Enhancing Investment Decisions in P2P Lending: An Investor Composition Perspective

Chunyu Luo¹², Hui Xiong¹, Wenjun Zhou³, Yanhong Guo¹, Guishi Deng¹

¹Faculty of Management and Economics, Dalian University of Technology, China
E-mail: luochy@gmail.com, guoyh@dlut.edu.cn, denggs@dlut.edu.cn
²Management Science and Information Systems Department, Rutgers University, USA
E-mail: hxiong@rutgers.edu
³Department of Statistics, Operations and Management Science, University of Tennessee, USA
E-mail: wzhou7@gmail.com

ABSTRACT

P2P lending, as a novel economic lending model, has imposed new challenges about how to make effective investment decisions. Indeed, a key challenge along this line is how to align the right information with the right people. For a long time, people have made tremendous efforts in establishing credit records for the borrowers. However, information from investors is still under-explored for improving investment decisions in P2P lending. To that end, we propose a data driven investment decision-making framework, which exploits the investor composition of each investment for enhancing decisions making in P2P lending. Specifically, we first build investor profiles based on quantitative analysis of past performances, risk preferences, and investment experiences of investors. Then, based on investor profiles, we develop an investor composition analysis model, which can be used to select valuable investments and improve the 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 investor composition can help evaluate the profit potential of an investment and the decision model based on investor composition can help investors make better investment decisions.

Categories and Subject Descriptors
H.2.8 [Database Management]: Database Applications—Data Mining

General Terms
Algorithms, Design, Experimentation

Keywords
P2P Lending, Investor Profile, Investor Composition

1. INTRODUCTION

As an emerging market for investment, P2P lending has imposed new challenges in decisions making. While tremendous efforts have been made to establish credit records for borrowers, the information from investors remains under-explored for improving investment decisions in P2P lending. Indeed, in the field of behavioral finance [29, 2], it has been found that investors make different investment decisions due to differences in knowledge structures, information sources, cognitive development, and behavioral preferences. For many-to-many investment relationships, such as stock market and P2P lending market, different investors can have different investment preferences and different investees (i.e., stocks, P2P loans, etc.) could be funded by different types of investors. In other words, the investor composition of an investee—the group of investors who invest on this investee—can 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.

As a matter of fact, composition analysis, or structure analysis, has been an important and commonly used method in finance and economics. For example, the portfolio composition [23] evaluates a portfolio of assets, each having different risks and returns, to reduce risk for a given level of return. Similarly, the capital structure [24] has been considered as an important indicator of a firm’s financial position. In ad-
diction, the board composition [19] has been considered as one of the important factors on a firm’s overall performance. However, since P2P lending is an emerging investment field, the research methods and types of data being collected are still under development. As a result, the problem of how to exploit investor composition for enhancing investment decisions is still under-explored.

The P2P lending marketplace can be modeled as a bipartite graph with investors and loans on either side. This graph represents many-to-many investment relationships between investors and loans. Therefore, each loan may be funded by multiple lenders, and each lender invests on many different loans. In this paper, we exploit such relationships for analyzing investor composition, based on which we build a data-driven investment decision-making framework for enhancing investment decisions in P2P lending.

Furthermore, we perform the quantitative analysis of the past performances, risk preferences, and investment experiences of investors and build investor profiles including various segments of investors with different characteristics. Then, based on investor profiles, we develop an investor composition analysis model which can be used to estimate the profit potential of each investment. Finally, we demonstrate the effectiveness of the proposed investment composition analysis through extensive experiments on a real-world P2P lending dataset. Experiment results reveal that the investor composition can effectively indicate investment values, and the investor composition model can improve the profit potential of investment in P2P lending.

Overview. The rest of this paper is organized as follows. An overview of related work is provided in Section 2. We define the bipartite investment network and analyze investor profiles in Section 3. Section 4 provides an investor composition analysis model to quantitatively analyze investor composition. In Section 5, an investor composition driven investment decisions model is proposed. Section 6 presents experiments with real-world data for validating the effectiveness of the proposed investment decision model. Finally, we conclude the work in Section 7.

2. RELATED WORK

P2P lending, as a novel economic lending model, has been studied extensively in the literature in recent years. For instance, Hulme and Wright [15] provided an in-depth treatment of the social lending subject from multiple perspectives. Freedman and Jin [10] have investigated whether social networks solve the information asymmetry problem in peer-to-peer lending. They found that the estimated return of group loans is lower than those of non-group loans due to lender’s learning and the elimination of group leader rewards. Wang et al. [33] provided an overview of the concept of people-to-people lending, a relatively new e-commerce phenomenon that has the potential to radically change the structure of the loan segment of the financial industry. Laver and Jeffrey [25] developed a borrower decision aid system, which helps the borrowers quantify their strategic options, such as starting interest rate, and the amount of loan to request. The role of financial intermediaries on the P2P online market was analyzed by Berger and Gleichner [3], which demonstrates that the recommendation of a borrower significantly enhances credit conditions, and the intermediary’s bid on a credit listing has a crucial impact on the resulting interest rate. Benjamin and Robert [6] build a theoretical framework for the evaluation and design of community reputation systems. Katherine and Sergio [17] examined the behavior of lenders and find that, while there exists high variance in risk-taking between individuals, many transactions represent sub-optimal decisions on the part of lenders. Garman et al. [12] made the case that lenders’ search for listings to bid on is an integral part of the lenders’ behavior. Ryan et al. [28] analyzed the correlation between variables and regression models using real-world data.

Traditional loan evaluation techniques also relate to P2P lending when we study from the investor’s perspective. One goal of these techniques is to distinguish good loan applications from bad applications. These techniques can be roughly categorized into five groups: statistical models, operational research methods, artificial intelligence techniques, hybrid approaches, and ensemble models [8]. For instance, statistical models include logistic regression [35, 32], linear discriminate analysis [26], k-nearest neighbor (KNN) [5], classification tree [9], Markov chain [11] and survival analysis [31]. Operations research methods include linear programming [22] and quadratic programming [4]. Artificial intelligence techniques include neural networks [16], Support Vector Machines (SVMs), Genetic algorithms [7] and genetic programming [14]. Hybrid approaches include fuzzy systems and neural networks [21], fuzzy systems and support vector machines [34] and neural networks and multivariate adaptive regression spines [18]. For ensemble models, the neural network ensemble is a typical example [36].

As a novel financial model that involves diverse elements, P2P lending renders lots of research opportunities and challenges that are currently under-explored. Specially, we view P2P lending as a many-to-many investment network, and study investment decisions from the investor’s point of view. In this paper, we exploit investor composition as an investment decisions aid in P2P lending.

3. INVESTOR PROFILES

In this section, we define the bipartite investment network and analyze investor profiles.

3.1 Investment Network

Many modern investment relations can be modeled as many-to-many relation. For example, in the stock market, most investors hold many different stocks and each stock is owned by many investors. In P2P lending, investors diversify their investment by allocating money on many different loans, and a loan can gain investments from many different investors. In this section, we use a bipartite network to describe the many-to-many investment relationship between investors and investees.

Bipartite network is widely used to model the relationship among two types of entities [13, 30, 20]. In P2P lending, we consider the following two different types of entities. **Investors** are the people who provide capital to selective loans, also known as “lenders”. **Investees** are the loans that need to be financed. In P2P lending, a borrower may submit multiple loan requests. Instead of considering borrowers for credit risk analysis, we consider each individual loan as a separate Investee, since each loan has a different set of investors available for composition analysis.

In the following, we describe the construction of the bipartite investment network with basic notations. Suppose there are m investors \( U = \{u_1, u_2, u_3, ..., u_m\} \) and n in-
vestees \( V = \{v_1, v_2, v_3, ..., v_n\} \). We can build a bipartite investment network \( G = \{U, V, E\} \), where \( U \) and \( V \) are vertices on the two sides, and \( E = \{e_{11}, e_{12}, ..., e_{ij}, ..., e_{mn}\} \) are the edges connecting them. Each edge, \( e_{ij} \), represents the amount investor \( u_i \) has lent to investee \( v_j \). Specially, \( e_{ij} = 0 \) if investor \( u_i \) has never lent to investee \( v_j \).

We define investors’ investment weight \( w_{ij} \) as the ratio of \( u_i \)'s investment on \( v_j \) to the total amount of \( u_i \)'s all investments in the investment network \( G = \{U, V, E\} \), as written in Equation (1):

\[
w_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}}
\]

Similarly, we also define investees’ investment weight \( \lambda_{ij} \) as the ratio of \( v_j \)'s investment in \( u_i \) to the total amount which \( v_j \)'s gain from all investors in the investment network \( G = \{U, V, E\} \), as written in Equation (2):

\[
\lambda_{ij} = \frac{e_{ij}}{\sum_{i=1}^{n} e_{ij}}
\]

An illustration of such bipartite graph and investment weights is shown in Figure 1 below.

![Figure 1: An Example Bipartite Investment Network.](image)

In P2P lending, based on investor’s investment data and investee’s performance data, we can construct a whole bipartite investment network that provides the basis to quantitatively analyze investor profile and investor composition.

### 3.2 Investor Profile Analysis

In order to analyze investor composition, we should first understand the investors’ behavior. In this section, we study how to quantitatively analyze investors’ past performance, risk preference, and investment experience in order to build investor profiles.

Based on past investment behavior of an investor, it is possible to build a profile to represent his/her proficiency in making successful investments. We consider an investor’s profile from three aspects: past performance, risk preference, and investment experience. In this section, we describe the computation formulas for synthesizing each of the three aspects into a numerical measure.

An investor in the P2P lending marketplace typically has multiple investments, each of which may have a different rate of return. Let \( R_j \) represent the rate of return of loan \( v_j \). Since loans can be paid back in full, or default with partial or no payment, the rate of return \( R_j \) can be negative. The past investment performance \( \bar{R}_i \) of investor \( u_i \) is measured by a weighted average of investment returns, as written in Equation (3):

\[
\bar{R}_i = \frac{\sum_{j=1}^{n} w_{ij} R_j}{\sum_{j=1}^{n} w_{ij}}
\]

where \( w_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}} \) is the ratio of \( u_i \)'s investment on \( v_j \) to the total amount of \( u_i \)'s all investments in the investment network \( G = \{U, V, E\} \).

Standard deviation is often used to quantify risk [23]. For each investor \( u_i \), we define investment risk preferences, \( P_i \), as the standard deviation of investment rates of returns, as written in Equation (4):

\[
P_i = \sqrt{\frac{\sum_{j=1}^{n} w_{ij} (R_j - \bar{R}_i)^2}{\sum_{j=1}^{n} w_{ij}}}
\]

It is reasonable to believe that the more past investments, the more experience an investor has acquired in the P2P lending marketplace. For each investor \( u_i \), we define his/her investment experience as the number of previous investments. In P2P lending, given an investment network, the degree of a node \( u_i \), denoted as \( E_i \), is the number of edges that have one end attached to the node. We can compute investment experience \( E_i \) using Equation (5):

\[
E_i = \# \{e_{ij} | v_{cij} \neq 0 \}
\]

where \# represents the cardinality of a set.

### 3.3 An Example of the Investor Profile

An example that computes an investor profile is shown in Figure 2. In this figure, investor \( u_1 \) has invested $500 in total on three loans \( v_1, v_2 \) and \( v_3 \). The invested amounts are $150 for \( v_1 \), $250 for \( v_2 \) and $100 for \( v_3 \), respectively. So the corresponding investment weights are \( w_{i1} = \frac{150}{150+250+100} = 0.3 \), \( w_{i2} = \frac{250}{150+250+100} = 0.5 \), and \( w_{i3} = \frac{100}{150+250+100} = 0.2 \). Suppose the loans’ rates of return are \( R_1 = 0.1, R_2 = 0.2 \) and \( R_3 = -0.4 \), respectively, we can calculate the past investment performance as \( \bar{R}_1 = 0.05 \). Similarly, we can find risk preference \( P_1 = 0.23 \) and experience \( E_1 = 3 \).

![Figure 2: An Example of the Investor Profile.](image)
on this investee, may exhibits certain characteristics, some of which can be quantified by statistics. In this section, we study the quantitative evaluation of investees by their composition of investors.

As the typical scenario in a bipartite network, an investor can invest across many different investees, and an investee can be funded by a number of different investors. The group of investors who invested on the same investee is called the investor composition of the investee. For investee \( v_j \), its investor composition \( C_j \) can be written as:

\[
C_j = \{ u_i | v_{ij} \neq 0 \}
\]  

An example of the investor composition is shown in Figure 3, where there are four investor composition \( C_1, C_2, C_3 \) and \( C_4 \) corresponding to four different investees \( v_1, v_2, v_3 \) and \( v_4 \). It is obvious that each investor composition contains a different combination of investors, who may have different past investment history.

![Figure 3: An Example of the Investor Composition.](image)

In following subsections, we study how to quantitatively analyze the investor composition for a investee, which can be applied for investment decisions in P2P lending.

4.1 Investor Composition Analysis

Since each investor has several aspects in their profiles, we consider these aspects accordingly when evaluating the investor composition of an investee. Also, we propose a combined measure for easier comparison across all investees.

4.1.1 Composition of Investor Performances

For each investee \( v_j \), we define composition of investor performance \( CR_j \) as the average past investment performance of investors who invest on this investee \( v_j \), weighted by the amount invested, as shown in Equation (7). We have

\[
CR_j = \frac{\sum_{i=1}^{m} \lambda_{ij} \bar{R}_i}{\sum v_{ij}}
\]  

where \( \lambda_{ij} = \frac{v_{ij}}{\sum_{i=1}^{m} v_{ij}} \) is the ratio of \( u_i \)'s investment in \( v_j \) to the total amount \( v_j \)'s gain from all investors in the investment network \( G = (U, V, E) \). \( \bar{R}_i \) is the past investment performance of investor \( u_i \).

Composition of investor performance \( CR_j \) is the average past investment performance of investors who invest on investee \( v_j \). An investee that gains investment from good investors who have good past investment performance is a more worthwhile to invest, because we believe a good investor will make more correct investment decisions.

4.1.2 Composition of Investor Risk Preferences

Viewing each investee as a portfolio of investors, we can evaluate the risk of an investee based on risk preferences of individual investors [23, 27]. For each investee \( v_j \), we define composition of investor risk preferences \( CP_j \) as the standard deviation of all investors’ risk preferences \( P_i \) who invest on this investee \( v_j \). For each investee \( v_j \), we can compute composition of risk preferences \( CP_j \) using Equation (8) below:

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

where \( J_{ik} \) is the correlation between investor \( u_i \) and \( u_k \).

We estimate \( J_{ik} \) according to similarity in past investments. For any two investors, the more proportion of investment on the same investee, the higher investment similarity they have. For example, in P2P lending, if two investors have invested on exactly the same set of loans, the correlation between them is 1; if they have never invested on any loan in common, their correlation is 0. Here, we compute \( J_{ik} \) using Jaccard coefficient between investors \( u_i \) and \( u_k \):

\[
J_{ik} = \frac{D_{ik}}{D_i + D_k - D_{ik}}
\]  

where \( D_i \) is the total number of past investments by investor \( u_i \), \( D_k \) is the total number of past investments by investor \( u_k \), and \( D_{ik} \) is the number of past investments that investors \( u_i \) and \( u_k \) have in common (i.e. on the same investee).

4.1.3 Investor Composition Score

We consider investor composition score as a comprehensive measure of return per unit of risk in an investment [29]. For each investee \( v_j \), we define score of investor composition \( CS_j \) as the ratio of composition of investor performance \( CR_j \) to composition of investor risk preferences \( CP_j \):

\[
CS_j = \frac{CR_j}{CP_j}
\]  

where \( CR_j \) is given in (7) and \( CP_j \) is given in (8).

Score of investor composition \( CS_j \) is a synthetic index that takes into account both the investor performance \( CR_j \) and risk preferences \( CP_j \). In P2P lending, a loan having a composition of high investor performance does not necessarily mean that it is a valuable loan, since higher risk might promise more returns than \( v_2 \)'s. We can see that \( CR_2 \) is much higher than \( CR_3 \). It means that the average past investment performance of \( v_2 \)'s investors is better than those of \( v_3 \)'s. If we were to decide whether to invest in \( v_2 \) or \( v_3 \) by following the investment decisions made by other investors, we would trust those having consistently better performance in the past. So in this case, it is reasonable to believe that \( v_2 \) promises more returns than \( v_3 \) according to investors’ “voting.” Also, we notice that the comprehensive risk of \( v_2 \), \( CP_2 \), is lower than that of \( v_3 \). With higher return and lower risk, investee \( v_2 \) has a higher investor composition score \( CS_2 \), which indicates that \( v_2 \) is more worthwhile to invest on.
5. INVESTMENT DECISIONS MODEL

The goal of investment decisions is to find the best subset of investors. In this section, we describe how to apply investor composition analysis for investment decisions.

5.1 Model Description

In P2P lending market, investors face many loans which can be chosen at the same time. A common task for the investors would be to find out the list of investment candidates. In our proposed model, the investors find candidates through investor composition analysis of the loans.

Figure 5 describes the computation and decisions process. The whole process can be divided into three phases. In the first phase, the goal is to build investor profiles. In this phase, past investment performance, risk preference and investment experience of investors are calculated to form the profiles of the investors. The second phase is investor composition analysis. Each investee will receive a score by investor composition. Finally, the last phase is to aid the investment decisions by identifying top investees as the candidate set.

There are two input datasets in the model. DataSetH is the past investment transactions that include all investors’ past investment behaviors, which is used for building investor profile. DataSetN is the current investment status that contains information of all loans waiting to be funded. The investment decisions model will find out the investment candidates based on investment composition analysis.

In the process of building investor profiles, based on past investment data, BuildNet(DataSetH) builds a past investment network to gain the rate of return of loan $R_i$, investors’ investment weight $w_{ij}$, investors’ investment times $D_i$, and same investment times between investors $D_{ij}$. Then for each investor, we calculate the past investment performance $\bar{R}_i$, the investment risk preferences $P_i$ and the investment experience $E_i$ using Equations in section 6.2. Deposit the Jaccard correlation coefficient $J_{ij}$ will be used in process of investor composition analysis, we have to calculate $J_{ij}$ based on investors’ past investment data using Equation (9).

In the process of investor composition analysis, based on current investment data, BuildNet(DataSetN, $\alpha$, $\beta$) is used to construct the current investment network and gain investees’ investment weight $\lambda_{ij}$. DataSetN is current investment data and parameters $\alpha$, $\beta$ are used to filter investors and loans. The parameter $\alpha$ is used to filter investors by keeping those with more investments than $\alpha$. Accordingly the parameter $\beta$ is used to filter loans with more fulfillment rate than $\beta$ and less than 1. Furthermore, for each investor, we calculate composition of investor performance $CR_j$, composition of investor performance $CR_j$ and score of investor composition $CS_j$ using Equations in section 4.1.

In the process of investment decisions, we rank all loans according to score of investor composition $CS_j$ from the highest to the lowest, then choose the top $\gamma$ proportion of investees as the candidate set.

5.2 Description of Parameters

Parameter $\alpha$ is the minimum requirement on the investor’s experience, or more specially, the number of investments in historical data. This is reasonable since the more investments in the past, the better we can estimate and predict future investment returns, due to reduced variability given more observations.

The parameter $\beta$ represents the ratio of acquired amount of the investee to the total amount requested. To make use of the investor composition driven investment decisions model, one constraint has to be that the investee’s requested amount has already been partially financed. In other words, there are investors who have already invested some money on the investee, but not all the requested amount has been
fulfilled. Obviously, we have $0 \leq \beta \leq 1$. The larger $\beta$, the more trustable the expert preference will be, yet the smaller investment opportunity will remain.

The parameter $\gamma$ is the size of proportion of candidate set to all investees. In the process of prediction, investment decisions model uses certain proportion of the most preferred investees to be the candidate set. The choice of $\gamma$ could be specified by the users according to size and variability of their investments.

The choices of parameters $\beta$ and $\gamma$ are usually determined by the users according to their investment needs and experiences. In the following, we discuss the adaptive choice of $\alpha$ based on the statistical foundations.

We formulate the choice of the minimum requirement on the investor’s experience as a sample size problem. As we know, for random variable $X$, in order to estimate the mean $\mu$, using the sample mean $\bar{x}$ from $n$ random samples from a distribution with variance $\sigma^2$, by central limit theorem, the confidence interval for given confidence level would be $(\bar{x} - 3\sigma/\sqrt{n}, \bar{x} + 3\sigma/\sqrt{n})$, where $B = u_{\alpha}\sigma/\sqrt{n}$. Typically $u_{\alpha} = 1.96$ with confidence level 0.95. Suppose that the success or failure of an investment (i.e. the loan being paid or default) by an investor follows a Bernoulli distribution with success probability $p$, we observe his investments for $n$ times. We can compute the confidence interval with $B = u_{\alpha}\sqrt{p(1-p)/n}$.

Furthermore, we define the investment reliability of investor as:

$$R = 1 - B/p = 1 - u_{\alpha}\sqrt{1-p/p \times n}.$$\hspace{1cm} (11)

The relationship between the minimum number of investments $\alpha$ and the investor reliability $R$ can be illustrated in Figure 6. It can be seen that the larger $\alpha$, the smaller proportion of investors will be selected. In that situation, since the minimum investment experience standards are higher, fewer investors will be selected, resulting in a smaller set of higher quality investors with respect to reliability. Thus we say that the choice of parameter $\alpha$ has to consider a tradeoff between investors’ reliability and the number of investors to be selected when constructing the current investment network. Such considerations provide basic guidelines when building a investment decisions model in practice.

![Investor Experience and Reliability](image)

Figure 6: Investor Experience and Reliability.

6. EXPERIMENTAL RESULTS

In this section, we validate our model by showing empirical results on real-world P2P lending datasets. In the following, we will describe the data, show the distributions of the statistics, analyze the effectiveness of using investor composition as an indicator, and compare the performance of our model against baselines. We find that investor composition can effectively indicate investment value and in turn, help investors make better decisions.

6.1 Experiment Data

The experiments are based on the public dataset from Prosper.com\(^1\), which includes six relational data tables. The Members and Groups data tables contain users’ basic registration information and member groups. The Credit Profile table contains borrowers’ personal credit information. The Listings table contains information about loans request. The Loans and Bids tables are the most important for modeling, and are described in more detail below.

**Loans** The loans table contains all information about loans, such as loan amount, interest rate and payment status (i.e. paid, late, or default). This table is the most important to evaluate the performance of a loan.

**Bids** The bids table contains the time and amount of each investment by investors on each loan. We can learn each investor’s number of investments and amount invested on each loan. We can also know whether each loan has been fulfilled. This information is the basis to build a whole investment network.

In order to verify the effectiveness of our model, we only use investments whose payment information is available. In this dataset, there are 7968 loans in DataSetH for building investor profiles, and 4781 loans in DataSetN for computing investor composition.

6.2 Empirical Statistics

It is well known in behavior finance that investors will make different investment decisions because of differences in knowledge structures, information sources, cognitive development, and behavioral preferences. Therefore, we first show that different investors have different profiles in P2P lending.

![Scatter Plot of Investor Profiles](image)

Figure 7: Scatter Plot of Investor Profiles.

Different aspects of investor profiles, i.e. past performance, risk preference and investment experience, are shown in Figure 8. Using the techniques mentioned in Section 3, we can show the distribution of these quantitative measures. It is not surprising to find that investment experience has a skewed distribution, where there are fewer investors having a large number of past investments. In order to show stable statistics on each investor’s past performance and risk,
we show these statistics only for investors having at least 20 past investments. Figure 7 shows the scatter plot of investor’s performance versus risk. The data points seem to lie on a curved shape, and we can tell that the best performing investors have lower risk than many others.

Furthermore, since each loan can be funded by different investors, we expect investor composition of different loans to have different statistical characteristics, as shown in Figure 9. It is interesting to see that the composition risk of a loan (Figure 9(b)) is generally lower than the risk of an individual investor (Figure 8(b)). This reflects the idea of reducing risk by diversification.

Similar to investor’s risk and return, we show a scatter plot of each loan’s risk versus return in Figure 10. In the figure, we can see that most of the loans’ risk levels are minimized at a given performance level, and there are much fewer loans with higher risk.

### 6.3 Investor Composition as An Indicator

In this subsection, we show that the score of investor composition, as studied in Section 4.1, can indicate the performance of loans in P2P lending.

![Figure 8: Distribution of Investor Profiles.](image)

(a) Past Performance (Experience > 20)  
(b) Risk Preferences (Experience > 20)  
(c) Experience

![Figure 9: Distributions of Investor Composition of Loans.](image)

(a) Composition of Performance  
(b) Composition of Risk  
(c) Composition Score

![Figure 10: Scatter Plot of Composition Scores.](image)

Figure 11: The Default Rate by Investor Composition Scores.

We first group loans based on their score of investor composition, and compare the default rate of loans within each group (Figure 11). We can tell a consistent decreasing trend in the proportion of loans that are default, when the investor composition score increases. This shows intuitively that a higher investor composition score might indicate a lower probability that the loan will become default.

On the other hand, we divide loans into two groups: the paid group and the default group, and then compare the distribution of investor composition scores, as shown in Figure 12 and Table 1. The box-plot shows that the investor composition scores of paid loans is overall higher than those of default loans. As shown in Table 1, the p-value is below 0.001, indicating that the difference is significant.

### 6.4 Decisions Model by Investor Composition

In Section 5, we proposed an investment decisions model based on investor composition analysis to help investors make better investment decisions. In this subsection, we find em-
Figure 12: Distributions of Investor Composition Scores by Loans Status.

Table 1: Significance Test between Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t-statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid</td>
<td>-0.871</td>
<td>0.5027</td>
<td>24.95</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Default</td>
<td>-1.248</td>
<td>0.5097</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 13: A Comparison of Rate of Return.

In P2P lending, our investment decisions model identifies a ranking of loans, from best loans to worst ones, according to the scores of investor composition. Investors can then choose the top ones as the candidate set. In Figure 13, we compare the rate of return by our model against two baselines. One baseline is the average rate of return by investing on all loans. We can see that without selectively picking good quality loans from the whole pool of all loans, we expect to have a negative rate of return. The other hand, we compare with candidate selection by logistic regression model, which is the most commonly used on personal loans risk analysis [8, 32]. Based on default probability computed by logistic regression and interest rate, loans are ranked by decreasing expected rate of return. We find that with different values of γ, where γ is the percentage of top loans to invest on, the candidates chosen by our model have consistently higher rate of return than logistic regression.

We compare the performance of our model with different values of the parameter α, which is the minimum investor experience. Experiment results in Figure 14 show the performance of model with α = 20 or α = 30 is better than one with α = 0 or α = 90. As discussed in Section 5.2, the choice of parameter α is a tradeoff between investors’ reliability and the number of investors who will be chosen to construct current investment network.

Figure 14: Return Rates with Different α.

Furthermore, we compare the performance of model with different values of the parameter β, as shown in Figure 15. For a loan, the parameter β represents the ratio of acquired amount to the total amount requested. It is make sense that for larger β, our model prefers a larger number of investors when analyzing investor composition, which results in better performance on the rate of return.

Overall, as shown in the experiments with different parameters, our model gains consistently higher rate of return than the average rate of return in the market.

7. CONCLUSION

In this paper, we designed a data driven investment decision-making framework, which exploits the investor composition of each investment for enhancing decisions making in P2P lending. Specifically, we first built investor profiles based on quantitative analysis of past performances, risk preferences, investment experiences and reliability of investors. Then, based on investor profiles, we developed an investor composition analysis model, which can be used to select valuable investments and improve the investment decisions. Experimental results on real-world P2P lending data revealed that the investor composition model could effectively indicate the investment value and this investor composition model can significantly improve the investment performances.

8. ACKNOWLEDGMENTS

This work was supported in part by Natural Science Foundation of China (No. 70972059, 70890080, 70890083), National Science Foundation (No. CCF-1018151) and Fund of Ministry of Education of China (No. 10YJC630065).
9. REFERENCES