

CITY UNIVERSITY OF HONG KONG
香港城市大學

**Risk Forecast Model for Online P2P
Lending Platform**
P2P 網絡借貸平臺風險預測模型研究

Submitted to
College of Business
商學院
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Business Administration
工商管理學博士學位

by

Zhang Daoshun
张道舜

July 2016
二零一六年七月

摘要

近年來，P2P 網絡借貸發展極為迅速，這使得對借款人的風險評估和預測變得尤為重要。儘管以往的研究從“硬信息”（即款項信息、借款人自身特點等）和“軟信息”（社會關係、社會網絡資本等）等多方面利用計量經濟學方法來評估 P2P 網絡借貸借款人的風險，但專門針對學生貸款平臺，挖掘影響學生風險因素的研究相對缺乏，研究逾期行為的研究更少。本研究針對學生 P2P 借款平臺，利用實際二手數據，從款項層面和單筆還款層面分別考察逾期和違約兩個階段的風險影響因素，基于決策樹分析的方法提出風險預測模型，并用另一批實際數據對模型進行檢驗。同時，我們利用梯度提升決策分析樹和計量模型對預測效果進行進一步驗證。最後，基于模型結果我們分別計算了本文所提出的風險預測模型所帶來的平臺資本損益。我們發現許多區別于傳統文獻研究的變量在風險預測方面發揮著重要作用。例如，學生院校和家鄉城市的距離、學生院校城市和家鄉城市的經濟水平差距、學生同父母的關係、學生的學歷、住房條件、性別等因素都會直接影響到學生的逾期和壞賬風險。利用我們所提出的風險預測模型能有效提升 P2P 借貸平臺的風險評估水平，并帶來很大的正向的資本收益。本研究在研究對象（“拍來貸”學生 P2P 網絡借貸平臺）和結合學生背景的影響因素、逾期和違約兩階段分析以及利用數據挖掘的方法論等方面具有較強的理論貢獻。此外，本研究在用戶風險預測和信用評估、P2P 借貸平臺損益分析和運營策略等方面具有很強的實踐意義。

ABSTRACT

In recent years, online P2P lending grows remarkably, which highlights the significance of risk evaluation and risk forecast of borrowers. Although past studies on risk evaluation in online P2P lending platform were from many aspects of hard information (i.e. loan information, borrower characteristics etc.) and soft information (i.e. social relationship, social network capital etc.) with econometric method, few research has paid attention to mining influential factors of risk regarding to student lending platform, and study considering delinquency behavior is rare. Focusing on student P2P lending platform and using actual second-hand data, this study investigated the influential factors of risk on borrowers' delinquency and default behaviors on the level of loan and level of monthly payment respectively. We proposed the risk forecast model with decision tree analysis, and adopted another real dataset to test the forecast model. Meanwhile, we used gradient boost decision tree (GBDT) and econometric model to further validate the forecast model. Finally, based on the proposed forecast model, we conducted capital gain and loss analysis for online P2P lending platform. We found that many factors that are different from past studies were important on student borrower risk prediction. For example, the distance between school (university or college) and hometown, the economic gap between school (university or college) and hometown, the relationship between student borrowers and their parents, housing condition, education level, gender will have direct impact on student risk (delinquency or default behaviors). The proposed forecast model will effectively improve the risk evaluation of online P2P lending platform, and bring large positive capital revenue. This study has several theoretical contributions on research object (i.e. student online P2P lending) and student-related risk factors, two-stage risk forecast on borrowers' delinquency and default behaviors, as well as adoption of data mining methodology. Moreover, the study has strong implications for online P2P lending platform on borrower risk prediction and credit assessment, P2P lending cost-benefit analysis and platform operation.

目錄

摘要.....	i
ABSTRACT.....	ii
Qualifying Panel and Examination Panel	iii
目錄.....	iv
表目錄.....	vii
圖目錄.....	ix
第一章 引言.....	1
1.1 研究背景.....	1
1.1.1 P2P 網絡借貸的概念與發展.....	1
1.1.2 P2P 網絡借貸的重要性.....	8
1.1.3 P2P 網絡借貸的風險與信用管理.....	9
1.2 研究問題與意義.....	15
1.2.1 研究問題.....	15
1.2.2 研究意義.....	17
1.3 主要研究方法.....	18
1.4 論文結構安排.....	19
第二章 國內外研究綜述.....	21
2.1 P2P 網絡借貸行業的發展	21
2.1.1 P2P 網絡借貸的興起.....	21
2.1.2 國外 P2P 網絡借貸的發展.....	23
2.1.3 國內 P2P 網絡借貸的發展.....	24
2.2 P2P 網絡借貸的模式與機制	28
2.2.1 P2P 網絡借貸模式.....	28
2.2.2 P2P 網絡借貸機制.....	30
2.3 P2P 網絡借貸的用戶行爲	31
2.3.1 借款成功率、借款利率以及違約概率的影響因素	31
2.3.2 用戶的羊群效應及信號傳遞	34
2.4 P2P 網絡借貸的風險控制	35
2.4.1 用戶借款風險因素研究：硬信息與軟信息	35
2.4.2 P2P 網絡借款平臺的風險管控.....	35
2.5 傳統金融行業的相關研究.....	37
2.6 現有文獻研究評述.....	39
第三章 研究理論基礎.....	40
3.1 一般壓力理論.....	40
3.1.1 一般壓力理論的主要概念	40
3.1.2 壓力類型與壓力分類	41
3.1.3 壓力源與負面情緒	43
3.1.4 一般壓力理論的理論模型	45
3.2 道德心理模型.....	46
3.2.1 道德心理模型的產生	46
3.2.2 道德心理四過程模型	47

第四章 研究方法介紹.....	52
4.1 決策樹方法.....	52
4.1.1 決策樹算法的基本思想.....	52
4.1.2 決策樹的構建過程.....	53
4.1.3 決策樹的性能評估.....	55
4.1.4 常見的決策樹算法.....	57
4.2 梯度提升決策樹方法.....	66
4.2.1 梯度提升決策樹的基本思想.....	66
4.2.2 Boosting 方法.....	67
4.2.3 梯度提升 (Gradient Boosting).....	68
4.2.4 Treelink 模型.....	68
4.2.5 梯度提升算法的模型評價.....	69
4.3 面板二值響應模型.....	69
4.3.1 標準 Logistic 模型.....	70
4.3.2 修正的 Logistic 模型.....	70
4.3.3 基于面板數據的二值響應模型.....	72
4.3.4 二值響應模型的評價.....	75
第五章 研究對象及樣本信息描述.....	76
5.1 研究平臺.....	76
5.2 樣本選擇與數據準備.....	77
5.3 變量定義.....	77
5.3.1 變量選取及定義.....	77
5.3.2 變量相關性.....	81
5.4 樣本描述性統計.....	84
5.4.1 借款人基本信息描述性統計.....	85
5.4.2 款項信息描述性統計.....	101
5.4.3 借款人還款信息描述性統計.....	106
第六章 P2P 網絡借貸平臺用戶逾期風險預測模型.....	115
6.1 基于款項數據的用戶逾期風險預測模型.....	115
6.1.1 逾期風險預測模型結果.....	115
6.1.2 逾期風險預測模型準確率.....	116
6.1.3 逾期風險預測模型檢驗.....	117
6.2 基于單筆還款數據的用戶逾期風險預測模型.....	118
6.2.1 逾期風險預測模型結果.....	118
6.2.2 逾期風險預測模型準確率.....	119
6.2.3 逾期風險預測模型檢驗.....	120
6.3 用戶逾期風險預測模型的進一步驗證.....	120
6.3.1 基于款項數據的用戶逾期風險預測模型進一步驗證.....	121
6.3.2 基于單筆還款數據的用戶逾期風險預測模型進一步驗證.....	123
6.4 結論與討論.....	125
第七章 P2P 網絡借貸平臺用戶壞賬風險預測模型.....	127
7.1 基于款項數據的用戶壞賬風險預測模型.....	127
7.1.1 壞賬風險預測模型結果.....	127
7.1.2 壞賬風險預測模型準確率.....	128

7.1.3 壞賬風險預測模型檢驗	129
7.2 基于單筆還款數據的用戶壞賬風險預測模型	129
7.2.1 壞賬風險預測模型結果	129
7.2.2 壞賬風險預測模型準確率	131
7.2.3 違約風險預測模型檢驗	131
7.3 用戶壞賬風險預測模型的進一步驗證	132
7.3.1 基于款項數據的用戶壞賬風險預測模型進一步驗證	132
7.3.2 基于單筆還款數據的用戶壞賬風險預測模型進一步驗證	134
7.4 基于用戶壞賬風險預測模型的平臺經濟損益分析	136
7.5 結論與討論	139
第八章 研究總結與展望	141
8.1 研究結論總結	141
8.2 研究創新與貢獻	141
8.2.1 研究的理論貢獻	141
8.2.2 研究的實踐貢獻	142
8.3 研究不足與展望	142
8.3.1 研究不足	142
8.3.2 未來展望	143
參考文獻	145

表目錄

表 1 本文的结构框架.....	20
表 2 国外 P2P 网络借贷平台模式比较.....	28
表 3 国内 P2P 网络借贷平台模式总结.....	29
表 4 部分借款成功率和违约率影响因素总结.....	33
表 5 终止决策树构建的各类方法.....	54
表 6 决策树的性能评估指标.....	55
表 7 ID3 算法的优点和不足总结.....	60
表 8 C4.5 算法的不足总结.....	63
表 9 款项信息相关变量.....	78
表 10 借款人基本信息相关变量.....	79
表 11 借款人行为相关变量.....	80
表 12 自变量与因变量（单笔是否逾期）的 Spearman 相关系数.....	81
表 13 自变量与因变量（单笔是否坏账）的 Spearman 相关系数.....	82
表 14 自变量与因变量（款项是否逾期）的 Spearman 相关系数.....	83
表 15 自变量与因变量（款项是否坏账）的 Spearman 相关系数.....	83
表 16 用户款项是否逾期预测模型规则.....	116
表 17 款项逾期风险预测模型训练分类准确率.....	116
表 18 款项逾期风险预测模型训练分类检验结果.....	117
表 19 用户单笔还款是否逾期预测模型规则.....	118
表 20 单笔还款逾期风险预测模型训练分类准确率.....	119
表 21 单笔还款逾期风险预测模型训练分类检验结果.....	120
表 22 梯度提升决策树预测款项逾期风险影响因素重要性占比.....	122
表 23 梯度提升决策树款项逾期风险预测准确率.....	122
表 24 面板 logit 模型回归结果（因变量：款项是否逾期）.....	123
表 25 梯度提升决策树预测单笔还款逾期风险影响因素重要性占比.....	124
表 26 梯度提升决策树单笔还款逾期风险预测准确率.....	124
表 27 面板 logit 模型回归结果（因变量：单笔还款是否逾期）.....	125

表 28	用户款项是否坏账预测模型规则.....	127
表 29	款项坏账风险预测模型训练分类准确率.....	128
表 30	款项坏账风险预测模型训练分类检验结果.....	129
表 31	用户单笔还款是否坏账预测模型规则.....	130
表 32	单笔还款坏账风险预测模型训练分类准确率.....	131
表 33	单笔还款坏账风险预测模型训练分类检验结果.....	132
表 34	梯度提升决策树预测款项坏账风险影响因素重要性占比.....	132
表 35	梯度提升决策树款项坏账风险预测准确率.....	133
表 36	面板 logit 模型回归结果（因变量：款项是否坏账）	133
表 37	梯度提升决策树预测单笔还款坏账风险影响因素重要性占比.....	134
表 38	梯度提升决策树单笔还款坏账风险预测准确率.....	135
表 39	面板 logit 模型回归结果（因变量：单笔还款是否坏账）	135
表 40	坏账风险预测模型所带来的平台经济损益.....	137
表 41	坏账风险预测模型所带来的平台经济损益（逾期款项）	138
表 42	坏账风险预测模型所带来的平台经济损益（特殊款项）	138

圖目錄

图 1 P2P 网贷市场产业链（发展模式）	3
图 2 P2P 网络借贷平台的发展年表	5
图 3 2015 年 P2P 网络借贷平台区域分布	7
图 4 P2P 网络借贷平台融资情况	8
图 5 P2P 网络借贷问题平台类型分布	11
图 6 P2P 网络借贷借款人还款行为过程	16
图 7 一般压力理论解释违约行为.....	45
图 8 ID3 核心算法流程图	60
图 9 Boosting 过程示例	67
图 10 Boosting 的公式表示	68
图 11 梯度提升算法.....	68
图 12 Treelink 模型的学习过程	69
图 13 借款学生院校及居住省份分布（训练样本）	86
图 14 借款学生院校及居住省份分布（预测样本）	87