A Novel Risk Assessment Scheme and Practice for Peer-to-Peer Lending

Fei Tan  
Department of Computer Science  
New Jersey Institute of Technology  
ft54@njit.edu

Xiuurui Hou  
Department of Computer Science  
New Jersey Institute of Technology  
xxh256@njit.edu

Jie Zhang  
Adobe Systems  
jiezhan@adobe.com

Zhi Wei  
Department of Computer Science  
New Jersey Institute of Technology  
zhiwei@njit.edu

Zhenyu Yan  
Adobe Systems  
wyan@adobe.com

Shih-Chuan Weng  
Department of Computer Science  
New Jersey Institute of Technology  
sw464@njit.edu

ABSTRACT

In this paper, we explore and address a fundamental problem underlying the return on investment for peer-to-peer lending: the assessment of two competing risks charge-off and pre-payment in funded loans. We first propose a unified competing risks grading framework to integrate the two risks of loans both qualitatively and quantitatively. Afterwards, an end-to-end deep learning approach is introduced to solve this problem by breaking it down into multiple binary classification sub-problems, which are amenable to both feature representation and risks learning. Particularly, we leverage deep neural networks to jointly solve these sub-tasks. Both offline and online experiments on real-world loan data show that our methodology is able to achieve a competitive predictive and appealing investment performance by learning both risks properly.

KEYWORDS

Deep Neural Networks, Competing Risks, Peer-to-Peer Lending, Return on Investment

1 INTRODUCTION

Peer-to-Peer (P2P) lending has become a fast-growing new channel of financing over the past decade. The corresponding platforms have been made available in an array of countries including United States1 (Lending Club, Prosper), China2 (Yirendai), United Kingdom3 (Zopa) although their specific working mechanisms may be divergent.

Most loans in P2P lending platforms have fixed terms that are fully amortized over a fixed number of monthly payments, e.g., 36 months or 60 months. In other words, the principal of the loan is paid down over the life of the loan according to an amortization schedule, typically through equal payments. At the same time, investors will receive interest as return monthly. A loan may become charged-off when there is no longer a reasonable expectation of further payments. This is a major risk for an investor that may lead to a potential loss of principle. Borrowers are also allowed to pre-pay full loan amounts and are not subject to any fees or penalties. For such a case, an investor has no loss in principle but will receive less interest compared to a fully executed loan.

We can see that charge-off and pre-payment are two competing events that will both influence the return of a loan. They represent two exclusive ending statuses of a loan, charge-off v.s. pre-payment. Their impacts on return are different and should be distinguished properly. To this end, we propose a novel risk/return scheme that integrates loan ending status with the actual life of the loan. To be specific, we assume that the ending status pre-paid is less bad than the ending status charge-off, and that, with the same ending status, the larger the number of payments received, the better. We then further transform this problem into multiple binary classification sub-problems [4, 9]. Given a risk category \( k \), a binary classifier can be learned associated with whether the category of a loan is less than \( k \) or not. The risk category of new loans can then be easily predicted according to a series of trained binary classifiers. An architecture of deep neural networks with multiple outputs of risk category is then developed for the above binary classification problems. All binary classifiers are trained collectively in this manner so that intermediate feature representations can be explored interactively.

2 RELATED WORK

There are few published works on the modeling of both competing events and the event time in P2P lending. One important research direction in P2P lending is the risk assessment. The common strategy is to group loans into two clusters based on the charge-off risk. Afterwards, large numbers of classifiers are leveraged to conduct classification learning [3, 6, 16]. More recently, work [15] considered fully-funded probability, charge-off risk and winning-bid probability simultaneously and proposed a multi-objective portfolio optimization approach. These works focused on the overall charge-off risk but ignored the event time of charge-off and the risk of pre-payment. Regarding the general-purpose competing

2http://yirendai.investorroom.com
3https://www.zopa.com
events, the popular methods cause-specific [8] and proportional hazards survival analysis [7] are utilized widely. Survival analysis, however, doesn’t emphasize differential impacts of two risks on the investment return by just modeling them as covariates. Our proposed assessment can exploit such difference by taking them as response variables explicitly.

3 METHODOLOGY

3.1 Competing Risks Assessment

Let D represent the original dataset with N loans \( \{(x_i, s_i, t_i)\}_{i=1}^N \), where \( x_i \in X = \mathbb{R}^d \) is the input feature vector and \( s_i \in S = \{0, 1\} \) is the status with 0 and 1 being full-payment and charge-off, respectively. It is noted that fully paid loans involve both pre-payment and scheduled payment as shown in Fig. 1. As with the risk of charge-off, the pre-payment is another risk existing in investments since less interests can be secured. Lastly, we have the total received payment times of loan \( i \) noted as \( t_i \in T = \{0, 1, \ldots, T-1, T\} \) where \( T \) is the official scheduled term of a loan.

In order to model both status and payment times of loans, we propose a risk grading rule \( g : S \times T \rightarrow Y \) as follows:

\[
y_i = \begin{cases} 
    t_i, & s_i = 1 \\
    t_i + T - 1, & s_i = 0 
\end{cases}
\]

where \( y_i \in Y = \{0, 1, \ldots, K-1, K\} \) and \( K = 2T - 1 \). Accordingly, we have the corresponding transformed dataset \( O = \{(x_i, y_i)\}_{i=1}^N \).

As per usual, we adopt \(<\) as the grading relation and risk grades follow \( 0 < 1 < \ldots < K-1 < K \). This scheme considers two aspects as mentioned early: 1) Generally speaking, fully-paid loans are more preferable for investors as compared to charged-off loans, the former are thus assigned higher grades than the latter; 2) In regards to loans of the same status, the more monthly payments are received, the more desirable loans are.

3.2 Deep Learning Approach

3.2.1 Conversion from Grading Categories to Multiple Binary Outputs. The assessment of loans with \( K \) grading categories can be converted to \( K \) binary classification problems. Concretely speaking, given grading dataset \( O = \{(x_i, y_i)\}_{i=1}^N \), the specific training data for binary classifier \( k \) is \( G^k = \{(x_i, y^k_i)\}_{i=1}^N \) where \( y^k_i \) indicates whether the category of sample \((x_i, y_i)\) is less than category \( k \) or not. Formally, \( y^k_i \) is defined as follows:

\[
y^k_i = \begin{cases} 
    1, & y_i \geq k \\
    0, & \text{otherwise} 
\end{cases}
\]

One interpretation is that if the category of a data sample is \( y_i = k \), it can be grouped into categories \( \{0, 1, \ldots, k-1\} \) as well. In this case, the final target vector is \( z = (1, \ldots, 1, 0, 0) \), where \( z_i (0 \leq i \leq k) \) is set to 1 and the remaining elements are zeros. In this manner, the final predicted probability vector is thus expected to share the property that \( z_i(i \leq k) \) approaches 1 and \( z_i(i > k) \) is close to 0.

3.2.2 Deep Neural Networks. We then utilize deep neural networks to conduct a series of binary classification. As shown in Fig. 2, the developed networks have an output layer of \( K \) grading categorical nodes. The sigmoid function is adopted as the activation function of the output layer. The main idea is to enable output probability of different classifiers to be estimated independently without constraints with each other. The corresponding loss is the widely-used binary cross entropy function:

\[
\text{loss} = - \frac{1}{N} \sum_{i=1}^N \sum_{k=0}^K y^k_i \log f_k(x_i) + (1 - y^k_i) \log(1 - f_k(x_i))
\]

We denote the probability of output node \( k \in \{0, 1, \ldots, K\} \) as \( f_k(\cdot) \in [0, 1] \) and the corresponding binary classifier as \( f^k(\cdot) \in [0, 1] \). In particular, \( f^k(\cdot) = 1 \) if \( f_k(\cdot) > 0.5 \) and 0 otherwise. Since deep neural networks with competing risks (events) representation are developed here, we call it cRDN for brevity.

3.2.3 Predictive Inference. The risk category for a newly given \( x_i \) can be estimated as

\[
h(x_i) = \sum_{k=0}^K f^k(x_i)
\]

When the binary classifiers \( f_k(\cdot) \) are consistent, i.e., \( f_0(x_i) \geq f_1(x_i) \geq \ldots \geq f_K(x_i) \), the equivalent representation of \( h(x_i) \) is

\[
h(x_i) = \min \{k : f_k(x_i) = 1\}
\]

Regarding final statuses (charge-off and full-payment), the estimation formula is naturally derived as follows:

\[
p(x_i) = 1 - f_T(x_i)
\]

where \( p(x_i) \) is the overall charge-off probability across the whole scheduled lifespan of a loan.
4 EXPERIMENT (OFFLINE)

In this section, we report empirical offline evaluation results of the proposed model on the dataset from the well-known P2P lending platform Lending Club.

4.1 Experimental Data and Preprocessing

We download loan data as of Quarter 4, 2016\(^4\) from Lending Club. To assess the performance of these loans exactly and track their return over time and alleviate the bias accurately, we select those completed loans of the 36-month term for study (i.e., charge off and full payment). After cleaning data, we have a total of 370, 243 loans for study. To tune hyper-parameters for the proposed framework and perform the modeling evaluation, we randomly split the whole dataset into training and test datasets. One part of the training dataset is further held out as validation dataset. Finally, the ratio among training, validation and test data is 6 : 2 : 2 as shown in Table 1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Charge off</th>
<th>Full Payment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>34,847</td>
<td>187,332</td>
<td>222,179</td>
</tr>
<tr>
<td>Validation</td>
<td>11,621</td>
<td>62,432</td>
<td>74,053</td>
</tr>
<tr>
<td>Test</td>
<td>11,617</td>
<td>62,394</td>
<td>74,011</td>
</tr>
</tbody>
</table>

4.2 Experimental Results

4.2.1 Evaluation Metrics. To evaluate the predictive and investment performance of the proposed risk assessment scheme, we consider two-fold comparisons: (a) The area under Receiving Operating Curves (AUC@ROC) and Precision-Recall curves (AUC@PR)\(^5\) for the binary classification of charged-off and fully-paid loans. (b) Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for the risk grading prediction\(^2\).

4.2.2 Baseline Algorithms. We compare our proposed framework with the following schemes: (1) Lending Club (LC) has its own grading and evaluation system. Each issued loan is assigned a grade from A to G with a matched interest rate. (2) Logistic regression (LR) is frequently utilized for the risk evaluation in the credit scoring and peer-to-peer lending study\(^6,\, 16\). (3) A multi-class deep neural networks without grading constraints (mcDNN) with loss function multinomial cross-entropy and activation function soft-max. (4) Competing risks based survival analysis\(^1,\, 8\) is recently applied to the credit scoring\(^12,\, 14\) (CRSA). The cause-specific model with lognormal and exponential distributions is adopted here for the charge-off and pre-payment risks. The basic R routine package ‘CFC’\(^11\) is employed\(^6\).

4.2.3 Experiment Setting and Hyper-Parameter Tuning. There are a total of 4 fully connected hidden layers with 200 nodes in each one. Leaky rectified linear unit with gradient $\alpha = 0.001$ for negative inputs\(^10\) is adopted as the activation function of hidden layers. The dropout rates of 4 hidden layers are 0.5, 0.5, 0.4, 0.4, respectively. The batch size is 128. The maximum epoch is 500 with the early stopping of 50 epochs.

4.2.4 Performance and Analysis. The comparison results are shown in Table 2. Overall, our model can present a sound performance gain over baselines. For both AUC@ROC and AUC@PR, crDNN is better than others in terms of discriminating charged-off loans from fully-paid ones. It is also demonstrated that such superiority is more evident for AUC@PR, which is more amenable to the case of class imbalance\(^5\). Also, RMSE and MAE amenable to the fine-grained evaluation are further provided for the comparison. It is noted that neither LR nor LC is able to provide the time-to-event prediction. Thus, we mainly compare crDNN with its variant mcDNN and CRSA. The tremendous dominance of crDNN can be observed as compared to mcDNN and CRSA when the grading order is taken into account. Naturally, all methods outperform LC as they all conduct model learning based on the grading system of Lending Club.

5 PRELIMINARY PRACTICE (ONLINE)

In order to assess the performance of the proposed risk assessment scheme, we perform the real-time online testing procedure. The widely used metric net annualized return (NAR)\(^7\) is utilized here to monitor the overall investment performance as follows:

$$NAR = \left(1 + \frac{\sum_{i=1}^{N} \text{Interest}_i + \text{Late Fee}_i - \text{Service Charge}_i}{\sum_{i=1}^{N} \text{Principal}_i} \right) \left(1 + \frac{\sum_{i=1}^{N} \text{Net Recoveries}_i - \text{ChargeOffs}_i}{\sum_{i=1}^{N} \text{Principal}_i} \right)^{12} - 1$$

where $i$ is the recurring monthly period for any period from month 1 to month N. The numerator is equal to interest received, plus late fees received, minus the 1% service fee paid on month $i$. If a corresponding loan is in charge-off status, we subtract the corresponding scheduled principal payment amount of the Note from the numerator. If a portion of the charged off loan is recovered, we add the amount recovered, net of fees, to the numerator. The denominator is the outstanding principal amount of the Note at the beginning of that monthly payment period.

Specifically, we choose only loans of grades B and C to fund as they are the most popular loans in investments of LendingClub. We here design two experiments for the comparison. One is the cohort of loans based on the expected return on investment of the proposed scheme (model). The other one is cohorts of loans selected by hashing the Note ID and performing the random sampling (control).

\(^{4}\)https://www.lendingclub.com/info/download-data.action
\(^{6}\)https://cran.r-project.org/web/packages/CFC/CFC.pdf
\(^{7}\)https://www.lendingclub.com/public/lendersPerformanceHelpPop.action
Furthermore, we adjust the ratio between loans involved in two different experiments by tuning the sampling probability. We perform our online comparison experiments of investment starting from August 2017. The statuses of funded loans in two experiments are reported in Table 3. Overall, our proposed algorithm can achieve better return on investment. For example, the charge off proportion of our method is only 1.04%, which is much lower than that of the control method (5.88%). Furthermore, the evolution of NAR from August 2017 to May 2018 are also reported in Fig. 3, which delivers the dynamic overview of the investment performance of algorithms. It is noted that NAR is 0 on August 2017 as no interest could be received at the beginning for the monthly payment loans. NARs of the model and control are close to each other on the early months since just a few loans are funded. The difference of NAR between two cohorts are becoming salient and then being stable as more loans are funded.

Table 3: Status Statistics of Funded Loans as of May 8, 2018.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Model</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charged Off</td>
<td>2 (1.04%)</td>
<td>2 (5.88%)</td>
<td>4 (1.76%)</td>
</tr>
<tr>
<td>Current</td>
<td>142 (73.58%)</td>
<td>28 (82.35%)</td>
<td>170 (74.89%)</td>
</tr>
<tr>
<td>Fully Paid</td>
<td>18 (9.33%)</td>
<td>2 (5.88%)</td>
<td>20 (8.81%)</td>
</tr>
<tr>
<td>In Funding</td>
<td>1 (0.52%)</td>
<td>0 (0%)</td>
<td>1 (0.44%)</td>
</tr>
<tr>
<td>In Grace Period</td>
<td>2 (1.04%)</td>
<td>1 (2.94%)</td>
<td>3 (1.32%)</td>
</tr>
<tr>
<td>In Review</td>
<td>1 (0.52%)</td>
<td>0 (0%)</td>
<td>1 (0.44%)</td>
</tr>
<tr>
<td>Issued</td>
<td>24 (12.44%)</td>
<td>1 (2.94%)</td>
<td>25 (11.01%)</td>
</tr>
<tr>
<td>Late (31-120 days)</td>
<td>3 (1.55%)</td>
<td>0 (0%)</td>
<td>3 (1.32%)</td>
</tr>
<tr>
<td>Total</td>
<td>193</td>
<td>34</td>
<td>227</td>
</tr>
</tbody>
</table>

Figure 3: The real-time NAR evolution of the proposed method against the control from August 2017 to May 2018.

Figure 4: The screenshots of the preliminary web-based application.

To facilitate the automation of the investment operation for potential users, we have also been developed a preliminary web based application as shown in Fig. 4.

6 DISCUSSION

The main purpose of this study is not to beat other state-of-the-art classification algorithms. It can be instead regarded as an alternative approach to modeling competing risks with differential impacts compared to the classical survival analysis. Our work aids in unleashing the power of machine learning algorithms for modeling time-to-event loan data in P2P lending.

In section 5, we adopt NAR as the evaluation metrics of online experiments. The adjusted NAR is actually recorded in the dashboard of Lending Club, which is usually lower than NAR due to the adjustments of expected defaults and late payments. The portfolio optimization or selection is an important research direction, which has also received certain attention in P2P lending study [15]. However, regarding the consideration of risks, only the overall risk of charge-off is considered. Our work will contribute to the further study of portfolio optimization, which remains the topic of future research. The coding scheme here is questionable, but the in-depth exploration has been performed and then offline results have been submitted for review [13]. No online experimental results are available so far. In practice, we will continue to employ our basic approach to loans of other lending platforms like Prosper and Yirendai. More importantly, we will devote more efforts to develop and improve our application to popularize our risk assessment scheme in finance.

7 CONCLUSION

In this paper, we try to assess competing risks for time-to-event loan data by imposing the fine-grained grading constraints on deep neural networks. In particular, two competing risks charge-off and pre-payment and their corresponding event time are simultaneously incorporated into the framework. We apply our approach to the large-scale loan data released by Lending Club. The empirical evaluation demonstrates the outperformance of the proposed approach compared to classical competing risks survival analysis in terms of risks prediction. The online investment experiments further demonstrate the feasibility and appealing property of the proposed scheme.
REFERENCES


