IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 11, 2021, accepted July 29, 2021, date of publication August 9, 2021, date of current version August 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103510

Finding the Next Interesting Loan for Investors on a Peer-to-Peer Lending Platform

CHICHANG LIU[1,2], YIDONG MING[1,2], YINGYUAN XIAO[1,2], WENGUANG ZHENG[1,2], AND CHING-HSIEN HSU[3], (Senior Member, IEEE)

[1]Engineering Research Center of Learning-Based Intelligent System, Ministry of Education, Tianjin University of Technology, Tianjin 300384, China
[2]Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China
[3]College of Information and Electrical Engineering, Asia University, Taichung 41354, Taiwan

Corresponding authors: Yingyuan Xiao (yyxiao@tjut.edu.cn) and Wenguang Zheng (wenguangz@tjut.edu.cn)

This work was supported in part by Tianjin ''Project + Team'' Key Training Project under Grant XC202022, in part by the National Natural Science Foundation of China under Grant 61702368, and in part by the Natural Science Foundation of Tianjin under Grant 18JCQNJC00700.

ABSTRACT With the development of the mobile Internet, a peer-to-peer(P2P) online lending platform has become increasingly popular in the financial market, and it attracts a massive number of users. The task that helps investors find potential loans for improving the funding success rate has become a major challenge for lending platforms. However, the traditional recommendation schemes rarely take into account the challenges, such as the timeliness of loans (i.e., when a loan funding is completed or expired, it will no longer recruit investment), the common cold start problem (continuously releasing new loans is a common phenomenon), and the loans' potential default risk. Considering the above characteristics, we propose a deep learning model based on a sequence of the incremental matrix factorization technology (DeepSeIMF). First, the cold start problem of loans can be effectively solved by designing an incremental matrix factorization model based on the time series. Then, a neural network is used to provide investors with personalized investment recommendation services based on risk assessment. Finally, the model performance is systematically evaluated based on a large-scale real-world dataset. The experimental results demonstrate the effectiveness of our solution.

INDEX TERMS P2P lending, recommender system, matrix factorization, deep learning.

I. INTRODUCTION

P2P lending platforms enable individuals to directly obtain loans from other individuals by removing the middlemen. They provide borrowers with an alternative to the traditional banks or a rate better than that offered by banks and provide investors with an optional wealth management scheme. According to the Forbes estimates in 2015, the amount of funds obtained through crowdfunding worldwide was $34 billion. The average annual financing of investment was $30 billion. The World Bank report predicts that the total amount will exceed $96 billion in 2025, and the proportion of Asia in this investment will significantly increase.[1] In December 2019, the cumulative number of users on the P2P lending platform has reached 2 million in the United States of America, and the cumulative transaction amount has reached $15 billion.[2] In a P2P lending platform, a borrower first initiates a loaning request. The platform then verifies the information of the borrower and performs a risk assessment of the loan (loaning request). Finally, the platform carries out a split auction on the amount of the loan that has passed the risk assessment. When the cumulative bid amount of all investors on the bidding loan equals the borrower's loan amount, the auction is completed. The entire business process is shown in Figure 1. The bidding process of many lending platforms follows the ''all-or-nothing'' rule, that is, when the cumulative bidding amount of all investors on a bidding loan equals its target amount, the loan bids successfully, otherwise it fails, and the platform returns the bidding funds to the investors. As the above-mentioned rules are adopted by the lending platforms, not only do the bidding funds on the completed loans fail to bring revenue to investors, but

The associate editor coordinating the review of this manuscript and approving it for publication was Sathish Kumar.

[1]http://www.morganstanley.com/ideas/p2p-marketplace-lending.

[2]https://www.prosper.com/

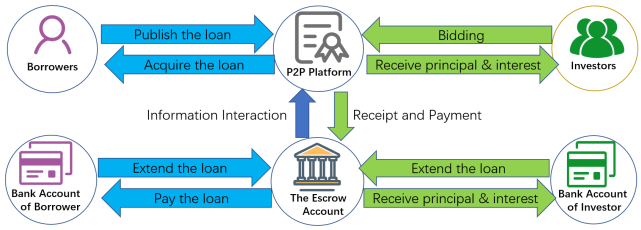

FIGURE 1. Schematic of the business process on the P2P platform.

also they reduce the overall transaction success rate of the platform. Therefore, the critical component for the success of funding communities is to recruit enough investments. However, the loan success rate is estimated between 20% and 40%, and the failed loans whose cumulative bidding amount is less than 20% exceed 60% [1], [2] in the P2P lending market. In addition, for quite a long time, relevant studies devoted to improving the success rate have been rather limited in the literature.

Fortunately, with the accumulation of large-scale user behavior data in P2P lending platforms, many data-driven studies focusing on risk evaluation [3]–[5] [6], [7], fundraising analysis [8], and lending or bidding behavior [9], [10] have been conducted. For example, a regression model was adopted by Iyer *et al.* [11] in order to evaluate whether lenders in the P2P lending market can use the borrower information to determine the creditworthiness on prosper data. Especially, Zhao *et al.* [12] not only used a regression model to evaluate the risk of loans but also integrated the loan risk into the recommendation system that inspired our research. However, how to find the next interesting loan for investors based on the loan risk and their investment performance is still largely unexplored areas. Thus, in this paper, we attempt to resolve this problem by transforming it into a recommendation problem. Therefore, the timeliness of loans, the sparseness of loans' data, and heterogeneous factors related to the loan risk need to be addressed because of the characteristics of the platform and its operational mechanism. The timeliness of loans shows that loans need to be recommended to investors as soon as possible in their life cycle, because it will be meaningless to recommend a loan to investors once the funding of the loan fails. The data sparseness of loans is a common phenomenon as new loans continue to emerge with time. Therefore, the recommendations must be good for this kind of data. The heterogeneous factors indicate a new method that must be adopted in order to deal with it.

In order to address the above-mentioned issues, we present a focused study on improving the recommender performance in the P2P lending platform, that is, we aim to predict which loan will be invested by an investor when he decides to invest. Specifically, a DeepSeIMF model is proposed to help investors find their potential loans. The *Deep* part of the DeepSeIMF model is flexible and could integrate the heterogeneous features of loans. The *SeIMF* part of the DeepSeIMF model is able to track the timeliness and sparseness of loans. Further, a combined approach is proposed to enhance the performance of recommendations. The main contributions of this work are as follows:

1) An incremental matrix factorization model based on time series is designed, which records the investors' changing preferences over time. The most important aspect of this model is that it uses a progressive matrix factorization to deal with the rapidly growing data on a lending platform.
2) A deep learning network is used to obtain the risk factors for lending items. Through feature interaction learning, some effective implicit combination features are learned, which makes the recommendation consider both investment habits and the potential risk of loans.
3) Extensive experiments are conducted on a real dataset. The experimental results show that our approach is effective and outperforms the state-of-the-art approaches.

The rest of the paper is organized as follows. Section II introduces the risk identification and investment recommendation in a P2P lending platform. Section III presents a detailed description of the proposed model. Details of the experiments carried out for validating the performance of our model are presented in Section IV. Section V presents a summary of this work and conjectures for future research directions.

II. RELATED WORK

For P2P lending platforms, accurately recommending loans to potential investors according to their investment preferences is directly related to the ability of the investors' assets to benefit and indirectly promote their growth and development. In this section, we briefly introduce the related work that can be divided into two categories, namely, risk identification and investment recommendation.

A. RISK IDENTIFICATION IN A P2P LENDING PLATFORM

In the financial sector, an investor's risk-taking capability is generally measured by the investor's utility function [13]. Under uncertain environmental conditions, investors generally take a decision that maximizes their expected utility, which implies that the investors prefer projects that are suitable for their own risk tolerance [14]. The objective of risk identification is to find out those loans that have a high probability of defaults or borrowers with low credit scores. Then, certain measures are taken to reduce the bad debt ratio of the transaction from the root cause and improve the investment income for the investors. Most of the previous studies have focused on the risk prediction of loans, and the corresponding model can be roughly formulated as follows: $R : M((f_1^1, \ldots, f_1^m), (f_2^1, \ldots, f_2^m), \ldots, (f_n^1, \ldots, f_n^m)) \rightarrow (S_1, S_2, \ldots, S_n)$ [15]. In this model M, f_i^j represents the jth attribute of the loan v_i, and S_i represents the risk evaluation value of v_i, which is generally expressed as a probability of a default score or the default probability. In this kind of model, researchers either find out a large number of f_i^j attribute characteristics that affect the prediction result or design a new risk assessment model, M, in order to improve the accuracy. For example, the personal information of a borrower and the information of loan are considered important indicators in assessing whether the borrower has a defaulting tendency [1], [16]. The narrative data is an important data source of risk assessment in P2P lending and some studies have centered on this [17], [18]. Some studies have also observed that people with higher credit ratings have lower default risk and a better borrowing rate [19]. Conversely, borrowers with low credit ratings cannot get loans from the platform or obtain loans at a higher borrowing rate [20]. In addition, some researchers have focused on investigating the relationship between social relationships and borrower performance. Previous studies by Chen *et al.* [21] and Krumme and Herrero [22] have shown that the borrowers' rich social networks can effectively reduce the default probability. Moreover, when a lender is a borrower's friend, there are almost no possibilities of defaults in the loan [23]. In the absence of sufficient information, a lot of research efforts have been put into developing risk prediction models. These models include both traditional models, such as generative analysis [24], logistic regression [12], [25], and K-nearest neighbor algorithms [26], and machine learning algorithms, such as support vector machines [27], neural network algorithms [28], decision trees [29], and so on. There are many studies focusing on ensemble credit scoring models of P2P lending. Random forests [30], gradient boosting decision tree [31], and heterogeneous ensemble models [32], [33] have already been considered in credit risk evaluation in P2P lending.

Unfortunately, many studies have been devoted to predicting the risks of loans, and few of them apply the predicted results in intelligent services.

B. INVESTMENT RECOMMENDATION

Investment recommendation is not a new terminology. However, a personalized investment recommendation can be called a kind of intelligent service because it can effectively reduce the choice of investors. At present, in the field of P2P online lending platforms, an investment recommendation can be divided into an intelligent recommendation and an unintelligent recommendation. The second one primarily refers to using the same investment strategy to provide recommendation solutions for all investors. For example, in order to help lenders build an optimal portfolio, Guo *et al.* [7] have proposed a portfolio optimization model with boundary constraints. Given a target rate of return, the model can provide investment decisions with minimal risk. Nevertheless, the model can only provide the optimal solution under the assumption that the number of loans at all levels (risk levels) is available for investment at any point in time. Furthermore, Ren and Malik [34] estimated the regularity of the number of loans at different time periods through statistics and analysis of historical data, and then they used this regularity to predict the probability of the loan number at all levels in the future. They modified Guo's assumption that investors will cost all their funds when making investment decisions because they believe that there might be more worthwhile loans in the future, and thus the investors reserve a part of their funds for subsequent loans. In addition, some researchers have analyzed the recommendation issues from the perspective of econometrics [35] and [36]. For example, Zhu *et al.* and [36] proposed an investment recommendation framework based on risk assessment and maximizing the residual value theory. Their model first evaluated the loan's risk and then combined the risk assessment and investment recommendation based on the residual value theory in economics in order to achieve high-yield, low-risk investment solutions for investors.

The method of providing personalized recommendation services to investors is called an intelligent service. In recent years, with the advancements in big data technology, an increasing number of research studies [37], [38] have been conducted in this field. Chang *et al.* and [39] have developed a friend network model in order to capture the mutual influence of investment behavior among investors, excavate the investment behavior and influence factor, and consider them as the indicator variable of investment interest prediction. Based on the above-mentioned indicator variables, the model then generated a candidate recommendation list. Zhang *et al.* [40] have used an improved random-walk algorithm in order to dynamically find potential investors for loans based on the financing progress and investor's investment behavior. This method only considered the impact of the user's recent investment behavior within 15 days and ignores the risk impact of the loans on investment activities. Zhao *et al.* [12] have restructured the investment decision-making process for different investors. In his paper, a user-based collaborative filtering (UCF) method was applied for generating a recommendation investment list for each investor. Then, the investment share of each loan in the recommended list was optimized according to the portfolio theory. Their model considered the differences in the current investment status of different investors and proposed a personalized recommendation strategy based on risk management. However, the above-mentioned intelligent recommendation solutions do not consider the short life cycles of loans. In other words, timeliness is not considered in these recommendation algorithms. Although there are a few studies that focus on time-aware recommendation algorithms [41], [42], they rarely consider the short life cycles of the recommended items. For example, Zhang *et al.* [41] integrated the time factor into the recommendation system by adding a time factor to the items in the recommendation list of a candidate. Ji *et al.* [42] adopted a parallel bidirectional LSTM with an attention mechanism in order to extract the short-term temporal dynamics from complex and nonlinear user-item interactions, and a time evolution model was also used to capture the long-term temporal dynamics effects. Nevertheless, these factors are not for an item that has a short product life cycle. The cold start problem caused by the lending platform continuously releasing new loans also does not attract considerable attention. However, in lending platforms, it is a common phenomenon to continuously issue new loans, which requires adopting a strategy to overcome the cold start problem caused by this phenomenon. Therefore, the recommendation system proposed in this work requires addressing these two problems satisfactorily.

III. DESCRIPTION OF OUR PROPOSED MODEL

In this section, we first formally define the studied problem. Then, we explain the DeepSeIMF model, including the framework.

A. PROBLEM STATEMENT

Consider a loan set $I = \{i_1, i_2, i_3, \ldots, i_k, i_{k+1}, \ldots, i_n\}$, whose elements are sorted into a sequence based on their creation time, where i_k denotes the loan initiated in a P2P lending platform at the time k. Let $U = \{u_1, u_2, u_3, \ldots, u_m\}$ denote the set of investors. Based on the appearance time of the investment records on a lending platform, the following matrix can be formulated, as shown in Table 1. "1" in Table 1 indicates that an investor has invested in the corresponding loan and "?" denotes that an investor has not invested in the corresponding loan. Our objective is to predict whether an investor should invest in a loan, which has not been invested by him/her, and decide whether to provide the user with an investment recommendation for the loan based on the prediction result. By analyzing the above problem description, we can conclude that there are two major difficulties. The first is that i_k might be a new loan that is just released, i.e., the columns are all "?". The second is that i_k will expire with time.

TABLE 1. Investment records.

	i_1	i_2	i_3	$\ldots$	i_k	i_{k+1}	i_{k+2}	$\ldots$
u_1	?	1	?	$\ldots$	?	?	1	$\ldots$
u_2	1	?	1	$\ldots$	1	?	?	$\ldots$
u_3	1	?	?	$\ldots$	?	?	1	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

B. DeepSeIMF FRAMEWORK

Actually, a personalized recommendation can be solved as a binary classification problem [43], [44]. However, the traditional collaborative filtering recommendation models are not suitable for the P2P lending recommendation because of the short life cycles of loans, the sparseness of interaction data, and other factors such as the default risk of loans, investment experience of investors, and so on. For modeling a complex situation, we propose a DeepSeIMF model to deal with such problems. As illustrated in Figure 2, the *DeepSeIMF* model primarily contains three components, namely, *SeIMF*, *Extraction*, and *Combination*. Specifically, the *SeIMF* component proposes a sequence-based increment matrix factorization model for tracking the timeliness and sparsity; the extraction of heterogeneous features that are relevant to the default risk of loans is performed in the second component; and the *Combination* component is designed for combining the results provided by the above two components.

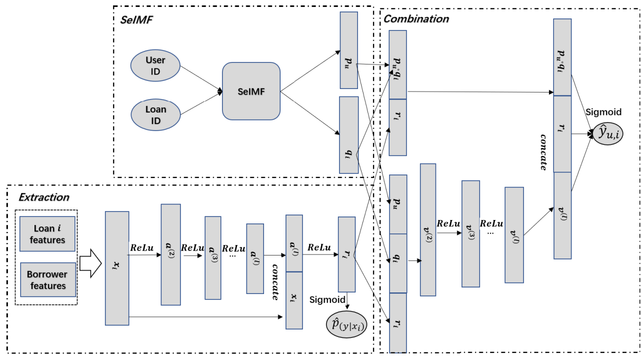

FIGURE 2. Schematic of the DeepSeIMF model.

C. INCREMENTAL MATRIX FACTORIZATION

1) MATRIX FACTORIZATION (MF)

During the Netflix Awards, one of the most popular methods for solving the collaborative filtering problem is based on matrix factorization [45]. Let d represent the size of a user's latent vector, $\mathbf{P} \in R^{(|U| \times d)}$ represent the user latent matrix, where one row of $\mathbf{P}$ represents a latent vector p_u, and $\mathbf{Q} \in R^{(|I| \times d)}$ represent the loan latent matrix, where one row of $\mathbf{Q}$ represents a latent vector q_i. MF predicts the feedback of u on i according to the inner product of p_u and q_i, that is, $\hat{y}_{ui} = p_u q_i^T$. Let $L(u, i)^{(1)}$ be the loss function.

$$L^{(1)} = \sum (y_{ui} - p_u q_i^T)^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \tag{1}$$

$\|\cdot\|$ represents the standard Euclidean norm, and the term after the parameter λ is a Tikhonov regular term used for avoiding overfitting.

$$(\mathbf{P}, \mathbf{Q}) = \arg\min_{p,q} \Rightarrow \begin{cases} \dfrac{\partial L^{(1)}}{\partial p_u} = -2\left(y_{ui} - p_u q_i^T\right) + 2\lambda p_u \\ \dfrac{\partial L^{(1)}}{\partial q_i} = -2\left(y_{ui} - q_i p_u^T\right) + 2\lambda q_i \end{cases} \tag{2}$$

Then, using stochastic gradient descent (SGD) learning method, and let the learning rate be 2η, we can get the following formulae:

$$\begin{cases} p_u \leftarrow p_u + \eta\left(\left(y_{ui} - p_u q_i{}^T\right) q_i - \lambda p_u\right) \\ q_i \leftarrow q_i + \eta\left(\left(y_{ui} - p_u q_i{}^T\right) p_u - \lambda q_i\right) \end{cases} \tag{3}$$

2) ASSUMPTION AND TRANSFORMATION TO MF

The MF model has been proven to have high accuracy and scalability [46]. However, since it is a batch learning method, it naturally requires batch processing based on static datasets. In other words, the loan set and the user set are fixed. Once the feedback data expand to a certain extent, the only way to find out new patterns is to redevelop the recommender. However, in the field of the P2P lending platform, it is normal for the new loans to appear constantly. Although the gradual accumulation of training data is a natural process, we always expect that once an investor invests in a loan, this feedback can quickly spread into our recommendation results. In other words, our recommendation algorithm requires to capture the investors' investment behavior on time in order to improve the recommendation system performance. Based on the above-mentioned problems, we design an incremental decomposition matrix factorization model in order to gradually learn from new feedback data. Before introducing this model in more details, it is necessary to introduce our assumption and transformation to MF. In contrast to the traditional scoring matrix, the matrix formulated in this work only has "1" and "?". According to the current p_u, q_i, the probability of predicting a user's investment behavior on a loan can be expressed as $\hat{y}_{ui} = \sigma\left(p_u q_i^T \middle| \mathbf{P}, \mathbf{Q}, \Theta_\sigma\right)$, where Θ_σ is a function that maps $p_u {q_i}^T$ to [0, 1]. However, there can be problems with only positive samples. Therefore, we propose an assumption that investors should not invest in loans that have already defaulted. In other words, if a loan is defaulted, all subsequent investors should not invest in the loan. For example, if the loan i_3 is defaulted, all "?" in this column should be replaced with "0'". With this reasonable assumption, the positive sample problem is solved. Although $L(u, i)^{(1)}$ can be explained by assuming that the observations follow a Gaussian distribution, it is not suitable for processing implicit data [47]. In this work, a new measure is used. Let $p(\mathbf{y})$ be the probability that all lending behaviors are correctly predicted.

$$p(\mathbf{y}) = \prod_{(u,i)\in\mathbf{y}^+} \sigma\left(p_u {q_i}^T\right) \prod_{(u,i)\in\mathbf{y}^-} \left(1 - \sigma\left(p_u {q_i}^T\right)\right) \quad (4)$$

$\mathbf{y}^+$ is a set and each element of this set represents an investor investing in a loan. $\mathbf{y}^-$ is the exact opposite of $\mathbf{y}^+$. A negative logarithm of (4) is taken to get the loss function. Let $L(u, i)^{(2)} = -\log p(y(\mathbf{x}))$ and $\sigma(x) = sigmoid(x)$, then

$$\begin{aligned} L^{(2)} = & \sum_{(u,i)\in\mathbf{y}^+} y_{ui} \log sigmoid(p_u {q_i}^T) \\ & - \sum_{(u,i)\in\mathbf{y}^-} (1 - y_{ui})(1 - \log sigmoid(p_u {q_i}^T)) \end{aligned} \quad (5)$$

$$\frac{\partial L^{(2)}}{\partial p_u} = -(y_{ui} - sigmoid(p_u {q_i}^T) q_i \quad (6)$$

Like (6) the following can be attained:

$$\frac{\partial L^{(2)}}{\partial q_i} = -(y_{ui} - sigmoid(p_u {q_i}^T) p_u \quad (7)$$

Then, according to (6) and (7):

$$\begin{cases} p_u \leftarrow p_u + \eta\left(\left(y_{ui} - sigmoid(p_u {q_i}^T)\right) q_i\right) \\ q_i \leftarrow q_i + \eta\left(\left(y_{ui} - sigmoid(p_u {q_i}^T)\right) p_u\right) \end{cases} \quad (8)$$

Finally, the SGD method is used to analyze the latent factor vectors of all investors and loans, and then the learned p_u and q_i can be used to predict the investment behavior of the investor u for the loan i.

3) SEQUENCE INCREMENT MATRIX FACTORIZATION (SeIMF)

From Equation (8), we find that during the iterative update process, p_u and q_i are continuously covered. When there are new investment records, in order to obtain the latest values of p_u and q_i, we require to train the entire training set once, and this training is very time-consuming. Moreover, in the MF process, the training set is disordered before training by applying the SGD algorithm. As a result, it is impossible to obtain p_u corresponding to the time factor. However, the historical investment behavior affects the current investment decisions over time in the P2P lending recommendations. For example, the recent investment behaviors affect the investors' current investment decisions more than the previous ones because the recent loans may form an investment portfolio with the current investment. In order to obtain the recent investment impacts, a sequence-based strategy is adopted to update p_u and q_i.

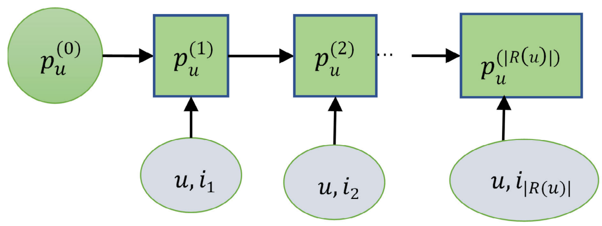

FIGURE 3. Update process for p_u.

Let $R(u) = \{(u, i_1), (u, i_2), \ldots, (u, i_{|R(u)|})\}$ represent the set of investment records of investor u on a lending platform and let $R(i) = \{(u_1, i), (u_2, i), \ldots, (u_{|R(i)|}, i)\}$ represent the set of all investment records that occurred for the loan i. A $\{(u, i_1), (u, i_2), \ldots, (u, i_{|R(u)|})\}$ sequence can be obtained by ordering $R(i)$ according to the appearance time of its elements. If the gradient descent method is used to update p_u, the process can be described by Fig. 3 and (9).

$$p_u^k = p_u^{k-1} + \eta(y_{u,k-1} - \sigma(p_u^{k-1} q_{k-1}^T)) q_{k-1}, 0 < k < |R(u)| \quad (9)$$

It can be found from Equation (9) that as the calculation advances, it becomes increasingly difficult for the previous calculation results to affect p_u (here $p_u = p_u^{(|R(u)|)}$). At the same time, the irrational investment behavior of a similar investor for a certain loan (when the loan defaults, this investment behavior can be considered irrational) also demonstrates the same effect as the rational behavior on the final recommendation list. As the MF model essentially follows the concept of collaborative filtering, it is believed that similar users are more identical because of the rational behavior, which means that in the recommendation process loans invested by investors with poor investment experience cannot be recommended to other investors. In addition, the impact of the past investment behavior of an investor on the current investment decisions will gradually decrease over time. In order to characterize this in the model, the effect function of the past investment of an investor on the current investment decisions can be expressed as follows: $f(t) = e^{(-\delta t)}$, where δ is the time sensitivity coefficient and t represents the time difference between two investments. Furthermore, $\lambda(t_m) = f(t_m)/\sum_{(k=1)}^{(|R(u)|)} f(t_k)$ indicates that the effect of investor u's (u, i_m) investment behavior on the current investment decision is affected the proportion of their all investment behavior effects. Then, the following expression is obtained:

$$p_u = p_u^{(|R(u)|)'} = \sum_{m=0}^{|R(u)|} \lambda(t_m) p_u^{(m)} e^{-\gamma r_{(u,i)}} \tag{10}$$

where $r_{(u,i)}$ indicates whether the investor u's behavior for the loan i is rational. If it is irrational, $r_{(u,i)} = 1$, otherwise $r_{(u,i)} = 0$. γ indicates the penalty sensitivity coefficient for the irrational behavior. Figure 4 illustrates the process of updating p_u. This work does not intend to obtain the calculation method for q_i as in Equation (10); first, because an investor does not have a life cycle, and second, for a specific investment loan i, all investment behaviors corresponding to it are relatively related to time. Thus, the approach to update a loan's latent vector is still $q_i \leftarrow q_i + \eta(y_{ui} - \sigma(p_u {q_i}^T))p_u$; however, here $p_u = p_u^{(k)}$ indicates the current latent vector of the investor u.

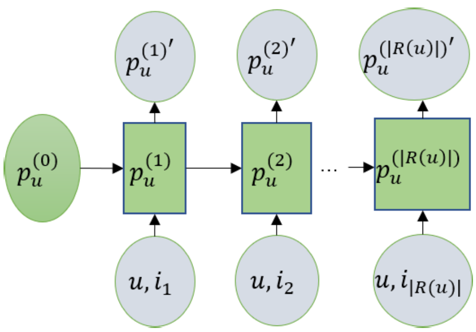

FIGURE 4. Update process for p_u.

D. EXTRACTION OF LOANS' DEFAULT RISK FACTOR

As described in the previous section, an incremental MF has been used to allow the recommendation system to respond to new investment records in time. However, the above-recommended solution cannot solve the following problem: when $R(i) = \phi$, that is, the loan i has just been released, and no investors have invested in it. We cannot update the value of q_i by using the above-described method, and it will not be possible to predict the investor's investment behavior with respect to it. Therefore, this work proposes a solution to take into account the default risk factors of a loan when q_i is missing. In addition to the investment preferences, we assume that the default risks of a loan will be considered when rational investors make an investment decision. For example, investors prefer to invest in loans initiated by borrowers with high credit ratings.

In the field of the P2P lending platform, due to the sparseness and heterogeneity of the lending data, it has become increasingly difficult to process such data by using the models that employ the traditional algorithms [48]. However, neural network technology is considered good for processing such data. Therefore, this work uses a neural network for assessing the investment risk of loans. In contrast to the risk assessment model used by the previous research scholars, this work uses a *Wide & Deep* neural network [49]. This model can combine the memorization ability of the linear model with the generalization ability of the deep neural network model. For the risk assessment task of loans, the *Wide* part of the model can be defined as follows:

$$p(y|\boldsymbol{x})_w = \sigma(\boldsymbol{wx} + b) \tag{11}$$

where $p(y|\boldsymbol{x})_w$ represents a default risk of the loan i, $\boldsymbol{w}$ represents the model parameter, b represents the model deviation, and $\boldsymbol{x} = [x_1, x_2, \ldots, x_d]$ represents the attribute value vector of the loan i. Equation (11) shows the lack of ability of the LR model to analyze nonlinear features. In order to improve the nonlinear modeling ability of the LR model, the feature combinations generated by $\beta_k(x) = \prod_{i=1}^{d} x_i^{c_{ki}}$, where $c_{ki} \in \{0, 1\}$, can be used in this model. Therefore, $\boldsymbol{x}' = [\boldsymbol{x}, \boldsymbol{\beta}(\boldsymbol{x})]$ can be considered as the input features of the LR model.

The *Deep* part of the model can be defined as follows:

$$x^{(l+1)} = ReLu(W^{(l)}x^{(l)} + b^{(l)})) \tag{12}$$

where l denotes the layer depth and ReLU denotes the activation function, and $x^{(l)}$, $W^{(l)}$, $b^{(l)}$ are the output, weight, and bias of the lth layer, respectively, and $x^{(0)} = \boldsymbol{x}$. For the categorical type of the input data, an embedding operation can be performed for mapping the sparse data into a dense space.

The above-mentioned two models can be combined as shown in Equation (13):

$$p(\hat{y}|\boldsymbol{x}) = \sigma(\boldsymbol{w}^{(l+1)} \begin{bmatrix} \boldsymbol{x}' \\ \boldsymbol{x}^{(l+1)} \end{bmatrix} + b^{(l+1)}) \tag{13}$$

Let $P = \prod_{x \in X^+} p(\hat{y}|\boldsymbol{x}) \prod_{x \in X^-} (1 - p(\hat{y}|\boldsymbol{x}))$ be the probability that all loans are correctly predicted. $\boldsymbol{X}^+$ represents a set of positive samples (here default loans are used as positive samples) and $\boldsymbol{X}^-$ represents a set of negative samples. As the target of the model is to maximize P, let's define the loss function as follows:

$$loss = -\frac{1}{N} \sum_{i=1}^{N} (y \log p(\hat{y}|\boldsymbol{x}) + (1 - y) \log(1 - p(\hat{y}|\boldsymbol{x})) \tag{14}$$

To create a cost-sensitive logistic loss [50], the original $p(\hat{y}|\boldsymbol{x})$ is replaced with

$$p(\hat{y}|\boldsymbol{x}) = \frac{1}{1 + e^{-2\delta f(x) - 2\eta}} \tag{15}$$

when $\delta = \frac{C(0,1)+C(1,0)}{2}$, and $\eta = \frac{1}{2} \log \frac{C(1,0)}{C(0,1)}$. The performance of this model is shown in Table 2. So far, in this section, the description of the risk assessment of a loan is presented. In order to incorporate the risk factors of a loan into the

TABLE 2. Features used in experiment.

Model	Accuracy	Precision
AdaBoost	0.782	0.485
Decision Tree	0.693	0.312
Gaussian Naive Bayes	0.634	0.313
Linear Discriminant Analysis	0.784	0.503
Logistic Regression	0.782	0.450
Random Forest	0.776	0.444
Wide&Deep	0.785	0.547

recommendation system, we make appropriate modifications to the Wide and Deep model. Let

$$p(\hat{y}|\boldsymbol{x}) = \sigma(\boldsymbol{h}^T ReLu(\boldsymbol{W}^{(l+1)} \begin{bmatrix} \boldsymbol{x}' \\ \boldsymbol{x}^{(l+1)} \end{bmatrix} + \boldsymbol{b}^{(l+1)})) \tag{16}$$

We define

$$\boldsymbol{r}_i = ReLu(\boldsymbol{W}^{(l+1)} \begin{bmatrix} \boldsymbol{x}' \\ \boldsymbol{x}^{(l+1)} \end{bmatrix} + \boldsymbol{b}^{(l+1)}) \tag{17}$$

Let $\boldsymbol{r}_i$ be the explicit risk vector of the loan i. In this method, the high-dimensional data related to a loan's risk is compressed into a limited dimensional space, which realizes the compression and extraction of the heterogeneous features of loans.

E. COMBINATION

The values of $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ can be obtained by an incremental MF and the default risk vector $\boldsymbol{r}_i$ of the loan i can be obtained from Equation (17). Then, the following question arises: how can we fuse $\boldsymbol{p}_u$, $\boldsymbol{q}_i$, and $\boldsymbol{r}_i$ in our model? First, the straightforward solution is to connect them as shown in Equation (18) and then input the connected result into a DNN model.

$$z_1 = \omega(\boldsymbol{p}_u, \boldsymbol{q}_i, \boldsymbol{r}_i,) = \begin{bmatrix} \boldsymbol{p}_u \\ \boldsymbol{q}_i \\ \boldsymbol{r}_i \end{bmatrix} \tag{18}$$

where $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ are the outputs of the SeIMF model and $\boldsymbol{r}_i$ is the explicit risk vector of the loan i. When $R(i) = \phi$, q_i cannot be obtained through the SeIMF model. Without loss of generality, we obtain $\boldsymbol{q}_i = \frac{1}{|\boldsymbol{I}_s|}\sum_{k=1}^{|\boldsymbol{I}_s|} \boldsymbol{q}_k, \boldsymbol{I}_s = \{i_t | i_t \in \boldsymbol{I}, t \geq k+1, \boldsymbol{q}_{i_t} \neq \boldsymbol{0}\}$. In other words, q_i is the weighted average of all loans that are currently being funded on the platform. On inputting the connected result into a logistic model, we obtain the following expression:

$$\hat{y}_{u,i} = \sigma(\boldsymbol{w}z_1 + b) \tag{19}$$

where $\hat{y}_{u,i}$ is the predicted value of the investor u's investment behavior for the loan i. This combination is called GSeIMF.

Second, considering the interaction between investors and loans, $\boldsymbol{p}_u$, $\boldsymbol{q}_i$, and $\boldsymbol{r}_i$ can be combined as expressed in Equation (20):

$$z_2 = \omega(\boldsymbol{p}_u, \boldsymbol{q}_i, \boldsymbol{r}_i,) = \begin{bmatrix} \boldsymbol{p}_u \circ \boldsymbol{q}_i \\ \boldsymbol{r}_i \end{bmatrix} \tag{20}$$

Here $\boldsymbol{p}_u \circ \boldsymbol{q}_i = [p_{u1}q_{i1}, p_{u2}q_{i2} \ldots p_{uD}q_{iD}]^T$. Then, using z_2 as the input of the DNN model, we get:

$$\begin{aligned} \hat{y}_{u,i} &= \sigma(f(\boldsymbol{w}^{(l)}\boldsymbol{v}^{(l)} + b^{(l)}) \\ \boldsymbol{v}^{(l)} &= f(\boldsymbol{W}^{(l-1)}\boldsymbol{v}^{(l-1)} + \boldsymbol{b}^{(l-1)}) \\ &\ldots \\ \boldsymbol{v}^{(2)} &= f(\boldsymbol{W}^{(1)}z_2 + \boldsymbol{b}^{(1)}) \end{aligned} \tag{21}$$

Inspired by Equation (16), Equations (19) and (21) can be combined to obtain the expression provided in Equation (22) so that our model has a certain memorization ability as well as a certain generalization ability. This combination is called DeepSeIMF.

$$\hat{y}_{u,i} = \sigma(\boldsymbol{w} \begin{bmatrix} \Phi(z_2) \\ z_1 \end{bmatrix} + b) \tag{22}$$

Here $\Phi(z_2) = \boldsymbol{v}^{(l)}$ The above process is illustrated in Figure (2). The result of $\boldsymbol{p}_u \circ \boldsymbol{q}_i$ is connected to the default risk vector of the loan i as the Wide input of the Wide and Deep model and $\boldsymbol{p}_u$, $\boldsymbol{q}_i$, and $\boldsymbol{r}_i$ are connected as the Deep input of this model. At this point, we have the final model. $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ are obtained by inputting the Ids of the investor u and the loan i into the SeIMF model, and the explicit risk vector, $\boldsymbol{r}_i$, is obtained by inputting the detailed information of the loan i and the information of its initiator into Equation (15). Then, the three vectors are inputted into Equation (21) in order to obtain the probability that the investor u will invest in the loan i. All the funded loans are processed according to the above process when recommending a candidate investment loan for an investor u. In this way, the probability that an investor u will invest in every loan can be obtained. By ranking these probability values, a final top-K recommendation list of investors can be generated.

IV. EXPERIMENT

In this section, we describe the experiments that are carried out for evaluating the performance of our approach. We introduce the experimental dataset, pre-processing procedure, and the experimental settings and report the experimental results.

A. DETAILS OF THE DATASET

The experimental dataset is collected from the Prosper platform. This dataset contains all records of loans, lenders, and pledges (i.e., bids in this dataset) of nearly six years on this platform. We primarily use three tables of these data for our experiments. The "Listing" table provides the features of the listed loans and how these loans ended (expired or succeeded). The "Member" table contains member information, such as addresses and credit levels. Note that all details of both borrowers and lenders are recorded in the "Member" table. The "Bid" table that is used to link the "Listing" and the "Member" tables in order to obtain information about who pledges to a certain loan. In our experiments, we only use those loans that have at least two pledges. Table 3 shows the statistics of this experimental dataset. Figure 5 shows the results of the statistical analysis available on the prosper

TABLE 3. Dataset description.

#Lending items	#Borrowers/ lenders	#Bidding records	Dataset size
371,896	1,309,510	9,638,888	2.3G

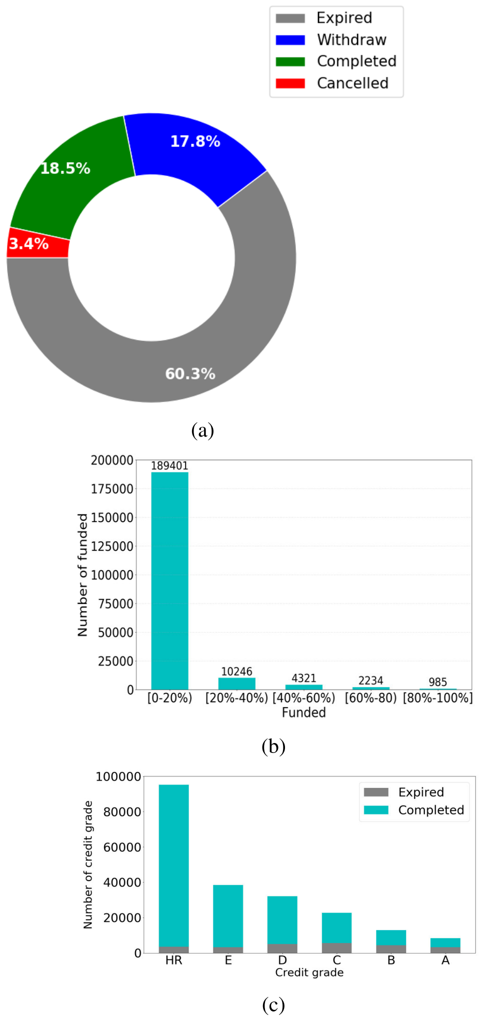

FIGURE 5. Statistical analysis.

platform. As shown in Figure 5(a), only 18.5% of the loans (completed) received enough investment on time, and at least 60.3% of the loans expired because they did not receive enough money. We analyze the amount financed for loans that failed until the expiry time. As shown in Figure 5(b), the sparseness is obvious in the dataset, with more than 60% of the failed loans receiving less than 20% of their financing target. Therefore, finding potential investors in a timely manner is crucial for loans and should be completed as soon as possible. Figure 5(c) shows the relationship between the credit rating and the success rate of a loan. As the investor's credit rating is closely related to the default risk of the loan, we can also indirectly determine the relationship between the default risk and the success rate of loans. We use oversampling to deal with our data imbalance. The features used in our method are detailed in Table 4.

TABLE 4. The features used in our method.

Feature Name	Description
UserID	Identification number of the investor
ListingKey	Identification number of the loan
ListingCreationDate	The beginning time of the transaction
CreditGrade, ProsperRating	Credit rating reflects the credit rating of customers
Term	The ultimate repayment period
LoanStatus	Loan status
ListingCategory	The listing category selected by the borrower at the time of listing
BorrowerAPR	The annual interest rate of the borrower of the loan
BorrowerRate	Loan interest rate
BorrowerState	Place of the borrower
Occupation	Occupation of the borrower
EmploymentStatus	Employment status
IsBorrowerHomeowner	A borrower will be classified as a homeowner if there is a mortgage in his credit file or if he provides documents confirming that he is the homeowner.
AmountDelinquent	Current arrears in US dollars
IncomeRange	Annual income range of the borrower
LoanMonthsSinceOrigination	The number of months after the loan was launched
LoanOriginationDate	The date on which the loan was initiated
PercentFunded	The percentage of listed capital
Investors	The number of investors financing the loan

B. PRE-PROCESSING

There is no environment available for conducting an online test. However, as the experimental data contain the detailed timestamp of each investment record, the funding process can be restored based on the timestamps, and we can simulate the real-world loans and the online recommendation procedure on the time axis.

From the Prosper dataset, it can be observed that most of the loans have a short funding period, usually not more than 15 days. Thus, $i_t^{(end)}(i_t^{(end)} = i_t^{(start)} + 15\ days)$ can be considered as the end time of the loan i, where $i_t^{(start)}$ is the time of the first investment of the loan i. We can divide the dataset by time, use the data before May 2008 as the training set, and use the remaining data as the test set. The training set can be used to complete the calculation of p_u and the learning of the models (14) and (21). Consider that an investor u is ready to invest at the time t and the set of all loans that accepts investment at the time t is represented by $\boldsymbol{S}_t = \{i | i \in \boldsymbol{I}, i_t^{(start)} < t < i_t^{(end)}\}$. We can obtain $\boldsymbol{p}_u$ by u's id and $\boldsymbol{q}_i$ by i's id ($i \in \boldsymbol{S}_t$) and input the detailed information of i into Equation (15) in order to obtain $\boldsymbol{r}_i$. Further, the three values obtained above can be fed into the already trained model (11), and the probability of the investment behavior of the investor u for the loan i can be obtained. We perform the above operations on each element in $\boldsymbol{S}_t$ to obtain their values and sort the values to obtain the top-K recommendation list $L(k)$. The SeIMF model is used to update $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ on time to prepare the next recommendation task, once the investor has completed the investment. Due to the addition of new investment records, we again need to train the model (11), when the data expand to a certain extent and finally slide forward with the time window until the test is completed.

C. EXPERIMENTAL MEASUREMENTS

1) METRICS

We evaluate the above-described methods in terms of the hit rate (HR) and the normalized discounted cumulative gain (NDCG). The HR and NDCG are two widely used metrics for the recommendation [47], [51]. The HR and NDCG are defined as follows:

$$HR@K = \frac{\sum_{n=1}^{N} h_n}{N}, NDCG@K = \frac{\sum DCG@K}{N} \quad (23)$$

where K is the length of the top-K list, and N is the total number of the test records. Before each investment, our model generates the loan list $L(k)$ of a candidate. If the loan $i \in L(k)$ is invested by an investor, then $h_n = 1$, otherwise $h_n = 0$. $DCG@K = \frac{2^{rel_l}-1}{\log_2(l+1)}$, $rel_l \in \{0, 1\}$, where l indicates the position where the loan i appears on the loan list $L(k)$; if i does not appear on the loan list $L(k)$, then $rel_l = 0$, otherwise $rel_l = 1$.

2) BASE LINE

We compared our DeepSeIMF methods (SeIMF, GSeIMF, and DeepSeIMF) with the following methods:

- **ItemPop** [52]. Items are ranked by their popularity based on the number of investments made in them. This is a non-personalized method to rank the recommendation performance.
- **UCF** [12]. The loan items, in which other investors with similar investment preferences have invested, are recommended to the investors. Due to the requirement of the situation, when calculating the investors' similarity, the similarity requires to be updated as the time window slides forward.
- **MF** [53]. Completion of (4) is estimated. The values of $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ also require to be updated as the time window slides forward.

TABLE 5. The overall performance of models.

Model	HR@10	NDCG@10
ItemPop	0.189	0.032
UCF	0.201	0.061
MF	0.251	0.089
SeIMF	0.307	0.104
GSeIFM	0.363	0.141
DeepSeIFM	**0.451**	**0.167**

D. PERFORMANCE COMPARISON

Table 5 shows the experimental results of the different methods in terms of the two evaluation metrics. We can observe that the DeepSeIMF model achieves the best performance on the dataset, SeIMF and GSeIMF work better than other models, indicating that our method is effective.

Figure 6 shows the performance of the HR@10 and NDCG@10 with respect to the number of predictive factors that are equal to the length of p_u. For UCF, we test different neighbor sizes and report the best performance.

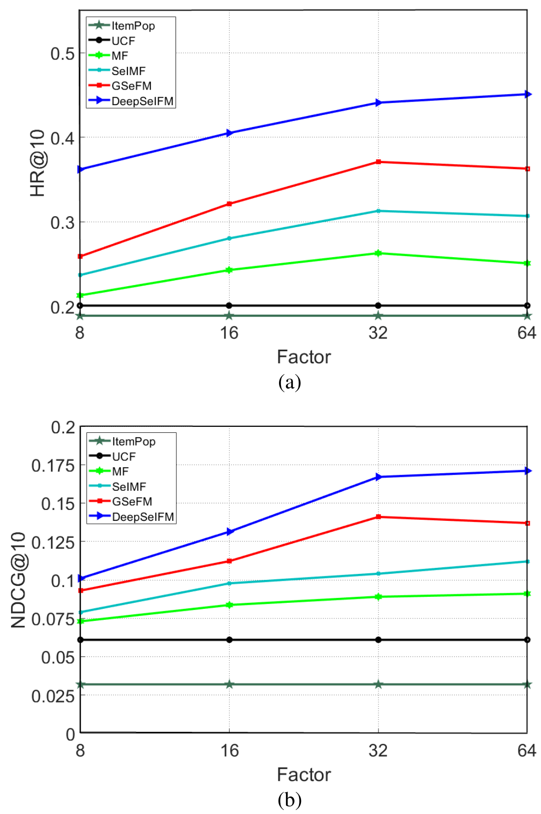

FIGURE 6. Performance of HR@10 and NDCG@10 *w.r.t.* the number of predictive factors on the prosper dataset.

We can observe that the DeepSeIMF model achieves the best performance on the dataset, which significantly outperforms the state-of-the-art method of the MF (on average, the relative improvement over MF is 14.3%). Even with a small predictive factor of 8, the DeepSeIMF model substantially outperforms the MF with a large factor of 64. The HR@10 and NDCG@10 of our proposed models and MF all increase as the factor increases. The recommendation performance also shows the same trend, that is, DeepSeIMF > GSeIMF > SeIMF > MF. This is because SeIMF takes into account time and experience, GSeIMF considers the default risk of a loan, and DeepSeIMF uses a neural network to improve its performance.

It can also be observed from Figure 7 that the HR@K and NDCG@K of all the above-mentioned models increase as K increases. This is because as the length of the candidate recommendation list increases, the probability that a loan invested by an investor will appear in the list also increases. From the figure, it can also be observed that the polyline of the DeepSeIMF model is always above other models. This indicates that our recommendation model exhibits better performance.

The size of the sliding window directly affects the number of loans being used for testing, and these are likely to form an investment portfolio with loans that have been tested before. Eventually, its influence will be observed on the HR@10 and NDCG@10. Figure 8 shows the performance

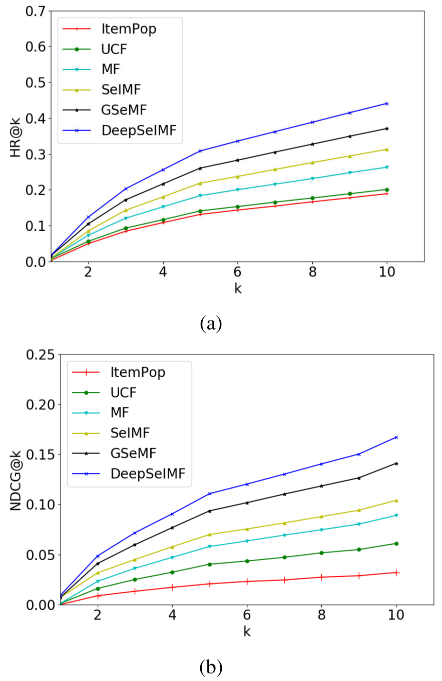

FIGURE 7. Evaluation of Top-K item recommendation where K ranges from 1 to 10 on the prosper dataset.

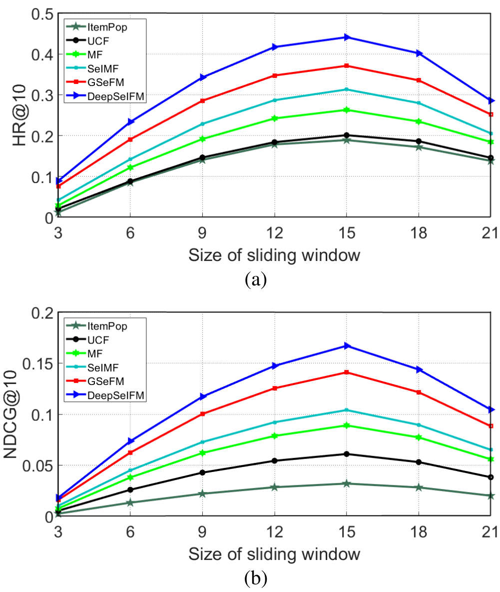

FIGURE 8. Performance of HR@10 and NDCG@10 *w.r.t.* the size of the sliding window on the prosper dataset.

of the HR@10 and NDCG@10 with respect to the size of the sliding window. The performance of all models increases with the increasing size of the sliding window. However, when the size of the sliding window increases to a certain level, the performance of all models no longer increases. Similar to Figure 7, the polyline corresponding to the DeepSeIMF model is still always above other models as shown in Figure 8.

In order to verify the performance of the models in dealing with the sparse data, we conducted another experiment.

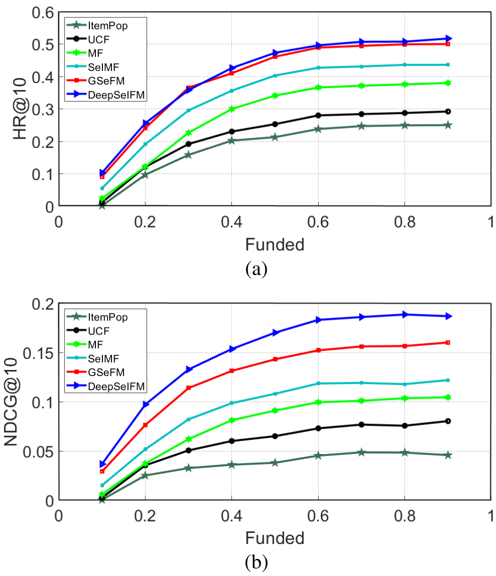

FIGURE 9. Performance of HR@10 and NDCG@10 *w.r.t.* the funded percentage of loans on the prosper dataset.

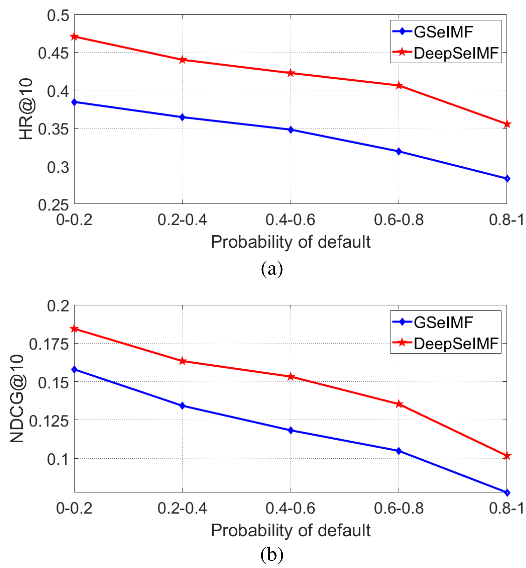

FIGURE 10. Performance of the HR@10 and NDCG@10 *w.r.t.* the probability of the default risk on the prosper dataset.

In Figure 9, the abscissa represents the funds that have been raised. From the figure, it can be observed that the DeepSeIMF model can achieve better performance as compared to other models. It should be noted that the recommended performance reduces slightly when the value of the pledge is too high. We speculate that investors wish to invest in such loans, but the target money has already been raised.

In order to verify the impact of the default risk of loans on the recommendation system, we carry out another experiment. In Figure 10, the abscissa represents the probability of default risks. From this figure, we can determine that as the probability of default risks increases, the HR@10 and NDCG@10 decrease. This shows that the default risk factors directly affect the recommendation results. When a loan is of high quality (the default risk of the loan is low), the recommendation algorithm on loan exhibits relatively better performance. Therefore, the recommendation performance of the recommendation system can be improved by enhancing the prediction performance of the default risk of the loan.

V. CONCLUSION

In this article, a holistic study on finding the next interesting investment loan for investors in P2P lending platforms is presented. In particular, a DeepSeIMF approach, which combines both collaborative filtering and the default risk of a loan, is proposed. We design a SeIMF model in order to capture the influence as the behavior of investors changes and adapted a Wide and Deep neural network in order to estimate the default risk of a loan and extract the default risk factor. Furthermore, a procedure to combine the results of the SeIMF model with the default risk of a loan is designed to improve our recommendation performance. In order to evaluate our approach, we carry out extensive experiments by using the Prosper data. The analysis and experimental results demonstrate the effectiveness of our solutions. Specifically, in addition to the great improvement in the recommendation performance, the effectiveness of dealing with the sparse data is clear from Figure 9, especially when the funded percent is low. The default risk factor is also proved useful for a better recommendation result, as shown in Figure 10.

In the future, we will apply our approach in other similar applications and evaluate its performance by using other datasets. We will also devise a marketing mechanism for loan platforms to increase the success rate of transactions.

REFERENCES

[1] M. Herzenstein, R. Andrews, U. Dholakia, and E. Lyandres, "The democratization of personal consumer loans? Determinants of success in online peer-to-peer lending communities," Work. Paper, 2008, vol. Paper.ssrn.com.

[2] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*. 2011, pp. 73–105.

[3] M. Klafft, "Online peer-to-peer lending: A lenders' perspective," in *Proc. Int. Conf. E-Learn., E-Bus., Enterprise Inf. Syst., E-Government*. EEE, 2008, pp. 371–375.

[4] C. Luo, H. Xiong, W. Zhou, Y. Guo, and G. Deng, "Enhancing investment decisions in P2P lending: An investor composition perspective," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2011, pp. 292–300.

[5] A. Byanjankar, M. Heikkila, and J. Mezei, "Predicting credit risk in peer-to-peer lending: A neural network approach," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 719–725.

[6] C. Serranocinca, B. Gutierreznieto, and L. Lopezpalacios, "Determinants of default in P2P lending," *PLoS ONE*, vol. 10, no. 10, pp. 1–22, 2015.

[7] Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong, "Instance-based credit risk assessment for investment decisions in P2P lending," *Eur. J. Oper. Res.*, vol. 249, no. 2, pp. 417–426, Mar. 2016.

[8] M. Herzenstein, R. L. Andrews, U. M. Dholakia, and E. Lyandres, "The democratization of personal consumer loans? Determinants of success in online peer-to-peer loan auctions," *Bull. Univ. Delaware*, vol. 15, no. 3, pp. 274–277, 2008.

[9] D. Shen, C. Krumme, and A. Lippman, "Follow the profit or the herd? Exploring social effects in peer-to-peer lending," in *Proc. IEEE 2nd Int. Conf. Social Comput.*, Aug. 2010, pp. 137–144.

[10] S. Ceyhan, X. Shi, and J. Leskovec, "Dynamics of bidding in a P2P lending service: Effects of herding and predicting loan success," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 547–556.

[11] R. Iyer, A. I. Khwaja, E. Luttmer, and K. Shue, "Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending?" *SSRN Electron. J.*, vol. 15242, no. rwp09-031, 2009.

[12] H. Zhao, L. Wu, Q. Liu, Y. Ge, and E. Chen, "Investment recommendation in P2P lending: A portfolio perspective with risk management," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 1109–1114.

[13] E. Weber and R. Milliman, "Perceived risk attitudes: Relating risk perception to risky choice," *Manage. Sci.*, vol. 43, pp. 123–144, Feb. 1997.

[14] R. Calcagno and C. Monticone, "Financial literacy and the demand for financial advice," *J. Banking Finance*, vol. 50, pp. 363–380, Jan. 2015.

[15] H. Zhao, Y. Ge, Q. Liu, G. Wang, E. Chen, and H. Zhang, "P2P lending survey: Platforms, recent advances and prospects," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 6, pp. 72.1–72.28, 2017.

[16] M. Brodsky, E. C. George, C. Antonelli, and M. Shtaif, "Loss of polarization entanglement in a fiber-optic system with polarization mode dispersion in one optical path," *Opt. Lett.*, vol. 36, no. 1, pp. 43–45, 2011.

[17] C. Jiang, Z. Wang, R. Wang, and Y. Ding, "Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending," *Ann. Oper. Res.*, vol. 266, nos. 1–2, pp. 511–529, Jul. 2018.

[18] Y. Xia, L. He, Y. Li, N. Liu, and Y. Ding, "Predicting loan default in peer-to-peer lending using narrative data," *J. Forecasting*, vol. 39, no. 2, pp. 260–280, Mar. 2020.

[19] M. Lin, "Peer-to-peer lending: An empirical study," in *Proc. AMCIS Doctoral Consortium*, San Francisco, CA, USA, 2009, pp. 1–7.

[20] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending," *Appl. Econ.*, vol. 47, nos. 1–3, pp. 54–70, 2015.

[21] D. Chen, H. Hu, H. Lou, and W. Yong, "Herding behavior in online microloan markets: Evidence from China," in *Proc. WHICEB*, 2015.

[22] K. A. Krumme and S. Herrero, "Lending behavior and community structure in an online peer-to-peer economic network," in *Proc. Int. Conf. Comput. Sci. Eng.*, vol. 4, Aug. 2009, pp. 613–618.

[23] S. Freedman and G. Z. Jin, "The information value of online social networks: Lessons from peer-to-peer lending," *Int. J. Ind. Org.*, vol. 51, pp. 185–222, Mar. 2017.

[24] A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems," *Neural Comput. Appl.*, vol. 19, no. 8, pp. 1165–1195, Nov. 2010.

[25] G. Dong, K. K. Lai, and J. Yen, "Credit scorecard based on logistic regression with random coefficients," *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 2463–2468, 2010.

[26] L. Bin, X. Feng, and C. Zhong, "A business oriented risk assessment model," *J. Comput. Res. Develop.*, vol. 48, no. 9, p. 1634, 2011.

[27] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 820–831, Dec. 2005.

[28] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen, "Using neural network rule extraction and decision tables for credit-risk evaluation," *Manage. Sci.*, vol. 49, no. 3, pp. 312–329, Mar. 2003.

[29] H. Zhao, Q. Liu, G. Wang, Y. Ge, and E. Chen, "Portfolio selections in P2P lending: A multi-objective perspective," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 2075–2084.

[30] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Syst. Appl.*, vol. 42, no. 10, pp. 4621–4631, Jun. 2015.

[31] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Syst. Appl.*, vol. 78, pp. 225–241, Jul. 2017.

[32] W. Li, S. Ding, Y. Chen, and S. Yang, "Heterogeneous ensemble for default prediction of peer-to-peer lending in China," *IEEE Access*, vol. 6, pp. 54396–54406, 2018.

[33] Y. Xia, J. Zhao, L. He, Y. Li, and M. Niu, "A novel tree-based dynamic heterogeneous ensemble method for credit scoring," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113615.

[34] K. Ren and A. Malik, "Recommendation engine for lower interest borrowing on peer to peer lending (P2PL) platform," 2019, *arXiv:1907.11634*. [Online]. Available: http://arxiv.org/abs/1907.11634

[35] L. Zhang, X. Wu, H. Zhao, F. Cheng, and Q. Liu, "Personalized recommendation in P2P lending based on risk-return management: A multi-objective perspective," *IEEE Trans. Big Data*, early access, May 8, 2020, doi: 10.1109/TBDATA.2020.2993446.
[36] Z. Mengying, Z. Xiaolin, and W. Chaohui, "Investment recommendation based on risk and surplus in P2P lending," *J. Comput. Res. Develop.*, vol. 53, no. 12, p. 2708, 2016.
[37] Z. Li, X. Wang, J. Li, and Q. Zhang, "Deep attributed network representation learning of complex coupling and interaction," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106618.
[38] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, and F. Xia, "Community-diversified influence maximization in social networks," *Inf. Syst.*, vol. 92, Sep. 2020, Art. no. 101522.
[39] W. Changxuan, Y. Yun, J. Tengjiao, L. Xiping, L. Guoqiong, and L. Dexi, "Personalized investment recommendation in P2P lending considering friend relationships and expected utilities of investors," *J. Comput. Res. Develop.*, vol. 55, no. 10, p. 2307, 2018.
[40] H. Zhang, H. Zhao, Q. Liu, T. Xu, E. Chen, and X. Huang, "Finding potential lenders in P2P lending: A hybrid random walk approach," *Inf. Sci.*, vol. 432, pp. 376–391, Mar. 2018.
[41] F. Zhang, Q. Liu, and A. Zeng, "Timeliness in recommender systems," *Expert Syst. Appl.*, vol. 85, pp. 270–278, Nov. 2017.
[42] Q. Ji, X. Shi, and M. Shang, "A deep temporal collaborative filtering recommendation framework via joint learning from long and short-term effects," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw. (ISPA/BDCloud/SocialCom/SustainCom)*, Dec. 2019, pp. 959–966.
[43] X. Zhang, "Building personalized recommendation system in e-commerce using association rule-based mining and classification," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 7, Aug. 2007, pp. 4113–4118.
[44] B. Wu, L. Qi, and X. Feng, "Personalized recommendation algorithm based on SVM," in *Proc. Int. Conf. Commun., Circuits Syst.*, 2007, pp. 951–953.
[45] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 426–434.
[46] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Comput.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
[47] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.
[48] H. Zhao, B. Jin, Q. Liu, Y. Ge, E. Chen, X. Zhang, and T. Xu, "Voice of charity: Prospecting the donation recurrence & donor retention in crowdfunding," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1652–1665, Aug. 2020.
[49] H. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, and R. Anil, "Wide & deep learning for recommender systems," in *Proc. 1st Workshop Deep Learn. Recommender Syst.*, 2016, pp. 7–10.
[50] Y. Xia, C. Liu, and N. Liu, "Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending," *Electron. Commerce Res. Appl.*, vol. 24, pp. 30–49, Jul./Aug. 2017.
[51] X. He, Z. He, J. Song, Z. Liu, Y.-G. Jiang, and T.-S. Chua, "NAIS: Neural attentive item similarity model for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2354–2366, Dec. 2018.
[52] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," 2012, *arXiv:1205.2618*. [Online]. Available: https://arxiv.org/abs/1205.2618
[53] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2016, pp. 549–558.

CHICHANG LIU received the bachelor's degree in information management and information system from Tianjin University of Technology, in 2019, where she is currently pursuing the master's degree in computer science and technology. She is also doing research on the application of internet finance. Her research interests include personalized recommendation systems, and data mining and applications based on internet finance. During her undergraduate studies, she won scholarships three times.

YIDONG MING received the B.S. degree in information management and information system and the M.S. degree in computer science from Tianjin University of Technology, Tianjin, China, in 2018 and 2021, respectively. He currently presides over the Tianjin Graduate Innovation Project. His research interests include recommender systems, data mining, and internet finance-based applications. During his undergraduate period, he received the First Prize of Tianjin University Students Mathematics Competition, in 2016, and the Excellent Award of Tianjin University Students Innovation and Entrepreneurship Competition, in 2017.

YINGYUAN XIAO received the Ph.D. degree in computer science from the Huazhong University of Science and Technology, China, in 2005. He is currently a Professor with the School of Computer Science and Engineering, Tianjin University of Technology, China. He was a Visiting Scholar with the School of Computing, National University of Singapore, from March 2009 to April 2010. His research interests include personalized recommender systems, advanced databases, and machine learning. He has published more than 120 journal and conference papers in these areas, including IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, *Future Generation Computer Systems*, *Engineering Applications of Artificial Intelligence*, *Information Processing Letters*, *Journal of Classification*, *Soft Computing*, WWW, DASFAA, ICWS, and DEXA. He served as the Program Chair for WAIM 2013 International Workshop and a Program Committee Member for a number of international conferences, including APWeb, APSCC, FSKD, and IEEE CloudCom.

WENGUANG ZHENG received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, Sichuan, China, in 2010, the M.S. degree from the National University of Singapore, in 2011, and the Ph.D. degree from the University of New South Wales, Sydney, Australia, in 2016. He is currently an Associate Professor with the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China. His research interests include computer architecture, embedded systems, and machine learning. He has published journal and conference papers in these areas, including LCTES, TACO, and TECS.

CHING-HSIEN HSU (Senior Member, IEEE) is currently the Chair Professor and the Dean of the College of Information and Electrical Engineering, Asia University, Taiwan, and a Professor with the Department of Computer Science and Information Engineering, National Chung Cheng University. He has been acting as the author/coauthor or an editor/co-editor of ten books from Elsevier, Springer, IGI Global, World Scientific, and McGraw-Hill. His research interests include high-performance computing, cloud computing, parallel and distributed systems, big data analytics, and ubiquitous/pervasive computing and intelligence. He has published 200 articles in top journals, such as IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Services Computing, ACM TOMM, IEEE Transactions on Cloud Computing, IEEE Transactions on Emerging Topics in Computing, IEEE Systems Journal, and *IEEE Network*, top conference proceedings, and book chapters in these areas. He is a fellow of the Institution of Engineering and Technology.