

面向互联网金融平台的违约风险量化模型

白云歌^{a,b,c}, 郭炳晖^{a,b,c}, 米志龙^{a,b,c}, 郑志明^{a,b,c}

(北京航空航天大学 a. 数学与系统科学学院; b. 数学信息与行为教育部重点实验室;
c. 大数据与脑机智能高精尖中心, 北京 100191)

摘 要: 针对互联网金融平台上的借贷违约行为, 结合典型相关分析模型与复杂网络特征提取方法, 建立违约风险量化模型, 并利用 ROC 曲线与 AUC 值对模型效果进行评价。将该模型应用于某互联网借贷平台实际交易数据中, 并与决策树模型进行比较分析, 结果表明, 该模型能够有效反映用户违约特征与复杂网络特征的关联效应, 其 AUC 值能达到 0.85 左右, 且具有更强的稳定性与鲁棒性。

关键词: 互联网金融; 典型相关分析; 复杂网络; 中心性; 风险量化

中文引用格式: 白云歌, 郭炳晖, 米志龙, 等. 面向互联网金融平台的违约风险量化模型[J]. 计算机工程, 2018, 44(12): 108-114.

英文引用格式: BAI Yunge, GUO Binghui, MI Zhilong, et al. Quantitative model of default risk for Internet financial platform[J]. Computer Engineering, 2018, 44(12): 108-114.

Quantitative Model of Default Risk for Internet Financial Platform

BAI Yunge^{a,b,c}, GUO Binghui^{a,b,c}, MI Zhilong^{a,b,c}, ZHENG Zhiming^{a,b,c}

(a. School of Mathematics and System Science; b. Key Laboratory of Mathematics Information and Behavior Ministry of Education;
c. Big Data and Brain Machine Intelligence High-grade Center, Beihang University, Beijing 100191, China)

[Abstract] Aiming at the loan default behavior on the Internet financial platform, combining the canonical correlation analysis model and the complex network feature extraction method, the quantitative model of default risk is established, and the ROC curve and AUC value are used to evaluate the effect of the model. The model is applied to the actual transaction data of an Internet lending platform and compared with the decision tree model. Results show that the proposed model can effectively reflect the correlation effect between user default features and complex network features, the AUC value of the model can reach about 0.85, and it has stronger stability and robustness.

[Key words] Internet finance; canonical correlation analysis; complex network; centrality; risk quantification

DOI: 10.19678/j.issn.1000-3428.0048789

0 概述

伴随着互联网技术的不断发展, 传统的民间资本交易、中小企业融资和个人借贷等线下业务在互联网金融平台上不断发展壮大。在具有金融中介性质的互联网金融平台中, 如何通过网络及数据分析实现对用户的信用评价, 是目前广大互联网金融机构关注的核心技术难题, 也是当今的一个研究热点^[1]。根据网络小额贷款的特点以及其与传统征信系统的不同点去建立合适的违约风险评价系统, 是减少不良交易、促进平台发展的重要途径。

信用评价与风险预测是解决上述问题的 2 个方向。文献[2]使用统计中的判别分析来评价贷款并提出 FICO 信用评分体系。近年来, 随着神经网络模型的建立, 较多研究者使用机器学习方法来解决信用评价与违约预测问题^[3-4]。但是, 使用机器学习方法解决信用问题存在一定局限性, 当样本不足时其无法进行训练, 且效果良好的模型往往复杂且解释性不强, 这限制了分析人员对模型的把控与干预能力。

现有研究以及传统方法更注重对指标的研究^[5], 原因是在传统金融信用评价体系中, 多数是金融机构对用户的单向评价, 而在网络借贷关系中, 用

基金项目: 国家自然科学基金重大项目(11290141); 国家自然科学基金(11401017, 11671025)。

作者简介: 白云歌(1993—), 男, 硕士研究生, 主研方向为复杂网络与数据分析; 郭炳晖(通信作者), 讲师、博士; 米志龙, 硕士研究生; 郑志明, 教授、博士。

收稿日期: 2017-09-25

修回日期: 2017-11-01

E-mail: baiyg@buaa.edu.cn

户通过借贷而构成一个复杂网络并进行互相评价,因此,用户在复杂网络中的特征显然也与用户信用相关。文献[6]研究通过手机通讯录构建静态网络。复杂网络指标,即对网络中节点的评价,也是该领域的热门研究内容^[7]。其中,各文献根据不同的侧重点来研究节点的重要性,如文献[8]基于网络局部结构,文献[9-10]基于节点在网络中所对应路径的属性,文献[11-13]基于网络中影响力的传播,即网络中基于迭代的指标等。加入用户的网络属性作为分析违约风险的特征,也是一个新的尝试。

本文在对具有金融中介性质的某互联网金融平台海量用户数据进行分析的基础上,引入复杂网络结构特征作为分析变量以增加分析维度,然后建立使用典型相关分析的用户信用评价量化模型。将基于社交网络的互联网借贷关系与用户发生债务违约的行为相关联,给出用户信用评分策略及相应算法,并通过在实际平台数据集上的对比实验,验证该模型在风险评价维度、违约概率预测等方面的性能。

1 风险量化模型

在P2P互联网借贷平台中,用户的交易是小额且频繁的,因此,其具有独有的优势,即拥有足量的历史交易数据可以用来构建每个用户的属性以及用户间的关系。

本文构建的风险量化模型如图1所示,通过历史交易记录提取借贷特征与用户网络特征,结合违约状态进行分析得到量化模型,并通过AUC值对模型进行评价和反馈修正。

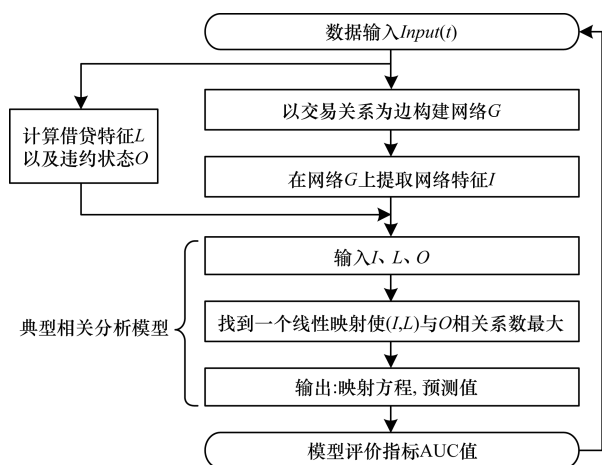


图1 风险量化模型流程

在原始数据输入中,每个记录都具有多个属性,用来描述一笔交易的过程,如借入方、借出方、交易金额、债务开始时间等。选择 $[T-t, T]$ 时间区间内

的原始数据 $Data$ 作为网络构建的依据:

$$Input(t) = \{(T-t) < Data < T\} \quad (1)$$

将借贷数据中每个用户对应网络中的一个节点,每条借贷记录构成网络中的一条边,借出方和借入方分别对应边的起始点和结束点,以此构建借贷网络 $G(V, E)$ 与邻接矩阵 A 。在网络 $G(V, E)$ 中, V 和 E 分别为节点和边的集合。 $A = \{a_{ij}\}$, 其中, v_i 和 v_j 连接时 $a_{ij} = 1$, 否则 $a_{ij} = 0$ 。

通过在构建的借贷网络上应用复杂网络节点重要性排序方法,得出用户在网络结构中对不同含义的地位,以及用户交易对象中所含与逾期用户相关的指标:度中心性(入度、出度),接近度中心性^[9],介数中心性^[10],特征向量中心性,PageRank 值^[12],HITS 算法权威值与枢纽值^[13]。同时,提取用户历史借贷记录中包含违约的记录,并将违约交易的若干属性来代表用户的历史违约情况,构成基于历史的用户信用情况以及目标预测区间内用户的违约情况 O 。

应用典型相关分析方法找到一组线性变换使得 (I, L) 经投影到一维空间时与 O 的相关系数达到最高,利用这种映射得到一个通过 (I, L) 预测 O 值的模型:

$$Predict = f(I, L) \quad (2)$$

式(2)是一个二分类模型,因此,通过绘制 ROC 曲线以及计算 AUC 值的方法来评价该模型。AUC 值计算方法如下:

$$AUC = ROC(Predict, O) \quad (3)$$

因模型 f 的输入 (I, L) 还受到参数 t 的影响,令 M 为 t 的参数,求得对应于 f 的 AUC 值的模型,结合数据优化得到最佳模型的时间区间为:

$$t_0 = \operatorname{argmax} M(Input(t)) \quad (4)$$

2 基于典型相关分析与复杂网络特征的风险量化算法

传统的风险量化算法更注重用户自身属性,在P2P网络借贷中考虑用户在网络中的状态,是本文对传统方法进行的改进。关于网络构建数据的选择,本文通过结合AUC值对模型的评价完成一个反馈过程,以此得到最佳网络构建数据选择区间。

2.1 模型输入数据

风险量化模型所使用的属性定义如下:

1) 借入方 (*Debtor*): 交易中借入方的用户 ID。

2) 借出方 (*Creditor*): 交易中借出方的用户 ID。

3) 交易金额 (*Principal*): 该笔交易中起始本金。

4) 债务开始时间 (*DebtStartTime*): 交易达成时间。

5) 债务结束时间 (*DebtEndTime*): 交易达成时设定的债务偿还时间。

6) 逾期利息 (*OverdueInterest*): 未按债务结束时间偿还债务所生成的额外逾期利息值, 若按时偿还或债务未到期, 则此项为 0。

以上 6 项历史交易记录特征为本文模型的输入内容:

$$Input(t) = \{Data((T-t) < DebtStartTime < T)\}$$

2.2 特征提取算法

网络特征 I 的各项指标定义如下:

1) 节点 v_i 的度定义为与该节点直接相连的邻居数量 k_i 。为比较不同网络中节点的度, 归一化度中心性 (DC) 为:

$$DC(I) = \frac{k_i}{n-1} \quad (5)$$

其中, n 为网络 G 的总节点数, $n-1$ 为最大可能的度。对于有向网络, 还可以分别考虑节点的入度和出度。

2) 节点 v_i 的接近度中心性 (CC) 定义为从 v_i 到其他所有节点的最短路径距离平均值的倒数:

$$CC(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}} \quad (6)$$

其中, d_{ij} 表示节点 v_i 与 v_j 的最短路径。接近度中心性可以理解为信息在网络中平均传播长度的逆。

3) 节点 v_i 的介数中心性 (BC) 定义为:

$$BC(i) = \frac{2}{(n-1)(n-2)} \sum_{i \neq s, i \neq t, s \neq t} \frac{g_{st}^i}{g_{st}} \quad (7)$$

其中, g_{st} 为节点 v_s 和 v_t 之间的最短路径数量, g_{st}^i 为节点 v_s 和 v_t 之间的最短路径中通过 v_i 的数量。介数中心性表现节点在最短路径传播中所占的比例, 即信息流的可控性。

4) 特征向量中心性表示一个节点的影响不仅由它的邻居节点数目来确定, 还由每个相邻节点的影响来确定。节点 v_i 由 x_i 表示其重要性为:

$$x_i = c \sum_{j=1}^n a_{ij} x_j \quad (8)$$

其中, c 是比例常数。一般情况下, $c = 1/\lambda$, λ 是邻接矩阵 $A = \{a_{ij}\}$ 的最大特征值。

5) PageRank (PR) 值表示从由网页之间的关系构成的网络上, 通过随机行走来区分不同网站的重要性。每个节点获得一个单位的 PR 值后这些 PR

值沿其指向链接的邻居均匀分布。节点 v_i 在 t 步骤的 PR 值为:

$$PR_i(t) = \sum_{j=1}^n a_{ji} \frac{PR_j(t-1)}{k_j^{\text{out}}} \quad (9)$$

其中, k_j^{out} 表示节点 v_j 的出度。如果所有节点的 PR 值达到稳定, 则上述迭代停止。

6) HITs 算法考虑网络中每个节点的 2 个属性: 权威性和枢纽性。权威性和枢纽性互相加强, 一个好的枢纽指向许多权威, 许多枢纽共同指向的必是好的权威节点。在有向网络中, 节点的权威得分等于指向该节点的所有节点的枢纽得分总和, 而节点的枢纽得分等于由该节点指向的所有节点的权威得分总和。在拥有 n 个节点的网络中, 分别用 $a_i(t)$ 和 $h_i(t)$ 表示 t 时间节点 v_i 的权威得分和枢纽得分。初始时刻, 所有节点的 2 个分值均为 1。节点 v_i 的权威得分和枢纽得分计算方法分别为:

$$\begin{aligned} a'_i(t) &= \sum_{j=1}^n a_{ji} h_j(t-1) \\ h'_i(t) &= \sum_{j=1}^n a_{ij} a'_j(t) \end{aligned} \quad (10)$$

每次迭代后, 对每个节点的 2 个得分进行标准化:

$$a_i(t) = \frac{a'_i(t)}{\|a'(t)\|}, h_i(t) = \frac{h'_i(t)}{\|h'(t)\|} \quad (11)$$

其中, $\|x\| = \sqrt{\sum_i (x_i)^2}$ 。

7) 平均邻居度: 节点的邻居节点度值的平均。

8) 一阶邻居违约数: 节点邻居中曾经逾期的邻居数量。

9) 一阶邻居违约率: 节点邻居中曾经逾期的邻居数量与节点邻居总数的比值。

违约状态 O 由研究目标时刻到之后 1 个月这一阶段时间内用户是否逾期得出。对于每个用户, 若该时间区间内有交易发生逾期, 则该目标值为 1, 否则为 0。

2.3 典型相关分析方法

典型相关分析是反映 2 组变量间相关性的多元统计分析方法, 其在所有线性组合中, 找到 2 组变量通过线性组合达到的最好的相关关系组合, 并以此实现通过一组变量的数值预测另一组变量数值的目的^[14]。

如图 2 所示, 本文研究变量 $X = (I, L)$ 与 $Y = O$ 间的关系。

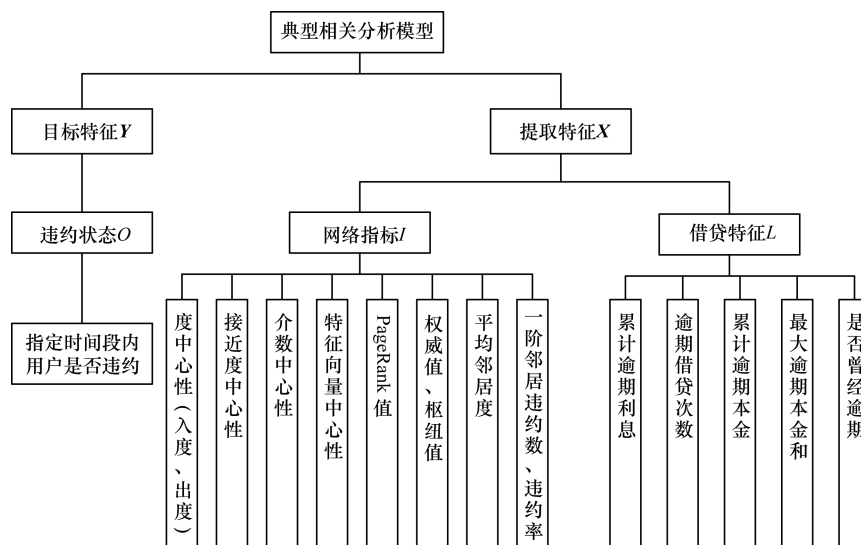


图2 典型相关分析模型中的2类变量分析

X 、 Y 线性组合为 $a^T X$ 、 $b^T Y$ 。目标为找到向量 a ，使得指数 $a^T X$ 、 $b^T Y$ 间的关系能被量化且易于解释，即找到最合适的映射向量 a 、 b 使 2 个指数间相关关系最大化(因为此模型中 Y 为一维变量，所以只考虑 X 的线性组合或考虑 b 为常映射)：

$$\rho(a) = \rho_{a^T X b^T Y}, \max_a a^T \Sigma_{XY} b \quad (12)$$

定义 $K = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$, $N_1 = K K^T$, $N_2 = K^T K$, γ_i 和 δ_i 是 N_1 和 N_2 的标准特征向量。则有：

$$a_i = \Sigma_{XX}^{-1/2} \gamma_i, b_i = \Sigma_{YY}^{-1/2} \delta_i \quad (13)$$

a_i 、 b_i 为典型相关向量，也是以上最大化问题的解。因此，得出风险量化模型为：

$$\text{Predict} = a^T (I, L) \quad (14)$$

2.4 效果评价

对于一个二分类问题而言，ROC 若分为正类和负类，则会出现 4 种情况：实为正预测为正 (True Positive, TP)，实为正预测为负 (False Negative, FN)，实为负预测为正 (False Positive, FP)，实为负预测为负 (True Negative, TN)。真正类率 TPR 代表判别出的正类占总正类的比例，假正类率 FPR 代表判别出的负类占有所有负类的比例。TPR 和 FPR 计算公式如下：

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (15)$$

根据模型给出的评分设定阈值并作为二分的依据。每一个阈值可以算出一组对应的 (FPR, TPR) ，将这 2 个值的组合作为坐标绘入图像。阈值最大时对应 $(0, 0)$ ，全部实例被划为负类，随着阈值减小，划分为正实例的越来越多，错分现象也会增多，即 (FPR, TPR) 增大，直到阈值达到最小时得到坐标 $(1, 1)$ ，理想目标点为 $(0, 1)$ 。ROC 曲线示意图如图 3 所示，随机猜测模型结果对应图中的虚线。

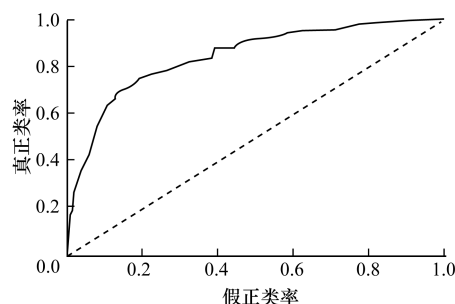


图3 ROC曲线示意图

在图 3 中，ROC 曲线下的面积即为 AUC 值，AUC 值是一个概率值，当随机挑选一个正样本与一个负样本时，现有分类算法根据计算得到的 Score 值将该正样本排在负样本前面的概率值即 AUC 值。AUC 值越大，分类算法越有可能将正样本排在负样本前，即能够更好地分类。

对于本文模型， O 值由选定目标时刻到后 1 个月内交易记录决定， I 、 L 值由选定时刻之前时间区间内交易记录决定，且每一组 I 、 L 、 O 可以计算一个 ROC 曲线及 AUC 值。AUC 值越大，说明通过模型建立的分器效果越好。

用户违约情况与用户网络特征情况如表 1、表 2 所示。

表1 用户违约特征基本属性以及其与违约情况的相关系数

违约特征	平均值	标准差	中位数	违约相关系数
累计违约利息	66.144 0	533.906 0	0	0.237 2
逾期借贷次数	3.642 7	23.599 1	0	0.252 6
累计逾期本金	45 954.080 0	24 844.100 0	0	0.295 0
最大逾期本金	8 320.140 0	27 974.100 0	0	0.261 6
历史逾期状态	0.213 8	0.410 0	0	0.391 2

表 2 用户网络特征基本属性以及其与违约情况的相关系数

网络特征	平均值	标准差	中位数	违约相关系数
度中心性	0.004 953	0.007 989	0.001 594 0	0.043 5
入度中心性	0.002 477	0.004 530	0.000 638 0	0.103 3
出度中心性	0.002 477	0.004 234	0.000 638 0	-0.028 3
特征向量中心性	0.003 815	0.017 442	0.000 005 0	0.038 5
PageRank 值	0.000 319	0.000 430	0.000 154 0	0.120 1
平均邻居数	12.642 800	14.290 700	8.000 000 0	-0.006 1
HITs 权威值	0.000 319	0.001 059	0.000 006 0	0.046 0
HITs 枢纽值	0.000 319	0.001 180	0.000 002 0	0.017 8
介数中心性	0.000 418	0.001 896	0.000 000 1	-0.001 0
接近度中心性	0.092 268	0.067 890	0.123 740 0	-0.057 6
一阶邻居违约数	3.593 300	5.800 100	1.000 000 0	0.034 3
一阶邻居违约率	0.247 200	0.271 100	0.200 000 0	-0.055 0

从表 1、表 2 可以看出,从单个特征自身与违约情况的相关系数来看,在历史违约中提取的相关特征具有明显的相关性。而网络特征中只有入度中心性与 PageRank 值有一定的相关性,其他特征基本没有明显的相关性,且因为违约用户在全部用户中占比较小,所以所有个人违约相关特征中位数为 0。

2.5 模型最佳参数选取

基于模型效果来分析最佳构成网络所使用的记录区间。网络构建最佳数据区间如图 4 所示,在使用 1 个月~7 个月数据时 AUC 值随时间选取增长而提高,在 7 个月之后几乎趋于稳定并在 12 个月达到峰值。数据区间为 1 个月时明显效果最差,此时使用的数据量较少,很难反映出用户的实际信用及行为特点。随着选取数据区间的扩大,模型效果越来越好,但数据区间扩大到一定程度后趋于稳定且使用全量数据模型效果反而会降低。原因是实际网络中存在动态变化,一味增大数据量而忽视时间的衰减效应,是不合理的。本文认为 12 个月即一年间的信用及交易记录是用户的最佳信用评价区间,即选择数据区间在 12 个月左右能够达到模型效果最佳值。

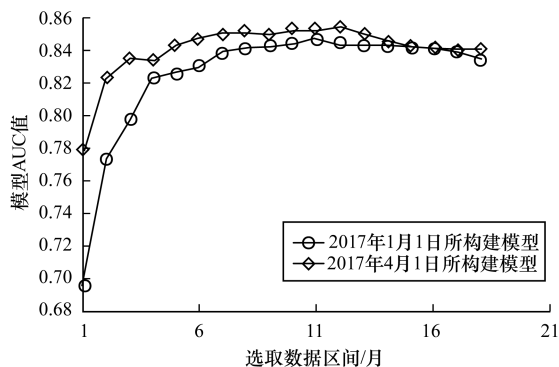


图 4 模型 AUC 值与选取数据区间的关系

将原数据进行归一化后,由 2017 年 4 月 1 日的 12 个月数据模型分析得到的各特征权重情况如表 3 所示。由表 3 可以看出,用户历史的违约情况在风险预测中影响最大,而在逾期相关属性中,最大逾期

本金与逾期次数对未来逾期影响较小,累计属性的影响较大。网络特征在模型中有较大的影响,其中,特征向量中心性、出度中心性、接近度中心性有显著的负向权重。

表 3 典型相关模型特征权重

特征	特征权重
历史逾期状态	0.708 8
HITs 权威值	0.253 5
累计逾期本金	0.239 6
入度中心性	0.179 6
累计违约利息	0.148 0
一阶邻居违约数	0.135 2
PageRank 值	0.133 6
HITs 枢纽值	0.133 4
逾期借贷次数	0.093 4
平均邻居数	0.057 1
特征向量中心性	-0.362 1
出度中心性	-0.298 5
接近度中心性	-0.155 5
一阶邻居违约率	-0.075 1
度中心性	-0.056 3
介数中心性	-0.045 6
最大逾期本金	-0.017 7

3 实验结果与分析

本次实验的数据为某 P2P 网贷公司中 3 345 个用户间的 929 403 条借贷记录数据。选定 2017 年 4 月 1 日前 12 个月的数据作为初始研究对象。实验数据的基本属性情况如表 4 所示。

表 4 实验数据基本属性

数据	总数	违约数
用户数	3 138	671
交易数	725 433	11 431

经模型计算出的风险用户为 755 人,其风险量化分布如图 5 所示。以所有研究特征均为 0 的空白用户风险评价价值作为分割,高于该值的用户定义为风险用户,将风险用户的预测值进行归一化。

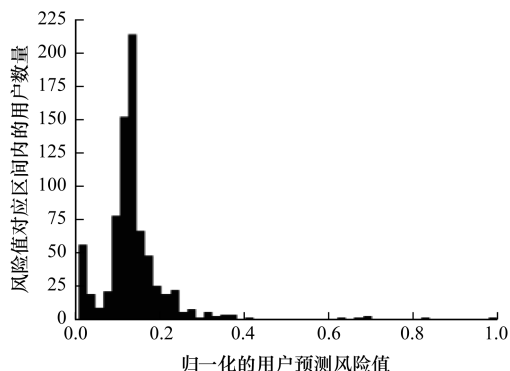


图 5 风险用户风险值柱状分布

2017 年 5 月构建模型的预测权重以及 2017 年 4 月模型权重减去 2017 年 5 月权重所得差值的情况如表 5 所示。由表 5 可以看出,对不同时间区间分析而得到的模型,其权重有一些差异,但模型中指标对结果的影响方向以及指标的权重排序是近似的。

表 5 不同时间模型的特征权重以及差值情况

特征	权重	差值
累计逾期本金	0.113 3	0.126 3
累计违约利息	0.049 9	0.098 1
出度中心性	-0.381 9	0.083 4
PageRank 值	0.063 1	0.070 5
HITs 权威值	0.194 9	0.058 5
度中心性	-0.106 1	0.049 7
历史逾期状态	0.660 9	0.047 8
平均邻居数	0.031 7	0.025 3
入度中心性	0.173 0	0.005 8
特征向量中心性	-0.364 4	0.002 3
一阶邻居违约数	0.311 4	-0.176 1
逾期借贷次数	0.208 1	-0.114 6
最大逾期本金	0.067 7	-0.085 4
接近度中心性	-0.078 0	-0.077 5
介数中心性	0.008 3	-0.054 0
HITs 枢纽值	0.180 0	-0.046 6
一阶邻居违约率	-0.032 2	-0.042 9

前文所提到的 6 个典型特征在不同时间的分布情况如图 6 所示。由图 6 可以看出,不同时间的特征分布差异并不大,即本文中特征的分布对时间

具有较强的鲁棒性。本文模型与决策树模型^[15]的训练结果对比如图 7 所示。其中,横坐标代表该次应用测试距离模型生成的时间。由图 7 可以看出,本文模型表现更稳定,更具有鲁棒性,模型在使用训练数据分析生成后,应用在之后的数据集上仍然有较好的效果,随时间推延模型效果衰减缓慢。决策树模型对训练集能够得到很好的效果,但将该模型应用在之后的未知情况时,预测效果降到 70% 左右,出现了过拟合的现象。鉴于在实际场景应用中,预测风险用户并采取相应措施是本文的主要目的,因此,可以认为本文模型在此场景下优于决策树模型。

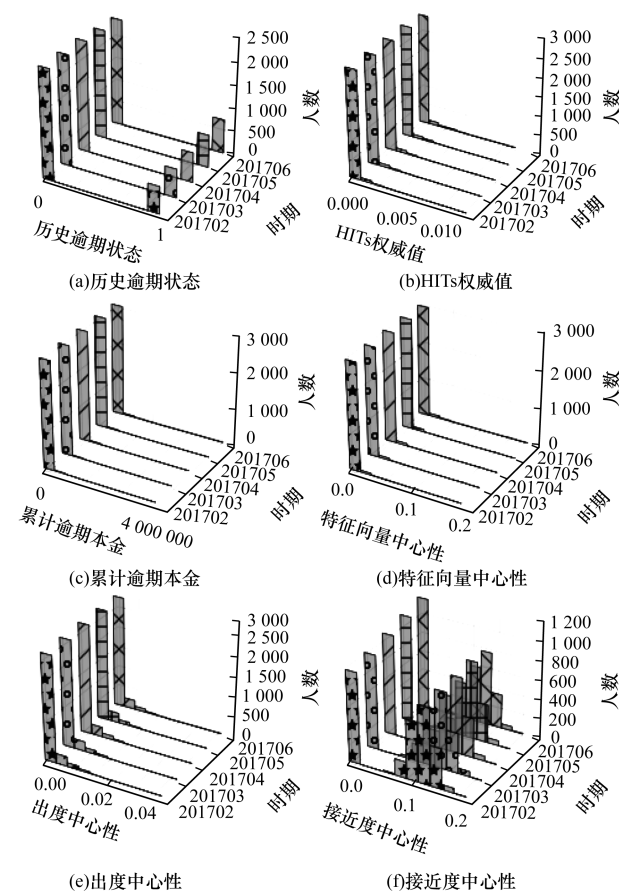


图 6 网络特征分布随时间的变化情况

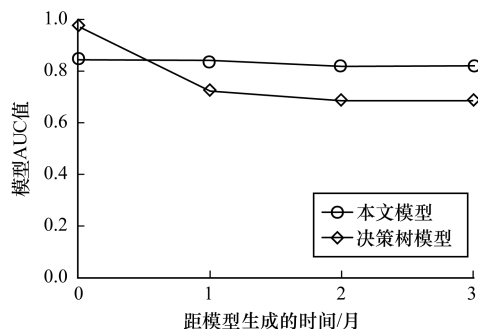


图 7 2 种模型训练结果对比

4 结束语

违约是互联网金融平台中常见的不良行为,不仅会使交易中的一方蒙受损失,对平台中用户的交易倾向也会产生不利影响。本文基于典型相关分析模型及用户复杂网络特征提取方法,对用户建立合理的违约风险量化及预测模型,并对用户的个人特征及网络特征对违约风险的影响进行分析与评价。分析结果表明,用户的一些网络特征对于违约的影响要大于其个人特征。在实际互联网平台运营过程中,对于预测风险较高的用户进行更多的关注与审核,可以有助于对平台违约率的控制。但本文模型的数据集仍存在一定局限性,下一步将对此进行改进并扩大特征的提取与选择范围,以取得更好的预测效果。

参考文献

- [1] 刘新海,丁伟. 大数据征信应用与启示——以美国互联网金融公司 ZestFinance 为例[J]. 清华金融评论, 2014(10):93-98.
- [2] 张权,罗嗣汉. 机器学习在金融大数据风险建模中的应用[EB/OL]. [2017-09-10]. <http://www.36dsj.com/archives/42843>.
- [3] HUANG C L, CHEN M C, WANG C J. Credit scoring with a data mining approach based on support vector machines[J]. Expert Systems with Applications, 2007, 33(4):847-856.
- [4] WANG G, MA J, HUANG L, et al. Two credit scoring models based on dual strategy ensemble trees[J]. Knowledge-based Systems, 2012, 26:61-68.
- [5] 石澄贤,陈雪交. P2P 网贷个人信用评价指标体系的构建[J]. 常州大学学报(社会科学版), 2016, 17(1):80-85.
- [6] 林国强,赵毅鸣,况青作,等. 基于复杂网络和机器学习的 P2P 用户违约预测[J]. 北京师范大学学报(自然科学版), 2017, 53(1):24-27.
- [7] LÜ L, CHEN D, REN X L, et al. Vital nodes identification in complex networks[J]. Physics Reports, 2016, 650:50-63.
- [8] CHEN D B, LÜ L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4):1777-1787.
- [9] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1978, 1(3):215-239.
- [10] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1):35-41.
- [11] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1):113-120.
- [12] BRIN S. The anatomy of large-scale hypertextual Web search engine[J]. Computer Networks, 2012, 56(18):3825-3833.
- [13] KLEINBERG J M. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM, 1999, 46(5):604-632.
- [14] HARDLE W, SIMAR L. 应用多元统计分析[M]. 陈诗一,译. 北京:北京大学出版社, 2011.
- [15] 李航. 统计学习方法[M]. 北京:清华大学出版社, 2012.
- [6] RONG Zhiwen. Software testing method researching of cyclomatic complexity[J]. Software Industry and Engineering, 2012, 6(1):28-34.
- [7] VINJU J J, GODFREY M W. What does control flow really look like? eyeballing the cyclomatic complexity metric[C]//Proceedings of IEEE International Working Conference on Source Code Analysis and Manipulation. Washington D. C., USA: IEEE Press, 2012:154-163.
- [8] JAKHAR A K, RAJNISH K. Measure of complexity for object-oriented programs: a cognitive approach[M]. Berlin, Germany: Springer, 2016.
- [9] ANTINYAN V, STARON M, SANDBERG A. Evaluating code complexity triggers, use of complexity measures and the influence of code complexity on maintenance time[J]. Empirical Software Engineering, 2017, 22(6):3057-3087.
- [10] ANSARI F J. A GUI tool for computation of cyclomatic complexity of a program[C]//Proceedings of National Conference on Architecture, Software Systems and Green Computing. Berlin, Germany: Springer, 2012.
- [11] BHATTI H R. Automatic measurement of source code complexity[D]. Lulea, Sweden: Lulea University of Technology, 2011.
- [12] 陈梦云,高建华. 基于圈复杂度的静态测试用例排序方法[J]. 计算机应用与软件, 2016, 33(1):1-3.
- [13] 陈梦云. 基于圈复杂度和调用次数的测试用例排序方法[D]. 上海:上海师范大学, 2015.
- [14] 谭浩强. C 程序设计[M]. 4 版. 北京:清华大学出版社, 2010.
- [15] 张海藩. 软件工程导论[M]. 北京:清华大学出版社, 2008.
- [16] 牟永敏,杨志嘉. 基于函数调用路径的软件实现与设计一致性验证[J]. 中国科学:信息科学, 2014, 44(10):1290-1304.
- [17] 牟永敏,李慧丽. 基于函数调用路径的测试用例优先级排序[J]. 计算机工程, 2014, 40(7):242-246.
- [18] 周希. 基于静态分析的程序控制流图生成工具的设计与实现[D]. 广州:中山大学, 2013.

编辑 吴云芳

编辑 顾逸斐

(上接第 107 页)