Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: https://www.tandfonline.com/loi/tjor20

Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending

Haomin Wang, Gang Kou & Yi Peng

To cite this article: Haomin Wang, Gang Kou & Yi Peng (2020): Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending, Journal of the Operational Research Society, DOI: 10.1080/01605682.2019.1705193

To link to this article: https://doi.org/10.1080/01605682.2019.1705193

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 12 May 2020.

Submit your article to this journal

Article views: 979

View related articles

View Crossmark data

Citing articles: 3 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjor20

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2019.1705193

THE OPERATIONAL RESEARCH SOCIETY
Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS

Check for updates

Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending

Haomin Wang[a], Gang Kou[b] and Yi Peng[a]

[a]School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China; [b]School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China

ABSTRACT

Online peer-to-peer (P2P) lending is a new form of loans. Different from traditional banks, lenders provide loans to borrowers directly through P2P platforms. Since many P2P loans are unsecured personal loans, credit rating of loans is vital to control default risk and improve profit for lenders and platforms. Standard binary classifiers are inappropriate in P2P lending because there are multiple credit classes and misclassification costs vary largely across classes in P2P lending. Though there are a few works that studied cost-sensitive classifiers in P2P lending, none of them have analyzed this issue from the perspective of multi-class classifications and measured misclassification costs of different credit grades using real losses and opportunity costs. The objective of this paper is to model credit rating in P2P lending as a cost-sensitive multi-class classification problem. We proposed a misclassification cost matrix for P2P credit grading with a set of equations and models to calculate the costs. An experiment using publicly available data from Lending Club was conducted to validate the usefulness of the proposed misclassification cost matrix. The results showed that the cost-sensitive classifiers can significantly reduce the total cost, which is essential for the survival and profitability of P2P platforms.

ARTICLE HISTORY
Received 9 January 2019
Accepted 11 December 2019

KEYWORDS
Cost-sensitive learning; P2P lending; credit rating; misclassification cost matrix; data mining

1. Introduction

In the past decade, online peer-to-peer (P2P) lending, as a popular form of personal loan, has emerged in credit market. It transfers traditional way of face-to-face personal loans through online services (Bachmann et al., 2011). P2P lending is an electronic marketplace where individual lenders provide loans to individual borrowers. It is pervasive, convenient, efficient, and low-cost without the involvement of traditional financial institutions (Guo, Zhou, Luo, Liu, & Xiong, 2016).

Since the first lending platform Zopa was established in UK in February 2005, an increasing number of P2P lending platforms, such as Prosper, Smava, and Lending Club, have been developed all around the world (Ge, Feng, Gu, & Zhang, 2017) and accumulated data and management experiences. Comparing with traditional banking systems, P2P lending has some characteristics. First, P2P platforms facilitate transactions by connecting borrowers and lenders directly. Borrowers fill in electronic loan application forms, including amounts, terms, purposes, and personal information (such as age, job, address, and credit card). Platforms provide available financial situations and credit histories of borrowers to lenders, who will decide whether to grant a loan and an interest rate. Platforms use various approaches to help lenders set interest rates. Some platforms carry out an auction at which a borrower set her/his maximum interest rate and lenders give their bids (Galloway, 2009). Another approach is to assign interest rates automatically using borrowers' credit grades, which are calculated based on borrowers' characteristics (Collier & Hampshire, 2010). Generally, better credit grades are associated with lower interest rates. Second, P2P lending platforms charge service fees for transactions (Klafft, 2008), instead of charging borrowers higher interest rates than the cost of the money as traditional financial institutions. P2P lending process benefits both borrowers and lenders. While borrowers can borrow money at lower costs than traditional financial institutions, lenders can make more money than putting their money in banks. This benefit comes with the risk of borrowers' defaulting on the loans because many P2P loans are unsecured personal loans and most lenders have little knowledge about credit risk management (Xia, Liu, & Liu, 2017).

To control default rates and risks, P2P lending platforms built classification models to evaluate credit risks of loans and borrowers and suggest appropriate interest rates for loan applications. The quality of credit classification models is vital to the credit risk

CONTACT Yi Peng pengyi@uestc.edu.cn University of Electronic Science and Technology of China, Chengdu, China

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

management and sustainability of P2P lending platforms. Using experiences from financial institutions, P2P lending platforms adopt and develop classification algorithms to categorize borrowers into different credit grades based on their characteristics and credit history, and recognize potential borrowers who are likely to default (Lessmann, Baesens, Seow, & Thomas, 2015; Florez-Lopez & Ramon-Jeronimo, 2014; Marqués, García, & Sánchez, 2013).

Though it is a common practice in traditional credit rating to use standard cost-insensitive binary classification algorithms (Li, Kou, Peng, & Shi, 2017; Morente-Molinera, Mezei, Carlsson, & Herrera-Viedma, 2017), such as logistic regression, neural networks, and decision trees (Butaru et al., 2016; Luo, Wu, & Wu, 2017), they are not appropriate in P2P lending for the following reasons. First, there are more than two classes of credit grades in P2P lending and each credit grade implies a certain level of risk. Thus multi-class classification should be considered in P2P credit grading. Second, P2P loan data are imbalanced. The number of samples in different credit grades varies dramatically. For instance, the number of ideal borrowers in the best grade or high-risk borrowers in the worst grade is much smaller than the other grade groups. Third, misclassification costs are not uniform across classes in P2P lending. In general, the cost of classifying a loan with bad credit as a good one is usually greater than classifying a good one as bad (Chen, Ribeiro, & Chen, 2016). In a multi-class credit-grading scenario, classifying a sample of grade C into grade A is more costly than classifying B into A. Therefore, standard cost-insensitive multi-class classification, in which all errors have the same cost, is not suitable for credit rating in P2P lending.

Cost-sensitive multi-class classifiers fit well for credit rating in P2P lending. Cost-sensitive classifiers were developed for imbalanced data classification (Elkan, 2001; Hu et al., 2015; Sun, Shang, & Li, 2014). Various cost-sensitive classifiers have been proposed for credit rating (Bahnsen, Aouada, & Ottersten, 2015; Chao & Peng, 2018; Marqués et al., 2013; Sahin, Bulkan, & Duman, 2013). The goal of cost-sensitive classifier is to minimize total costs measured by a misclassification cost matrix (Guan, Yuan, Ma, Khattak, & Chow, 2017), which is not only necessary but also important for cost-sensitive classification problems.

Though there are a few works in P2P lending (Xia et al., 2017; Xu, Chen, & Chau, 2016) that studied cost-sensitive classifiers, none of them have analyzed this issue from the perspective of multi-class classifications and measured misclassification costs of different credit grades using real losses and opportunity costs associated with P2P lending. How to measure the misclassification costs of different credit grades is a useful but understudied problem. Serrano-Cinca and Gutiérrez-Nieto (2016) showed that loan profitability outperformed loan default probability in P2P lending, which proved the importance of considering both interest rates and the probability of default in P2P credit scoring.

Misclassification costs are losses of lenders' earnings due to misclassifying credit grades of loans. It equals to the difference between the return of a loan when it is correctly classified and the return of a loan when it is misclassified as other credit grade. The difference can be one of the following situations: (1) If a loan is classified to a better credit grade with a lower interest rate, the risk to default of the loan is underestimated and the interest rate of the loan is set lower than it should be, which means that the interest maybe insufficient to cover the risk that the lender bears. The lender will lose potential returns that they could have gotten, including an unpaid risk that the borrower should pay for the higher-risk loan. (2) If a loan is classified to a worse credit grade with a higher interest rate, borrowers might be scared away or it may increase their chance to default, which causes opportunity costs and financial losses to lenders.

The objective of this paper is to propose a multi-class cost matrix that measures misclassification costs of P2P credit grading by considering real losses and opportunity costs associated with P2P lending. We developed a set of equations and models to calculate misclassification costs. The parameters in the proposed equations and models are designed to calculate the cost matrix and support P2P lending platforms' operations. A case study using data from Lending Club is conducted to demonstrate the performances of the proposed cost matrix using several well-known cost-sensitive classifiers. The results show that the proposed cost matrix can not only reveal the sources of losses caused by misclassifications, but also reduce the total costs for real-world P2P platforms, which is better than cost-insensitive classification algorithms.

The rest of this paper is organized as follows. Section 2 reviews related works. Section 3 proposes an abstract structure of credit grades, and misclassification costs which measure real financial losses in P2P lending. Section 4 analyzes the range of parameters in the misclassification cost matrix and explains their managerial implications. Section 5 conducts an experiment using data from Lending Club. Section 6 concludes the paper with limitations and future research directions.

2. Related works

The goal of most classifiers is to maximize accuracy and minimize misclassifications. Various

Table 1. Cost matrix proposed by Beling et al. (2005).

		Prediction	
		Positive	Negative
Actual	Positive	0	int_r
	Negative	*Lgd*	0

classification methods have been proposed for credit rating and risk management (Santana, Lanzarini, & Bariviera, 2018; Huang & Kou, 2014; Kou, Peng, & Wang, 2014; Lanzarini, Villa Monte, Bariviera, & Jimbo Santana, 2017; Peng, Wang, Kou, & Shi, 2011; Wu & Kou, 2016;). Standard classifiers treat the costs of misclassifications the same, which is not true in real credit risk management (Fiore, De Santis, Perla, Zanetti, & Palmieri, 2017; Tapkan, Özbakır, Kulluk, & Baykasoğlu, 2016). Many researches support the use of cost-sensitive classifiers in credit rating. Sahin et al. (2013) proposed a cost-sensitive decision tree approach with varying misclassification costs. It is successfully used in credit card fraud detection to decrease financial losses. Alejo, García, Marqués, Sánchez, and Antonio-Velázquez (2013) improved the Multilayer Perceptron neural network using three misclassification cost functions and can be used to improve the prediction effectively in credit rating. Bahnsen, Aouada, and Ottersten (2014, Bahnsen et al., 2015) suggested example-dependent cost-sensitive methods and proposed logistic regression and decision trees for credit scoring.

Misclassification cost can be described by a cost matrix $\boldsymbol{C} = (c_{ij})_{n\times n}$, where c_{ij} indicates the cost due to misclassifying an instance of class i as class j, and n is the number of classes (Domingos, 1999). In credit rating, the measurement of misclassification costs in $\boldsymbol{C}$ is not only a basic component of cost-sensitive classification, but also vital for high quality credit rating. Real financial indicators, like profit-based or financial loss-related measures, are well aligned with the objectives in credit rating (Maldonado, Bravo, Lopez, & Perez, 2017; Serrano-Cinca & Gutiérrez-Nieto, 2016; Verbraken, Bravo, Weber, & Baesens, 2014). Beling, Covaliu, and Oliver (2005) set the cost of a false negative to a loan's interest rate charged to the customer int_r, the cost of a false positive to the loss given default Lgd, and both the costs of true positive and true negative are set to zero. Following this notation, this paper regards default loans as negative instances and good loans as positive instances. Table 1 shows the cost matrix (Beling et al., 2005).

Hand, Whitrow, Adams, Juszczak, and Weston (2008) proposed a cost matrix (Table 2) for credit card fraud detection. It represents the costs of misclassification by the administrative cost Ca, which is related to analyzing the transactions and contacting

Table 2. Cost matrix proposed by Hand et al. (2008).

		Prediction	
		Positive	Negative
Actual	Positive	0	C_a
	Negative	100*C_a	C_a

Table 3. Cost matrix proposed by Bahnsen et al. (2013).

		Prediction	
		Positive	Negative
Actual	Positive	0	C_a
	Negative	A_{mti}	C_a

Table 4. Example-dependent cost matrix proposed by Bahnsen et al. (2014).

		Prediction	
		Positive	Negative
Actual	Positive	0	r_i+C_a
	Negative	$Cl_i{\cdot}Lgd$	0

card holders. In the cases of false negative and true negative, the associated costs both equal to C_a because the card holder will have to be contacted. However, in the case of false positive, due to the fact that frauds are not detected, the cost is defined as a hundred times C_a.

Bahnsen, Stojanovic, Aouada, and Ottersten (2013) pointed out a limitation of the above cost matrices. Since losses of different frauds range from a few to a large amount, it is unrealistic to assume constant cost in false positive. In Bahnsen et al. (2013), a new cost matrix (Table 3) is proposed as a better representation of the actual costs, where the cost of false positive is replaced by the amount Amt_i of the transaction i.

Bahnsen et al. (2014) proposed a cost matrix (Table 4) with example-dependent varying misclassification costs. For every borrower i, the costs of correct classifications are zero, and the cost of false positive is the losses if borrower i defaults which is proportional to his credit line Cl_i. The cost of false negative is r_i plus C^a, where r_i is the profit that can be earned from a good borrower, and C^a is related to uncertainty of next alternative borrower.

Xia et al. (2017) proposed a cost-sensitive boosted tree loan evaluation model to discriminate potential default borrowers in P2P lending. The model considers the imbalanced misclassification cost (shown in Table 5), based on the assumption that the cost of misclassifying a default borrower is larger than that of misclassifying a good one, i.e. C(0, 1) > C(1, 0). It adopts a cost matrix with two classes (Good and Default).

Existing studies in cost-sensitive classification and cost matrices in credit rating focused on binary classification problems and traditional credit risk

Table 5. Cost matrix in Xia et al. (2017).

		Predicted	
		Good	Default
Observed	Good	0	C(1, 0)
	Default	C(0, 1)	0

applications. Few, if there's any, analyzed this issue from the perspective of multi-class classifications and in the context of P2P lending.

In cost-sensitive multi-class classification, the corresponding cost matrix $\boldsymbol{C} = (c_{ij})$ is n-dimensional (Kou, Ergu, Lin, & Chen, 2016) where n is number of classes. For any $i, j = 1, 2, \ldots, n$, the element c_{ij} indicates the misclassification cost that is caused by misclassifying an instance of class i as class j. The objective of this study is to determine the values of misclassification costs by incorporating profit losses and other indirect costs in P2P lending. A high-quality n-dimensional cost matrix can help to improve the performances of classifiers and reduce default losses in P2P lending.

3. Misclassification cost measures

This section analyzes the risks and profits in P2P lending, and proposes a multi-class misclassification cost matrix for P2P lending.

3.1. Modeling risks and profits in P2P lending

P2P lending platform provides services to lenders and borrowers to facilitate their transactions, and charges a proportion of profits earned by successful repaid loans as an essential part of services fee. One of the main services is to provide credit information of borrowers and suggestions (including credit rating and appropriate interest rates) to lenders. In fact, platforms and lenders share profits and risks.

For lenders, borrowers' default would cause loses. In this paper, loss given default (Lgd) (Schuermann, 2004) is used to measure the proportion of a lender's loss when a borrower defaults on a loan. Although Lgd is different for different loans, it is usually treated as a constant number for all loans to simplify models (Bahnsen et al., 2014; Beling et al., 2005). Lenders' profits come from the interests of investments if borrowers fully repaid.

For platforms, default loans impact their profits directly by causing the loss of service fees and indirectly by damaging their reputations in quality of services on credit rating and setting interest rates. Thus, accurate credit rating is the foundation of healthy and profitable operations of P2P platforms. Consequently, an important requirement of credit rating in P2P lending is to measure the profitability of a loan (Xia et al., 2017), and the amount of loss in profit for lenders should be considered in misclassification cost matrix.

Table 6. Structure of credit grades in P2P lending.

Credit grades	Probability of default (PD)	Interest rate (I)
1(A)	PD_1	I_1
2(B)	PD_2	I_2
3(C)	PD_3	I_3
…		
n	PD_n	I_n

P2P platforms usually classify borrowers into multiple credit grades and determine hierarchical interest rates for them. We construct a structure of credit grades to measure the profits and losses in P2P lending (Table 6), based on the assumption that different credit grade i is associated with a certain probability of default (PD) PD_i. Therefore it is assigned an interest rate I_i to reflect the corresponding credit risk. Credit rating classifies loans into these grades, where '1(A)' refers to the best credit, and 'n' refers to the worst credit risk. Generally, the worse the credit grade, the higher probability of default and the higher interest rate. That is, for any $i, j = 1, 2, \ldots, n,\ i > j$, so that $PD_i > PD_j$ and $I_i > I_j$.

When a lender lends a loan of credit grade i ($i = 1,\ 2,\ \ldots,\ n$) with PD_i probability of default and I_i interest rate, the lenders' expected return is:

$$\begin{aligned} ER_i &= (1 - PD_i) \cdot (1 + I_i) + PD_i \cdot (1 - Lgd) \\ &= 1 + I_i - PD_i \cdot (I_i + Lgd) \end{aligned} \tag{1}$$

where Lgd is the loss given default which indicates the average loss rate of money when a borrower defaulting on the loan.

3.2. Misclassification cost matrix

Based on Table 6, we propose a misclassification cost matrix $\boldsymbol{C}$ for credit rating in P2P lending (Table 7), where the cost of correct classification is zero, i.e. $c_{ii} = 0,\ i = 1,\ 2,\ \ldots,\ n$.

Misclassification cost matrix $\boldsymbol{C}$ can be decomposed into two blocks $\boldsymbol{C^1}$ and $\boldsymbol{C^2}$, that is $C = \begin{pmatrix} 0 & C^2 \\ C^1 & 0 \end{pmatrix}$. $\boldsymbol{C^1}$ is the lower triangular submatrix corresponding to the loans whose predicted credit grades (j) are better than their actual credit grades (i), i.e. $i > j$, and $\boldsymbol{C^2}$ is the upper triangular submatrix corresponding to the loans whose actual credit grades (i) are better than their predicted credit grades (j), i.e. $i < j$.

3.2.1. Lower triangular submatrix C^1: Prediction (j) better than actual (i)

In the lower triangular matrix $C^1 = (C^1_{ij})_{i>j}$, C^1_{ij} is the cost of misclassifying a loan in grade i as grade j. As a result of misclassification, $PD_i > PD_j$ and $I_i > I_j$,

Table 7. Misclassification cost matrix ***C*** for credit rating in P2P lending.

		Prediction			
		1(A)	2(B)	…	n
Actual	1(A)	0	c_{12}	…	c_{1n}
	2(B)	c_{21}	0	…	c_{2n}
	…	…	…	0	…
	n	c_{n1}	c_{n2}	…	0

which means a borrower pays a lower interest rate than he/she should. Lower interest rate may affect borrowers' potential defaults to a certain extent, which has been supported by theoretical and empirical evidences (Edelberg, 2006; Serrano-Cinca, Gutierrez-Nieto, & López-Palacios, 2015). Specifically, the relationship between interest rate and risk of default is positive in P2P lending. To describe this kind of changes, we revise the PD in a linear function related with the difference of interest rates:

$$PD'(j|i)_{i>j} = PD_i + \beta \cdot (I_i - I_j) \tag{2}$$

where β is a revised coefficient of PD, indicating how the PD of a borrower changes with different interest rates. Obviously, $\beta < 0$.

$PD'(j|i)$ is the actual probability of default, after misclassifying a loan in grade i to grade j. The misclassification leads to a change of expected return for the lender:

$$ER'(j|i)_{i>j} = 1 + I_j - PD'(j|i) \cdot (I_j + Lgd) \tag{3}$$

The misclassification also leads to a loss in return and the cost is formulated as:

$$\begin{aligned} C^1_{ij} &= ER_i - ER'(j|i) \\ &= (I_i - I_j) \cdot [1 - PD_i + \beta \cdot (I_j + Lgd)] \end{aligned} \tag{4}$$

3.2.2. Upper triangular submatrix C^2: Actual (i) better than prediction (j)

In the upper triangular matrix $C^2 = (C^2_{ij})_{i<j}$, $PD_i < PD_j$ and $I_i < I_j$, which means a borrower pays a higher interest rate than he/she should. This misclassification may lead to a withdrawal of loan application with a certain probability because the borrower can't tolerate the high interest rate. This probability of application withdrawal is related to the difference of interest rates $I_i - I_j$. To simplify the model, a linear function is used to calculate the probability that a borrower of credit grade i will give up application when misclassified as credit grade j (worse than grade i).

$$\text{Prob}_{give-up}(j|i)_{i<j} = \alpha \cdot (I_i - I_j) \tag{5}$$

where α is borrowers' churn rate (Zhu, Baesens, Backiel, & Vanden Broucke, 2018). It indicates how the probability $\text{Prob}_{give-up}(j|i)$ varies with the difference of interest rates $I_i - I_j$.

On the other hand, the probability of a borrower accepting interest rate I_j caused by misclassification is $1 - \text{Prob}_{give-up}(j|i)$. In this case, the PD of the borrower increases with the higher interest rate. Equation (6) is used to calculate the actual PD $PD'(j|i)_{i<j}$ when $i<j$.

$$PD'(j|i)_{i<j} = PD_i + \beta \cdot (I_i - I_j) \tag{6}$$

For a lender who accepts the misclassified loans when $i<j$, the expected return is:

$$ER'(j|i)_{i<j} = 1 + I_j - PD'(j|i) \cdot (I_j + Lgd) \tag{7}$$

Thus, the misclassification cost C^2_{ij} can be measured as:

$$C^2_{ij} = ER_i - (1 - \text{Prob}_{give-up}(j|i)) \cdot ER'(j|i) \tag{8}$$

It is rewritten as:

$$\begin{aligned} C^2_{ij} = {} & (1 - \text{Prob}_{give-up}(j|i)) \cdot (ER_i - ER'(j|i)) \\ & + \text{Prob}_{give-up}(j|i) \cdot ER_i \end{aligned} \tag{9}$$

On the right hand side of Equation (9), the first term indicates the loss in profit from transactions, and the second term $\text{Prob}_{give-up}(j|i) \cdot ER_i$ indicates the opportunity cost caused by application withdrawals due to misclassifications.

This model does not consider the idle investment due to borrowers' termination and the uncertainty of next alternative borrower, which were discussed in Bahnsen et al. (2014). If a borrower terminates the application, the transaction will not happen, and the money of a lender invested will be lent to an alternative borrower. Credit rating for the alternative borrower is also a classification problem using the proposed misclassification cost matrix.

In summary, based on a structure of credit grades in P2P lending, we designed a misclassification cost matrix to measure real business losses, which are calculated using Equations (4) and (9). In addition, we analyzed the parameters (Lgd, β, and α) from the perspective of both cost-sensitive classification and operations in P2P lending.

4. Parameters analysis

4.1. Loss given default: Lgd

We adopt a quadratic programming to deduce loss given default (Lgd) backward from observed PDs, interest rates, and actual return rates:

$$\begin{aligned} \min_{Lgd} \quad & \sum_{i=1,2,\ldots,n} (ER_i - AR_i)^2 \\ s.t. \quad & 0 < Lgd < 1 \end{aligned} \tag{10}$$

where ER_i indicates the expected returns of lenders who lend to borrowers in credit grade i. AR_i is the actual return rate of investing loans to borrowers in credit grades i, which considers potential losses due to defaults and is calculated by P2P platforms. The Lgd value is calculated by minimizing the sum of squared differences between expected returns ER_i and actual returns AR_i on different grades. This fits the realistic environment and reflects the average percentage of losses due to defaults.

4.2. PD's revised coefficient: β

β is PD's revised coefficient in Equation (2). It indicates how the PD of a borrower changes with different interest rates she/he bears. There are two reasonable assumptions about the revised PD in practice. First, the higher the interest rate is, the higher the probability that a borrower defaults. Second, a lower interest rate will not completely convert PDs of lower-credit-grades borrowers into those with better-credit-grades. In other words, when $i > j$, even if a borrower of credit grade i lowers her/his PD from PD_i to $PD'(j|i)$, $PD'(j|i)$ is still larger than PD_j due to the lower interest rate I_j. That is, $PD_j < PD'(j|i) < PD_i$. According to Equation (2), it is equivalent to:

$$PD_j < PD_i + \beta \cdot (I_i - I_j) < PD_i$$

because $I_i - I_j > 0$,

$$\frac{PD_j - PD_i}{I_i - I_j} < \beta < 0, \text{ for any } i, j = 1, 2, \ldots, n \text{ and } i > j \tag{11}$$

In the view of cost matrix, for any $i, j = 1, 2, \ldots, n$, if $i > j$, then $C^1_{ij} > 0$. According to Equation (4), we get:

$$\begin{aligned}&(I_i - I_j) \cdot [1 - PD_i + \beta \cdot (I_j + Lgd)] > 0\\&1 - PD_i + \beta \cdot (I_j + Lgd) > 0\\&\beta > \frac{PD_i - 1}{I_j + Lgd}, \text{ for any } i, j = 1, 2, \ldots, n \text{ and } i > j\end{aligned} \tag{12}$$

Moreover, for any $i, j, k = 1, 2, \ldots, n$, if $i > j > k$, the cost due to misclassifying a loan of credit grade i into k should be larger than misclassifying it as j. That is, $C^1_{ij} < C^1_{ik}$ so

$$\begin{aligned}&(I_i - I_j) \cdot [1 - PD_i + \beta \cdot (I_j + Lgd)] < (I_i - I_k)\\&\quad \cdot [1 - PD_i + \beta \cdot (I_k + Lgd)]\\&(I_i - I_j) \cdot \beta \cdot (I_j + Lgd) - (I_i - I_k) \cdot \beta\\&\quad \cdot (I_k + Lgd) < (I_j - I_k) \cdot (1 - PD_i)\end{aligned}$$

If $(I_i - I_j) \cdot (I_j + Lgd) - (I_i - I_k) \cdot (I_k + Lgd) > 0$,

$$\beta < \frac{(I_j - I_k) \cdot (1 - PD_i)}{(I_i - I_j) \cdot (I_j + Lgd) - (I_i - I_k) \cdot (I_k + Lgd)},$$

that is, β is less than a positive number, which have been covered by $\beta < 0$ in formula (12).

If $(I_i - I_j) \cdot (I_j + Lgd) - (I_i - I_k) \cdot (I_k + Lgd) < 0$,

$$\begin{aligned}&\beta > \frac{(I_j - I_k) \cdot (1 - PD_i)}{(I_i - I_j) \cdot (I_j + Lgd) - (I_i - I_k) \cdot (I_k + Lgd)},\\&\text{for any } i, j, k = 1, 2, \ldots, n \text{ and } i > j > k\end{aligned} \tag{13}$$

Similarly, because $C^1_{jk} < C^1_{ik}$,

$$\begin{aligned}&\beta > \frac{(I_j - I_k) \cdot (1 - PD_j) - (I_i - I_k) \cdot (1 - PD_i)}{(I_i - I_j) \cdot (I_k + Lgd)},\\&\text{for any } i, j, k = 1, 2, \ldots, n \text{ and } i > j > k\end{aligned} \tag{14}$$

4.3. Borrowers' churn rate: α

When a loan of credit grade i is misclassified into grade j ($\forall i, j = 1, 2, \ldots, n,\ i < j$), a borrower will give up his loan application with a probability $\text{Prob}_{give-up}(j|i)$. α is borrowers' churn rate which indicates how the probability $\text{Prob}_{give-up}(j|i)$ varies with the difference of interest rates $I_i - I_j$. Naturally, $0 < \text{Prob}_{give-up}(j|i) < 1$ so that:

$$\frac{1}{I_i - I_j} < \alpha < 0, \text{ for any } i, j = 1, 2, \ldots, n \text{ and } i < j \tag{15}$$

We analyze the parameter α in the view of cost matrix. Then α holds:

$$C^2_{ij} > 0, \text{ for any } i, j = 1, 2, \ldots, n \text{ and } i < j \tag{16}$$

$$C^2_{ij} < C^2_{ik}, \text{ for any } i, j, k = 1, 2, \ldots, n \text{ and } i < j < k \tag{17}$$

$$C^2_{jk} < C^2_{ik}, \text{ for any } i, j, k = 1, 2, \ldots, n \text{ and } i < j < k \tag{18}$$

Specifically, the cost of misclassifying a bad loan into a better one is larger than the other way around. That is,

$$C^2_{ij} < C^1_{ji}, \text{ for any } i, j = 1, 2, \ldots, n \text{ and } i < j \tag{19}$$

where C^1_{ji} indicates the misclassification costs of misclassifying a loan (of grade j) to a better credit grade i, and C^2_{ij} indicates the misclassification costs of misclassifying a loan (of grade i) to a worse credit grade j.

5. Experiment: A case study on lending club

5.1. Data collection

To validate the proposed measure of misclassification costs in a real P2P platform, the empirical study utilizes data collected from Lending Club (Lending Club, 2017), which is the largest P2P lending platform in U.S and the data is publicly available.

We collected the data of loans on Lending Club from the first quarter of 2016 to the third quarter of

Table 8. Number of instances in lending club data.

Period	# of instances
2016Q1	133889
2016Q2	97856
2016Q3	99122
2016Q4	103548
2017Q1	96781
2017Q2	105451
2017Q3	122701
Total	759348

2017. The total number of loans in this data set is 759348 (as shown in Table 8). There are 128 features for each loan in the original data, including loan characteristics (such as loan amount, term, and purpose), borrowers' financial situation (such as annual income and home ownership), and credit history (such as FICO score, the number of inquiries, the number of open credit lines, and incidences of delinquency). In particular, a variable named LC grade, which is the credit grade for loans assigned by Lending Club, is used as the class label for training and testing classification algorithms in the experiment.

The Lending Club data has been used in existing P2P lending researches (Serrano-Cinca & Gutiérrez-Nieto, 2016; Xia et al., 2017). Serrano-Cinca and Gutiérrez-Nieto (2016) provided some statistics of the data, such as the proportions of loans in different credit grades, the proportions of default, and the mean of interest rates for each grade. Table 9 summarizes the structure of credit grades of the Lending Club data, where the probability of default is estimated using the proportion of default loans to all loans in the grade.

Table 9 supports our assumptions of the proposed models. First, there are seven credit grades (A–G) implemented on Lending Club, where A is best and G is worst. It confirms the use of multi-class classification in P2P lending. Second, the proportions of loans vary from 33.60% to 0.15%, which means that the data are highly imbalanced. Third, worse credit grades are associated with higher PDs and interest rates, which is a basic assumption of our model to measure misclassification costs using real losses in P2P lending.

Table 9. Structure of credit grades on lending club data.

Grades	% of loans	Probability of default (PD)	Interest rate (I)
A	32.33%	6.28%	7.42%
B	33.60%	11.54%	11.33%
C	19.97%	15.56%	13.94%
D	10.85%	18.46%	16.15%
E	2.56%	20.65%	17.78%
F	0.55%	25%	19.27%
G	0.15%	33.87%	21.03%

Table 10. Actual annualized returns provided by lending club*.

Grades	A	B	C	D	E	F/G
Annualized return	4.38%	5.21%	5.35%	3.96%	2.99%	−1.5%

*The data is available at: https://www.lendingclub.com/info/demand-and-credit-profile.action.

5.2. Setting parameters and managerial implications

The interest rates in P2P lending are set based on the business environment, which can be measured using parameters, such as Lgd, β, and α. In the experiment, we set the values of parameters using the observed PDs and interest rates in Table 9. The parameters not only reflect the operating environment of Lending Club, but also provide guidelines for interest rates adjustments.

Furthermore, Lending Club provides actual average annualized returns for lenders, shown in Table 10. According to formula (10), the parameter Lgd is approximately 0.5 (0.50543), which means that about 50% of principal in investment will be lost if default happens.

To satisfy inequalities in (11)–(14), the range of parameter β is solved as $-0.1116 < \beta < 0$. To calculate the proposed model, we set β as the median of the range, that is $\beta = -0.0558$. Thus, the revised PD on Lending Club, shown in Table 11, can be estimated using Equation (2). Bold numbers in Table 11 indicate the inherent PDs for loans of credit grades. In fact, fluctuations of interest rates caused by misclassification can affect borrowers' default. The experiment proved that inherent credit ratings have more effects on the probability of default than the fluctuations of interest rates. This makes sense in P2P lending, where loans are normally small in size and repayment of unsecured microloans is more dependent on the willingness to pay than the ability to pay.

To satisfy inequalities in (15)–(19), the range of parameter α is $-1.2916 < \alpha < -0.8412$. We set α as the median of the range, this is $\alpha = -1.0664$. It indicates that a borrower's probability of terminating a loan application increases about 1.1% if the interest rate rises by 1%.

In the experiment, we calculate the parameters by analysis of models based on observed structure including probability of default and interest rates. In fact, the parameters are determined by behaviors of borrowers, which can be measured through investigation. It provides a guidance to check and adjust interest rates. If there's a big difference between the values of parameters calculated and the results obtained from the real practice, the interest rates on P2P lending platforms should be adjusted, because PDs are given and generally stable.

5.3. Cost matrix

According to Equations (4) and (9), we calculate the misclassification cost matrix $\boldsymbol{C}$ (Table 12), which is

Table 11. Revised PD on lending club.

		Prediction						
		A	B	C	D	E	F	G
Actual	A	**0.0628**	0.0650	0.0664	0.0677	0.0686	0.0694	0.0704
	B	0.1132	**0.1154**	0.1169	0.1181	0.1190	0.1198	0.1208
	C	0.1520	0.1541	**0.1556**	0.1568	0.1577	0.1586	0.1596
	D	0.1797	0.1819	0.1834	**0.1846**	0.1855	0.1863	0.1873
	E	0.2007	0.2029	0.2044	0.2056	**0.2065**	0.2073	0.2083
	F	0.2434	0.2456	0.2470	0.2483	0.2492	**0.2500**	0.2510
	G	0.3311	0.3333	0.3347	0.3360	0.3369	0.3377	**0.3387**

Table 12. Misclassification cost matrix on lending club.

		Prediction						
		A	B	C	D	E	F	G
Actual	A	0	0.0089	0.0166	0.0241	0.0303	0.0365	0.0443
	B	0.0333	0	0.0073	0.0144	0.0203	0.0262	0.0336
	C	0.0530	0.0211	0	0.0070	0.0128	0.0184	0.0256
	D	0.0684	0.0376	0.0172	0	0.0056	0.0112	0.0182
	E	0.0789	0.0489	0.0291	0.0123	0	0.0055	0.0124
	F	0.0851	0.0568	0.0380	0.0222	0.0106	0	0.0070
	G	0.0856	0.0608	0.0443	0.0304	0.0202	0.0109	0

the main contribution of this paper. Elements in the cost matrix measure the loss of lenders' expected return if a loan is misclassified. For example, in the cost matrix $\boldsymbol{C}$, $c_{21} = 0.0333$ means that a lender who provides a loan with a certain amount *amt* will loss 0.0333**amt* in her/his expected return when a loan of credit grade B is misclassified as grade A. Meanwhile, the revenue of the platform reduces because of the misclassification of credit grades, where the costs should be minimized in cost-sensitive credit rating.

The largest cost in Table 12 is 0.0856, which happens when class G was misclassified as class A. It is align with the reality because this misclassification assigns the lowest interest rate to the highest credit risk group. And the smallest cost is 0.0055, which appears when class E is misclassified into F.

5.4. Sensitivity analysis

The parameters α and β were set as the median value of the corresponding ranges in the experiments. In reality, there is uncertainty on the parameters due to borrowers' behaviors. This section uses sensitivity analysis to study how the uncertainty of input parameters affects the cost matrix in the output.

To facilitate observations of the cost matrixes under different parameters, cosine similarity is adopted to compare them with the cost matrix $\boldsymbol{C} = (c_{ij})$ shown in Table 12, where parameters are taken as median (α_0, β_0). Given any $-1.2916 < \alpha_1 < -0.8412$ and $-0.1116 < \beta_1 < 0$, the corresponding cost matrix $D = (d_{ij})$ can be obtained. Then the cosine similarity between C and D is calculated as follows:

$$\text{Cosine Similarity}(C, D) = \frac{\sum c_{ij} \cdot d_{ij}}{\sqrt{\sum c_{ij}^2} \cdot \sqrt{\sum d_{ij}^2}} \tag{20}$$

In other words, the cost matrix C in Table 12 is considered as benchmark in calculation of cosine similarity. Figure 1 shows the similarity result when the parameters vary in their ranges. Obviously, the similarity is equal to 1 when α and β are median, because it is benchmark. To observe more details of the result, we take the parameters α and β respectively at 9 equidistant points within their ranges. Table 13 lists the cosine similarities on these discrete values of parameters. The element at the center of Table 13 indicates the situation of benchmark (Table 12). The lowest similarity appears at the corner of the range. When both α and β are taken as the maximum in their ranges, the fact that similarity equals 0.9451 still guaranteed a small difference between the corresponding cost matrix and benchmark.

The sensitivity analysis for the parameters shows that the proposed method used to calculate cost matrix is robust in the presence of uncertainty of parameters. Thus, it is reasonable to set the parameters as the median value of the ranges.

5.5. Cost-sensitive credit rating

After calculated the misclassification cost matrix, we can use cost-sensitive credit rating to classify the loans into credit grades (A, B, C, D, E, F, or G), using the collected Lending Club data.

Data preprocessing was performed first. We deleted some null features and irrelevant features (such as title and other date-type features). The number of features reduced from 128 to 74. We randomly sample 10% instances from 759348 loans to train and test classifiers.

In this experiment, we selected four well-known classifiers (Lessmann et al., 2015) and the corresponding Meta-cost-sensitive classifiers to compare their performances. The four classifiers are C4.5, Random Forest, Logistic Regression, and SVM. Then we adopted a Meta approach proposed by Domingos (1999) to make the four classifiers cost-sensitive by wrapping a cost-minimizing procedure around it. The proposed cost matrix is used to guide credit rating in cost-sensitive classification.

For experimental setup, the cost-insensitive and cost-sensitive classifiers were evaluated by 10-folds cross validation using Weka software (Frank, Hall, & Witten, 2016). The performances of classifiers were evaluated using accuracy, total cost, average cost, and cost saving rate. They were computed using the confusion matrix $\boldsymbol{N}$ (Table 14) in classification result and the cost matrix proposed in Section 3.

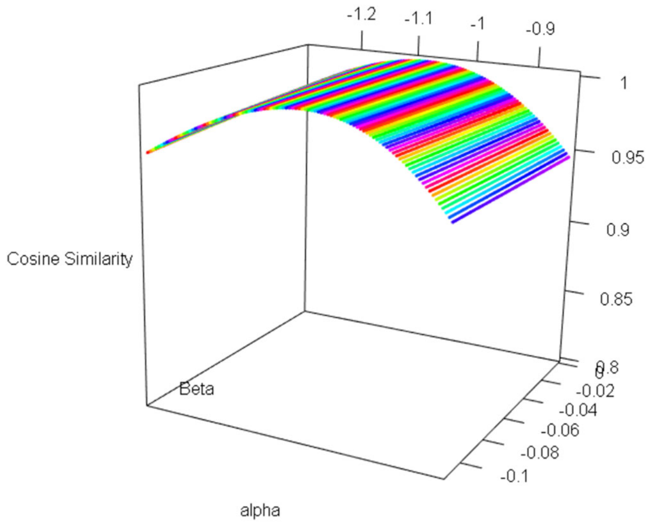

Figure 1. Sensitivity analysis for cosine similarity of cost matrixes in terms of parameters uncertainty.

Table 13. Cosine similarities of cost matrixes when the parameters are discrete values.

Alpha \ Beta	−0.1116	−0.0977	−0.0837	−0.0698	−0.0558	−0.0419	−0.0279	−0.0140	0.0000
−1.2916	0.9619	0.9635	0.9651	0.9666	0.9681	0.9695	0.9708	0.9721	0.9734
−1.2353	0.9758	0.9770	0.9783	0.9794	0.9805	0.9816	0.9826	0.9836	0.9845
−1.1790	0.9874	0.9883	0.9891	0.9899	0.9906	0.9914	0.9920	0.9926	0.9932
−1.1227	0.9958	0.9963	0.9967	0.9971	0.9975	0.9978	0.9981	0.9984	0.9986
−1.0664	0.9998	0.9999	1.0000	1.0000	**1.0000**	1.0000	1.0000	0.9999	0.9998
−1.0101	0.9982	0.9979	0.9977	0.9974	0.9971	0.9968	0.9965	0.9962	0.9959
−0.9538	0.9897	0.9892	0.9887	0.9883	0.9878	0.9873	0.9868	0.9863	0.9859
−0.8975	0.9732	0.9727	0.9722	0.9716	0.9711	0.9706	0.9701	0.9696	0.9691
−0.8412	0.9483	0.9478	0.9474	0.9470	0.9466	0.9463	0.9459	0.9455	0.9451

Table 14. Confusion matrix $\boldsymbol{N}$*.

		Prediction			
		A	B	…	G
Actual	A	n_{11}	n_{12}	…	n_{17}
	B	n_{21}	n_{22}	…	n_{27}
	…	…	…		…
	G	n_{71}	n_{72}	…	n_{77}

*n_{ij} indicates the number of instances of ith grades classified into jth grades.

Table 15. Classification results using proposed cost matrix.

Classifier	Accuracy	Total cost	Average cost	Cost saving rate
C4.5	79.29%	328.1648	0.0043	13.12%
Cost-sensitive C4.5	77.96%	285.112	0.0038	
Random forest	57.18%	730.5374	0.0096	25.48%
Cost-sensitive RF	40.66%	544.4132	0.0072	
Logistic regression	61.12%	565.1794	0.0074	21.37%
Cost-sensitive LR	48.44%	444.4158	0.0059	
SVM	52.95%	808.4923	0.0106	11.94%
Cost-sensitive SVM	48.48%	711.933	0.0094	

Accuracy is the percentage of correctly classified loans (Kou, Lu, Peng, & Shi, 2012) and computed as $Accuracy = \sum_i n_{ii} / \sum_{i,j} n_{ij}$. Total cost is the sum of the costs of misclassified instances, and $Total\ cost = \sum_{i,j} n_{ij} \cdot c_{ij}$. Average cost is cost per instance, that is, $Average\ cost = Total\ cost / \sum_{i,j} n_{ij}$. Cost saving rate is the rate of decreased cost after using cost-sensitive classifiers.

The classification results were summarized in Table 15. Since traditional standard classifiers focus on optimizing the accuracy and ignoring the relationship between classes and different misclassifications, they have higher accuracies than their cost-sensitive counterparts, and the cost-sensitive classifiers reduced the total costs dramatically, which is a major concern of P2P lending platforms and lenders.

It is really difficult, if not impossible, to achieve the highest accuracy and the lowest total cost simultaneously. Cost-sensitive classification normally sacrifices accuracy for lower total cost (Wang, Kou, & Peng, 2018). The objective of cost-insensitive classifiers is to maximize the total accuracy. But none of them can reach 100% accuracy in multi-class classification. Although misclassified instances may cause different costs according to the cost matrix, cost-insensitive classifiers treat them the same. This is the reason that cost-insensitive classifiers have higher accuracy and higher total misclassification cost. Cost-sensitive classifiers, on the other hand, try to minimize total costs caused by misclassification, and some misclassification errors with low costs are compromised to achieve this goal when training the classifiers. Thus, cost-sensitive classifiers reduce total cost of classification, but have lower accuracy than cost-insensitive classifiers. As shown in Table 15, the classifiers which achieved the highest accuracy and the lowest total cost are C4.5 and cost-sensitive C4.5, respectively. While the accuracy of cost-sensitive C4.5 is 1.33% lower than C4.5, cost-sensitive C4.5 reduced 13.12% total cost, compared to C4.5.

Since the proposed cost matrix measured the real losses of lenders, the total cost is more practical than the accuracy. Compared with the corresponding standard classifiers, the four cost-sensitive classifiers reduced 13%, 25%, 21%, and 12% total costs, respectively (shown as the column "Cost saving rate" in Table 15). In order to connect the classification results and business reality in P2P lending, the average cost indicates how much money is lost in a loan, on average, due to potential misclassification of credit grades when a classification algorithm is used. For example, if standard C4.5 is used to grade a loan, potential misclassification cost will cause 0.43% less returns for lenders, comparing to 100% accurate classification. And this cost dropped to 0.38% if cost-sensitive C4.5 is used.

In summary, using cost-sensitive classifiers for credit rating is able to reduce losses effectively in P2P lending, which is measured by the cost matrix proposed in this paper. It also shows the need to evaluate and develop more credit rating approaches with respect to minimizing financial losses for P2P lending platforms.

6. Conclusion and discussion

Online peer-to-peer (P2P) lending platforms provide convenient and low costs lending option to individuals and small businesses. Since many P2P loans are unsecured personal loans, the quality of credit risk classification is vital to P2P lending platforms. Traditional cost-insensitive binary classification is not appropriate in P2P lending because there are more than two credit classes in real-life P2P lending and misclassification costs are not uniform across classes. Cost-sensitive classifiers try to minimize total costs measured by misclassification cost matrixes. How to measure misclassification costs of different credit grades in P2P lending is a useful but understudied issue.

The objective of this paper is to model credit rating in P2P lending as a cost-sensitive multi-class classification problem. We first proposed a misclassification cost matrix for P2P credit grading that takes into account real losses and opportunity costs associated with P2P lending. A set of equations and models were developed to calculate the costs in the misclassification cost matrix. Then we analyzed the parameters in the proposed equations and models from the perspective of both cost-sensitive classification and business operations.

To validate the proposed misclassification cost matrix, an experiment using publicly available data from Lending Club was conducted. The results showed that standard classifiers have higher accuracy than cost-sensitive counterparts, but the cost-sensitive classifiers significantly reduced the total costs, which is essential for the survival and profitability of P2P lending platforms.

One of the limitations of this work is that it did not take administrative cost into consideration. The administrative cost is an important part of the total cost in traditional bank loans. We did not include the administrative costs in the total cost calculation because Lending Club does not provide such information. Since the administrative cost in P2P lending, comparing with traditional bank loans, is insignificant, the reduced total cost calculated in the experiment can still be used as a useful reference for P2P lending platforms.

This paper defined two variables, revised PD and a borrower's probability of give-up, as linear function related with different interest rates. The actual situations may be more complex than these assumptions. One of the future research directions is to investigate how changes in interest rates affect human behaviors on default and giving up a loan. Though this paper focuses on misclassification costs in credit rating, the basic idea of the proposal is not necessarily restricted to the financial domain and can be applied to a wide selection of areas. For instance, misclassification costs of type I and type II errors are quite different in medical diagnosis. The difficulty lies in the definition of misclassification cost matrix for medical diagnosis. Another future research direction is to generalize the misclassification cost matrix for other multi-class cost-sensitive classification problems.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by grants from the National Natural Science Foundation of China (#U1811462, #71771037, #71971042, #71725001, and #71910107002).

References

Alejo, R., García, V., Marqués, A. I., Sánchez, J. S., & Antonio-Velázquez, J. A. (2013). Making accurate credit risk predictions with cost-sensitive mlp neural networks. In *Management intelligent systems* (pp. 1–8). Heidelberg: Springer.

Bachmann, A., Becker, A., Buerckner, D., Hilker, M. K. M., Lehmann, M., & Tiburtius, P. (2011). Online peer-to-peer lending-a literature review. *Journal of Internet Banking and Commerce*, *16*(2), 1–18.
Bahnsen, A. C., Aouada, D., & Ottersten, B. (2014). Example-dependent cost-sensitive logistic regression for credit scoring. In *2014 13th International Conference on Machine Learning and Applications (ICMLA)* (pp. 263–269). IEEE.
Bahnsen, A. C., Aouada, D., & Ottersten, B. (2015). Example-dependent cost-sensitive decision trees. *Expert Systems with Applications*, *42*(19), 6609–6619. doi:10.1016/j.eswa.2015.04.042
Bahnsen, A. C., Stojanovic, A., Aouada, D., & Ottersten, B. (2013). Cost sensitive credit card fraud detection using Bayes minimum risk. In *2013 12th International Conference on Machine Learning and Applications (ICMLA)* (Vol. 1, pp. 333–338). IEEE.
Beling, P., Covaliu, Z., & Oliver, R. M. (2005). Optimal scoring cutoff policies and efficient frontiers. *Journal of the Operational Research Society*, *56*(9), 1016–1029. doi:10.1057/palgrave.jors.2602021
Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., & Siddique, A. (2016). Risk and risk management in the credit card industry. *Journal of Banking & Finance*, *72*, 218–239. doi:10.1016/j.jbankfin.2016.07.015
Chao, X., & Peng, Y. (2018). A cost-sensitive multi-criteria quadratic programming model for imbalanced data. *Journal of the Operational Research Society*, *69*(4), 500–516. doi:10.1057/s41274-017-0233-4
Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: A recent review. *Artificial Intelligence Review*, *45*(1), 1–23. doi:10.1007/s10462-015-9434-x
Collier, B., & Hampshire, R. (2010). Sending mixed signals: Multilevel reputation effects in peer-to-peer lending markets. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (pp. 197–206). ACM. doi:10.1145/1718918.1718955
Domingos, P. (1999). Metacost: A general method for making classifiers cost-sensitiv. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 155–164). ACM. doi:10.1145/312129.312220
Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of Monetary Economics*, *53*(8), 2283–2298. doi:10.1016/j.jmoneco.2005.09.001
Frank, E., Hall, M. A., & Witten, I. H. (2016). *The WEKA workbench. Online appendix for "Data mining: Practical machine learning tools and techniques"* (4th ed.). Burlington, MA: Morgan Kaufmann.
Elkan, C. (2001). The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*. (Vol. 17, No. 1, pp. 973–978). Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.
Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2017). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448–455. doi:10.1016/j.ins.2017.12.030
Florez-Lopez, R., & Ramon-Jeronimo, J. M. (2014). Modelling credit risk with scarce default data: On the suitability of cooperative bootstrapped strategies for small low-default portfolios. *Journal of the Operational Research Society*, *65*(3), 416–434. doi:10.1057/jors.2013.119
Galloway, I. (2009). Peer-to-peer lending and community development finance. *Community Investments*, *21*(3), 19–23.
Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, *34*(2), 401–424. doi:10.1080/07421222.2017.1334472
Guan, D., Yuan, W., Ma, T., Khattak, A. M., & Chow, F. (2017). Cost-sensitive elimination of mislabeled training data. *Information Sciences*, *402*, 170–181. doi:10.1016/j.ins.2017.03.034
Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, *249*(2), 417–426. doi:10.1016/j.ejor.2015.05.050
Hand, D. J., Whitrow, C., Adams, N. M., Juszczak, P., & Weston, D. (2008). Performance criteria for plastic card fraud detection tools. *Journal of the Operational Research Society*, *59*(7), 956–962. doi:10.1057/palgrave.jors.2602418
Hu, Y., Feng, B., Mo, X., Zhang, X., Ngai, E. W. T., Fan, M., & Liu, M. (2015). Cost-sensitive and ensemble-based prediction model for outsourced software project risk prediction. *Decision Support Systems*, *72*, 11–23. doi:10.1016/j.dss.2015.02.003
Huang, Y., & Kou, G. (2014). A kernel entropy manifold learning approach for financial data analysis. *Decision Support Systems*, *64*, 31–42. doi:10.1016/j.dss.2014.04.004
Klafft, M. (2008). Peer to peer lending: Auctioning microcredits over the internet. In *Proceedings of the International Conference on Information Systems, Technology and Management*. Dubai: IMT.
Kou, G., Ergu, D., Lin, C., & Chen, Y. (2016). Pairwise comparison matrix in multiple criteria decision making. *Technological and Economic Development of Economy*, *22*(5), 738–765. doi:10.3846/20294913.2016.1210694
Kou, G., Lu, Y., Peng, Y., & Shi, Y. (2012). Evaluation of classification algorithms using MCDM and rank correlation. *International Journal of Information Technology & Decision Making*, *11*(01), 197–225. doi:10.1142/S0219622012500095
Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences*, *275*, 1–12. doi:10.1016/j.ins.2014.02.137
Lanzarini, L. C., Villa Monte, A., Bariviera, A. F., & Jimbo Santana, P. (2017). Simplifying credit scoring rules using LVQ + PSO. *Kybernetes*, *46*(1), 8–16. doi:10.1108/K-06-2016-0158
Lending Club. (2017). Data from: Lending club statistics [Dataset]. Retrieved from https://www.lendingclub.com/info/download-data.action.
Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. doi:10.1016/j.ejor.2015.05.030
Li, T., Kou, G., Peng, Y., & Shi, Y. (2017). Classifying with adaptive hyper-spheres: An incremental classifier based on competitive learning. In IEEE Transactions on Systems, Man, and Cybernetics: Systems (pp. 1–12). IEEE.
Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. *Engineering Applications of Artificial Intelligence*, *65*, 465–470. doi:10.1016/j.engappai.2016.12.002
Maldonado, S., Bravo, C., Lopez, J., & Perez, J. (2017). Integrated framework for profit-based feature selection

and SVM classification in credit scoring. *Decision Support Systems*, *104*, 113–121. doi:10.1016/j.dss.2017.10.007

Marqués, A. I., García, V., & Sánchez, J. S. (2013). On the suitability of resampling techniques for the class imbalance problem in credit scoring. *Journal of the Operational Research Society*, *64*(7), 1060–1070. doi:10.1057/jors.2012.120

Morente-Molinera, J. A., Mezei, J., Carlsson, C., & Herrera-Viedma, E. (2017). Improving supervised learning classification methods using multi-granular linguistic modelling and fuzzy entropy. *IEEE Transactions on Fuzzy Systems*, *25*(5), 1078–1089.

Peng, Y., Wang, G., Kou, G., & Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, *11*(2), 2906–2915. doi:10.1016/j.asoc.2010.11.028

Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, *40*(15), 5916–5923. doi:10.1016/j.eswa.2013.05.021

Santana, P. J., Lanzarini, L., & Bariviera, A. F. (2018). Fuzzy Credit Risk Scoring Rules using FRvarPSO. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *26*(Suppl. 1), 39–57. doi:10.1142/S0218488518400032

Schuermann, T. (2004). What do we know about loss given default?

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems*, *89*, 113–122. doi:10.1016/j.dss.2016.06.014

Serrano-Cinca, C., Gutierrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS One*, *10*(10), e0139427doi:10.1371/journal.pone.0139427

Sun, J., Shang, Z., & Li, H. (2014). Imbalance-oriented SVM methods for financial distress prediction: A comparative study among the new SB-SVM-ensemble method and traditional methods. *Journal of the Operational Research Society*, *65*(12), 1905–1919. doi:10.1057/jors.2013.117

Tapkan, P., Özbakır, L., Kulluk, S., & Baykasoğlu, A. (2016). A cost-sensitive classification algorithm: BEE-Miner. *Knowledge-Based Systems*, *95*, 99–113. doi:10.1016/j.knosys.2015.12.010

Verbraken, T., Bravo, C., Weber, R., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, *238*(2), 505–513. doi:10.1016/j.ejor.2014.04.001

Wang, H., Kou, G., & Peng, Y. (2018). Cost-sensitive classifiers in credit rating: A comparative study on P2P lending. In *2018 7th International Conference on Computers Communications and Control (ICCCC)* (pp. 210–213). IEEE. doi:10.1109/ICCCC.2018.8390460

Wu, W., & Kou, G. (2016). A group consensus model for evaluating real estate investment alternatives. *Financial Innovation*, *2*(1), 8. doi:10.1186/s40854-016-0027-8

Xia, Y., Liu, C., & Liu, N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, *24*, 30–49. doi:10.1016/j.elerap.2017.06.004

Xu, J., Chen, D., & Chau, M. (2016). Identifying features for detecting fraudulent loan requests on P2P platforms. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)* (pp. 79–84). IEEE.

Zhu, B., Baesens, B., Backiel, A., & Vanden Broucke, S. K. (2018). Benchmarking sampling techniques for imbalance learning in churn prediction. *Journal of the Operational Research Society*, *69*(1), 49–65. doi:10.1057/s41274-016-0176-1