Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)

DeepCredit: Exploiting User Cickstream for Loan Risk Prediction in P2P Lending

Zhi Yang,[§*] **Yusi Zhang,**[§*] **Binghui Guo,**[†] **Ben Y. Zhao,**[‡] **Yafei Dai**[§]
[§]Peking University, [†] Beihang University, [‡]University of Chicago.
{yangzhi, zhangyusi, dyf}@pku.edu.cn, guobinghui@buaa.edu.cn, ravenben@cs.uchicago.edu

Abstract

Peer-to-peer (P2P) lending or crowdlending, is a recent innovation allows a group of individual or institutional lenders to lend funds to individuals or businesses in return for interest payment on top of capital repayments. The rapid growth of P2P lending marketplaces has heightened the need to develop a support system to help lenders make sound lending decisions. But realizing such system is challenging in the absence of formal credit data used by the banking sector. In this paper, we attempt to explore the possible connections between user credit risk and how users behave in the lending sites. We present the first analysis of user detailed clickstream data from a large P2P lending provider. Our analysis reveals that the users' sequences of repayment histories and financial activities in the lending site, have significant predictive value for their future loan repayments. In the light of this, we propose a deep architecture named DeepCredit, to automatically acquire the knowledge of credit risk from the sequences of activities that users conduct on the site. Experiments on our large-scale real-world dataset show that our model generates a high accuracy in predicting both loan delinquency and default, and significantly outperforms a number of baselines and competitive alternatives.

Introduction

Online P2P lending has grown significantly over the recent years according to the alternative finance industry reports across the world. In 2015, $2.3 billion of finance was originated through P2P lending in the UK alone, helping cement online marketplaces as part of the financial mainstream. For the US, P2P lending remained the largest alternative finance market segment with $21 billion recorded in 2016. In China, by the end of 2011, 50 P2P lending providers were reported to be operating and the number had climbed to over 2400 providers by the end of 2016, with more than $100 billion outstanding loans, which makes China's P2P lending sector the largest in the world.

Compared with the traditional financial loans, P2P platforms cut off the role of financial intermediaries such as banks. Thus, in addition to providing a media for trading, these platforms also have the responsibility to manage risk and protect the profits of lenders. Unfortunately, the lending platforms are hardly able to get formal credit information used by banks, which makes the risk assessment in these platforms more difficult. Most lending sites rely on the credit scores provided by third-part agencies, such as FICO scores (Corporation 2017) from Fair Isaac Corporation and VantageScore (VantageScore Solutions 2017). Although applicants with higher credit scores are less likely to default, credit scores cannot directly and accurately predict individual loan defaults. Also in certain countries such as China, there's no unified credit scoring system even in conventional commercial banks.

The main contribution of this paper is that we propose a new risk-assessment strategy, which we call the *clickstream*-based model, to infer the risk of a loan from the activities that users conduct on the site. A clickstream is the sequence of HTTP requests made by a user to the P2P lending site, where the majority of which are financial activities with specific monetary objectives (such as borrow/lend, repay, deposit/withdraw and money transfer). Through studying the unique clickstream data from one of the largest P2P platforms in China, we uncover that the risk for a user's upcoming loan is highly correlated with his repayment/debt history (e.g., past payment outcomes and unpaid debt) and the types and sequence of his recent financial activities indicating cash flow. For instance, the activity transitions indicating a high demand for money and outward transfer of money from the lending platform can be signal for the risk.

The informativeness of the clickstream data provides a new opportunity to tailor risk assessment to the individual loans. However, modeling such a fine-grained user behavior dynamics brings up several challenges:

- **Temporal dynamics:** The risk of individual loans depends on not only the long-term financial profile of the user but also short-term financial distress. It is important to explore the time information in clickstream to identify certain impending risk conditions.
- **Information integration:** An effective strategy is needed to explore feature interactions between the repayments on loans and financial activity sequences which contain the complementary information reflecting a person's financial state over time (e.g., debt and cash flow dynamics).
- **Ground-truth missing:** Although a user's recent repay-

[*]These authors contributed equally to this work.

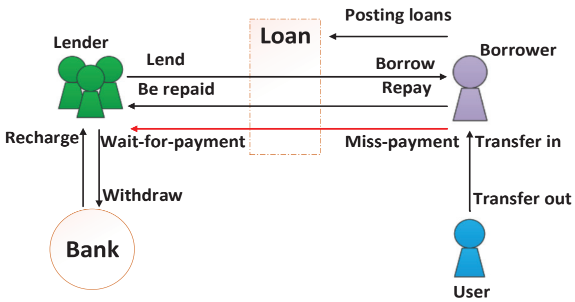


Figure 1: Illustration of financial activity types.

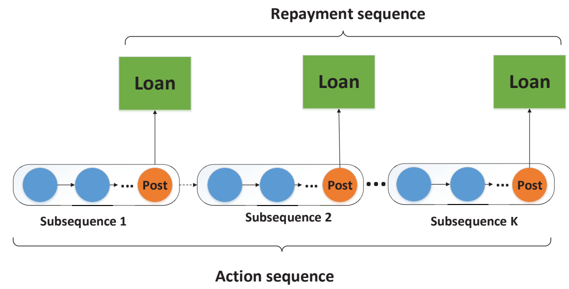


Figure 2: Illustration of action sequence partition

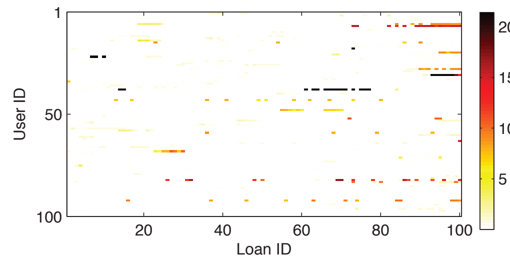


Figure 3: The repayment delay sequence of delinquent users, where each point (i, j) gives the repayment delay (in week) for j_{th} loan of user i.

ment history provides strong sign of his ability and willingness to repay, when we make prediction on a loan, we cannot known the repayment outcomes of previous loans which are not yet due for payment.

To address the aforementioned challenges, we propose a new Long Short-Term Memory (LSTM) variant, time-aware LSTM (TLSTM), to better capture a user's short-term and long-term dynamics at the same time. By utilizing the proposed TLSTM unit, we propose a novel deep architecture, named DeepCredit. It uses a two-layer TLSTM where the repayment and financial activity sequences are separately encoded by the first-layer ones, and then integrated together by the second-layer one. This hierarchical architecture can efficiently learn the complementary information of two types of sequences over time. Meanwhile, the model also uses feed-forward neural networks (FCN) to concurrently generate predictions on past loans as remedy for the incomplete ground-truth. These intermediary predictions can also be thought of as a teacher supervising for effectively integrate the repayment and activity information at the corresponding step.

We conduct experiments on large-scale real dataset showing that, in terms of area under the curve (AUC), the DeepCredit can achieve 0.87 and 0.90 predication accuracy on delinquency and default, respectively. Further, we find the prediction errors of our model are highly acceptable. In the case of false negative, our model tend to miss less serious delinquency with few days late in the repayment, which are more tolerable. In the case of false positive, our model tend to incorrectly classify the loan paid by incurring another with a higher interest rate. Despite these loans were repaid on time, the corresponding users actually become riskier in the near future due to struggling with more debt.

Data Analysis

In this section, we examine the feasibility of using clickstream data gathered from the P2P lending sites to predict the risk of individual loans. For a given loan, lenders are especially concerned with the borrower's willingness to repay and ability to repay, respectively. So we characterize the loan risk in terms of delinquency (late payment) and default (capital loss). In particular, a loan becomes *delinquent* when the payment is not made by the due date, and goes into *default* when the payment has been missed for an extended period of time. We use the Peer-to-Peer Finance Association (P2PFA)'s definition of a default (120+ days delinquency).

Clickstream Data

Our study is based on the complete clickstreams for 10K users from an anonymous platform, which is one of the largest P2P lending provider in China. The users in our dataset are selected at random from the population who have posted at least one loan in March, 2017. Each click is characterized by a timestamp, an anonymized userID, an activity and related variables. The activity describes the action the user is undertaking. For example, the "post" activity corresponds to a posting-loan request, with the variables of loan amount, interest rate and period. Since we are attempting to the evaluate the financial status of users, we restrict our attention to those financial activities involving the money-amount variable. With this focus, our dataset summarizes 4,881,056 financial activities for these 10K users, of which 712,857 are loan requests.

Figure 1 illustrates the typical financial activity. A user u can *post* a loan request and *borrow* money from lenders who are willing to *lend*. When the loan comes due, the borrower can *repay* the loan, or otherwise, trigger the *miss-payment* action. In addition, a user can directly *transfer* in/out money from/to other users without necessarily being associated to a loan. Also, a users can *recharge* the money into his platform account from his bank accounts for investment or repayment purpose, and can *withdraw* money from platform account into bank account. So a clickstream is able to capture a user's

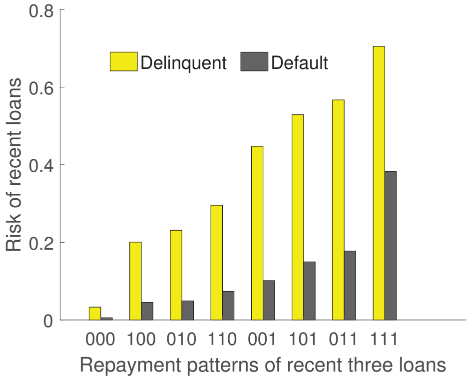


Figure 4: The correlation of past loan repayments with future repayment, where $0/1$ indicates an on-time repayment/delinquency on the corresponding loan.

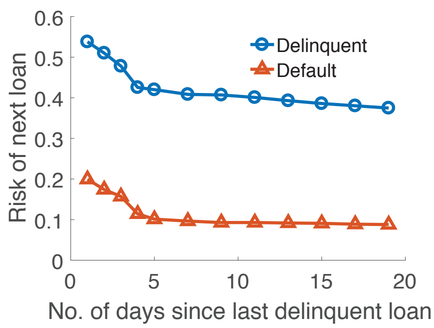


Figure 5: The effect of the temporal factor on loan risk.

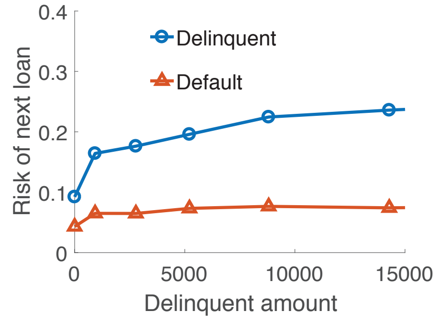


Figure 6: The effect of the Delinquent amount (RMB) on loan risk.

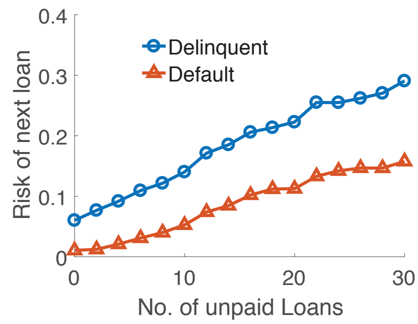


Figure 7: The effect of the unpaid loans on loan risk.

cash flow and debt dynamics (e.g., unpaid debt and overdue loan amount) which is informative for user financial state and potential risk.

In order to analyze the correlation between action patterns and individual loan risk, we first investigate how the borrower has performed on loan repayment in the past. Also, for each user, we partition the clickstream into a set of subsequences S, where each $\mathbf{s_k} \in S$ contains the activities a user engages in during the period between posting $(k-1)_{th}$ loan and k_{th} loan, as shown in Figure 2. We choose *post* action as the partition point because we need to predict a user's risk right after he posting a new loan. This partition helps us to examine if there exists differences across the sequences of user activities before they post loans with different final outcomes (e.g., on-time repay or delinquency).

Repayment Sequences

To gain a basic understanding, we pick 100 users that have delinquent loans in the past. For each user, we look at his repayment delays on past loans arranged in ascending order of the loan posting time, as shown in Figure 3. We see that delinquent loans tend to be *bursty* in the sense that there is a local concentration of multiple late repayments in the sequence. An intuitive explanation of this burstiness is that users are unlikely to repay several temporal-close loans in the status of financial distress. The figure also shows that loans with relatively longer repayment delay are more likely to be concentrated, implying that users are more difficult to recover from severe financial distress. The delinquency concentration highlights the feasibility of using the payment history and associated delinquency variables to predict future loan repayments.

Delinquency pattern. The delinquency pattern of a user can be viewed as a sequence consisting the binary delinquency outcomes for his previous loans arranged in ascending order of their posting time. To examine the correlation between the delinquency pattern and the risk on the next loan, Figure 4 shows the probability of finding users with the same delinquency pattern across the past three loans missing the repayment on the next loan. We see that both recentness and frequencies of past delinquencies indicate future repayments. The new loans posted by users with more recent delinquencies are more likely to go into delinquency and default (inertia). Intuitively, the late payments on recent loans implies that users are experiencing financial problems, leading to continued risk during the follow-up period.

Time since last delinquency. To examine the effect of temporal dynamics, Figure 5 shows the user risk given the time since his last delinquency. Not surprisingly, we see the more recent the user's last delinquency, the greater the risk. However, we also notice that the risk probability decreases slow-

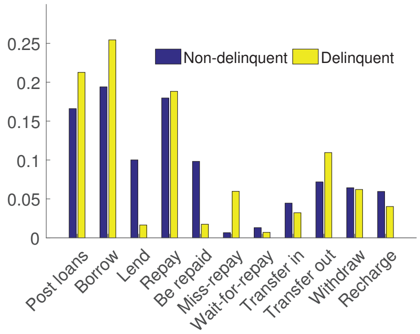


Figure 8: The fraction of each activity in each type of subsequence of activities.

ly after the initial rapid decline, which implies that the past delinquency has long-term effects.

*Debt Dynamics.*We next examine the effect of current debt (including unpaid debts and delinquent loan amount) on loan risk. Figure 6 and Figure 7 show the risk of the next loan grows along with the current delinquent and unpaid debt. Here the probability is computed by the fraction of delinquent loans among those with the similar debt amount. We see that the debt-strapped users tend to have high risk, implying that more serious state in financial distress can have long-lasting repercussions.

Activity Sequences

For action sequence data, we first classify each subsequence $\mathbf{s_k}$ into "delinquent" or "non-delinquent" group based on whether the k_{th} loan will be delinquent or not finally, and analyze the inter-group differences. We focus on delinquency which includes the case of default.

Frequencies. Figure 8 shows the distribution of activities in each group, and we see that *posting loans*, *borrow* and *repay* are the three most prevalent activities in both groups. But we still observe several apparent differences. First, users *borrow* more frequently before posting delinquent loans than before posting non-delinquent ones, indicating a higher demand of money. By contrast, users *lend* more frequently before posting non-delinquent loans, which means they have good financial status at that period. Second, we see that users have more activities with the outward transfer of money from his account before delinquent loans, such as *transfer out* and *withdraw*, whereas performing less activities with the inward transfer of money, such as *be-repaid*, *transfer in* and *recharge*. Finally, users have more *miss-payment* activities before delinquent loans, indicating the higher likelihood of experiencing financial stress. The above differences can be signal of risk.

Sequences. To understand differences in the activity ordering, Figure 9 shows the *ratio* of 2-gram transition probability in the non-delinquent subsequence to that in the delinquent ones for all pairs of activities, where we take the logarithm of the ratio for better plotting. A color pixel at (x, y) represents the log-ratio of the transition probability from activity y in the horizontal axis to activity x in the vertical axis.

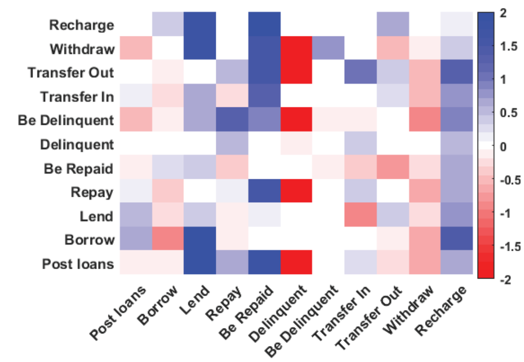


Figure 9: The ratio of activity transition probability before non-delinquent loans to that before delinquent ones.

The color blue indicates this transition is more common in activity subsequence prior to posting non-delinquent loans and color red indicates the opposite. By looking the transitions of user activities, the differences between activity subsequences prior to posting delinquent and non-delinquent loans become more apparent. The more activity transitions associated with demand for money and outward transfer of money can be signals for the higher risk of delinquency, such as *post*&*post*, *post* &*withdraw* and *withdraw*&*transfer out*, etc, indicating that the user may fall into financial crisis and need money urgently. By contrast, the more transitions associated with investment purpose or inward transfer of money can be signal for the opposite, such as *recharge*&*lend*, *withdraw*&*lend* and *recharge*&*be-repaid*, etc, indicating a good economic situation.

DeepCredit Model

In this section, we introduce our *DeepCredit* model, which takes the sequences of previous repayments and financial activities as input and generates the delinquent/default probability for the new loan before it is funded as output. The proposed architecture consists of the following components.

Input Layer

Given a new loan posted by a user, we use the repayment sequence $\mathbf{P} = \{\mathbf{p}_1, \ldots, \mathbf{p}_T\}$ capturing repayments across the most recent T loans (including the new loan itself) and the activity subsequences $\{\mathbf{A}_i | i = 1 \ldots T\}$ capturing actions before posting each of these loans as input. The repayment sequence and activity subsequences are arranged in ascending order of the loan posting time and action time, respecitvely (as illustrated in Figure 2).

In the repayment sequence, each $p_i \in P$ is a vector contains the following variables related to a loan i:

- *Loan features* including the loan amount, interest rate and loan term in days.
- *Delinquency feature* indicating the binary delinquency outcome for the loan.

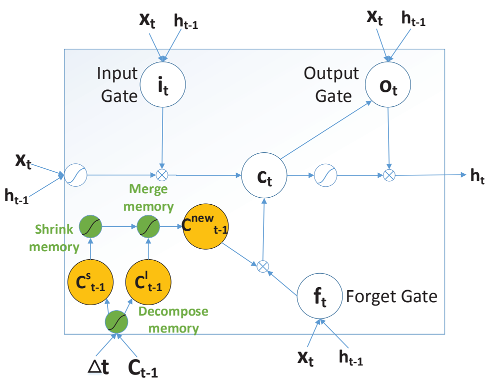


Figure 10: The architecture of time-aware LSTM (TLSTM).

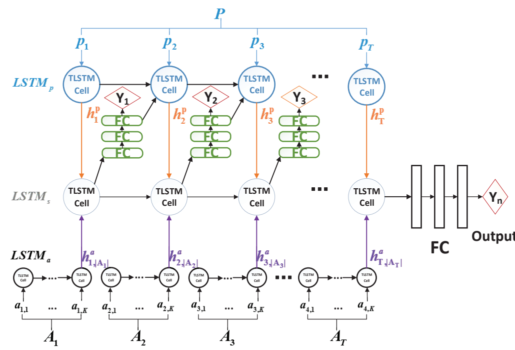


Figure 11: The architecture of H-LSTM model.

- *Debt features* including the unpaid debts and the delinquent loan amount at that time.
- *Temporal feature* gives the time gap between this loan and the previous loan.

Also, any $\mathbf{a}_{i,j} \in \mathbf{A}_i = \{\mathbf{a}_{i,1}, \ldots, \mathbf{a}_{i,|\mathbf{A}_i|}\}$ is a vector containing the activity type, involved money amount and the time interval to the previous activity.

Time-aware LSTM (TLSTM)

Recurrent Neural Networks (RNN) models learn to map input sequences to output sequence via a continuous vector-valued intermediate hidden state. The most basic RNN are difficult to train due to the so-called vanishing and exploding gradient problem. To tackle this problem, the more complex Long Short-Term Memory (LSTM) RNN were designed. However, a standard LSTM unit has the implicit assumption of uniformly distributed elapsed time between the elements of a sequence. However, in P2P lending, the distribution of action time intervals is highly irregular varying from seconds to months. So the intervals are important to capture the relations of user actions, e.g. actions with short time intervals tend to be related and indicate user's short-term risk.

There are several studies in the literature studying the time LSTM. For example, the authors of (Zhu et al. 2017) adds time gates to decay the past cell memory based on the time intervals. But this model decay all the previous cell memory, which is not suitable in P2P lending risk prediction, where we do not want to lose the global risk profile of the user. For example, the past delinquency of a user should not be discarded entirely over time, as the past delinquency has long-term effects (see Figure 5).The prior work (Baytas et al. 2017) split the cell memory in a linear way, and use a pre-set function to adjust the memory afterwards. Although this model can solve the above problem, the pre-set adjust function and linear split are not well adapted to the data and thus limit the capacity of the model,besides, after decomposing the previous memory, it lacks an effective combination method to get the new overall memory.

We propose a new time-aware LSTM variant, TLSTM, which is illustrated in Figure 10 and its mathematical expressions is given below:

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)\\
g_t &= \tanh(W_{xg}x_t + W_{hc}h_{t-1} + b_g)\\
c_{t-1}^{short} &= \tanh(W_{short_c}c_{t-1} + w_{short_t}\Delta t + b_{short_c})\\
c_{t-1}^{long} &= \tanh(W_{long_c}c_{t-1} + w_{long_t}\Delta t + b_{long_c})\\
c_{t-1}^{short_{new}} &= \tanh(w_{shrink}\Delta t + b_{shrink})c_{t-1}^{short}\\
c_{t-1}^{new} &= \tanh(w_{short}^{new}c_{t-1}^{short_{new}} + w_{long}^{new}c_{t-1}^{long} + b_{merge})\\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t\\
h_t &= o_t \odot \tanh(c_t)
\end{aligned} \tag{1}$$

where $\mathbf{W}_{x*}$ is the transformation matrix from the input to LSTM states, $\mathbf{W}_{h*}$ is the recurrent transformation matrix between the recurrent states $\mathbf{h}_t$, and Δt is the elapsed time between x_{t-1} and x_t.

In every step, the TLSTM cell decomposes the previous memory c_{t-1} into the short-term memory c_{t-1}^{short} and the long-term memory c_{t-1}^{long} based on Δt. Then the short-term memory is discounted according to Δt and we get the new short-term memory $c_{t-1}^{short_{new}}$. Intuitively, given a larger Δt, a more portion of memory could be allocated to the short-term memory for discounting. Finally, we use a nonlinear function to merge the $c_{t-1}^{short_{new}}$ with c_{t-1}^{long} as new previous memory c_{t-1}^{new}, which is feed into the following LSTM process. All the decomposing, discounting and merging functions are learned in the training process through Back Propagation Through Time (BPTT). By doing so the information contained in the memory of previous time step is adjusted so that it can discount the short-term effect, meanwhile maintaining the long-term effect. We denoted the above equation (1) as TLSTM(.) function.

Hierarchical TLSTM Model (H-TLSTM).

Next, we propose our H-TLSTM model to incorporate both repayment and action sequence data, as illustrated in Figure 11.

In the first-level TLSTM layer, our model uses a TLSTM function to map the repayment sequence:

$$\mathbf{h}_i^p = \text{TLSTM}\left(\mathbf{h}_{i-1}^p, \mathbf{p}_i\right), \forall i \in \{1, \ldots, T\}, \tag{2}$$

where $\mathbf{h}_i^p \in \mathcal{R}^D$ can be considered as a representation of the input sequence from repayment $\mathbf{p}_1$ to $\mathbf{p}_i$.

Meanwhile, in order to alleviate the long-range dependencies in sequence modeling, our model uses an independent TLSTM for each activity subsequence, so we have:

$$\mathbf{h}_{i,j}^a = \text{TLSTM}\left(\mathbf{h}_{i,j-1}^a, \mathbf{a}_{i,j}\right), \forall \mathbf{a}_{i,j} \in \mathbf{A}_i, \tag{3}$$

where the last hidden state $\mathbf{h}_{i,|\mathbf{A}_i|}^a$ summarizes the information about all the activities in the subsequence $\mathbf{A}_i$.

Then the model adopts another TLSTM to integrate the first-level representations of the past repayments and activities in a higher-level perspective. This allows the model to efficiently capture interactions among the features of two sequences across multiple loans, and thus yielding more accurate predictions. In particular, we feed the concatenation of the first-level hidden vectors into state-level LSTM layer,

$$\mathbf{h}_i = \text{TLSTM}\left(\mathbf{h}_{i-1}^s, \left[\mathbf{h}_i^p; \mathbf{h}_{i,|\mathbf{A}_i|}^a\right]^\top\right), \forall i \in \{1, \ldots, T\}, \tag{4}$$

where $\mathbf{h}_i^s$ represents a summary of repayment and activity histories up to the i_{th} loan, and the final hidden state $\mathbf{h}_T^s$ can be considered as a compact representation of the whole input history.

After the above two-stage encoding, the model is able to track the change of financial state with the complementary information from both sequences. We feed the final hidden state $\mathbf{h}_T^s$ of this high-level LSTM into FC layer to predict the risk of the current loan (i.e., the T_{th} loan).

Complementary Layer (CL)

As we have discussed in section , the delinquency outcomes of previous loans (especially those adjacent to the current loan) have significant predictive value. However, when we make prediction on a loan of a user, we cannot known whether some loans he posted previously will be delinquent or not (i.e., delinquency outcome) if they are not yet due for payment, referred to as undue loans. For example, Figure 12 shows the cumulative distribution of the number of undue past loans when a user posts the new loan in our dataset. We see that in most cases, the ground-truth of loan delinquency outcome in the repayment sequence is incomplete. Thus, simply ignoring the delinquency information of undue past loans can impair the prediction accuracy on the current loan.

Our approach to this problem is to reuse the model's predictions on previous loans as a remedy for their incomplete delinquency information. To do so, an straightforward strategy is to directly replace the unobserved delinquency outcomes of previous undue loans in the repayment sequence with their predicted delinquent probabilities. But we find this strategy is quite inefficient due to the need of model retraining and error amplification . In particular, as the predictions of previous loans are unavailable during the training initially, we need to arrange a user's undue loans in time

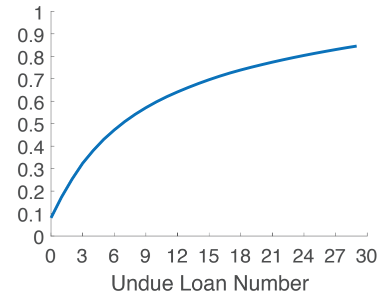


Figure 12: The CDF of the number of undue loans when a new loan is posted.

ascending order, and get their predicted probabilities in the one-by-one manner. The retraining procedure should be repeated until the predictions on all undue loans are obtained and thus could be up to 30 times according to Figure 12. Moreover, each retraining procedure only uses limited loan cases for training, which might incur a relatively large error and increase the error in subsequent prediction.

To address the above limitations, we add H-TLSTM model a complementary layer (CL) containing multiple fully-connected feed-forward neural networks (FCN) to predict the risk of current and past loans concurrently. As illustrated in Figure 11, for each previous loan i, we use a FCN which takes the corresponding hidden state $\mathbf{h}_i^s$ of the state-level T-LSTM as input and outputs the delinquent probability Y_i of loan i. To boost our prediction, we feed each delinquency probability Y_{i-1} into the encoding of repayment sequence by rewriting the equation (2) as:

$$\mathbf{h}_i^p = \text{TLSTM}\left(\mathbf{h}_{i-1}^p, [\mathbf{p}_i; Y_{i-1}]^\top\right), \forall i \in \{1, 2, \ldots, T\}, \tag{5}$$

where $\mathbf{h}_i^p$ contains the summary of delinquency probabilities of previous $i-1$ loans and can be a remedy for their incomplete delinquency information.

In this way, the model avoids retraining and error amplification due to concurrent predictions where the full repayment and activity sequences are exploited even for the prediction of each past loan. Also, these complementary predictions can be thought of as a teacher supervising for learning the integration of repayment and activity states at the corresponding step.

Training

Once a user posts a new loan (before it is funded), our model predicts if the repayment delay d on this loan is larger than a certain threshold δ, i.e., $Pr(d > \delta)$. For example, the model predicts the delinquency and default if we set threshold $\delta = 0$ and $\delta = 120$ days, respectively.

Recall that a user's recent T loans are arranged in ascending order of posting time and T_{th} loan is the one newly posted (prediction target). Let $\boldsymbol{\theta}$ represent the parameters of the deep neural network. Given the input repayment sequence $\mathbf{P}$ and activity subsequence set $\mathbf{A} = \{\mathbf{A}_i | i = 1 \ldots T\}$, the task is specified to find a function $f(\mathbf{P}, \mathbf{A}, \boldsymbol{\theta}, \delta)$ to predict the repayment outcomes of these T loans. Note here the predictions on the previous $T-1$ loans are used as complementary information for the prediction of the T_{th} loan.

Let L be the log loss function and o_i is the repayment ground-truth of i_{th} loan. Our model parameters are learned by minimizing the following loss function:

$$
\begin{aligned}
J(\boldsymbol{\theta}) &= L(f(\mathbf{P}, \mathbf{A}, \boldsymbol{\theta}, \delta)_T, o_T) \\
&\quad + \frac{\alpha \sum_{i=1}^{T-1} L(f(\mathbf{P}, \mathbf{A}, \boldsymbol{\theta}, \delta)_i, o_i)}{T-1} + \lambda ||\boldsymbol{\theta}||^2,
\end{aligned} \tag{6}
$$

where $f(\mathbf{P}, \mathbf{A}, \boldsymbol{\theta}, \delta)_i$ gives the model's prediction on the i_{th} $(i \leq T)$ loan, and α captures how the model weighs the relative importance of prediction accuracy on previous loans.

Evaluation

In this section, we first introduce the settings of our experiments, and then show the experiment results with further analysis on prediction errors and model parameters.

Experiment Setup

We perform the evaluation on the large-scale dataset we analyzed before, where we remove the loan requests which have not get funded at all. It leaves us with a total of 586,957 funded loans, of which 78,652 (13.4%) are delinquent and 28,174 (4.8%) are default. We separate out the user clickstream data into a training set and a test set in terms of users, i.e., 80% users contribute their sequences for training and the remaining 20% for testing. In this way, we ensure that the historical information of users in the test dataset is not used in the training. We conduct 5-fold cross validation on the training dataset and test the model against the test dataset.

The implementation [1] is completed using Tensorflow. All LSTM networks contain 256-dimensional hidden states, and every FCN has three layers, containing 256, 128, 64 nodes, respectively. We empirically set other parameters, including the learning rate, momentum, and minibatch size, which are 0.001, 0.9 and 128, respectively. We discuss the setting of parameter T (the length of repayment sequences) and weight parameter α in loss function (6) in Section . We evaluate the performance of the prediction based on area under ROC curve (AUC) metric, which can better handle class imbalance in our dataset.

Baselines

In addition to compare with conventional methods such as gradient boosting decision tree (GBDT) or hidden Markov model (HMM), we also evaluate the importance of each design choice by comparing our complete DeepCredit model (H-TLSTM + CL) with those variants removing one component or disabling one type of input data. The compared methods are summarized as follows:

- **H-TLSTM:** The model uses hierarchical TLSTM Model, with the complementary layer (CL) removed
- **H-LSTM+CL:** The model replaces TLSTM with traditional LSTM..

[1] Code:http://net.pku.edu.cn/p2p/doku.php?id=p2p:deepcredit

Table 1: Experimental results

	AUC	
Model	delinquent	default
DeepCredit (H-TLSTM + CL)	**0.870**	**0.901**
H-TLSTM	0.843	0.868
H-LSTM+CL	0.830	0.866
TLSTM (activity)	0.815	0.847
TLSTM (repay)	0.811	0.841
XGB	0.804	0.836
HMM	0.744	0.788
LR	0.720	0.756

- **TLSTM (activity):** The single TLSTM model only takes the activity sequence as input.
- **TLSTM (repay):** The single TLSTM model only takes the repayment sequences as input.
- **XGB:** The implementation of a tree boosting system with XGBoost, where we take the each variable in the elements of repayment and activity sequences as the model feature.
- **HMM:** We use a hidden Markov model (HMM) based approach which takes both the activity and repayment sequences as observed data. For each posting loan action, we label it as non-delinquent or delinquent (default) based on the repayment on the loan, then we use Baum-Welch algorithm to train the model and predict the label of upcoming requests of posting loans.
- **Logistic Regression:** We use Logistic Regression (LR) as a representative of the linear model, with the same input setting as XGB.

Prediction Performance

Experimental results are shown in Table 1. We see that the proposed DeepCredit achieves the best performance on the prediction of both individual delinquencies and defaults, with the AUC scores of 0.87 and 0.90, respectively. The results demonstrate the feasibility of using behavior dynamics collected from the P2P platform itself to predict the risk of individual loans. More interestingly, we see that the model yields a higher accuracy in default prediction than in delinquency prediction. As default indicates a serious delinquency, behavioural signals are probably more apparent before the borrower goes into default.

It is also found that DeepCredit outperforms HMM ,XGB and LR significantly. The underlying reason is that all the three baselines are not able to fully explore the complementary information and long range dependencies over the entire sequence histories, especially the LR model, when dealing with such variety of complex patterns, the linear transformation limits its capability. By contrast, our model uses TLSTM models which allows an effective integration of different sequences and capture the long-term dependencies between inputs due to their powerful hidden representation.

Among the DeepCredit variations, the improvement of DeepCredit over H-TLSTM and H-LSTM+CL reveals the potential benefit of leveraging intermediate predictions with

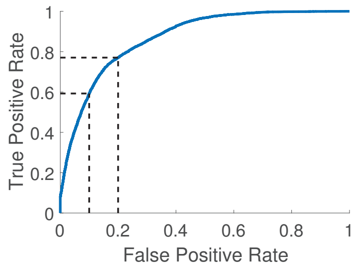


Figure 13: Two typical thresholds.

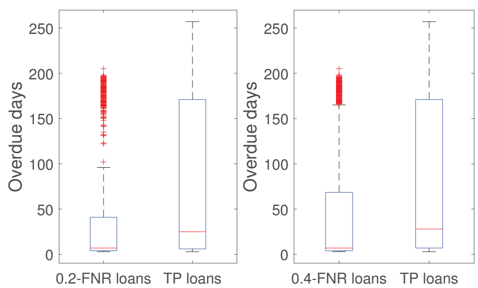
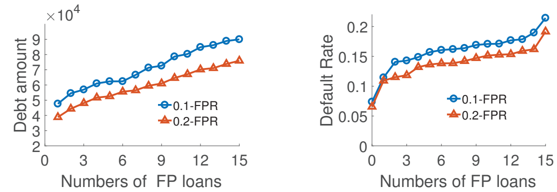
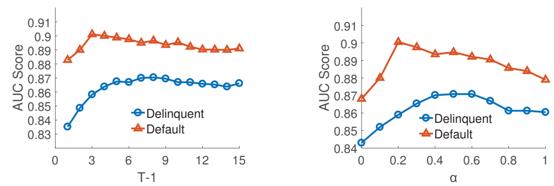


Figure 14: Delinquent days of false-negative and true-positive loans, respectively.



(a) Debt amount (b) Default rate

Figure 15: (a) debt amount and (b) default rate with different numbers of false-positive loans.



(a) History length (b) Weight

Figure 16: Predicting performance over (a) the number of past loans looked back and (b) weight on past loans.

multiple FCN networks and time-aware LSTM cells respectively. The results also demonstrate that both the repayment and financial activity sequences play important roles in the prediction. Recall the financial volume of P2P lending can be billions of dollars, an improvement in risk prediction accuracy of just a few percentage points can involve large amounts of money. Thus, all of our designs are the necessary components of the model.

Prediction Error Analysis

We now examine the loans which our model tends to wrongly predict, in order to better understand the causes of errors. We focus primarily on delinquency which includes the default case. Figure 13 shows the ROC curve for delinquency prediction. To define the prediction error, we choose two typical classification thresholds which gives 0.1 and 0.2 false positive rates (FPR), with the corresponding 0.4 and 0.2 false negative rates (FNR), respectively.

We first examine the false negatives (FN), i.e., delinquent loans incorrectly identified as non-delinquent. Figure 14 shows the delinquent days for FN loans and true-positive (TP) loans. We can see that the delinquent days of FN loans are much smaller than those of TP ones. For example, given a 0.1-FPR threshold, the delinquent days of half of FN loans are only within one week, and there are almost no default FN loans. By contrast, almost all the TP loans are 7+ days past due, among which about 40% are default (120+ days delinquent). The measurement indicates that our model tends to miss less serious delinquency, where the signs of financial problems are relatively weak. However, the majority of missed delinquent loans could be tolerable and acceptable for the users given only a few days late in the payment.

We next examine the false positives (FP), i.e., non-delinquent loans incorrectly identified as delinquent. We find that users are more likely to perform the action pattern of "borrow $\cdots$ repay" when paying FP loans, this means that most FP loans were paid in a manner of robbing Peter to pay Paul, i.e., the user pays off a loan by incurring another. However, using one loan with a higher interest to pay another debt is obviously not an effective way of reducing debt, although it delays confronting the issue. Thus, despite FP loans were repaid on time, the users with these loans actually become riskier in the near future.

Figure 15(a) shows that, on average, the total amounts of debt of users increase rapidly with the number of FP loans. And figure 15(b) shows the final default rates (observed in our dataset) for users with different number of FP loans, which become much higher than the average default rate (about 0.06 in terms of users) as the number of FP loans increases. Given the large amount of debt in the repayment sequence and the risky patterns in the activity sequence, our model tends to wrongly predict these loans to be delinquent. Therefore, the FP of our model can help the platform to identify potential users struggling with debt, who have high risk in the near future.

Parameter Analysis

We first examine the effect of parameter T on prediction performance, which determines the number of past loans the model looks back. Ideally, we would like the parameter T to be as small as possible, so that training and running our model is computationally cheap. Figure 16(a) shows the prediction performance under different values of T. We see that the AUC scores in both predictions increase rapidly once the model begins to incorporate the information of few recent loans, demonstrating that reasonable prediction can be made based on limited history. Moreover, the figure also shows that looking back longer history is not always the best choice, which could incur slight performance penalty after including too many outdated signals.

An interesting observation here is that the default prediction seems to need even less historical information than delinquency prediction, as the former reaches the peak performance faster than the latter in the figure. An intuitive explanation is that default is usually caused by severe financial distress, where the recent behavior of the corresponding user involves more delinquencies and activities related to financial distress alleviation. This makes the signs of default risk relatively apparent. Also, a user cannot recovery from a default loan, whereas he could recover from a delinquency and to be delinquent again. So users have more complex delinquency patterns which also requires a longer history to learn.

We next examine how weight parameter α affects prediction performance, which gives the extent of bias toward the predictions on past loans once T is fixed. Figure 16(b) shows the effect of α when the values of T are fixed to 8 and 4 in delinquency and default predictions. As expected, we see that either too small or too large bias would incur accuracy loss. Also, the best value of α is relatively smaller in default prediction than that in delinquency prediction.

Related Work

In recent years, P2P lending has attracted many researchers from different backgrounds, such as sociologists and data scientists (Zhao et al. 2017). Risk assessment is one of the most concerning problems in P2P lending, since risk can be the most important factor affecting the decision making of lenders. Recall that the task of risk assessment can be formally defined as a assessment model which takes the features of loan (and user) as input and the estimated score that loan will repay in time as output. So the relevant works tackling this problem attempt to develop new assessment models or to extract new features for better assessment.

A lot of research has been done for risk assessment models, most of which adopt conventional classification models from machine learning field to assess the loan risk or borrower credit. For instance, some studies use linear model such as logistic regression to predict the loan risk (Ceyhan, Shi, and Leskovec 2011; Zhao et al. 2014; Guo et al. 2016), whereas others apply nonlinear transformation model with Random Forest (Malekipirbazari and Aksakalli 2015), artificial neural networks (Byanjankar, Heikkilä, and Mezei 2015; Zang, Qi, and Fu 2014) and gradient boosting decision tree (Zhao et al. 2016). There are also some relevant studies that are from the perspective of extracting assessment features. Luo et al. (Luo et al. 2011) evaluate the risk of an borrower based on risk preferences of historical individual lenders. Serrano-Cinca et al (Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios 2015) find that static features extracted from the loan's properties (e.g., rate, amount and purpose) and the associated borrower's properties (e.g., annual income, current housing situation, credit history, and indebtedness) affect the loan default in LendingClub, and the credit grade assigned by the P2P lending site is the most predictive factor of default. Emekter et al. (Emekter et al. 2015) also find that higher interest rates charged on the high-risk borrowers are not enough to compensate for higher probability of the loan default. Besides static features, Ceyhan et al. (Ceyhan, Shi, and Leskovec 2011) examine how the temporal dynamics of bidding behavior predicts the loan outcome. Zhao et al (Zhao et al. 2016) extract dynamic features (e.g., auction phase/time, past fully-funded loan percent and past default loan percent) from the temporal auction and incremental bidding lenders of a loan for better risk prediction.

However, most of the works focus on static features, with a few on simple dynamic features. The problem of leveraging complex behavior sequence data still needs further investigation which motivates this work. Furthermore, most of the methods are devoted to shallow models with feature engineering. One powerful approach to capture underlying structure in sequential data is Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber 1997) and Gated Recurrent Unit (GRU) (Chung et al. 2015), which have been applied to many areas including natural language processing (Sutskever, Vinyals, and Le 2014), speech recognition (Graves, Mohamed, and Hinton 2013), medical diagnosis (Suhara, Xu, and Pentland 2017) and mobile data processing (Yao et al. 2017). However, there have not been works which model user behavioral patterns with RNN for loan risk prediction problems. Exploiting the power of RNN along with the informativeness of user behavior dynamics is a new promising way for risk assessment in P2P lending.

Conclusion

To best of our knowledge, this is the first work to leverage clickstreams to predict the risk for individual loans in P2P lending. Our study uncovered a number of interesting findings related to the connection between loan risk and sequences of activities that users conduct on the site. Based on these findings, we present a novel DeepCredit model, which could efficiently infer user risks from their behavior dynamics. With the sophisticated deep architecture, our model is able to efficiently exploit the complementary information of different sequences and the feasibility of leveraging previous predictions to boot the current one. Our experiments on large-scale real-world dataset show that DeepCredit is able to achieve 87% and 90% accuracy (AUC values) in delinquency and default predictions, respectively. Our model provides the P2P lending system a function of self risk assessment, where it only needs to passively collect clickstream data from the platforms itself. The company which provides us dataset is very pleased with the initial results of our model. We are now collaborating with the company to implement our model in production, which enables the platform to adopt new and potentially more effective risk-management strategies.

Acknowledgement

We thank the anonymous reviewers for their comments. This work was supported by the National Basic Research Program of China (Grant No. 2014CB340400), the National Natural Science Foundation under Grant No. 61472009.

References

Baytas, I. M.; Xiao, C.; Zhang, X.; Wang, F.; Jain, A. K.; and Zhou, J. 2017. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 65–74. ACM.

Byanjankar, A.; Heikkilä, M.; and Mezei, J. 2015. Predicting credit risk in peer-to-peer lending: A neural network approach. In *IEEE Symposium Series on Computational Intelligence (SSCI)*, 719–725. IEEE.

Ceyhan, S.; Shi, X.; and Leskovec, J. 2011. Dynamics of bidding in a p2p lending service: effects of herding and predicting loan success. In *Proceedings of the 20th international conference on World wide web (WWW)*, 547–556. ACM.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2015. Gated feedback recurrent neural networks. In *International Conference on Machine Learning*, 2067–2075.

Corporation, F. I. 2017. Fico. [Online].

Emekter, R.; Tu, Y.; Jirasakuldech, B.; and Lu., M. 2015. Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics* 47(1):54–70.

Graves, A.; Mohamed, A.-r.; and Hinton, G. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*, 6645–6649. IEEE.

Guo, Y.; Zhou, W.; Luo, C.; Liu, C.; and Xiong, H. 2016. Instance-based credit risk assessment for investment decisions in p2p lending. *European Journal of Operational Research* 249(2):417–426.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Luo, C.; Xiong, H.; Zhou, W.; Guo, Y.; and Deng, G. 2011. Enhancing investment decisions in p2p lending: an investor composition perspective. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, 292–300. ACM.

Malekipirbazari, M., and Aksakalli, V. 2015. Risk assessment in social lending via random forests. *Expert Systems with Applications* 42(10):4621–4631.

Serrano-Cinca, C.; Gutiérrez-Nieto, B.; and López-Palacios, L. 2015. Determinants of default in p2p lending. *PloS One* 10(10).

Suhara, Y.; Xu, Y.; and Pentland, A. 2017. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*, 715–724. International World Wide Web Conferences Steering Committee.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 3104–3112.

VantageScore Solutions, L. 2017. Vantagescore. [Online].

Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; and Abdelzaher, T. 2017. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web (WWW)*, 351–360.

Zang, C.; Qi, M.; and Fu, Y. 2014. The credit risk assessment of p2p lending based on bp neural network. In *Proceedings of the 2014 International Conference on Industrial Engineering and Management Sciencee (IEMS)*.

Zhao, H.; Wu, L.; Liu, Q.; Ge, Y.; and Chen, E. 2014. Investment recommendation in p2p lending: A portfolio perspective with risk management. In *2014 IEEE International Conference on Data Mining (ICDM)*, 1109–1114. IEEE.

Zhao, H.; Liu, Q.; Wang, G.; Ge, Y.; and Chen, E. 2016. Portfolio selections in p2p lending: A multi-objective perspective. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2075–2084. ACM.

Zhao, H.; Ge, Y.; Liu, Q.; Wang, G.; and Enhong Chen, H. Z. 2017. P2p lending survey: Platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8(6).

Zhu, Y.; Li, H.; Liao, Y.; Wang, B.; Guan, Z.; Liu, H.; and Cai, D. 2017. What to do next: Modeling user behaviors by time-lstm. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 3602–3608.