Investment Recommendation in P2P Lending: A Portfolio Perspective with Risk Management

Hongke Zhao[†], Le Wu[†], Qi Liu[†], Yong Ge[‡], Enhong Chen[†*]
[†]University of Science and Technology of China, [‡]UNC Charlotte
E-mail:{zhhk,wule}@mail.ustc.edu.cn, {qiliuql,cheneh}@ustc.edu.cn, yong.ge@uncc.edu

***Abstract*—P2P lending is an online platform to make borrowing and investment transactions. A central question on these platforms is how to align the right products with the right investors, thus helping investors to make better decisions. Along this line, tremendous efforts have been devoted to modeling the credits of products and borrowers from an economic perspective. However, these global models are only exploratory in nature and are not practical. In this paper, we focus on the personalized investment recommendation by reconstructing the two steps for investment decision making: *what to buy* and *how much money to pay*. Specifically, we first generate a candidate investment recommendation list for each investor that tackles "what to buy" problem. In this process, we consider various unique properties of investment recommendation. Furthermore, according to the portfolio theory, we optimize the shares of each recommended candidate by incorporating the investments an investor currently holds, thus solving the "how much money to pay" problem. Finally, extensive experimental results on a large-scale real-world dataset show the effectiveness of our model under various evaluation metrics.**

I. Introduction

P2P lending or online social lending, is an Internet-based platform to borrow money from others. There are two main kinds of roles in this market: the *borrowers* who want to borrow money from others and the *investors* who lend their money to borrowers. P2P lending has become a fast growing investment with greater than 100% year over year growth [1].

As a trend, this rapid prevalence of P2P lending in industry has enabled new research opportunities with the availability of massive transaction data. Previously, many efforts have been devoted to modeling the credits of borrowers and products [4], [7], [13], [17]. These works try to provide a global model to assess the qualities of products and borrowers, thus helping investors to make safer decisions. However, all these works are only exploratory in nature and there is still a long road from the assessment to an investor's final decisions. Now let's reconstruct a rational and successful investment scene for an investor *Alice*. For each product in the market, *Alice* would first decide whether to buy it. If her answer is *no*, then this product will be discarded. Otherwise, she would put this product into a candidate list. After the first step, *Alice* gets a small candidate list consisting of products she is interested in. Then, she would determine how much money should be allocated to a certain candidate. To do so, *Alice* would also take the products she holds currently (context) into consideration. According to the above analysis, we can formalize each investor's decision for investment into two steps: "what to buy" and "how much money to pay".

To facilitate an investor's decisions for investment, an intuitive idea is to recommend the global optimal products to all investors that are modeled by classical product assessment models. However, as the financing amount of a product is limited, the global optimal products are difficult to invest successfully. Thus, instead of providing identical recommendations for all investors, a better idea is to generate *personalized investment recommendation* for each investor. However, there are some unique characteristics of investment recommendation, which distinguish it from traditional recommendations. First, traditional recommendations assume "a user would probably choose items that similar users like" [12]. In investment, as the goal of investors is to get as many returns as possible with their unique risk preferences, most investors would like to get recommendations from experienced investors rather than the beginners even though these beginners are more similar to them. What's more, some investors are usually willing to trust products in some particular *groups*; how to incorporate this information? Last but not least, there exists a *portfolio* [8] perspective in investment, i.e., investors often adjust their shares in an investment portfolio in order to reduce risks. Meanwhile, an investor may hold some investments currently, how to exploit these current investments each investor holds (context) to better allocate money to each candidate?

In this paper, we propose to make personalized investment recommendation with risk management in P2P lending. The design goal is to equip each investor with recommended personalized investment compositions that best match his/her personal preference and reduce the risk simultaneously. Specifically, we first propose to model the profiles of products and investors. Then, we tackle the "what to buy" problem by considering both the expertise and preference of investors. Next, based on the portfolio theory, we incorporate the current investment contexts of investors for consideration, and propose an optimization function that minimizes the risk of each investor as a weighted combination of candidate recommendations. Thus, the problem of "how much money to pay" is solved. Fig. 1 illustrates the framework of our approach.

II. Profile Modeling for Products and Investors

In this section, we introduce how to model the profiles of products and investors in terms of return and risk.

[*]Contact Author.

1550-4786/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDM.2014.104

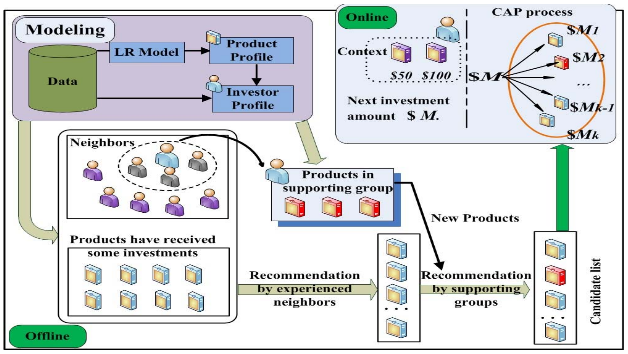


Fig. 1. The whole framework of our proposed approach.

In P2P lending, if a borrower wants to borrow money, a loan is created by him/her to solicit bids from investors by describing himself and the reason why he borrows money (e.g., for wedding). When an investor wishes to invest a loan, a bid is created by him describing how much money he/she wants to invest (e.g., \$50). In the soliciting period, if a loan has received enough investment money it will not receive investments any more. After the soliciting period, the loan begins its repayment period. However, if the loan can't receive enough money in time, it would be expired and the previous bids would also fail. For simplicity and consistence, we call these loans that can receive enough investments as products in this paper. At a time point, the products V can be grouped into two categories $V = V^f \cup V^n$, where $V^f = \{v_1^f, v_2^f, ..., v_{|V^f|}^f\}$ represents the completed products that have already finished the repayment periods and other processes. $V^n = \{v_1^n, v_2^n, ..., v_{|V^n|}^n\}$ are products that are still on sales or in their soliciting periods.

A. Bipartite Investment Network

An investor can diversify his/her investment to many products and a product would receive money from many investors. Bipartite graph is very suitable for modeling this many-to-many relationship of investment [7].

Specifically, an investment bipartite $G = (U, V^f, E)$ models the relationship between investors U and investment products V^f, associated with edges E connecting investors and their invested products in the past. If u_i has invested v_j^f, then $e_{ij} \in E$ represents the specific money he/she has invested in product v_j^f. For investor u_i, the weight to product v_j, denoted as α_{ij} is the proportion of money he/she invested in v_j^f (e_{ij}) divided by the total money he/she has already invested: $\alpha_{ij} = e_{ij} / \sum_j e_{ij}$. Fig.2 shows an example of a bipartite investment profile. For instance, $a_{12} = 42.86\%$ indicates u_1 has invested 42.86% of his/her money to v_2^f in the past.

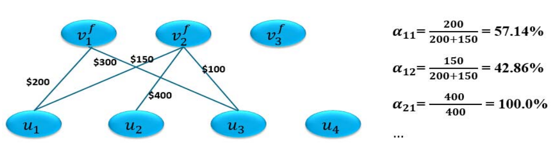


Fig. 2. An example of the bipartite investment network.

B. Product Profile

For a given product, *return* and *risk* are the two most important components for measuring its quality [7], [13]. In the soliciting period of a product, the borrower would put the lend rate which can be equaled to the expected return (denoted as P_j^v) of that product. However, the product's risk is hard to be profiled directly. Thus, we propose a prediction model for product risk estimation based on features of product.

To this end, we summarize two kinds of features: the product features (*Amount, LendRate, Term*) and the borrower features (*Credit, DebtToIncome*). With some historical training data, we can estimate the risks of products in V^n based on these features. Here, we adopt the logistic regression model due to its simplicity and relative high performance [4]. Specifically, for any $v_j \in V$, we denote the features of it as $\boldsymbol{v_j} = (1, v_{j,1}; v_{j,2}; ...; v_{j,d})$. Then the risk of v_j, denoted as R_j^v can be modeled as:

$$R_j^v = p(v_j) = \frac{\exp(\boldsymbol{\beta}^T \cdot \boldsymbol{v_j})}{1 + \exp(\boldsymbol{\beta}^T \cdot \boldsymbol{v_j})}, \quad (1)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_d)$ are the coefficients. We use the completed products V^f to learn $\boldsymbol{\beta}$. $\forall v_l^f \in V^f$, if it is paid by the borrower as announced in the soliciting period, it has no risk and the associated risk value $y_j^f = 0$. Else $y_j^f = 1$ denoting it is not paid in time. Given all these training data of products V^f, the logistic regression model learns the weight of $\boldsymbol{\beta}$ by maximum likelihood estimation [4].

Thus, product v_j's profile can be represented as a two-element vector $\overrightarrow{P_j^v} = [P_j^v, R_j^v]$ where the first term is the expected return and the second term is the estimated risk.

C. Investor Profile

Similar to the product profile, we also build an investor profile from two aspects: $\overrightarrow{P_i^u} = [P_i^u, R_i^u]$, where the first term characterizes the return expectation of this investor and the second is the risk preference. Generally, these investor profile terms of P_i^u can be defined as the weighted average of the corresponding profile terms of the products he/she has invested: $P_i^u = \sum_{j=1}^{|S_i^v|} \alpha_{ij} P_j^v, R_i^u = \sum_{j=1}^{|S_i^v|} \alpha_{ij} R_j^v$. where α_{ij} is the ratio of u_i's investment in v_j. $S_i^v \in V^f$ is the set of products that u_i has invested in the past.

Besides these two aspects, an investor's expertise is also valuable to judge his/her investment performance. A natural idea is that an investor's expertise is proportional to the money he/she has earned. We define the real return of an investor as: $E_i^u = \sum_{j=1}^{|S_i^v|} \alpha_{ij} R_j$, where R_j is the real return of v_j, and the specific form of investor's expertise will be defined as a function of E_i^u in the next section.

III. Personalized Candidate Investment Recommendation Generation

In this section, we solve the "what to buy" problem by generating a personalized candidate investment recommendation list for each investor.

A. Candidate Recommendation by Experienced Neighbors

User-based collaborative filtering (UCF) [11], [15] is widely used for personalized recommendation due to its simplicity and relative high performance. Generally, for a product v_j^n

which has been invested by a set of investors S_j^u, UCF evaluates its ranking value for investor u_i as shown in Equation (2),

$$r(u_i, v_j) = \frac{\sum_{u_k \in S_j^u} s(u_i, u_k)}{|S_j^u|}, \tag{2}$$

where $s(u_i, u_k)$ is the similarity between investor u_i and investor u_k. The larger the predicted ranking value, the more likely an investor would prefer this product. In Equation (2), we need to calculate investors' similarities. Usually, it can be computed by the Jaccard similarity [20] of the investors' historical investment behaviors, as shown in Equation (3).

$$s(u_i, u_k) = \frac{|S_i^v \cap S_k^v|}{|S_i^v \cup S_k^v|}, \tag{3}$$

where $S_i^v \cap S_k^v$ are products invested by u_i and u_k in common.

Different from traditional UCF, in P2P lending, as investors would like to get recommendations from experienced investors, experienced neighbors should play a higher weight for an investor's decision. To summarize, we have two factors that may determine an investor's final decision: similarity and expertise. Thus we propose the following predicted ranking score that considers both two aspects:

$$r(u_i, v_j^n) = \frac{\sum_{u_k \in S_j^u} s(u_i, u_k) F(u_k)}{|S_j^u|}, \tag{4}$$

where $F(u_k)$ is the expertise value of investor u_k. Intuitively, a investor u_k's expertise is a monotone increasing function with regard to his/her real return E_k^u in the past. Without loss of generality, we simply set the $F(u_k)$ as a sigmoid function:

$$F(u_k) = \frac{1}{1 + e^{-E_k^u/\gamma}}. \tag{5}$$

The expertise value ranges in (0,1). If $E_i^u > E_j^u$, then $F(u_i) > F(u_j)$. When γ is smaller, the curve is steeper and the experienced investors' effects will increase. Thus γ can be seen as a parameter that tunes the weight of similarity and the expertise of neighbors for predicted ranking values.

B. New Product Recommendation by the Supporting Groups

As shown in Equation (4), if a product v_j^n has been invested by only a few investors, i.e., the value of $|S_j^u|$ is small and then the recommendation will not work well. This phenomenon is termed as the cold-start problem [15], [6], and it is also very common in P2P lending market. For a new product, it is more difficult to solicit bids from investors because investors prefer to invest products that other investors have invested [3]. Luckily, there is a group effect in P2P lending [5]. Group is a set of borrowers that join together. Specifically, the *supporting groups* are the given groups that an investor pays special attention to, which means this investor believes these groups and is more likely to lend to these borrowers in his/her supporting groups. This group effect of the supporting groups can be used to alleviate the cold-start problem to some extent.

Specifically, for a new product v_l^n in investor u_i's one supporting group, we add it into a suitable position in his/her candidate list got by the previous recommendation. The position is determined by computing its similarities with the products in the candidate list. The similarity is computed by Equation (6), and thus we could put the new product before the product in candidate list that is most similar with it.

$$s(v_j^n, v_l^n) = Cos(\overrightarrow{P_j^v}, \overrightarrow{P_j^l}) = \frac{\overrightarrow{P_j^v} \cdot \overrightarrow{P_l^v}}{|\overrightarrow{P_j^v}| \cdot |\overrightarrow{P_l^v}|} \tag{6}$$

In summary, we generate a rough recommendation list for each investor, and then add some new products based on the *recommendation by supporting groups* into the list.

IV. Recommendation Optimization via Portfolio

In this section, we provide a *c*ontext-*a*ware *p*ortfolio optimization model (CAP) that assigns the right amount of money to each candidate, thus solving the "how much money to pay" problem. According to the portfolio theory [8], an investor would always adjust their shares in an investment portfolio in order to have lower risks. This classical theory instructs investors to diversify their investments on multiple products to provide the maximum future returns within their risk tolerances. Also, in P2P lending, an investor's risk preference is for an investment portfolio rather than for a single product. E.g., when an investor has invested several products of low risk values then he/she is likely to invest products of high returns. That is a simple illustration about the portfolio effect for an investor's further investments by considering the investments he/she currently holds.

A. Context Representation

In the real world, an investor will invest more than one product over a period of time. What's more, an investor's further investments will be influenced by his/her previous investments. As the current status of investments an investor holds is important for his/her further decision, we call it as the *context* of this investor. Specifically, the context of investor u_i (denoted as C_i) is the investment product(s) hold currently by him/her: $C_i = \{v_1 : M_1; v_2 : M_2; ...; v_{|C_i|} : M_{|C_i|}\}$, where M_j is the investment amount on product v_j, e.g., \$50.

B. CAP Process : Context-aware Portfolio Optimization

For a given total investment amount *M*, a portfolio Υ can be represented by a collection of products with a corresponding investment amount M_j assigned to each product v_j, i.e.,

$$\Upsilon = \{(v_j, M_j)\}, \quad s.t. \sum_{j=1}^{|\Upsilon|} M_j = M. \tag{7}$$

Here we first define the return of portfolio as $E[\Upsilon]$, which can be compute by

$$E[\Upsilon] = \sum_{j=1}^{|\Upsilon|} \frac{M_j}{M} \cdot P_j^v. \tag{8}$$

Also, we could define the risk of portfolio as $R[\Upsilon]$, which can be computed by the following function [8], [19],

$$R[\Upsilon] = \sum_{j=1}^{|\Upsilon|} ((\frac{M_j}{M})^2 (R_j^v)^2 + 2 \sum_{k=j+1}^{|\Upsilon|} \rho_{jk} \frac{M_j \cdot M_k}{M^2} \cdot R_j^v \cdot R_k^v), \tag{9}$$

where ρ_{jk} is the correlation coefficient between v_j^n and v_k^n. In fact, since the products in P2P lending are one-time and the correlation between different products is small, it's reasonable to approximately consider the correlation coefficient ρ among different products as 0, and meanwhile, it makes the portfolio optimization easier. Thus, $R[\Upsilon]$ can be simplified as:

$$R[\Upsilon] = \sum_{j}^{|\Upsilon|} (\frac{M_j}{M})^2 (R_j^v)^2. \quad (10)$$

For a given total investment amount M that an investor wants to invest and his/her investment context $C_i = \{v_1 : M_1, v_2 : M_2, ..., v_{j'} : M_{j'}, ..., v_{|C_i|} : M_{|C_i|}\}$, $\sum_{j'=1}^{|C_i|} M_{j'} = M'$, the goal of CAP is to recommend a portfolio with products and the optimal investment amount distribution for each product, which can be formalized as the following constraint optimization problem:

$$\begin{aligned} &\arg\min_{M} : R[\Upsilon] = \sum_{j}^{|\Upsilon|} (\frac{M_j}{M + M'})^2 (R_j^v)^2 \\ &S.t. : P_i^u = E[\Upsilon + C_i] = \sum_{j=1}^{|\Upsilon|} \frac{M_j}{M + M'} \cdot P_j^v + \sum_{j'=1}^{|C_i|} \frac{M_{j'}}{M + M'} \cdot P_{j'}^v \\ &\sum_{j=1}^{|\Upsilon|} M_j = M, \quad M_j > 0, \end{aligned} \quad (11)$$

where Υ is the recommendation portfolio, $\boldsymbol{M} = \{M_1, M_2, ..., M_j, ..., M_{|\Upsilon|}\}$ is the amount vector for products. The optimal solution of this constrained optimization problem can be easily obtained by traditional gradient method or the method proposed in [19]. *CAP* takes the context effect into consideration by optimizing the context of an investor and the recommendation candidate list simultaneously. When there are no products currently hold by an investor, the *CAP* directly optimizes the recommendation candidate list.

The utilities of *CAP* are mainly reflected in two aspects. First, *CAP* could **rerank** the products in the candidate list by taking an investor's context into consideration, and the final ranking of a product is based on its allocation amount [19]. Second, *CAP* suggests detailed allocation of money to each product, and help to **reduce risk** by diversified investment.

V. Experiments

In this experimental section, we will demonstrate: 1) the findings from profile modeling; 2) the performance of our model and other baselines.

A. Dataset Overview

The experimental dataset is download from Prosper.com[1]. We mainly use four tables of this dataset. The *Listing table* contains the product status and some basic credit features. The *Bid table* contains the specific time and the amount of money investors have paid for each product. These investment records are the basis to build investor profiles, product profiles and the investment bipartite network. The *Loan table* is used to evaluate the performance of a product. Finally, the *Member table* contains the information of the groups an investor supports. This dataset contains all the records in this platform from November, 2005 to the end of May, 2011. We only extract the investors that have at least two investment records. We use the products that have completed before May, 2008 and their associated bidding records as the training records (TrR), and the remaining records are used for test (TeR). Table I shows the basic statistics of the experimental dataset.

[1]http://www.prosper.com/tools/DataExport.aspx

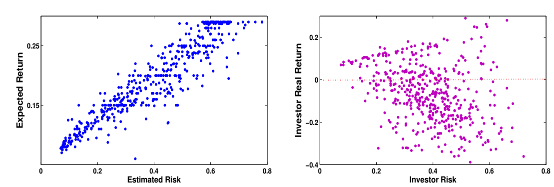
TABLE I
Data Statistics

#products	#investors	#records	#groups	#TrR	#TeR
19,077	34,210	2,616,877	4,025	2,099,482	517,395

B. Results of Profile Modeling

Fig. 3(a) shows the scatter plot of randomly sampled products with regard to their returns and risks (estimated by the model in Section II-B). We can see that the larger the predicted return, the larger the risk. The Pearson correlation coefficient between the risk and return of products is 0.909, indicating a highly positive correlation between each product's estimated return and risk. This is consistent with the basic investment theory in finance [2].

Since we can calculate the real return value E_i^u for each investor u_i. Thus, instead of plotting an investor's estimated return and estimated risk, we turn to plot each investor's real return v.s. his/her estimated risk, which is shown in Fig.3(b). We can see that some investors earned a lot while quite a range of investors even incurred financial loss. Also, Some investors are rather conservative while some others are willing to try high-risk products. The result of the Pearson correlation coefficient between investors' real return and the estimated risk is -0.305, indicating an investor's return is even negatively correlated with his/her risk preference. One possible reason is that, for some inexperienced investors, though they select relative low return product, they may also take high risk if the product is not of high quality. To further validate this hypothesis, we only keep the experienced investors ($E_i^u > 0$). The plots of these experienced investors are above the read line in Fig.3(b) and the correlation coefficient for these investors is 0.379. That is, for experienced investors, the estimated risk is proportional to the return, i.e., if they want to earn more, they could try higher risk products.



(a) Products (b) Investors

Fig. 3. Scatter Plot of Profiles. Randomly sampled 500 products/investors.

C. Benchmark Methods

We call the whole recommendation framework proposed as the *REC_G_P*. We also consider several variants of the REC_G_P: the *REC* model that provides investment recommendations by Equation (4); the *REC_G* model that further considers the new product recommendation by supporting groups compared to REC. Furthermore, we implement *REC_G_E* model. This model first gets the candidate list through REC_G and then calculates the weight of each recommended product that is proportional to the expected rate of return values E_j^v, where $E_j^v = P_j^v \cdot (1 - R_j^v)$.

Traditional Recommendation Methods. We implement user-based CF (UCF) [11] and item-based CF (ICF) [14] for investment recommendation. Besides, we calculate the similarity between the profiles of an investor and the products, and then recommend the most similar products to the investor. We call this model as *C*ontent-*B*ased (CB) model.

Non-personalized or Global Assessment Methods. For the popular global assessment models, we implement the wildly used logistic regression (LR) [4].

For fair comparison, in some baseline models, e.g., UCF, ICF, the weight of the money paid to each recommendation is proportional to the predicted likeness. The maximum value of candidate list size k is empirically set to be 50 as recommending a larger set of items to investors is useless. We set parameter $\gamma = 0.05$ (in Equation (5))[2].

D. Evaluation Metrics

We borrow the metrics from both recommendation area and economy area, and in order to better measure the quality of our model, we propose two new appropriate metrics to evaluate the overall performance.

Recommendation Metrics. We first evaluate these methods on precision and recall. Their definitions are: $precision(i) = \frac{L(k) \cap T(i)}{k}, recall(i) = \frac{L(k) \cap T(i)}{|T(i)|}$, where k is the candidate list size, *L(k)* is the candidate list, and *T(i)* is the true investment products set of investor u_i. These two metrics measure how the recommendations match an investor's decisions.

Economic Metrics. In investment, the return is the most important measurement. Thus, we also use the real returns (RR) of the products as one metric. Besides, the *sharpe ratio* (SR) [16] is often used to evaluate the robustness of a portfolio. These two metrics are defined as $RR(i) = \sum_{j=1}^{k} w_j \cdot R_j, SR(i) = RR(i)/\sqrt{\sum_{j=1}^{k} w_j \cdot (R_j - RR(i))^2}$, where R_j is the real return of product v_j, and w_j is the investment weight of product v_j in the recommendation portfolio.

Overall Metrics. We propose two new metrics to evaluate both the "personalization" and recommendation "quality". These two metrics suggest that the recommendations for an investor are meaningful only if the average return of the recommended portfolio is higher than the investor's own real return E_i^u. The first measure is denoted as *PAQ* and is defined using precision and return as follows:

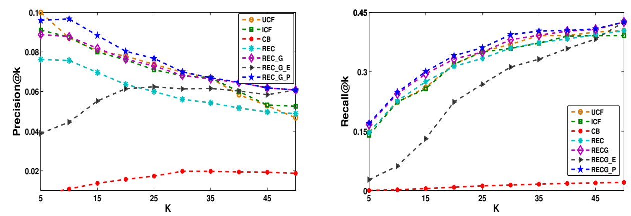
$$PAQ(i) = \begin{cases} \frac{L(k) \cap T(i)}{k} \cdot (RR(i) - P_i^u) & if: RR(i) > E_i^u \\ 0 & else \end{cases} \tag{12}$$

The other new metric is denoted as *RAQ*, which is defined using recall and return as follows:

$$RAQ(i) = \begin{cases} \frac{L(k) \cap T(i)}{|T(i)|} \cdot (RR(i) - P_i^u) & if: RR(i) > E_i^u \\ 0 & else \end{cases} \tag{13}$$

E. Experimental Results

Recommendation Performance. Fig.4(a) and Fig.4(b) show the results on precision and recall. Since the LR model always recommends the same global top-k products with the highest product qualities for every investor. Its results are nearly zero, so we do not report them. From these two figures, we can see that UCF and ICF show the similar performances, and REC's performance reduces a little since it weakens some inexperienced neighbor investors' influence. The REC_G approach gives improvements on both precision and recall compared with the REC. This is because REC_G considers the supporting group effect and solves the cold-start problem to some extent. REC_G_P performs the best since it not only takes advantage of REC_G, but also explores investors' contexts. The REC_G_E reranks the products only from the return perspective, so it reduces the recommendation performance compared to REC_G. As REC_G_E processes on the candidate list get from REC_G, it converges to REC_G when k is the maximum size of candidate list. Among all the models, CB shows the worst performance.



(a) The performance on precision. (b) The performance on recall.

Fig. 4. Recommendation Performance.

Economic Performance. The economic results are shown in Table II. LR always recommends the top-k products that have the highest predicted qualities to all investors only from the economic perspective. However, most invertors can not invest the same several target products successfully, so LR only provides a possible economic *upper bound*. In fact, REC_G_E can be treated as a local assessment model since the estimated return E_i^v of a product is obtained through LR directly. Thus, REC_G_E is more comparable. From Table II, we can see that UCF and ICF perform badly. CB even gives a result of negative returns. The returns of REC are more than 3.5%, and REC_G obtains the similar results compared with REC. REC_G_E can improve the returns especially when the portfolio size is small. Although REC, REC_G and REC_G_E can give better returns, their SR values decrease heavily when the portfolio size increases. That reflects they are not robust enough. From the perspectives of return and robustness, REC_G_P performs the best. Especially on returns, REC_G_P provides the returns of about 4.5%, which is better than REC_G_E.

Overall Performance. The results of *PAQ* and *RAQ* are shown in Fig.5(a) and Fig.5(b) respectively. The LR results are nearly zero since its performances on precision and recall are nearly zero. Our REC_G_P works the best, followed by the REC_G and REC. REC_G_E doesn't have a good performance especially when the portfolio size is small. UCF and ICF perform similarly and worse than others except CB.

VI. Related Works

P2P lending. Readers can refer to [2] for an overview of P2P lending. For P2P lending, many studies aimed at the product quality evaluation, such as [4], [17]. Besides, Luo et

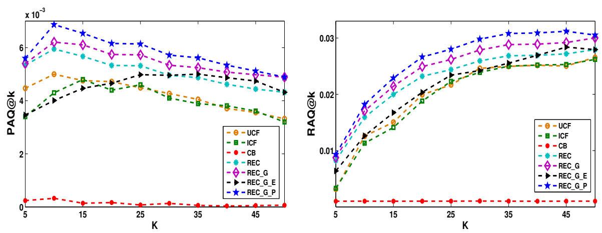
[2]We omit the parameter discussion due to the space limitation.

TABLE II
ECONOMIC PERFORMANCE.
(a) The performance on $RR@k$, the larger the better.

$Methods.\backslash K$	10	20	30	40	50
UCF	0.0062	0.0121	0.0129	0.0123	0.0128
ICF	0.0040	0.0137	0.0106	0.0084	0.0104
CB	-0.0783	-0.0732	-0.0837	-0.0901	-0.0890
REC	0.0387	0.0386	0.0369	0.0372	0.0385
REC_G	0.0389	0.0383	0.0348	0.0337	0.0343
REC_G_E	0.0482	0.0469	0.0424	0.0414	0.0347
REC_G_P	**0.0503**	**0.0474**	**0.0432**	**0.0440**	**0.0455**
LR(upper bound)	*0.0678*	*0.0625*	*0.0654*	*0.0675*	*0.0687*

(b) The performance on $SR@k$, the larger the better.

$Methods.\backslash K$	10	20	30	40	50
UCF	0.780	0.380	0.273	0.236	0.198
ICF	0.812	0.333	0.275	0.225	0.197
CB	Invalid	Invalid	Invalid	Invalid	Invalid
REC	1.094	0.551	0.428	0.421	0.363
REC_G	1.322	0.532	0.431	0.402	0.361
REC_G_E	7.179	3.069	1.444	0.646	0.367
REC_G_P	**8.856**	**6.117**	**5.012**	**4.423**	**4.083**
LR(upper bound)	*16.14*	*16.47*	*16.91*	*16.62*	*12.84*



(a) The performance on PAQ. (b) The performance on RAQ.

Fig. 5. Overall Performance.

al.[7] developed an investor composition model to measure the products. Except for the studies on product assessment, other research issues included social analysis [5], bidding behavior analysis [3], borrower decision optimization [10]. Works in this area tried to build a finer model to access the quality of products or borrowers, thus helping investors to make decisions with these global models. However, all these models are only preliminary in nature, and to the best of our knowledge, none has considered the personal aspects for investment recommendation.

Recommender system. Recommender system [12] provides suggestions of items that may interest users. Generally, recommendation techniques contain content-based recommendation [9] and collaborative filtering [15]. In collaborative filtering, user-based [11] and item-based [14] are the two most widely-used strategies. Besides, context-aware recommendation is a research hot spot recently. Context-aware recommender system explores various contextual information in order to provide more accurate recommendations. In some specific applications, many studies of context-aware recommendations have been conducted, such as tourist [6] and song recommendation [18]. However, in P2P lending area, there have been few works on personalized or context-ware analysis.

VII. CONCLUSION

In this paper, we provided a study of investment recommendation in P2P lending market. Our solution solved the "what to buy" and "how much money to pay" problems in investment decision process. Specifically, we first modeled the profiles of products and investors. Then we tackled the "what to buy" problem by modifying a traditional user-based collaborative filtering model. The proposed solution can discover the uniqueness of investment decision, thus generating a specific candidate recommendation list for each investor. Furthermore, by borrowing the portfolio theory, we proposed to further allocate the weight of each candidate by considering the investments an investor currently holds. Thus, the "how much money to pay" problem was also solved. The experimental results demonstrated the proposed framework was effective under various metrics.

Acknowledgement. This research was supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), the Fundamental Research Funds for the Central Universities of China (Grant No.WK0110000042) and the Anhui Provincial Natural Science Foundation (Grant No.1408085QF110).

REFERENCES

[1] Peer To Peer Lending, http://techcrunch.com/2012/05/29/peer-to-peer-lending-crosses-1-billion-in-loans-issued/.
[2] S. C. Berger and F. Gleisner. Emergence of financial intermediaries in electronic markets: The case of online p2p lending. *BuR-Business Research*, 2(1):39–65, 2009.
[3] S. Ceyhan, X. Shi, and J. Leskovec. Dynamics of bidding in a p2p lending service: effects of herding and predicting loan success. In *20th WWW*, pages 547–556. ACM, 2011.
[4] G. Dong, K. K. Lai, and J. Yen. Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*, 1(1):2463–2468, 2010.
[5] S. Herrero-Lopez. Social interactions in p2p lending. In *3rd Workshop on Social Network Mining and Analysis*, page 3. ACM, 2009.
[6] Q. Liu, Y. Ge, Z. Li, E. Chen, and H. Xiong. Personalized travel package recommendation. In *ICDM*, pages 407–416. IEEE, 2011.
[7] C. Luo, H. Xiong, W. Zhou, Y. Guo, and G. Deng. Enhancing investment decisions in p2p lending: an investor composition perspective. In *17th SIGKDD*, pages 292–300. ACM, 2011.
[8] H. Markowitz. Portfolio selection*. *The journal of finance*, 7(1):77–91, 1952.
[9] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.
[10] L. Puro, J. E. Teich, H. Wallenius, and J. Wallenius. Borrower decision aid for people-to-people lending. *Decision Support Systems*, 49(1):52–60, 2010.
[11] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *CSCW*, pages 175–186. ACM, 1994.
[12] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.
[13] E. Rosenberg and A. Gleit. Quantitative methods in credit management: a survey. *Operations research*, 42(4):589–613, 1994.
[14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *10th WWW*, pages 285–295. ACM, 2001.
[15] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.
[16] W. F. Sharpe. The sharpe ratio. *Streetwise–the Best of the Journal of Portfolio Management*, pages 169–185, 1998.
[17] Y. Wang, S. Wang, and K. K. Lai. A new fuzzy support vector machine to evaluate credit risk. *IEEE TFS*, 13(6):820–831, 2005.
[18] X. Wu, Q. Liu, E. Chen, L. He, J. Lv, C. Cao, and G. Hu. Personalized next-song recommendation in online karaokes. In *7th RecSys*, pages 137–140. ACM, 2013.
[19] W. Zhang, J. Wang, B. Chen, and X. Zhao. To personalize or not: a risk management perspective. In *7th RecSys*, pages 229–236. ACM, 2013.
[20] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Mobile app recommendations with security and privacy awareness. In *20th SIGKDD*, pages 951–960. ACM, 2014.