined in section 3.2. Then, we convert the composition information scores to investor composition distance dc_{ij} in Equation 9 and the investor weight wc_{ij} based on the kernel regression method as follows:

$$wc_{ij} = \frac{K(\frac{dc_{ij}}{hc})}{\sum_{i=1}^{n} K(\frac{dc_{ij}}{hc})} \qquad (10)$$

where, $K(\cdot)$ is a kernel function and hc the investor band-width. The investor bandwidth hc is chosen to minimize $CV(hb)$ in the following:

$$CV(hb) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}_{-i}(x_i) - y_i \right)^2 \qquad (11)$$

where, $\hat{f}_{-i}(x_i)$ is the leave-one-out estimation of $f(x_i)$.

Furthermore, two correlation coefficients (ρ_b and ρ_c) between the two information sources and the historical performance of loans are respectively computed. we define the correlation coefficient between the investor composition score CS and the historical performance R of loans as ρ_b in Equation (12):

$$\rho_b = \frac{\sum_{i=1}^{n} (p_i - \bar{p}_i)(R_i - \bar{R}_i)}{\sqrt{\sum_{i=1}^{n} (p_i - \bar{p}_i)^2 (R_i - \bar{R}_i)^2}} \qquad (12)$$

similarly, we define the historical performance R as ρ_c in Equation (13):

$$\rho_c = \frac{\sum_{i=1}^{n} (CS_i - \overline{CS}_i)(R_i - \bar{R}_i)}{\sqrt{\sum_{i=1}^{n} (CS_i - \overline{CS}_i)^2 (R_i - \bar{R}_i)^2}} \qquad (13)$$

The above two correlation coefficients are referred to as Pearson's Product Moment Correlation Coefficient, which measures the degree or strength of linear dependence relation between two variables [69]. In our case, the two coefficients present the importance of the two information sources, that two information sources will be given different weights when we integrate two kernel weights into a multi-kernel weight.

Next, we combine two correlation coefficients and two kernel weights into a weighted and standardized unified weight wm_{ij} for loan evaluation. We call the unified weights wm_{ij} as multi-kernel weights, as shown in Equation (14):

$$wm_{ij} = \frac{\rho_b * wb_{ij} + \rho_c * wc_{ij}}{\rho_b * \sum_{j=1}^{n} wb_{ij} + \rho_c * \sum_{j=1}^{n} wc_{ij}} \qquad (14)$$

finally, for a given loan L_i, based on the multi-kernel weight wm_{ij} and n past loans' observed return rate R_j, we can directly predict its return rate μm_i using the weighted average of the past loans performance in Equation (15):

$$\mu m_i = \sum_{j=1}^{n} wm_{ij} R_j \qquad (15)$$

similarly, we predict the risk σm_i of loan L_i using the weighted variance, shown in Equation (16):

$$\sigma m_i = \sqrt{\sum_{j=1}^{n} wm_{ij} \left(R_j - \mu m_i \right)^2} \qquad (16)$$

Overall, we combine both the borrower's credit information and investor composition score to develop a unified multi-kernel loan evaluation model. This model can quantitatively assess the return and risk of each loan, which will be used as two important input parameters of the investment decision model based on portfolio theory.

4.4 Investment decision algorithm based on portfolio theory

The investment decisions in P2P lending could be formulated as a constrained portfolio problem. In this subsection, we combine the multi-kernel loan evaluation model and portfolio selection to develop a unified investment decision model to help investors make better investment decisions. The investment decision process is described Figure 3.
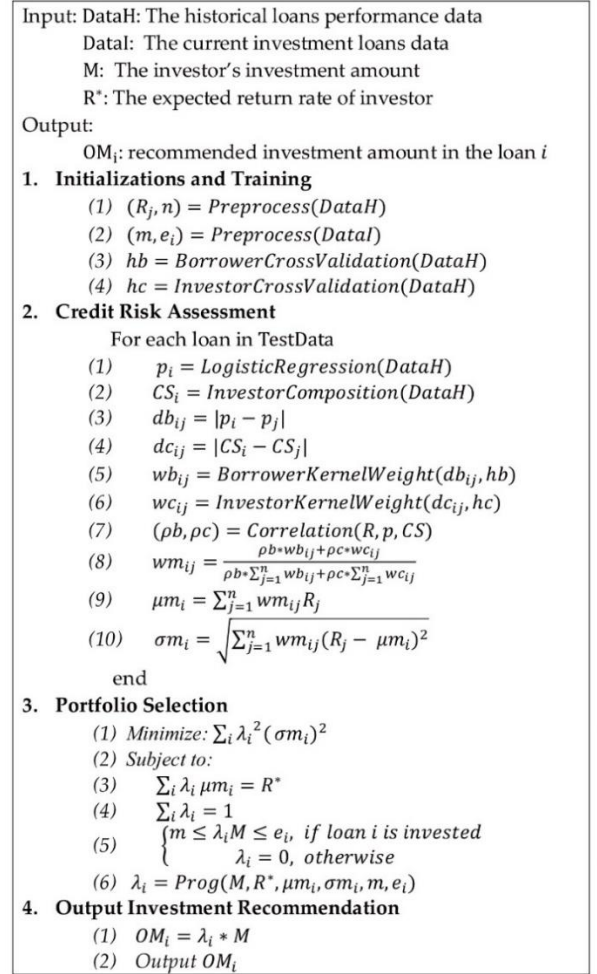


```
Input: DataH: The historical loans performance data
       DataI: The current investment loans data
       M: The investor's investment amount
       R*: The expected return rate of investor
Output:
       OM_i: recommended investment amount in the loan i
1.  Initializations and Training
    (1)  (R_j, n) = Preprocess(DataH)
    (2)  (m, e_i) = Preprocess(DataI)
    (3)  hb = BorrowerCrossValidation(DataH)
    (4)  hc = InvestorCrossValidation(DataH)
2.  Credit Risk Assessment
       For each loan in TestData
    (1)   p_i = LogisticRegression(DataH)
    (2)   CS_i = InvestorComposition(DataH)
    (3)   db_ij = |p_i - p_j|
    (4)   dc_ij = |CS_i - CS_j|
    (5)   wb_ij = BorrowerKernelWeight(db_ij, hb)
    (6)   wc_ij = InvestorKernelWeight(dc_ij, hc)
    (7)   (pb, pc) = Correlation(R, p, CS)
    (8)   wm_ij = (pb*wb_ij + pc*wc_ij) / (pb*Σ_{j=1}^{n} wb_ij + pc*Σ_{j=1}^{n} wc_ij)
    (9)   μm_i = Σ_{j=1}^{n} wm_ij R_j
    (10)  σm_i = √(Σ_{j=1}^{n} wm_ij (R_j - μm_i)^2)
       end
3.  Portfolio Selection
    (1)  Minimize: Σ_i λ_i^2 (σm_i)^2
    (2)  Subject to:
    (3)   Σ_i λ_i μm_i = R*
    (4)   Σ_i λ_i = 1
    (5)   { m ≤ λ_i M ≤ e_i, if loan i is invested
          {        λ_i = 0, otherwise
    (6)  λ_i = Prog(M, R*, μm_i, σm_i, m, e_i)
4.  Output Investment Recommendation
    (1)  OM_i = λ_i * M
    (2)  Output OM_i
```

Figure 3. Investment decision model

The model has two input data sets. *DataH* is a data set for historical loans that includes borrowers' credit information, investor investment information, loan amount requested and performance information. *DataI* is the current investment loans dataset that includes borrowers' credit information, investor investment information, and loan amount requested. The two input parameters regarding investors in this model are the investment amount M and investors' expected return rate R^*. The parameters are different from one investor to another, and therefore, the investors' expected return rate R^* can be considered a personal risk preference in portfolio selection. The process of investment decision involves loan evaluation in *DataI* based on the learning from loans in *DataH* and make investment decisions in loans in *DataI* for the given preference information of investors M and R^*.

During the initialization and training stages, we first gain the basic information from *DataH* and *DataI*, where R_j is the return on each historical loan, n is the number of historical

loans, e_i is the amount of each loan application and m is the minimum investment constraint in P2P lending market. Furthermore, we adopt cross-validation to calculate the optimized bandwidth hb and hc, which are the important parameters to compute kernel weights of borrower credit risk and investor composition in the following loan evaluation process.

During the loan evaluation, firstly, we adopt the logistic regression based on the borrower's credit information to predict the default probability of P_i. We then compute the investor composition score CS_i of each loan using the investor composition analysis model based on investor behavior information in *DataI*. Next, for each given loan L_i in *DataI*, we compute the default likelihood distances db_{ij} and dc_{ij} between L_i and each loan L_j in *DataH*. Then, we compute the borrower kernel weights wb_{ij} and the investor composition kernel weight wc_{ij} based on db_{ij}, dc_{ij} and the optimized bandwidth hb and hc using Equations (8) and (10). Furthermore, we compute multi-kernel weights wm_{ij} using Equation (14) based on wb_{ij}, wc_{ij}, correlation coefficients ρ_b and ρ_c using Equations (12) and (13). Finally, we evaluate the return μ_i and the risk σ_i based on the multi-kernel weights wm_{ij} and the performance information R_j of each historical loan in *DataH* using Equations (15) and (16).

In the portfolio selection and investment recommendation process, we input μ_i, σ_i, M, R^*, e_i, and m into the portfolio selection model and get the investment proportion λ_i in each loan. The output $M * \lambda_i$ is the investment amount recommendation.

5. EXPERIMENTS

In this section, we first introduce the real-world P2P lending dataset that was used in our experiment. Then, to validate the proposed model in this paper, extensive experiments will be performed on real-world datasets from the state-of-the-art P2P lending marketplace, by comparing our multi-kernel loan evaluation model to three baseline models with respect to investment performances and robustness.

5.1 Experiment data

The development of P2P lending market provides a large number of real P2P lending transaction data. Our experiments use data sets from the most advanced P2P lending platform, Prosper.com, with more than one million members and over 2.5 billion U.S. dollars in funded loans. Our dataset provided by Prosper.com consists of 4128 loans that span the time range from September 2007 to March 2008.

We group these variables into three categories and describe their meanings in Table 2. Variables X_1-X_6 are attributes about borrowers that are related to a borrower's credit profile, such as FICO score and debt-to-income ratio. Variable Y is the investor composition score. The two prediction variables are $\mu_i/\mu m_i$ and $\sigma_i/\sigma m_i$, which are computed in the loan evaluation models (μ_i and σ_i are the outputs of IOM, μm_i and σm_i are the outputs of MKM) and used as inputs in the investment decision model.

For our model, we validate the validity of the model using k-fold cross-validation. Specifically, we divide the Prosper dataset into *24* loan subsets, such that there are 172 loans in each subset.

Table 2. Category and description of variables

Abbr.	Category	Description
$X1$	Borrower	FICO credit score
$X2$	Borrower	The amount of the loan
$X3$	Borrower	The number of inquires
$X4$	Borrower	Debt-to-income ratio
$X5$	Borrower	The number of delinquencies
$X6$	Borrower	Home ownership
Y	Investor	Investor composition score
$\mu i/\mu mi$	Prediction	The return rate of the loans
$\sigma i/\sigma mi$	Prediction	The risk of the loans

5.2 Baseline models

In order to show the effectiveness of the multi-kernel loan evaluation model, we compare it to the other three baseline models with the above mentioned real-world data. The following is a detailed description of these models.

R-model is a rate-based model. According to the borrower's credit information, loans are divided into seven group, and each group has the same return and risk.

I-model is a kernel-based single source model that uses only borrower credit information. The performance of each loan is assessed on the basis of similarities with historical loans.

MA-model is a multi-attribute information fusion model, where all the loans are assessed using both borrower's credit information and investor composition information. The two information sources are combined at the step of probability prediction as the attributes of logistics regression. The detail of MAM is shown in the Figure 4.
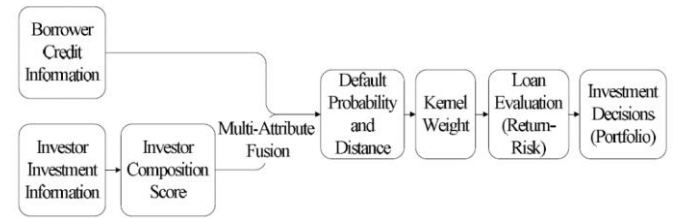

Figure 4. Method of MA-model

MK-model is a multi-kernel information fusion loan evaluation model, where the borrower's information and the investor composition information are proceeded separately and converted into kernel weights based on and correlation coefficients. The model combines the kernel weights based correlation coefficients of the information to a unified multi-kernel weight to evaluate each loan.

Once the original set of loans are randomly partitioned into k subsets, one subset is used as the testing set and the other as the training set. These steps for model comparison are as follows:

(1) Use the training model to train each model and to predict the return and risk of each loan in the test set.

(2) In each model, the portfolio algorithm is used to calculate the expected return based on the risk and return of the test set loan calculated in the first step.

(3) Compare the Sharpe ratio and return of the four models.

5.3 Experiment results

When carrying out the portfolio algorithm, we set the investment amount M=15000, the target rate of return R^* =

0.06 , and the risk-free rate of return is 0.025, which is equivalent to the average return from T-Bills at the same period of time.

Table 3 shows the investment return rate on each test subset, and the average investment return rate and Sharpe ratio for the Prosper dataset. It is obvious that the MK-model performs better than either the R-model, I-model or MA-model on both average investment return rate and Sharpe ratio.

Table 3. Rate of return from the optimal portfolio on the prosper dataset

Subset No.	R-moedl	I-model	MA-model	MK-model
1	0.0783	0.0787	0.0791	0.0785
2	0.0840	0.0761	0.0728	0.0786
3	0.0779	0.0820	0.0599	0.0796
4	0.0690	0.0765	0.0766	0.0780
5	0.0478	0.0513	0.0453	0.0811
6	0.0498	0.0727	0.0616	0.0862
7	0.0676	0.0761	0.0695	0.0826
8	0.0849	0.0752	0.0759	0.0810
9	0.0775	0.0756	0.0785	0.0836
10	-0.0443	0.0147	0.0300	0.0812
11	0.0878	0.0768	0.0835	0.0810
12	0.0878	0.0765	0.0847	0.0816
13	0.0373	0.0780	0.0747	0.0901
14	-0.0034	0.0701	0.0726	0.0837
15	0.0622	0.0710	0.0765	0.0753
16	0.0482	0.0549	0.0443	0.0493
17	0.0672	0.0798	0.0692	0.0800
18	0.0590	0.0717	0.0654	0.0626
19	0.0119	0.0447	0.0652	0.0746
20	0.0654	0.0738	0.0615	0.0655
21	0.0163	0.0424	0.0579	0.0682
22	0.0496	0.0708	0.0322	0.0531
23	0.0398	0.0310	0.0284	0.0380
24	0.0760	0.0207	0.0546	0.0486
Average	0.0541	0.0642	0.0633	0.0734
Std. Dev.	0.0322	0.0196	0.0166	0.0136
Sharpe Ratio	0.9019	1.9976	2.3080	3.5578

Finally, we find that compared with the other three models, MK-model performs better in improving investors' investment decisions. Specifically, each investor may have different investment amount and expectations on the return rate. We perform the portfolio algorithm based on Prosper data set using different input parameters. As the Table 4 shows, we consider *10* typical combinations of investment amounts M and investors' expected return rates R^*.

Table 4. Investors' choices of input parameters for portfolio selection

Group	Investment Amount (M)	Expected Return ($R*$)
1	$10,000	5.5%
2	$10,000	6.0%
3	$10,000	6.5%
4	$15,000	5.5%
5	$15,000	6.0%
6	$15,000	6.5%
7	$20,000	5.5%
8	$20,000	6.0%
9	$20,000	6.5%
10	Average Performance of 1-9	

In the case of each parameter combination, we use the portfolio algorithm to calculate the optimal allocation of funds (that is, the amount allocated to each test loan). However, the input of the portfolio algorithm (i.e., expected return and risk) is different and generated by four models. Finally, we calculate the actual return on investment and Sharpe ratio for all loan subsets, as shown in Figures 5 and 6.
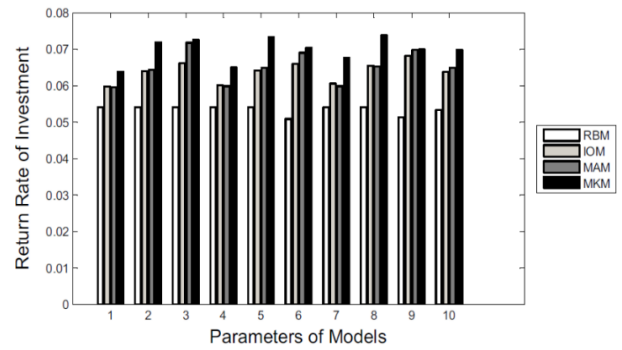

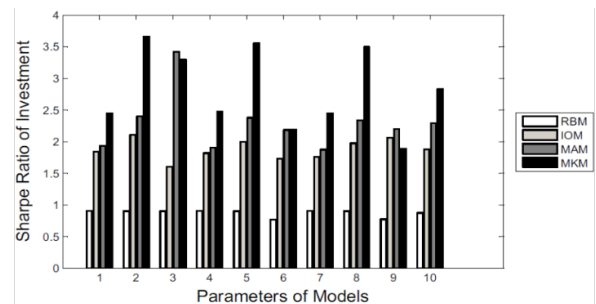**Figure 5.** Investment rates of return

Figure 6. Investment sharpe ratios

Figure 5 shows the performance comparisons of the Four models on the Prosper dataset. We can see that for nearly all parameter combinations, compared with R-model, I-model and MA-model, MK-model can generate higher average investment return and Sharpe ratio.

To sum up, our experiments indicates that the MK-model makes better investment decisions than other baseline models.

6. CONCLUSION

In this study, we propose an integrated loan evaluation model that exploits and combines both the borrower credit information and investor composition information using a multi-kernel algorithm for improving investment decisions in P2P lending. Specifically, from an investor's perspective, we exploit the investor behavior information to build an investor composition model for quantitative analysis of each loan value. From a borrower's perspective, we build a kernel-based credit risk model to quantitatively analyze the loan value based on the borrowers' credit information. We combine the two information sources from both borrowers and investors using the multi-kernel method to develop an integrated loan evaluation model. Furthermore, based on the integrated loan evaluation model, we transform investment decisions in P2P into portfolio optimization problems with boundary constraints to help investors make better investment decisions. To validate the proposed model, we perform extensive experiments on the real-world data from the world's largest P2P lending marketplace. Experimental results reveal that a multi-source loan evaluation model can effectively improve the evaluation accuracy of loan value and significantly

enhance investment performance compared to existing methods.

REFERENCES

[1] Wang, H., Greiner, M., Aronson, J. (2009). People-to-people lending: The emerging e-commerce transformation of a financial market. Value Creation in E-Business Management, 36: 182-195. https://doi.org/10.1007/978-3-642-03132-8_15

[2] Hulme, M., Wrigh, C. (2006). Internet based social lending: Past, present and future. Social Futures Observatory, 174(2): 1-115.

[3] Berger, S.C., Gleisner, F. (2009). Emergence of financial intermediaries in electronic markets: The case of online P2P lending. Business Research, 2(1): 39-65. https://ssrn.com/abstract=1568679

[4] Serrelis, E., Alexandris, N. (2010). A new paradigm for secure social lending. Next Generation Society, Technological and Legal Issues, 26: 373-385. https://doi.org/10.1007/978-3-642-11631-5_34

[5] Lin, M., Prabhala, N., Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. Management Science, 59(1): 17-35. http://dx.doi.org/10.2139/ssrn.1355679

[6] Liu, D., Brass, D.J., Lu, Y., Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. Mis Quarterly, 39(3): 729-742. https://doi.org/10.25300/MISQ/2015/39.3.11

[7] Faia, E., Paiella, M. (2010). P2P lending: Information externalities, social networks and loans' substitution. Social Science Electronic Publishing. https://ssrn.com/abstract=3028601

[8] Duarte, J., Siegel, S., Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. Review of Financial Studies, 25(8): 2455-2484. https://doi.org/10.1093/rfs/hhs071

[9] Lin, M., Viswanathan, S. (2015). Home bias in online investments: An empirical study of an online crowdfunding market. Management Science, 62(5): 1393-1414. http://dx.doi.org/10.2139/ssrn.2219546

[10] Chen, J., Zhang, Y., Yin, Z. (2012). Education premium in the online peer-to-peer lending marketplace: Evidence from the big data in China. The Singapore Economic Review, 63(1): 1-20. http://dx.doi.org/10.1142/S0217590818410023

[11] Doreitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., Kammler, J. (2016). Description-text related soft information in peer-to-peer lending-evidence from two leading European platforms. Journal of Banking & Finance, 64: 169-187. https://doi.org/10.1016/j.jbankfin.2015.11.009

[12] Wei, Z.Y., Lin, M.F. (2016). Market mechanisms in online peer-to-peer lending. Management Science, 63(12): 4236-4257. http://dx.doi.org/10.1287/mnsc.2016.2531

[13] Zhao, H., Qi, L., Zhu, H., Yong, G., Chen, E., Yan, Z., Du, J. (2017). A sequential approach to market state modeling and analysis in online P2P lending. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48(1): 21-33. https://doi.org/10.1109/TSMC.2017.2665038

[14] Redmond, U., Cunningham, P. (2013). A temporal network analysis reveals the unprofitability of arbitrage in the prosper marketplace. Expert Systems with Applications, 40(9): 3715-3721. https://doi.org/10.1016/j.eswa.2012.12.077

[15] Rigbi, O. (2013). The effects of usury laws: Evidence from the online loan market. Review of Economics and Statistics, 95(4): 1238-1248. https://doi.org/10.1162/REST_a_00310

[16] Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., Gorski, S. (2011). Peer to peer lending: the relationship between language features, trustworthiness, and persuasion success. Journal of Applied Communication Research, 39(1): 19-37. https://doi.org/10.1080/00909882.2010.536844

[17] Puro, L., Teich, J.E., Wallenius, H., Wallenius, J. (2010). Borrower decision aid for people-to-people lending. Decision Support Systems, 49(1): 52-60. https://doi.org/10.1016/j.dss.2009.12.009

[18] Wu, J., Xu, Y. (2011). A decision support system for borrowers' loan in P2P lending. Journal of Computers, 6(6): 1183-1190. http://dx.doi.org/10.4304/jcp.6.6.1183-1190

[19] Herzenstein, M., Dholakia, U.M., Andrews, R.L. (2011). Strategic herding behavior in peer-to-peer loan auctions. Journal of Interactive Marketing, 25(1): 27-36. https://doi.org/10.1016/j.intmar.2010.07.001

[20] Puro, L., Teich, J.E., Wallenius, H., Wallenius, J. (2011). Bidding strategies for real-life small loan auctions. Decision Support Systems, 51(1): 31-41. https://doi.org/10.1016/j.dss.2010.11.016

[21] Herzenstein, M., Sonenshein, S., Dholakia, U.M. (2011). Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. Journal of Marketing Research, 48(SPL): 138-149. http://dx.doi.org/10.2139/ssrn.1840668

[22] Serrano-Cinca, C., Guti_errez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. Decision Support Systems, 89: 113-122. https://doi.org/10.1016/j.dss.2016.06.014

[23] Mild, A., Waitz, M., Wockl, J. (2015). How low can you go? Overcoming the inability of lenders to set proper interest rates on unsecured peer-to-peer lending markets. Journal of Business Research, 68(6): 1291-1305. https://doi.org/10.1016/j.jbusres.2014.11.021

[24] Malekipirbazari, M., Aksakalli, V. (2015). Risk assessment in social lending via random forests. Expert Systems with Applications, 42(10): 4621-4631. https://doi.org/10.1016/j.eswa.2015.02.001

[25] Thomas, L. (2009). Consumer credit models: Pricing, profit and portfolios. Oxford University Press, USA.

[26] Rosenberg, E., Gleit, A. (1994). Quantitative methods in credit management: A survey. Operations research, 42(4):

589-613. https://doi.org/10.1287/opre.42.4.589

[27] Chatterjee, S., Barcun, S. (1970). A nonparametric approach to credit screening. Journal of the American Statistical Association, 65(329): 150-154. https://doi.org/10.1080%2F01621459.1970.10481068

[28] Feldman, D., Gross, S. (2005). Mortgage default: classification trees analysis. The Journal of Real Estate Finance and Economics, 30(4): 369-396. https://doi.org/10.1007/s11146-005-7013-7

[29] Xia, Y., Liu, C., Li, Y., Liu, N. (2017). A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. Expert Systems with Applications, 78: 225-241. https://doi.org/10.1016/j.eswa.2017.02.017

[30] Frydman, H., Kallberg, J., Kao, D. (1985). Testing the adequacy of markov chain and mover-stayer models as representations of credit behavior. Operations Research, 33(6): 1203-1214. http://dx.doi.org/10.1287/opre.33.6.1203

[31] Feng, X., Xiao, Z., Zhong, B., Dong, Y., Qiu, J. (2019). Dynamic weighted ensemble classification for credit scoring using markov chain. Applied Intelligence, 49(2): 555-568. https://doi.org/10.1007/s10489-018-1253-8

[32] Stepanova, M., Thomas, L. (2002). Survival analysis methods for personal loan data. Operations Research, 50(2): 277-289. http://dx.doi.org/10.1287/opre.50.2.277.426

[33] Bugera, V., Konno, H., Uryasev, S. (2002). Credit cards scoring with quadratic utility functions. Journal of Multi-Criteria Decision Analysis, 11(4-5): 197-211. http://dx.doi.org/10.1002/mcda.327

[34] Jensen, H.L. (1993). Using neural networks for credit scoring. Managerial Finance, 18(6): 15-26. http://dx.doi.org/10.1108/eb013696

[35] Xiao, H., Hou, Y., Cui, C. (2018). Evaluation of P2P lending borrowers' credit on BP artificial neural network. Operations Research and Management Science, 27(9): 112-118.

[36] Wang, Y., Wang, S., Lai, K. (2005). A new fuzzy support vector machine to evaluate credit risk. IEEE Transactions on Fuzzy Systems, 13(6): 820-831. https://doi.org/10.1109/TFUZZ.2005.859320

[37] Yu, L., Wang, S., Lai, K., Zhou, L. (2008). Bio-inspired credit risk analysis: Computational intelligence with support vector machines. Springer Verlag. https://doi.org/10.1007%2F978-3-540-77803-5

[38] Tian, Y., Yong, Z., Luo, J. (2018). A new approach for reject inference in credit scoring using kernel-free fuzzy quadratic surface support vector machines. Applied Soft Computing, 73: 96-105. https://doi.org/10.1016/j.asoc.2018.08.021

[39] Desai, V., Conway, D., Crook, J., Overstreet, G. (1997). Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. IMA Journal of Management Mathematics, 8(4): 323-346. https://doi.org/10.1093/imaman/8.4.323

[40] Huang, J., Tzeng, G., Ong, C. (2006). Two-stage genetic programming (2SGP) for the credit scoring model. Applied Mathematics and Computation, 174(2): 1039-1053. https://doi.org/10.1016/j.amc.2005.05.027

[41] Zhang, W., He, H., Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. Expert Systems with Applications, 121: 221-232.

[42] Malhotra, R., Malhotra, D. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research, 136(1): 190-211. https://doi.org/10.1016/S0377-2217(01)00052-2

[43] Lee, T.S., Chen, I.F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 28(4): 743-752. https://doi.org/10.1016/j.eswa.2004.12.031

[44] Hill, S., Ready-Campbell, N. (2011). Expert stock picker: The wisdom of (experts in) crowds. International Journal of Electronic Commerce, 15(3): 73-102. https://doi.org/10.2753/JEC1086-4415150304

[45] Wall, L. (1998). A European proposal for terms of reference in data fusion. International Archives of Photogrammetry & Remote Sensing, XXXII(7): 651-654.

[46] Valet, L., Mauris, G., Bolon, P. (2001). A statistical overview of recent literature in information fusion. IEEE Aerospace & Electronic Systems Magazine, 16(3): 7-14. https://doi.org/10.1109/IFIC.2000.862457

[47] Thakur, G. (2013). Deterministic bayesian information fusion and the analysis of its performance. Statistics, 3(4): 345-366. https://doi.org/10.1093/imaiai/iau009

[48] He, Y.J., Zeng, W.Q., Zeng, W.Y. (2011). Based on BP neural network multisensory optimization of data fusion technology to optimize. Microcomputer & Its Applications, 22: 1-20.

[49] Safari, S., Shabani, F., Simon, D. (2014). Multirate multisensor data fusion for linear systems using kalman filters and a neural network. Aerospace Science and Technology, 39: 465-471. https://doi.org/10.1016/j.ast.2014.06.005

[50] Rodriguez, R., Bedregal, B., Bustince, H., Dong, Y., Farhadinia, B., Kahraman, C., Martinez, L., Torra, V., Xu, Y., Xu, Z., Herrera, F. (2016). A position and perspective analysis of hesitant fuzzy sets on information fusion in decision making, towards high quality progress. Information Fusion, 29: 89-97. https://doi.org/10.1016/j.inffus.2015.11.004

[51] Manjunatha, P., Verma, A., Srividya, A. (2008). Multi-sensor data fusion in cluster based wireless sensor networks using fuzzy logic method. 2008 IEEE Region 10 and the third international Conference on Industrial and Information Systems, 1-6. https://doi.org/10.1109/ICIINFS.2008.4798453

[52] Xu, W., Yu, J. (2017). A novel approach to information fusion in multi-source datasets: A granular computing viewpoint. Information Sciences, 378: 410-423. https://doi.org/10.1016/j.ins.2016.04.009

[53] Basir, O., Yuan, X. (2007). Engine fault diagnosis based on multi-sensor information fusion using dempster-shafer evidence theory. Information Fusion, 8(4): 379-386. https://doi.org/10.1016/j.inffus.2005.07.003

[54] Lin, G., Liang, J., Qian, Y. (2015). An information fusion approach by combining multi-granulation rough sets and evidence theory. Information Sciences, 314: 184-199. https://doi.org/10.1016/j.ins.2015.03.051

[55] Nadaraya, E. (1964). On estimating regression. Theory of Probability and Its Applications, 9(1): 141-142. http://dx.doi.org/10.1137/1109020

[56] Nadaraya, E.A. (1965). On non-parametric estimates of density functions and regression curves. Teoriya Veroyatnostei i ee Primeneniya, 10(1): 199-203.

http://mi.mathnet.ru/eng/tvp458

[57] Pavlidis, P., Weston, J., Cai, J., Grundy, W.N. (2001). Gene functional classification from heterogeneous data. Proceedings of the Fifth Annual International Conference on Computational Biology, ACM, pp. 249-255. http://dx.doi.org/10.1145/369133.369228

[58] Ckriet, G., Deng, M., Cristianini, N., Noble, W. (2003). Kernel-based data fusion and its application to protein function prediction in yeast. Biocomputing, 300-311. https://doi.org/10.1142/9789812704856_0029

[59] Gonen, M., Alpaydin, E. (2011). Multiple kernel learning algorithms. Journal of Machine Learning Research, 12: 2211-2268.

[60] Schölkopf, B., Tsuda, K., Vert, J. (2004). Support vector machine applications in computational biology. Kernel Methods in Computational Biology, 71-92.

[61] Luo, C., Xiong, H., Zhou, W., Guo, Y., Deng, G. (2004). Enhancing investment decisions in P2P lending: An investor composition perspective. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 292-300. https://doi.org/10.1145/2020408.2020458

[62] Ravina, E. (2008). Love & loans: The effect of beauty and personal characteristics in credit markets. Social Science Electronic Publishing. http://dx.doi.org/10.2139/ssrn.1107307

[63] Pope, D.G., Sydnor, J.R. (2011). What's in a picture? Evidence of discrimination from prosper.com. Journal of Human Resources, 46(1): 53-92.

http://dx.doi.org/10.1353/jhr.2011.0025

[64] Gonzalez, L., Loureiro, Y.K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. Journal of Behavioral and Experimental Finance, 2: 44-58. https://doi.org/10.1016/j.jbef.2014.04.002

[65] Serrano-Cinca, C., Gutierrez-Nieto, B., Lopez-Palacios, L. (2015). Determinants of default in P2P lending. PloS one, 10(10): e0139427. https://dx.doi.org/10.1371%2Fjournal.pone.0139427

[66] Berger, A.N., Black, L.K. (2011). Bank size, lending technologies, and small business finance. Journal of Banking & Finance, 35(3): 724-735. https://doi.org/10.1016/j.jbankfin.2010.09.004

[67] Crone, S.F., Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. International Journal of Forecasting, 28(1): 224-238.
https://doi.org/10.1016/j.ijforecast.2011.07.006

[68] Guo, Y., Zhou, W., Luo, C., Liu, C., Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. European Journal of Operational Research, 249(2): 417-426. https://doi.org/10.1016/j.ejor.2015.05.050

[69] Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. Journal of Diagnostic Medical Sonography, 6(1): 35-39. https://doi.org/10.1177%2F875647939000600106