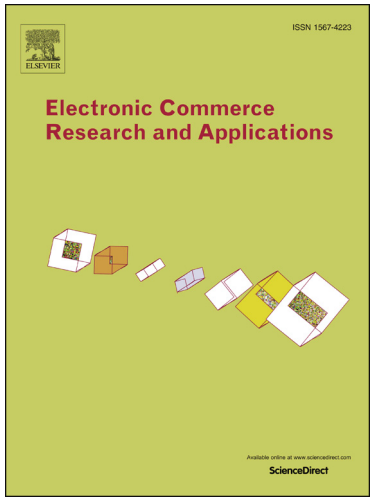
Accepted Manuscript



Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending

Yufei Xia, Chuanzhe Liu, Nana Liu

PII: S1567-4223(17)30049-2
DOI: http://dx.doi.org/10.1016/j.elerap.2017.06.004
Reference: ELERAP 715

To appear in: *Electronic Commerce Research and Applications*

Received Date: 8 December 2016
Revised Date: 20 June 2017
Accepted Date: 20 June 2017

Please cite this article as: Y. Xia, C. Liu, N. Liu, Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending, *Electronic Commerce Research and Applications* (2017), doi: http://dx.doi.org/10.1016/j.elerap.2017.06.004

Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending

Yufei Xia[a], Chuanzhe Liu[a*], Nana Liu[a]

[a] School of Management, China University of Mining and Technology, Xuzhou, Jiangsu 221116, PR China

Abstract: Loan evaluation is an effective tool for credit risk assessment in the peer-to-peer (P2P) lending market. However, most of the traditional loan evaluation models assume a balanced misclassification cost, which is far from business reality and leaves opportunities for investigating the cost-sensitive modelling approaches. Moreover, few existing studies consider the special characteristics of P2P lending. In this paper, we propose a cost-sensitive boosted tree loan evaluation model by incorporating cost-sensitive learning and extreme gradient boosting (XGBoost) to enhance the capability of discriminating potential default borrowers. Therefore, a portfolio allocation model that converts the portfolio optimization problem into an integer linear programming is proposed as a decision support system for unprofessional lenders. Unlike existing studies, which mainly evaluate the model through accuracy-based metrics, this study measures the expected profitability based on the annualized rate of return (ARR) and proposes a new metric (ARR curve) reflecting the ARR over different rejection rates. Two real-world P2P lending datasets are examined, and explanatory data analysis shows that explanatory variables differ in predicting the probability of default and profitability. Experimental results reveal that the area under the ARR curve is not a considerable metric in evaluating models from a profit perspective. Conventional cost-insensitive models may even lead to negative returns, whereas the proposed loan evaluation and portfolio allocation model are the best performing methods over the two datasets in terms of profitability.

Keywords: cost-sensitive learning, P2P lending, loan evaluation, extreme gradient boosting, portfolio allocation

[*] Corresponding author at: School of Management, China University of Mining and Technology, Xuzhou, Jiangsu 221116, PR China. Tel: +86 516 83590168.
Email address: xiareyoung@cumt.edu.cn (Yufei Xia)
rdean@cumt.edu.cn (Chuanzhe Liu),
liunana1004@163.com (Nana Liu).