

基于 PU learning 的信用卡交易安全监管研究

陈任峰, 朱鸿斌

(复旦大学计算机科学技术学院, 上海 200433)

摘要: 目前信用卡套现手段复杂多变、虚假交易形态层出不穷, 在仅有账户级套现标签数据的基础上, 信用卡套现管理面临着与客户交互过程难以获取其真实交易情况的业务痛点。为了探究一种精准的信用卡虚假交易监管方法, 以商业银行信用卡系统的套现账户交易标签数据为研究对象, 建立了基于 PU learning (positive-unlabeled learning) 的信用卡单笔交易安全识别模型。所提模型在样本数据标注中引入了间谍 (Spy) 机制, 随机抽取高可靠套现交易正样本 100 万笔及待标注的交易样本 130 万笔, 借助学习器预测结果分布对难以判别的非套现交易负样本进行标注, 以获取相对可靠的负样本标签 120 万笔。基于上述正样本及标注得到的负样本数据, 构建了信用卡客户属性信息、额度使用情况及交易偏好特征等 120 个候选变量, 通过变量重要性筛选得到入模变量近 50 个, 利用 XGBoost 二分类算法进行模型开发预测。结果显示, 所提模型对信用卡套现虚假交易的识别准确率为 94.20%, 群体稳定性指标 (PSI) 为 0.10%, 表明基于 PU learning 的单笔交易安全识别模型能够实现对信用卡虚假交易的有效监测。该研究改进了机器学习二分类算法在难以获取高精度样本标签数据场景下的模型判别性能, 为商业银行信用卡系统交易安全监控提供了新方法。

关键词: 套现交易数据监测; 信用卡系统安全监管; 半监督学习; PU learning

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023039

Research on credit card transaction security supervision based on PU learning

CHEN Renfeng, ZHU Hongbin

School of Computer Science, Fudan University, Shanghai 200433, China

Abstract: The complex and ever-evolving nature of credit card cash out methods and the emergence of various forms of fake transactions present challenges in obtaining accurate transaction information during customer interactions. In order to develop an accurate supervision method for detecting fake credit card transactions, a PU (positive-unlabeled learning) based security identification model for single credit card transactions was established. It was based on long-term transaction label data from cashed-up accounts in commercial banks' credit card systems. A Spy mechanism

收稿日期: 2022-12-25; 修回日期: 2023-03-27

通信作者: 陈任峰, sunmoonsh@126.com

基金项目: 中国工程院战略研究与咨询项目 (2023-33-14)

Foundation Item: Strategic Research and Consulting Project of the Chinese Academy of Engineering (2023-33-14)

引用格式: 陈任峰, 朱鸿斌. 基于 PU learning 的信用卡交易安全监管研究[J]. 网络与信息安全学报, 2023, 9(3): 73-78.

Citation Format: CHEN R F, ZHU H B. Research on credit card transaction security supervision based on PU learning[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 73-78.

was introduced into sample data annotation by selecting million positive samples of highly reliable cash-out transactions and 1.3 million samples of transactions to be labeled, and using a learner to predict the result distribution and label negative samples of non-cash-out transactions that were difficult to identify, resulting in 1.2 million relatively reliable negative sample labels. Based on these samples, 120 candidate variables were constructed, including credit card customer attributes, quota usage, and transaction preference characteristics. After importance screening of variables, nearly 50 candidate variables were selected. The XGBoost binary classification algorithm was used for model development and prediction. The results show that the proposed model achieve an identification accuracy of 94.20%, with a group stability index (PSI) of 0.10%, indicating that the single credit card transaction security identification model based on PU learning can effectively monitor fake transactions. This study improves the model discrimination performance of machine learning binary classification algorithm in scenarios where high-precision sample label data is difficult to obtain, providing a new method for transaction security monitoring in commercial bank credit card systems.

Keywords: cash transaction data monitoring, credit card system security supervision, semi-supervised learning, PU learning

0 引言

随着电商平台、微商二维码、智能POS机自动跳码等新型套现手段^[1-2]愈演愈烈,信用卡虚假交易行为日渐趋于隐蔽化、多样化、差异化^[3],通过人工对比信用卡账户近期交易及其历史行为特征的传统客户级侦测规则体系已难以应对监管部门对商业银行信用卡单笔资金交易监控的核查要求^[4-5]。如何精准挖掘新型套现风险特征、优化虚假交易监控模型体系,加强信用卡单笔交易追踪监控是满足银行业监管审计标准、有效防范信用风险及共债风险暴露、提升客户正常用卡体验的关键。

1 相关理论

1.1 信用卡套现概述

信用卡套现^[6]是指持卡人 not 通过 ATM 或柜台等正规渠道提取现金,而通过其他手段将卡中信用额度内的资金以现金的方式套取,同时不支付银行提现费用的行为。常见的套现特征^[7-8]有以下几个维度。

(1) 商户集中性交易

连续多个账单周期内信用卡交易均发生在少数固定商户。

(2) 高额度使用交易

连续多个账单周期内信用卡交易金额均接近卡片信用额度上限,且交易笔数较少。

(3) 时间集中性交易

连续多个账单周期内信用卡交易金额短期内

出现交易金额均接近卡片信用额度上限。

(4) 大额异常交易

持卡人财务状况恶化,急需大额周转资金或其他情况。

(5) 还款后即刷交易

持卡人连续多次在临近账单周期免息期末进行信用卡还款,并在还款后短期内发生接近卡片额度上限的大额交易。

(6) 小额多笔跳码交易

持卡人短时间内存在小额多笔循环套现还款、异地刷卡且额度使用率较高等交易。

为了尽早防控信用卡套现可能诱发的共债风险,商业银行在监管政策的指导下持续优化套现侦测方法并不断挖掘套现特征,通过对信用卡异常交易客户的引导管控及交互反馈,积累了少量高可信的虚假交易客户标签,但对如何判定非虚假交易客户并精准定位单笔异常交易仍有较大难度。由于套现目标定义的维度缺失,逻辑回归、XGBoost 等信贷风控领域常用二分类算法在其预测性能上失去优势甚至表现不佳。近年来,半监督学习在机器学习领域的应用愈发成熟,其利用少量标注样本和大量未标注样本进行分类识别的重要优势,可有效弥补难以量化定义的非套现数据标签,成为信用卡套现交易精准侦测的新型路径之一。

1.2 PU learning 算法简介

半监督学习^[9]介于传统监督学习和无监督学习之间,其思想是在标记样本数量很少的情况下,

通过在模型训练中引入无标记样本来避免传统监督学习在训练样本不足、学习不充分时出现性能退化的问题。PU learning^[10-12] (positive-unlabeled learning) 是半监督学习的一个研究方向, 指在只有正样本集 P 和无标记样本集 U 的情况下, 引入间谍 Spy 机制^[13-14], 通过两阶段训练标注得到可靠的正样本集 P 和负样本 RN, 从而达到优化二分类器的目的。Spy 机制的基本思想^[15]是从正样本集 P 中划分出一个子集 S , 将 S 中的样本放到无标记样本集 U 中, 从而得到新的正样本集 $P-S$ 和未标识样本集 $U+S$ 。使用 $P-S$ 作为正样本, $U+S$ 作为负样本, 训练二分类算法, 通过 Spy 样本的预测结果分布确定参数阈值 θ , 再对无标记样本集 U 进行划分得到可靠的反样本集合 RN。其中, 从 P 中划分子集 S 的数量比例一般为 15%^[16-17]。Spy 算法步骤^[18]如下。

1) RN 集合置空。

2) 从 P 中随机选取子集 S , 得到新的正样本集 $PS=P-S$ 和未标识样本集 $US=U+S$, 记 PS 中各个样本类别为 1, US 各个样本类别为-1。

3) PS 和 US 作为训练集, 用二分类算法训练得到一个分类器。

4) 使用子集 S 确定出一个概率阈值 θ 。

5) 对 US 中的每个样本 d 使用二分类算法计算其属于正类别的概率 pr , 如果 pr 小于阈值 θ , 则将其加入 RN 集合。

本文以套现账户交易数据为研究对象, 建立基于 PU learning 的信用卡单笔交易安全识别模型, 在样本数据标注中引入 Spy 机制, 借助机器学习二分类算法预测结果分布对难以人工判别的负样本进行二次标注。利用标注得到的正负样本再次进行二分类训练, 强化模型训练标签的可靠性, 为商业银行信用卡系统交易安全监控提供一定的参考价值和数据支持。

2 基于 PU learning 的信用卡单笔交易安全识别模型

2.1 模型目标定义

为了精准识别信用卡单笔虚假交易, 以信用卡客户一段时间内的交易明细作为研究对象, 围绕每笔交易是否为套现交易的核心目标, 将模型

定义如下。

1) 正样本: 观测期内信用卡套现交易;

2) 负样本: 观测期内信用卡非套现交易。

其中, 参考套现监管要求及行业监控规则, 模型观测期选取的观察点为近 6 个月账单。

2.2 样本筛选及标注

在已有套现账户交易标签 (正样本) 的基础上, 利用 PU learning 算法对未知标签交易样本进行标注, 挖掘出相对可靠的非套现交易标签 (负样本)。

2.2.1 正样本筛选

针对已有套现交易样本标签, 为保证模型正样本标签的高可靠性, 结合套现特征进行进一步筛选, 具体筛选条件如表 1 所示。

表 1 正样本筛选条件
Table 1 Positive sample screening conditions

序号	套现交易特征	正样本筛选条件
1	单一商户交易	观测月内所有交易
2	商户集中性交易	观测月内所有交易
3	高额度使用交易	观测月内所有交易
4	时间集中性交易	观测月内所有交易
5	大额异常交易	观测月内所有交易
6	还款后即刷交易	最大一笔还款 5 天内所有交易
7	小额多笔跳码交易	跳码当天所有交易

对较为显著的套现交易初步筛选后, 选择部分有代表性的交易明细作为正样本集 P 。

2.2.2 负样本标注

由于套现交易认定缺乏有效的外界反馈机制, 非套现交易的识别难度相对较高。对于难以辨别的非套现交易样本, 参照有 Spy 机制的 PU learning 算法流程, 对剔除正样本后的交易集合进行标注, 具体步骤如下。

1) 针对筛选后的正样本集 P 和其余待标注样本集构建二分类模型, 基于模型预测概率初步筛选出与正样本特征最不接近的数据集 U 。

2) 参照 1.2 节中的 Spy 算法步骤, 从正样本集 P 中随机选取 15% 的数据作为 Spy 样本集 S 。将 $P-S$ 样本标注为套现交易, $U+S$ 样本标注为非套现交易, 再次采用二分类模型训练, 考察集合 S 的概率分布, 将样本集 U 中所有概率值小于 S 中 5% 分位值的样本标注为非套现交易, 得到相对可靠的负样本集 N 。

基于此,通过二分类算法与 Spy 机制不断寻找相对显著的非套现交易,并将其纳入被标注的负样本集,最终获取与正样本差异性显著且特征类型损失度较低的可靠负样本。

2.3 特征工程

得到可靠的正负样本标签后,为了精准挖掘信用卡单笔交易的近期特征,一般采用短时特征显著的时间窗口(近 1 天、2 天、3 天、7 天、15 天)与常用统计指标(最大值、最小值、平均值、标准差)进行交叉衍生,基于业务理解构建出解释性较强的模型候选变量。

针对候选变量列表,选用 Random Forest 进行特征重要性筛选,并将重要性值较高的变量纳入模型。Random Forest 特征重要性计算方法如下。

1) 对每棵决策树,选择相应的袋外(OOB, out of bag)数据计算袋外数据误差,记为 err_{OOB1} 。Random Forest 是无偏估计,无须进行交叉验证或者单独测试集来获取测试集误差的无偏估计。

2) 随机对 OOB 数据的所有样本的特征 X 加入噪声干扰,再次计算袋外数据误差,记为 err_{OOB2} 。

3) 假设森林中有 M 棵树,则特征 X 的重要性为 $\sum(err_{OOB2}-err_{OOB1})/M$ 。加入随机噪声后,袋外数据的准确率大幅度下降,说明这个特征对于样本的预测结果有很大影响,进而说明其重要程度较高。

2.4 模型构建

模型构建是指利用标注筛选得到正样本集 P 和负样本集 N ,利用机器学习应用相对成熟的 XGBoost 二分类算法进行单笔套现交易模型的开发验证,并进一步迭代优化得到适用于信用卡交易系统监测实际应用的模型方案。

3 实验与分析

3.1 样本选择

以某合作机构信用卡部分交易数据为研究样本,选择时效性较强的 2022 年 3 月数据为实验样本,随机抽取正样本集 P 为 100 万笔,同期待标注样本集 U 为 130 万笔,通过层层筛选最终得到可靠负样本集 N 为 120 万笔,正负样本比为 1:1.2。

3.2 变量筛选

参照特征衍生逻辑,构建了信用卡客户属性信息、额度使用情况及交易偏好特征等 120 个候选变量,具体说明如表 2 所示。

表 2 候选变量说明
Table 2 Description of candidate variables

特征维度	变量描述	变量个数
属性信息	性别/年龄/学历/婚姻状态/...	13
额度使用	授信额度/额度使用率/...	2
交易偏好	交易金额/笔数/商户/时间/城市/...	105
总计		120

依据 Random Forest 特征重要性筛选,按照候选变量的重要性进行排序,最终选择重要性指标大于 0.005 的 46 个变量入模,其中排名前 10 的入模变量如表 3 所示。

表 3 入模变量说明
Table 3 Description of entry variables

序号	变量名称	变量重要性
1	近 15 笔交易总额	0.078 2
2	近 7 笔平均交易金额	0.063 1
3	近 7 天最大单日交易金额	0.052 2
4	近 10 天平均交易金额	0.049 8
5	授信额度	0.046 6
6	近 3 天平均交易金额	0.043 2
7	近 2 天交易金额	0.039 9
8	近 7 天平均交易金额	0.038 1
9	近 5 笔交易总额	0.035 4
10	近 3 天最大单日交易金额	0.031 2

3.3 结果验证

在通过 PU learning 算法标注得到正负样本及对应 46 个入模变量确定的基础上,本文进一步构建了基于 XGBoost 二分类算法的单笔交易识别模型。考虑到套现交易定义难以量化统一,不适用于传统机器学习二分类模型的区分度评价指标(如 KS、AUC 等),该模型准确率性能验证采用套现侦测专家多人复核的方式,对模型判别的信用卡客户交易明细样本进行抽样验证。根据多轮抽样验证数据表现,将预测概率 0.9 以上的交易样本判定为套现虚假交易。本次抽样信用卡账户为 100 个,正负交易样本数共计 1 000 笔,总体抽样准确率为 94.20%,具体验证结果如表 4 所示。

表4 模型性能验证——抽样准确率分布
Table 4 Model performance validation—sampling accuracy distribution

套现标签	抽样 账户数	抽样 交易数	预测 准确数	预测 准确率
套现交易	50	518	499	96.33%
非套现交易	50	482	443	91.91%
总计	100	1 000	942	94.20%

为了进一步考量相比不利用该模型的整体提升度,将上述 1 000 笔抽样信用卡账户交易输入该合作机构原套现侦测规则体系,经过 4 名业务专家独立进行交易复核,得到最终规则体系及人工复核的套现交易识别结果。将上述 4 名业务专家人工复核准确率取均值后与该模型准确率进行对比分析。

结果显示,不利用该模型的总体准确率为 72.50%,利用模型后准确率提升 29.93%,表明基于半监督学习 PU learning 的信用卡套现交易识别模型具有良好的判别能力,并且性能提升效果较为显著。

3.4 应用分析

为了考察模型后期应用的泛化能力,借助群体稳定性指标 (PSI, population stability index) 对开发训练时点 (T 月) 和跨时区时点 ($T+6$ 月) 的模型分布变化情况进行对比验证。PSI 计算步骤如下。

1) 根据下述公式计算各项概率分布区间的 Index, 即 $\text{Index} = (\text{训练时点的概率区间样本分布占比} - \text{跨时区时点概率区间样本分布占比}) \times \ln(\text{训练时点概率区间样本分布占比} / \text{跨时区时点的概率区间样本分布占比})$ 。

2) 将各项概率分布区间的 Index 加总求和得 PSI 值。PSI 值越小,代表两个分布之间的差异越小,模型越稳定。一般而言,当 PSI 小于 10% 时,表明模型没有变化或变化很小,整体稳定性优越。

考虑到跨时区时点交易量庞大,随机抽取其中 5% 的交易样本进行模型预测,并对训练时点和跨时区时点模型预测概率区间内的样本分布情况进行统计分析,具体如表 5 所示。结果显示,模型 PSI 为 0.10%,远小于代表模型稳定性良好的指标上限 10%,表明模型整体稳定性较好,可用于后续信用卡交易系统监控。

表5 模型应用验证——群体稳定性分布
Table 5 Model application validation—population stability index distribution

预测概率 区间	观察点 账户占比	跨时区 账户占比	概率区间 分布差值	概率区间 分布比值	Index
[0.9, 1.0]	14.34%	15.40%	-1.06%	0.93	0.08%
[0.8, 0.9]	1.79%	1.85%	-0.06%	0.97	0.00%
[0.7, 0.8]	1.29%	1.35%	-0.06%	0.96	0.00%
[0.6, 0.7]	1.19%	1.22%	-0.03%	0.98	0.00%
[0.5, 0.6]	1.15%	1.16%	-0.01%	0.99	0.00%
[0.4, 0.5]	1.23%	1.24%	-0.01%	0.99	0.00%
[0.3, 0.4]	1.49%	1.49%	0.00%	1.00	0.00%
[0.2, 0.3]	2.09%	2.08%	0.01%	1.00	0.00%
[0.1, 0.2]	3.86%	3.80%	0.06%	1.02	0.00%
[0.0, 0.1]	71.57%	70.41%	1.16%	1.02	0.02%
PSI					0.10%

4 结束语

本文为商业银行信用卡系统交易安全监控提出了基于半监督学习 PU learning 的模型方案。基于当前信用卡套现交易的主要特征,利用 Spy 机制分阶段从难以辨别的未标注样本中挖掘相对可靠的负样本标签,进一步优化了单笔虚假套现交易模型的分类预测性能。经验证,模型各方面指标表现稳定,交易识别准确率为 94.20%,PSI 为 0.10%。

综上,本文的研究改进了机器学习二分类算法在难以获取高精度样本标签数据场景下的模型判别性能,能够实现对信用卡虚假交易的有效监测,具有良好的应用前景。

参考文献:

- [1] 李杉杉. 新场景下商业银行信用卡套现审计思路与方法[J]. 审计观察, 2020(11): 38-43.
LI S S. Audit ideas and methods of commercial banks' credit card cash out under the new scenario[J]. Audit Observation, 2020(11): 38-43.
- [2] 林素云. 新型信用卡套现问题及对策研究[J]. 金融科技时代, 2016, 24(1): 53-56.
LIN S Y. Research on cash-out problems and countermeasures of new credit cards [J]. Financial Technology Era, 2016, 24(1): 53-56.
- [3] 吕晓明. 论我国信用卡套现的法律规制[D]. 北京: 中国政法大学, 2012.
LYU X M. On the legal Regulation of credit card cash out in China[D]. Beijing: China University of Political Science and Law, 2012.
- [4] 王永超, 杨朔. 信用卡消费资金流向监控的难点与可行路径[J].

- 中国信用卡, 2018, (8): 37-40.
- WANG Y C, YANG S. Difficulties and feasible ways to monitor the flow of credit card consumption funds[J]. China Credit Card, 2018, (8): 37-40.
- [5] 李丹丹. 个人信息保护与信用卡交易安全[J]. 经济研究导刊, 2014, (20): 297-298.
- LI D D. Personal information protection and credit card transaction security [J]. Economic Research Guide, 2014, (20): 297-298.
- [6] 郭从秀, 潘昱. 信用卡套现的识别和风险控制探析[J]. 中国信用卡, 2010, (20): 49-53.
- GUO C X, PAN Y. Identification and risk control of credit card cash[J]. China Credit Card, 2010, (20): 49-53.
- [7] 周文君. 信用卡套现行为模式及其法律规制[J]. 金融法苑, 2009, (2): 98-108.
- ZHOU W J. Credit Card cash-out behavior pattern and its legal regulation[J]. Journal of Financial Law, 2009, (2): 98-108.
- [8] 李峰. 基于第三方支付平台的信用卡套现流程、特征及监管建议[J]. 上海金融, 2014, (4): 108-109.
- LI F. Credit card cashing process, characteristics and regulatory suggestions based on third-party payment platform[J]. Shanghai Finance, 2014, (4): 108-109.
- [9] SHI Z, WANG N, KONG F, et al. A semi-supervised learning method of latent features based on convolutional neural networks for CT metal artifact reduction[J]. Medical Physics, 2022, (6): 48-50.
- [10] HSIEH C J, NATARAJAN N, DHILLON I S. PU learning for matrix completion[J]. Computer Science, 2014: 2445-2453.
- [11] REAVIS E A, FRANK S M, GREENLEE M W, et al. Neural correlates of context-dependent feature conjunction learning in visual search tasks[J]. Human Brain Mapping, 2016, 37(6): 2319-2330.
- [12] YANG P, LI X L, MEI J P, et al. Positive-unlabeled learning for disease gene identification[J]. Bioinformatics, 2012, 28(20): 2640.
- [13] LI T T, LV J, FAN W Y. Semi-supervised self-training positive and unlabeled learning based on new spy technology[J]. Journal of Computer Applications, 2019, 39(10): 2822-2828.
- [14] GAO B T, ZHAI Z G, LIU B. Research on biomedical named entity recognition algorithm in PU scene[J]. Technology of IoT & AI, 2019, 51(1): 22-28.
- [15] ZHENG H, YU H, HAO Y, et al. Distantly supervised named entity recognition with Spy-PU algorithm[C]//Proceedings of 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). 2021: 56-63.
- [16] KOLOSOV N, DALY M J, ARTOMOV M. Prioritization of disease genes from GWAS using ensemble-based positive-unlabeled learning[J]. European Journal of Human Genetics, 2021, 29(10): 1527-1535.
- [17] LI H, LIU B, MUKHERJE A, et al. Spotting fake reviews using positive-unlabeled learning[J]. Computacion Y Sistemas, 2015, 18(3): 467-475.
- [18] SHAO Y H, CHEN W J, LIU L M, et al. Laplacian unit-hyperplane learning from positive and unlabeled examples[J]. Information Sciences, 2015, 314(6): 152-168.

[作者简介]



陈任峰 (1974—), 男, 浙江诸暨人, 复旦大学博士生, 主要研究方向为隐私计算在金融行业的应用。



朱鸿斌 (1991—), 男, 浙江绍兴人, 博士, 复旦大学副研究员, 主要研究方向为金融科技与安全, 边缘人工智能和信号处理。