图神经网络在 12306 黑产用户挖掘的研究

郝晓培,朱建生,单杏花
(中国铁道科学研究院,北京 100081)

摘　要:随着铁路信息化技术的高速发展以及铁路互联网售票系统的不断优化完善,12306 已成为铁路客运主要的售票渠道,为旅客的出行带来了极大的便利,然而节假日部分线路供需仍存在巨大缺口,铁路客票销售市场存在巨大的牟利空间,从而也面临着网络黑色产业链的威胁。针对当前铁路 12306 互联网售票系统存在黑产用户抢票,倒票,囤票等问题,提出了兼顾旅客社会关系以及个体特征的黑产用户识别模型。首先基于旅客的历史购票及出行行为,从时间、空间等维度构建旅客个体特征,然后基于旅客的出行关系以及购票关系构建旅客社交网络,通过频率反映旅客社交关系强度,最后,采用图神经网络将节点个体特征以及邻居节点的特征信息线性表示为低维稠密的向量空间,将其最终旅客特征向量输入无核二次曲面支持向量机进行黑产用户识别。实验表明,综合考虑旅客社交关系以及旅客个体特征的黑产用户识别模型相对于只考虑个体特征的模型准确率有了显著的提高。

关键词:黑产用户;个体特征;社交网络;图神经网络;邻居节点;支持向量机
中图分类号:TP301.6　　**文献标识码**:A　　**文章编号**:1673-629X(2022)07-0185-06
doi:10.3969/j.issn.1673-629X.2022.07.032

Research on Graph Neural Network in 12306 Black Production User Mining

HAO Xiao-pei,ZHU Jian-sheng,SHAN Xing-hua
(China Academy of Railway Sciences,Beijing 100081,China)

Abstract:With the rapid development of railway information technology and the continuous optimization and improvement of the railway Internet ticketing system,12306 has become the main ticketing channel for railway passenger transportation,bringing great convenience to passengers. However,there is still a huge gap in supply and demand on some routes during holidays. There is a huge profit-making space in the railway ticket sales market,which also faces the threat of a black network industry chain. Regarding the current railway 12306 Internet ticketing system,there are problems such as black-produced users grabbing tickets,scalping tickets,and hoarding tickets. We propose a black product user identification model that takes into account the social relationship of passengers and individual characteristics. Firstly,based on the historical ticket purchase and travel behavior of passengers,the individual characteristics of passengers are constructed from the dimensions of time and space,and then the passenger social network is constructed based on the travel relationship and ticket purchase relationship of the passengers,and the intensity of the social relationship of the passengers is reflected by frequency. Finally,graph nerves are used. The network linearly expresses the individual characteristics of nodes and the characteristic information of neighboring nodes as a low-dimensional dense vector space,and inputs the final passenger feature vector into the coreless quadric support vector machine for black user identification. Experiments show that the black product user identification model that comprehensively considers the social relationship of passengers and the individual characteristics of passengers has a significant improvement in accuracy compared with the model that only considers individual characteristics.

Key words:black production users;individual characteristics;social networks;graph neural network;neighbor node;support vector machine

0　引　言

高速铁路以及铁路信息化技术的快速发展,为旅客的出行、购票等提供了极大的便利。目前铁路 12306 互联网售票系统售票量占全渠道的 80% 以上,日均售票量超过千万,已成为全球最大的票务系统。然而节假日运力资源紧张,旅客“一票难求”的情况依

收稿日期:2021-08-03　　**修回日期**:2021-12-07
基金项目:国家重点研发计划(2020YFF0304101)
作者简介:郝晓培(1990-),男,博士研究生,研究方向为大数据、机器学习;朱建生,研究员,研究方向为计算机应用技术;单杏花,研究员,研究方向为铁路计算机应用。

然存在,从而衍生出一批线上黄牛利用互联网法律法规不健全的漏洞以及高效的云资源,囤积大量黑产账户进行抢票、囤票、倒票等不正当的方式谋取利益。为了维护客运购票场景的公平公正以及系统的安全稳定,客运团队研发并上线了风控系统,采用实时行为分析以及基于行为特征进行有监督的用户分类的方式,识别出了大量的异常购票行为,进行不同策略的卡控,取得了明显的效果。不仅仅在铁路客运,其他互联网行业也存在黑产用户非法活动,故国内外针对黑产用户识别进行了大量的研究。在用户个体特征领域,周亮谨等人[1]基于购票流程中主要购票行为为用户特征,基于朴素贝叶斯分类算法构建行为分类器,实现异常用户识别,运行效率和准确率均满足需要。Moh 等人[2]构建 Twitter 用户的相关特征,构建特征矩阵计算用户信任度,从而识别异常用户。在社交网络领域,宋艳红[3]利用 G-N 社区发现算法以及粗糙集理论计算用户特征权重以及特征信用值,并构建用户可行度,从而识别异常用户。Din I U 等人[4]基于社交网络自身的拓扑结构,对类似社交网络中的异常邮件进行识别。仲丽君等人[5]介绍了当前基于社交网络进行异常用户识别的方式,包括分类、聚类、统计、信息论、混合、图等六大类,并对各类方法进行了对比。

基于个体特征的异常用户识别,忽略了个体间的社会关系,无法完整描述用户特征,社区发现也只关注局部的拓扑结构,没有很好利用每个节点的语义信息,均存在一定的局限性。为了进一步完善用户特征表示,该文在研究铁路客运售票业务的基础上,构造了旅客个体特征以及社交关系网络,利用图神经网络充分融合个体特征及社交关系特征,构建特征向量,以提高识别异常用户的准确率。

1 黑产用户识别模型总体方案

为了提高黑产用户识别准确率,该文提出了一种基于旅客个体特征及旅客社交关系的黑产用户识别模型,其总体流程如图 1 所示。

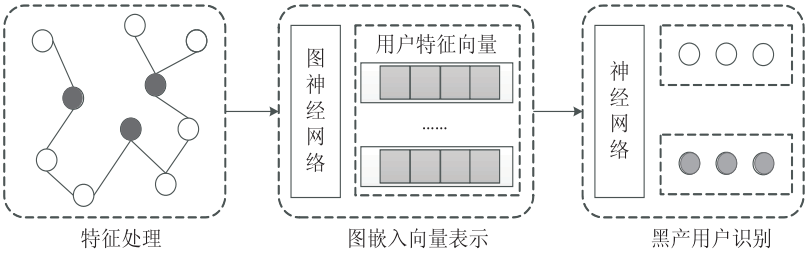

图 1 模型总体方案流程

主要包括特征构建、图嵌入向量表示以及黑产用户识别等。

(1)特征构建。

图数据中同时包含两部分信息:实体特征属性和实体之间的关系[6]。其中属性信息描述图中个体节点的固有特征,对应铁路旅客的出行及购票特征;结构信息描述了个体之间的关联性质,对应铁路旅客之间的购票关系以及同行关系[7]。

综合考虑铁路购票系统以及其他出行服务数据的复杂性、多样性以及安全性,实现了对不同系统,不同类型数据进行获取、转码、清洗、入库、关联等处理,从时间、空间等两个维度构建用户特征体系[8]。

铁路旅客之间的购票关系以及同行关系代表了铁路旅客的社交关系,通过分析具有铁路同出行社交关系的旅客具有相似度较高的出行特征,比如:常驻地、出行目的等,同时旅客的特征除了跟自身的购票及出行特征有关之外,还与其有同行关系或购票关系相关的旅客特征有关系,这里基于铁路购票及出行社交关系构建旅客社交关系网络,如图 2 所示。

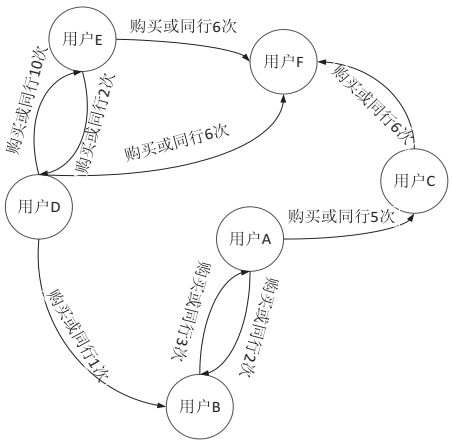

图 2 关系网络构建

(2)图嵌入向量表示。

基于旅客个体特征以及关系网络生成的数据组成了非欧氏数据集,无法采用传统的方式进行处理,而图神经网络是针对典型的不具备规则空间结构的非欧氏数据类型进行深度学习而发展起来的,可以挖掘出节点之间的关系特征[9],故这里采用图神经网络对旅客特征进行处理。其主要采用一定的方法对旅客社交网

络中的旅客节点进行向量表示,随着节点自身特征、邻居节点的特征的更新以及图形拓扑结构的变化得到最终的表示向量,该特征向量保留原始图的结构和节点属性信息。

(3)黑产用户识别。

该文主要采用无核二次曲面支持向量机作为黑产用户识别模型,该模型可在原空间中使用非线性二次曲面对样本进行直接分类,避免了核函数结构以及参数的选择,提高了模型的效率及可用性,该模型将图嵌入向量成特征向量作为样本的输入,实现黑产用户识别。

2　基于图神经网络的特征向量表示

本节主要基于旅客个体特征及关系网络通过图神经网络模型进行特征向量表示。其核心思想是通过学习一个对邻居顶点进行聚合表示的函数来产生目标顶点的 embedding 向量[10]。

(1)算法流程。

目前铁路 12306 互联网售票系统已有 6 亿注册用户,每年出行人次达到 30 亿,购票及同行关系复杂,其旅客社交网络组成了规模较大的图数据,其子图的节点数存在呈指数级增长的问题,同时也存在部分度非常大的超级节点,导致进行全图训练的时间代价、计算代价以及存储代价不可控。为了解决该问题,该文针对 GraphSAGE[11]模型进行训练,该模型从聚合邻居节点操作出发,对邻居节点进行抽样以控制实际运算节点的规模,同时防止邻居节点随机采样导致局部信息丢失。增加了特征初始化操作,以降低信息丢失率,其主要训练过程如下:

算法:改进 GraphSAGE 算法主要流程。

Input:样本集 B;图 $G=(V,E)$,其中 V 代表用户节点集合,E 代表节点之间的关系;层数 K;权重矩阵 $\boldsymbol{L}^{(k)}$,$\forall k \in \{1,2,\cdots,k\}$;非线性函数 σ;聚合函数 $\mathrm{Agg}^{(k)}$;邻居采样函数 $N^{(k)}:v \rightarrow 2^{v}$,$\forall k \in \{1,2,\cdots,k\}$

Output:输出所有节点的特征向量 $z_v, k \in B$

1. for $i = 1\cdots N$ do
2. for $j = 1\cdots b_{ih}$ do
3. $x_i^{'} = x_i + \dfrac{w_j}{\sum_{t=1}^{b_{ih}} w_t} \times x_j$
4. end
5. end
6. $B^{(k)} \leftarrow B$
7. for $k = K\cdots 1$ do
8. $B^{(k-1)} \leftarrow B$
9. for $u \in B^{(k)}$ do
10. $B^{(k-1)} \leftarrow B^{(k-1)} \cup N^{k}(u)$
11. end
12. end
13. $h_u^{(0)} \leftarrow x_v^{'}, \forall v \in B^{(0)}$
14. for $k = 1\cdots K$ do
15. for $u \in B^{(k)}$ do
16. $h_{N_{(u)}}^{(k)} \leftarrow \mathrm{Agg}^{(k)}(\{h_{u^{'}}^{(k-1)}, \forall u^{'} \in N^{k}(u)\})$
17. $h_u^{(k)} \leftarrow \sigma(W^{k}[h_u^{(k-1)} \| h_{N_{(u)}}^{k}])$
18. $h_u^{(k)} \leftarrow h_u^{(k)} / \| h_u^{(k)} \|_2$
19. end
20. end
21. $z_u \leftarrow h_u^{(k)}, \forall u \in B$

(2)特征值初始化。

如上述算法所示,为了防止对邻居节点采样导致局部信息丢失,算法的 1 ~ 5 行,遍历每一个节点,提前将节点的特征向量与它所有的邻居节点的特征向量按照固定的权重进行线性组合,使得每一个节点初始状态下已经包含周围邻居节点的一些信息,从而在采样初始阶段保留部分局部信息。

(3)邻居节点采样。

算法的 6 ~ 12 行,首先遍历出样本集 B 内参与中心节点聚合操作的所有 k 阶子图,并在这些节点上进行 K 次聚合操作的迭代运算,基本思路是:要获得某个中心节点第 k 层的特征,需要对第 $k-1$ 层的邻居进行采样,接着对 $k-1$ 层的每个节点采样其 $k-2$ 层的邻居节点,不断循环,直到采样完第 1 层的所有邻居位置。

(4)聚合操作。

算法的 13 ~ 20 行主要是对邻居节点进行聚合操作,其中第 15 行通过聚合函数对每个节点的邻居节点特征进行聚合,接着第 16 行对聚合后的邻居特征与中心节点的上一层特征进行拼接,输到单层网络中获得中心节点的特征向量,最后对计算好的特征向量进行归一化处理,以保证所有节点向量在相同的单位尺度上。

(5)参数学习。

该文主要采用无监督学习,即节点与其邻居具有类似的特征标识,没有直接相连的节点特征标识相差较大,损失函数如下:

$$Q(Z_u) = -\log(\sigma(Z_u^{\mathrm{T}} Z_v)) - R \cdot \mathbb{E}_{v_n \sim P_n(v)} \log(\sigma(Z_u^{\mathrm{T}} Z_{v_n})) \quad (1)$$

其中,Z_u 表示节点 u 的图特征向量表示,v 表示 u 节点的邻居节点,σ 指的是 sigmod 函数,表示样本数,$v_n \sim P_n(v)$ 表示负样本数。

3　基于二次曲面 SVM 模型的黑产用户识别

本节以图嵌入向量表示生成的特征向量为输入,

利用无核二次曲面支持向量机[12]以实现黑产用户识别,该算法基本原理如下:

(1)模型输入。

铁路用户样本集: $T = \{(Z_1,y_1),(Z_2,y_2),\cdots,(Z_m,y_m)\}$,其中 Z_i 表示第 i 个样本的特征向量, $Z_i = [z_i^1,z_i^2,\cdots,z_i^n]^{\mathrm{T}} \in R^n$, Z_i^j 表示第 i 个样本的第 j 个特征; y_i 表示第 i 个样本对应的标签值, $y_i = \{-1,1\}$, $i = 1,2,\cdots,m$,即当 y_i 为-1 时,该样本为黑产用户,反之为正常用户。

(2)模型输出。

$$g(z) = \frac{1}{2}Z^{\mathrm{T}}WZ + b^{\mathrm{T}}Z + c \tag{2}$$

其满足分类曲面 $g(z) = 0$ 可以将黑产用户与正常用户分开。

(3)优化函数。

模型的核心思想为构建二次曲面函数以实现样本点到分类曲面的相对集合距离最大,二次分类曲面主要由参数 $(\boldsymbol{W},\boldsymbol{b},c)$ 构成,其中 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{T}} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mm} \end{bmatrix} \in R^{n\times n}$, $\boldsymbol{b} = [b_1,b_2,\cdots,b_n]^{\mathrm{T}} \in R^n$, $c \in R$,因此针对样本点 Z_1 ,其对应的几何距离为: $\frac{1}{\| \boldsymbol{WZ}_i + \boldsymbol{b} \|^2}$,故所有样本点到分类曲面的距离之和为: $\sum_{i=1}^{m} \frac{1}{\| \boldsymbol{WZ}_i + \boldsymbol{b} \|^2}$,样本点被二次曲面正确分类函数为:

$$y_i(\frac{1}{2}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{WZ} + b^{\mathrm{T}}\boldsymbol{Z} + c) \geqslant 1,\ i = 1,2,\cdots,m \tag{3}$$

故二次曲面支持向量机优化问题可以表示为:

$$\min \sum_{i=1}^{m} \| \boldsymbol{WZ}_i + b^{\mathrm{T}}\boldsymbol{Z} \|^2$$

$$\text{s.t } y_i(\frac{1}{2}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{WZ} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{Z} + c) \geqslant 1,\ i = 1,2,\cdots,m \tag{4}$$

其中, $\max \sum_{i=1}^{m} \frac{1}{\| \boldsymbol{WZ}_i + \boldsymbol{b} \|^2}$ 等价为:

$\min \sum_{i=1}^{m} \| \boldsymbol{WZ}_i + b \|^2$ 。

(4)决策函数。

该模型的目标是将所有的正常用户样本点分布在 $g(z) = 1$ 的外侧,黑产用户在 $g(z) = -1$ 的外侧,并最大化所有样本到分类曲面的相对几何距离,其对应的分类所测函数为:

$$f(x) = \mathrm{sign}(g(z)) \tag{5}$$

假设参数最优解为: W^* , b^* , c^* ,则最终分类决策函数为:

$$f(x) = \mathrm{sign}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{W}^*\boldsymbol{Z} + \boldsymbol{b}^{*\mathrm{T}}\boldsymbol{Z} + c^*) \tag{6}$$

4 实验及结果分析

本模型基于目前铁路客运固定时间段的生产数据作为原始数据,构建用户基础特征以及关系网络。面对海量的交易信息,实验主要采用 Spark 进行基础特征处理,Spark GraphX[13]进行关系网络构建,最后基于 Spark 实现 GraphSAGE 以及支持向量模型。

4.1 实验数据及预处理

该文主要解决的是二分类问题,将黑产用户用-1 表示,正常用户用 1 表示。以现有的风控策略以及铁路客运用户画像系统为基础,随机抽取 100 个正常用户以及 100 个黑产用户作为种子样本,以整个旅客关系网络为基础,扩展出 100 万个样本作为测试数据,其中正常用户 897 813,黑产用户 102 187。为降低模型复杂度,主要挑选了 11 个基础特征,如表 1 所示,并对跨度较大的特征进行归一化处理。

表 1 旅客个体特征

特征编号	特征名称	类别	说明
1	是否学生	枚举	0:否,1:是
2	性别	枚举	0:男,1:女
3	年龄	数值	归一化处理
4	普通车购票次数	数值	归一化处理
5	动车组购票次数	数值	归一化处理
6	普通车乘车次数	数值	归一化处理
7	动车组乘车次数	数值	归一化处理
8	代购比例	数值	实际值
9	自购比例	数值	实际值
10	代购人数	数值	归一化处理
11	结伴比例	数值	实际值

首先将 100 万样本数据按照 1∶9 分为两部分,其中 90% 作为训练集,10% 作为测试样本,以验证模型的好坏。同时为了排除不同训练子集带来的统计误差,采用 k 折交叉验证(k-fold-cross-validation)方法[14],即将数据随机分为 k 组,依次将其中一组作为测试集,剩下的 $k-1$ 组作为训练组构建模型,基于本实验样本的数据量,随机将训练样本集分成 20 组。

在图嵌入向量表示模型中,基于铁路旅客购票关系以及同行关系构造的关系网络结构复杂,为了保证模型训练的效率,设 k 为 2,关系网络中每个用户节点可以最多根据其 2 跳邻接点的信息进行聚合学习。常见的聚合函数包括平均、GCN 归纳式、LSTM、pooling 聚合器。其中 LSTM 具有更强的表达能力,故该文主要采用 LSTM 聚合。

4.2　评价指标

分别从准确率、ROC 曲线、误伤率以及滞后性对模型进行对比分析。

准确率:即整个样本集预测结果的准确性,准确率越高越好。

ROC 曲线以及 AUC 值:ROC 曲线反映敏感性和特异性连续变量的综合指标;AUC 值即 ROC 曲线下方面积的大小,值越大代表分类器效果越好[15]。

误伤率:将正常用户识别为黑产用户的比例,比例越小越好。

滞后性:识别异常用户的耗时对比情况,耗时越短越好。

4.3　结果分析

该文主要采用两种对比方案进行实验,实验一:将原始特征向量,图嵌入表示生成的特征的向量分别作为二次曲面支持向量机进行对比;实验二:将该黑产用户识别模型与现有的风控策略模型进行对比。

4.3.1　实验一

该实验的目的是对比旅客个体原始特征与聚合相邻节点特征(下面简称融合特征)对相同模型性能的影响,分别对两类特征采用二次曲面支持向量机模型训练之后,分别对测试样本进行测试,同时对两个模型的准确率、ROC 曲线进行对比分析。

(1)准确率。

从表 2 可以看出,在相同的黑产用户识别模型中,基于融合特征的样本的准确率明显高于原始特征,同时融合特征的方差相对较小,具有更好的鲁棒性,且相对稳定,因此基于图嵌入向量表示生成的旅客特征更能完整地表达旅客的特征,对黑产用户的识别具有积极作用。

表 2　准确率对比　　%

类型	均值	方差	中位数	25% 分位数	75% 分位数
原始特征	84.7	3.38	84.9	81.7	86.6
融合特征	92.4	3.28	92.6	88.6	94.8

(2)ROC 曲线。

ROC 曲线和对应的 AUC 值是衡量分类模型性能以及能力的重要指标,是研究模型泛化能力的主要工具,实验一两个模型对应的 ROC 曲线以及 AUC 值如图 3 所示。

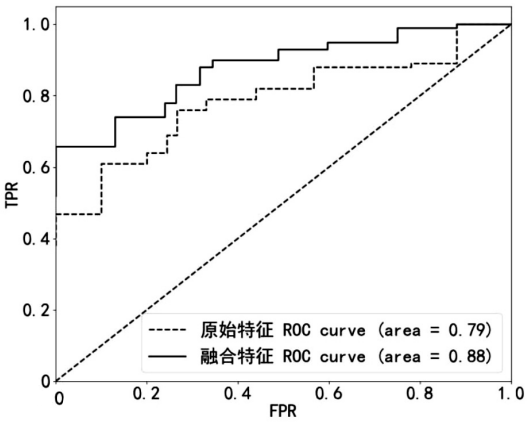

图 3　ROC 曲线与 AUC 值

如图 3 所示,横轴 FPR 表示假正率,纵轴 TPR 表示真正率,其中融合特征模型的 ROC 曲线完全覆盖原始特征的 ROC 曲线,同时融合特征模型的 AUC 值相对较大,因此融合特征模型的性能更优且泛化能力更强。

4.3.2　实验二

该实验的目的是对比现有风控策略模型与黑产用户识别模型,由于目前样本的训练集以及测试集的类别标签均是采用现在的风控策略进行标记,仅仅依靠针对测试集进行测试的准确率以及 ROC 无法体现黑产用户识别模型的性能。因此,测试样本改为随机抽取一周的互联网购票用户作为测试集,依次用两个模型进行异常用户标记,分别从误伤率以及滞后性两个方面进行分析。

(1)误伤率。

误伤率即将正常购票的用户识别为异常用户的比例,通过分析一个月的用户访问日志、投诉、行为分析等,发现黑产用户识别模型的误伤率降低了 10%。

(2)滞后性。

滞后性即识别出异常用户需要的时间。采用两个模型识别近半年的异常用户,并随机抽取 1 000 个异

常账户(编号为 1 到 1 000)进行识别时间分析,如图 4 所示(为方便显示 1 000 个账户随机抽取节点绘制),其中横坐标表示时间,纵坐标表示异常账户的编号。可以看出,文中模型的散点相对当前风控模型比较集中在左侧,经过统计,文中模型在随机抽取的异常账户中,超过 60% 识别时间有所提前,即现有的风控策略需要积累一定的用户行为特征才能够识别出异常用户,而文中黑产用户识别模型可以同时通过用户关系层面的关联挖掘潜在的黑产用户,在其进行异常操作之前进行相应的控制。

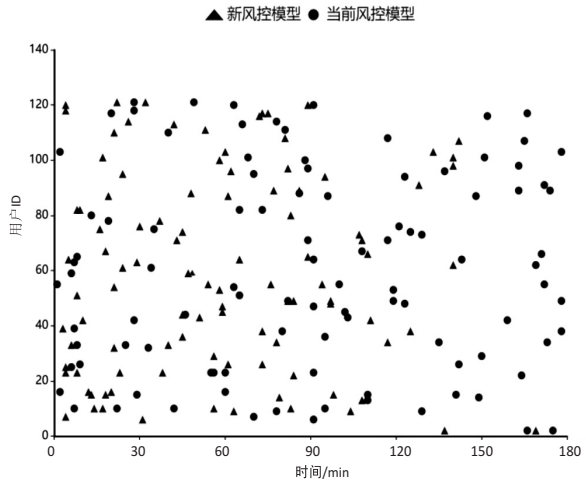

图 4 异常用户识别时间对比

5 结束语

分析了现有售票系统的数据特点,构建了用户实体特征以及实体之间关系特征,基于 GraphSAGE 模型进行邻居节点采样,将旅客个体特征与其相邻节点特征进行融合,生成最终的旅客特征向量,将其作为无核二次曲面支持向量机的输入进行黑产用户识别。实验证明融合个体特征及邻居节点特征生成的特征向量在相同的模型下,准确率及 AUC 值都表现较好,而且相对现有的风控策略模型,降低了误伤率,缩短了黑产用户识别的时间。

参考文献:

[1] 周亮瑾,阎志远,戴琳琳. 铁路互联网售票异常行为分类技术的研究与应用[J]. 中国铁道科学,2019,40(6):133-139.

[2] MOH T S, MURMANN A J. Can you judge a man by his friends? - enhancing spammer detection on the twitter microblogging platform using friends and followers[C]//International conference on information systems, technology and management. Bangkok:Springer,2010:210-220.

[3] 宋艳红. 社交网络中异常用户的识别与研究[D]. 长春:长春理工大学,2017.

[4] DIN I U, MASOOD F, AMMAD G, et al. Spammer detection and fake user identification on social networks[J]. IEEE Access,2019,7(1):1-14.

[5] 仲丽君,杨文忠,袁婷婷,等. 社交网络异常用户识别技术综述[J]. 计算机工程与应用,2018,54(16):13-23.

[6] 刘丽娇,陶俊才,肖晓军,等. 电信大规模社交关系网络图数据挖掘研究[J]. 电信科学,2015,31(1):23-31.

[7] 于 静,刘燕兵,张 宇,等. 大规模图数据匹配技术综述[J]. 计算机研究与发展,2015,52(2):391-409.

[8] 郝晓培. 基于大数据的铁路客运用户画像系统研究及应用[D]. 北京:中国铁道科学研究院,2018.

[9] 白 铂,刘玉婷,马驰骋,等. 图神经网络[J]. 中国科学:数学,2020,50(3):31-48.

[10] ZHANG Z. Semi-supervised hyperspectral image classification algorithm based on graph embedding and discriminative spatial information[J]. Microprocessors and Microsystems,2020,75:103070.

[11] YU B,ZHANG Y,XIE Y,et al. Influence-aware graph neural networks[J]. Applied Soft Computing, 2021, 104(6):107169.

[12] LUO J,FANG S C,DENG Z,et al. Soft quadratic surface support vector machine for binary classification[J]. Asia-Pacific Journal of Operational Research,2016,33(6):144-152.

[13] 孙 海. Spark 的图计算框架:GraphX[J]. 现代计算机,2017(9):120-122.

[14] 王 钰,赵晓艳,杨杏丽,等. 基于 K 折交叉验证 Beta 分布的 AUC 度量的置信区间[J]. 系统科学与数学,2020,40(9):1564-1577.

[15] 王彦光,朱鸿斌,徐维超. ROC 曲线及其分析方法综述[J]. 广东工业大学学报,2021,38(1):46-53.