计算机科学
COMPUTER SCIENCE

基于异构信息网络的信贷反欺诈研究

刘华玲, 张国祥, 王柳月, 梁华璧

引用本文

刘华玲, 张国祥, 王柳月, 梁华璧. 基于异构信息网络的信贷反欺诈研究[J]. 计算机科学, 2023, 50(11A): 221100173-9.

LIU Hualing, ZHANG Guoxiang, WANG Liuyue, LIANG Huabi. Study on Credit Anti-fraud Based on Heterogeneous Information Network [J]. Computer Science, 2023, 50(11A): 221100173-9.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于图嵌入的正交局部保持投影无监督特征选择

Orthogonal Locality Preserving Projection Unsupervised Feature Selection Based on Graph Embedding

计算机科学, 2023, 50(11A): 220900003-9. https://doi.org/10.11896/jsjkx.220900003

基于课程学习和图嵌入的协同推荐

Collaborative Recommendation Based on Curriculum Learning and Graph Embedding

计算机科学, 2023, 50(11A): 221100030-8. https://doi.org/10.11896/jsjkx.221100030

基于二部图表示的属性网络社区发现算法

Community Discovery Algorithm for Attributed Networks Based on Bipartite Graph Representation

计算机科学, 2023, 50(11): 107-113. https://doi.org/10.11896/jsjkx.221000226

面向移动应用评分推荐的多任务图嵌入深度预测模型

Multi-task Graph-embedding Deep Prediction Model for Mobile App Rating Recommendation

计算机科学, 2023, 50(9): 160-167. https://doi.org/10.11896/jsjkx.220700035

基于异构信息网络的最大影响力社区搜索

Maximum Influential Community Search in Heterogeneous Information Network

计算机科学, 2023, 50(8): 16-26. https://doi.org/10.11896/jsjkx.220600262

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221100173

基于异构信息网络的信贷反欺诈研究

刘华玲　张国祥　王柳月　梁华璧
上海对外经贸大学统计与信息学院　上海 201620

摘　要　近年来,移动终端设备的数字化程度陡升,信贷行业的欺诈行为呈现出动态发展、行为隐蔽和专业伪装等新特点,海量数据的跨量级增长为传统反欺诈算法的有效性和计算效率都带来了不小的挑战。因此,为了充分学习信贷场景中不同实体间的交互信息,降低算法计算消耗以使其适用于大规模图数据任务,提出了基于异构信息网络的特异群组挖掘算法 BKH-(Bron-Kerbosh-H-II),即首先针对源数据中的信贷实体及实体间的关系进行界定和分类,并将不同实体间的相似度作为关系权重,以此构建信贷异构信息网络,对该网络采取了两阶段的基于 H 图的极大团枚举算法,用于挖掘特异群组,最终通过局部特征工程修正划分得到潜在的欺诈群体,经实验证明,BKH-II 在 4 种评价指标上的准确度分别为 $NMI=0.983$,$NRI=0.96$,$F\text{-}score=0.943$,$Omega=0.95$,并表现出了良好的泛化性和较低的计算复杂性。

关键词:异构信息网络;信贷反欺诈;特异群组挖掘;社区发现;图嵌入
中图法分类号　TP391

Study on Credit Anti-fraud Based on Heterogeneous Information Network

LIU Hualing,ZHANG Guoxiang,WANG Liuyue and LIANG Huabi
School of Statistics and Information,Shanghai University of International Business and Economics,Shanghai 201620,China

Abstract　In recent years,the digitization of mobile terminal equipment has risen sharply,and fraudulent behaviors in the credit industry have shown new characteristics such as dynamic development,concealment of behavior,and professional camouflage. The cross-order growth of massive data has brought considerable challenges to the effectiveness and computational efficiency of traditional anti-fraud algorithms. Therefore,this paper aims to fully learn the interaction information between different entities in the credit scene,reduce the computational consumption of the algorithm to make it suitable for large-scale graph data tasks,and propose a specific group mining algorithm BKH-II(Bron-Kerbosh-H-II) based on heterogeneous information networks. First,defining and classifing the credit entities and the relationships between them in the source data,and using the similarity between different entities as the relationship weight to build a credit heterogeneous information network. A two-stage H-graph-based maximal clique enumeration algorithm is adopted for the network to mine unique groups. Finally,potential fraud groups are obtained through local feature engineering correction and division. Experiments prove that the accuracy of BKH-II on the four evaluation indicators is $NMI=0.983$,$NRI=0.96$,$F\text{-}score=0.943$,$Omega=0.95$,and shows good generalization and low computational complexity.

Keywords　Heterogeneous information network,Credit anti-fraud,Specific group mining,Community discovery,Graph embedding

1　引言

数字化技术的日趋成熟为交易场景电子化创设了基础,诸多新型金融产品的服务场景由线下转移至线上,这在无形之中增加了数字欺诈的风险。数字化金融场景下的信贷欺诈正呈现出动态、隐蔽且多样化的特点,其中欺诈识别是反欺诈的核心任务。在反欺诈研究中,欺诈行为定义为明显异于其正常行为模式的疑似行为,Shen 等[1]提出可通过一定范围内的规范行为区域来进一步判别疑似行为的合法性,从而予以识别[1]。然而,由于欺诈行为具有多变性、主体类型较多、数据获取难度较高且难以准确定义不同情境下的规范区域等困难,欺诈识别任务的挑战性进一步增加。

随着数字化程度的不断提高,海量的用户行为信息和属性信息成为了金融领域重要的生产要素,这些数据资源可支撑针对用户行为和用户间的联系进行更为深层次的挖掘和分析。基于大数据的用户行为分析挖掘,也将为金融反欺诈识别任务提供更高准确度的保证,针对欺诈行为模式进行学习和预测,进而能够为信贷金融体系的风险管控和监测预警提供支持。

在基于特异群组的异常群体检测中,由于欺诈数据与正常数据分布的不平衡性,特异群组相比网络整体来说所占比重较小,但同时小部分特异群组间的行为带有明显的相似性和集聚性。特异群组挖掘任务的目的便是从大规模群体中挖掘出带有异常行为关系的小部分群体。传统的聚类方法存在以下两方面不足:(1)由于数据的不平衡性,比重较大的正常群体将会迷惑聚类算法而忽视小部分异常群体,因此聚类

通信作者:刘华玲(liuhl@suibe.edu.cn)

结果往往不尽人意。(2)由于个人行为习惯的原因,使其也可能被识别为特异群组,传统方法通常有较高的误报率。实验结果证明了本文解决方案的可行性和优势。针对以上不足,本文采用特异群组挖掘技术,聚焦于关键少数对象,提出了针对群体反欺诈的识别算法,采取两阶段重叠社区划分模型,其中的主要创新点如下。

1)BKH-II(Bron-Kerbosh-H-II)模型创新性考虑了由于不良标签而引起的高欺诈率群体行为的发生,针对性与实时性强,在异构信息网络的基础上提前检测欺诈行为的发生,从而触发反欺诈预警。

2)极大团枚举法有效减少了图计算量,算法复杂度明显优于全图计算的其他模型。

本文第 2 章介绍了相关工作;第 3 章给出了相关概念和问题的定义;第 4 章详细阐述了具体的模型;第 5 章进行了实验验证;最后总结全文并展望未来。

本文的研究框架如图 1 所示。

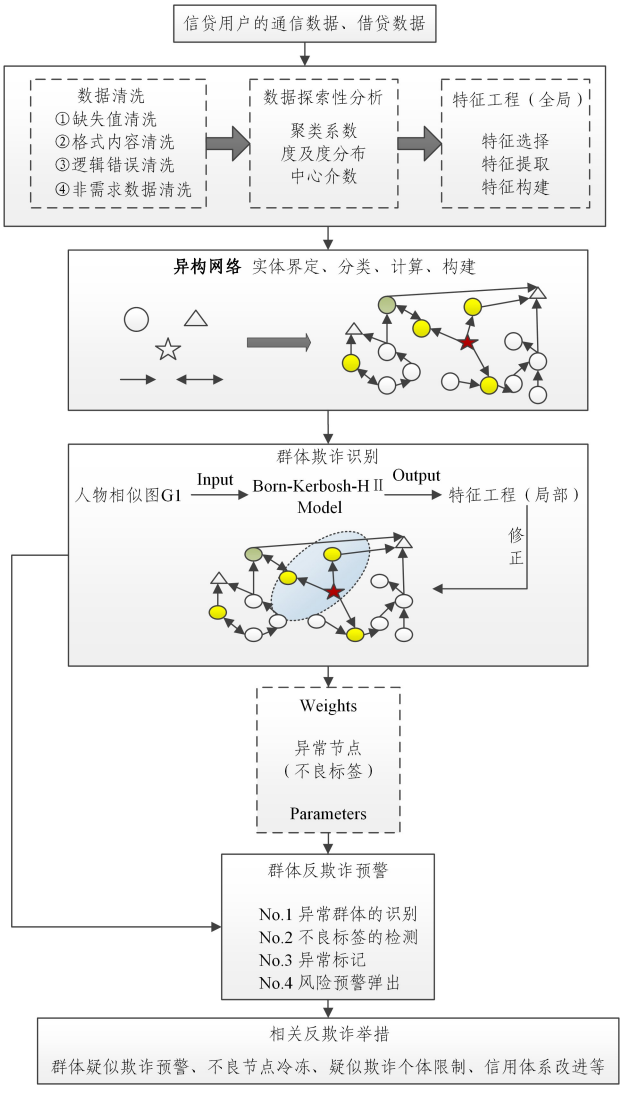


图 1 研究框架图

Fig. 1 Research framework

2 相关工作

国内外对于欺诈识别的研究根据反欺诈模型类型可分为 6 类,如图 2 所示。分别为反欺诈分类、聚类、离群点检测、模式挖掘、图挖掘和深度学习算法。

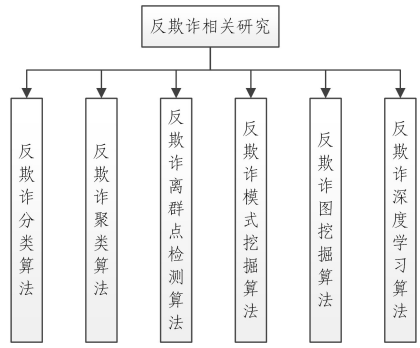


图 2 反欺诈算法分类

Fig. 2 Classification of anti-fraud algorithms

Lu 等[2]提出了基于类加权支持向量机的信用卡欺诈检测模型,该模型在支持向量机模型的基础上缩短模型的训练维数,在大幅缩短训练时长的同时提高了算法的精度。Nian 等[3]针对汽车保险的欺诈问题,提出了一种基于无监督光谱排序算法的欺诈检测模型 SRA,通过拉普拉斯矩阵计算非主特征量来生成对应的异常排序,具有较高的异常识别准确率。高维数据往往呈现出稀疏性、等距离分布、多样化的特点,许多属性信息多为不相关关系。为解决此类问题,Keller 等[4]提出了基于密度的检测方法 HiCS,该算法适用于具有高对比度子空间的数据场景,Nguyen 等[5]针对高维数据的维度诅咒和网格分辨率问题,提出了一种基于距离的离群点检测算法,该模型可处理高维数据中的一些局部属性信息。此类算法模型无需提供标记信息,可通过数据本身的内容信息进行识别。Bhattacharyya 等[6]通过分析信用卡交易场景下的相关欺诈模式,来助力金融机构更新和改变其风控规则[7]。

图挖掘算法的研究思路是利用图论相关的知识对数据进行建模分析,将数据行为抽象为实体和关系表示。Yan 等[7]针对稠密子图结构异常而难以度量的问题,提出了一种新的度量方法,该方法综合考虑了网络结构、时间变化以及节点和连边的属性信息,通过改进贪婪算法并将其引入至网络结构,来识别网络中异常的结构和所表示的欺诈者。He 等[8]通过研究发现,网络结构可自然地表示出各实体的属性信息及不同实体间的交互关系,这为进一步深挖欺诈行为中的复杂关系提供了可能,同时网络中参数的迭代计算也可充分考虑时间因素所导致的欺诈行为动态性的问题,这将为欺诈检测领域的研究提供新的研究方向。常用的静态图结构往往难以刻画时间动态性变化,Xie 等[9]提出的基于时空稀疏注意力的时空图挖掘算法能够对时空依赖关系建模,以捕获动态图特征[10]。

反欺诈深度学习算法是近年来反欺诈领域的研究重点之一,该类算法模型多依靠大量带有标签的数据作为训练集,并应用深度学习算法完成实际的欺诈检测任务。Roy 等[10]对深度学习模型中的拓扑结构展开了深入的研究,提出了一种基于信用卡交易场景下的深度学习参数调整框架,该框架对调整深度模型中关键参数的优化过程提供了有效的帮助。Cai 等[11]在多场景视角下训练 CNN 来自动提取特征信息,引入模糊犹豫集进行综合决策,以提高欺诈攻击检测的准确度。Wang 等[12]借助基于随机森林的序列向前搜索策略算法对行业欺诈交易行为进行检测,较好地解决了多数类样本误分类

问题。深度学习算法依赖于大量的训练数据和较高的计算资源，有较高的准确率和较低的实现难度。

3 异构信息网络的构建

3.1 问题分析

金融信贷数据包含了信贷用户属性、信贷行为特征及时序行为表现的 3 部分特征数据，传统的反欺诈算法会损失特征数据中的部分交互信息。图模型在对现实世界的表示中具备天然优势，网络中的节点、连边最大程度地描述了现实中实体及实体间关系的交互信息，辅以相关属性参数支撑了对于数据中隐藏的时序信息的挖掘，近年来图挖掘已成为解决大规模复杂数据的推理、发现的重要研究工具。

本文对源数据进行异构网络建模，旨在充分保留不同实体间的复杂交互信息。其中，网络的节点分别表示了信贷者、不良标签和信贷记录等不同实体，连边表示 3 种实体间的交互关系。本文针对欺诈行为发生主体的类型及各自带有的行为特点，结合并改进异构信息网络中的相关技术（如图嵌入、社区划分、子图挖掘、多层贝叶斯网络、图神经网络和极大团枚举算法等），借鉴文献[13]提出了针对于不同欺诈主体类型的金融信贷反欺诈算法。

异构信息网络的构建分为 3 个步骤：(1)节点实体界定、分类；(2)计算用户级别信贷记录相似度及连边权重，以此构建信贷用户信贷图 G；(3)计算图结构相关系数和可视化分析。

3.2 节点实体界定、分类

本节定义的信贷流程图具有高复杂性的数据结构，这对相似性的计算和社区划分提出了较大的挑战。表 1 列出了计算过程所涉及到的相关变量的符号表示和含义。

表 1 符号说明

Table 1 Explanation of symbols

符号	含义
D	信用得分
PO	用户属性
HA	信贷记录
P	用户
ρ	局部密度
γ	信贷 a 的临近信贷集合
λ	与其他具有更高密度的任何对象之间的最小距离
G	用户信贷记录图

定义 1（信贷记录图） 信贷记录图是一种包含两类节点和三类边的异构图。将信贷记录图表示为 $G=(V,E,W)$，V 是节点集合，E 是连边集合，W 是边的权重集合。

3.3 相似度计算及网络构建

对网络中的相似社区进行划分，需要计算每次信贷记录间的相似性。但是，信贷记录包含了类别标签信息和数值信息。用户画像视角下的相似性度量的计算式如下：

$$Sim(PO_r,PO_s)=\frac{\delta(PO_r^Y,PO_s^Y)*\left[\delta(PO_r^{\text{Education}},PO_s^{\text{Education}})+\dfrac{\min(PO_s^{\text{MS}}\cdot PO_r^{\text{MS}})}{\max(PO_r^{\text{MS}},PO_s^{\text{MS}})}\right]}{2} \tag{1}$$

其中，$\delta(X,Y)$为二值判断函数。如式(1)所示，根据 Lusseau 等[14]提出的用户相似度计算方法，本节首先对用户 PO_r 和 PO_s 的 Y 标签属性进行比较，若标签相同，则继续比较类别特征，若标签不同，则标记两者的相似度为 0。

$$HA_1=\{\text{Personas}_1,PO_{11},PO_{12},\cdots,PO_{1m}\} \tag{2}$$

$$HA_2=\{\text{Personas}_2,PO_{11},PO_{12},\cdots,PO_{1h}\} \tag{3}$$

其中，m，h 分别表示 HA_1 和 HA_2 所包含的用户画像数目，它们之间的相似度定义为：

$$Sim(HA_1,HA_2)=\frac{|Common(HA_1,HA_2)|}{|Ution(HA_1,HA_2)|} \tag{4}$$

其中，$|Common(HA_1,HA_2)|$和$|Ution\ (HA_1,HA_2)|$分别指的是 HA_1 和 HA_2 所拥有的共同用户画像的数量和总数量。

$$|\ Common(HA_1.\ HA_2)\ |=\sum Sim(PO_{1p},PO_{2q})*a_{pq} \tag{5}$$

其中，$Sim\ (PO_{1p},PO_{2q})$表示 HA_1 中第 p 个用户画像和 HA_2 中第 q 个用户画像之间的相似度，$\boldsymbol{A}=(a_{pq})_{|HA_1|*|H.A_2|}$ 是一个分配矩阵，可以通过求解下列凸优化问题来获得分配矩阵 $\boldsymbol{A}$。

$$\begin{aligned}&\arg\max_{\boldsymbol{A}} z=\sum_{\substack{p,q\in\mathbb{Z}\\ p\in[1,|HA_1|]\\ q\in[1,|HA_2|]}} Sim(PO_{1p},PO_{2q})*a_{pq}\\ &\text{s.t.}\ \sum_q a_{pq}\leqslant freq\ (PO_{1p})\\ &\quad\sum_p a_{pq}\leqslant freq\ (PO_{2q})\\ &\quad a_{pq}\geqslant 0\end{aligned} \tag{6}$$

其中，$freq\ (PO_{1p})$和 $freq\ (PO_{2q})$分别代表 HA_1 中第 p 个用户画像和 HA_2 中第 q 个用户画像出现的次数。最终的信贷记录 HA_1 和 HA_2 间的相似度计算如下：

$$Sim(HA_1,HA_2)=\frac{\sum_{pq}Sim(PO_{1p},PO_{2q})*a_{pq}}{|HA_1|+|HA_2|-\sum_{pq}Sim(PO_{1p},PO_{2q})*a_{pq}} \tag{7}$$

3.4 网络结构、统计特征分析及可视化

对原始数据进行图结构转化，对脏数据字段进行数据清洗，并删除因计算方式而产生的非实体节点和带有无效权重的连边，最终形成异构信息网络图的构建算法，如算法 1 所示。

算法 1 Credit behavior similarity graph construction

输入：Person set P＝{p_1，p_2，…，p_n}：Credit behavior record of each person in the time period

输出：Similarity Adjacency Graph Gp

```
1. Begin
2.   For i in [1,n−1] do
3.     For j in [1,n] do
4.       Obtain Joint Credit Behavior Set B{p_i,p_j} between p_i and p_j
5.       For each b{b_ik,b_jk,a_k,I_k} in B{p_i,p_j} Compute SJ{b_ik,b_jk,a_k,I_k}
6.     Weight set Sim{p_i,p_j}
7.   End for
8.  End for
9.   Vertex Set V=P
10.  Edge set E indicates the connected vertices are similar
11.  Construct Credit behavior similarity graph Gp =(V,E,W)
12.  Return Gp
13. End
```

针对网络结构的特性及相关参数进行探索性分析，网络基本结构的描述如表 2 所列。

表 2 网络结构特征表

Table 2 Characteristics of network structure

统计特征	规格	用途
节点数	62561	用户、记录异构实体
连边数	147878	三类连边
平均聚类系数	0.0055	社区密集程度
三元节点数	2024	子图特征
Fraction of closed triangles	0.001294	
最长最短路径	11	优化路径
90-percentile effective diameter	6.7	
权重	Sim	计算参数

为了进一步对网络结构展开分析和讨论，将上述构建的网络结构数据依次导入 Gephi 软件并设置 Node/Edge/Weight 这 3 类关键网络结构参数，从而生成初始的网络图。考虑到图运算的算法复杂性和计算规模的挑战，旨在更清晰地展示网络的结构特性，本文借鉴 Zhao[15] 提出的 ForceAtlas2 布局算法，对可视化网络图进行布局运算。该算法融合了 Barnes Hut 近似算法、节点度决定性斥力因素、自调整的整体与局部迭代速度等技术。相比 Force Atlas 算法，第二代算法具有更快的运行速度，且可处理更大规模的图结构。算法运行时，节点与节点之间将会相互排斥，存在连边的两个节点将会相互吸引。算法参数的设置包括惯性＝0.5，斥力强度＝200，吸引强度＝10，最大位移量＝20，自动稳定强度＝80，自动稳定敏感性＝0.2，重力＝20，开启吸引力分布，开启曲尺寸调整，运算速度为 10。

此外，本文参考 Gregory 等[16] 提出的发现重叠社区结果的算法，计算了网络的平均度、平均加权度、图密度，并同时利用 PageRank 算法计算了网络图中每个实体节点的 PR 值，以此 PR 值和 Lable 标签填充节点颜色并设置节点大小，以权重值设置连边粗细和颜色，最终形成的可视化网络图如图 3 所示。

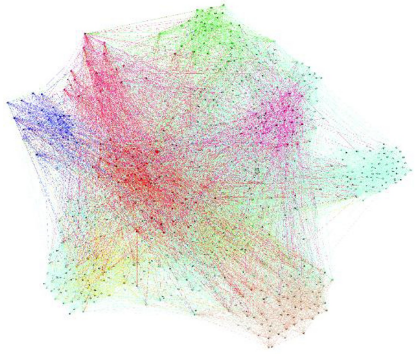


图 3 信贷异构信息网络整体可视化

Fig. 3 Overall visualization of credit heterogeneous information network

通过采用相似度权重作为网络结构大小及颜色的设置因素，可以看出在图网络中存在着不同的颜色分块，我们将这种现象称之为网络的社区结构。相同着色的节点簇聚集到了网络的同一区域，可以看出图中的颜色分布存在明显聚集和重叠区域。为进一步看清局部网络结构的详情，将分布视角拉至某单一颜色区域，如图 4 所示，可以观察到局部结构中某些大节点与小节点相互连接交叉形成了局部特性，可将其理解为信贷行为数据中存在一定的相似性集聚表现，也存在一定的社区重叠现象。

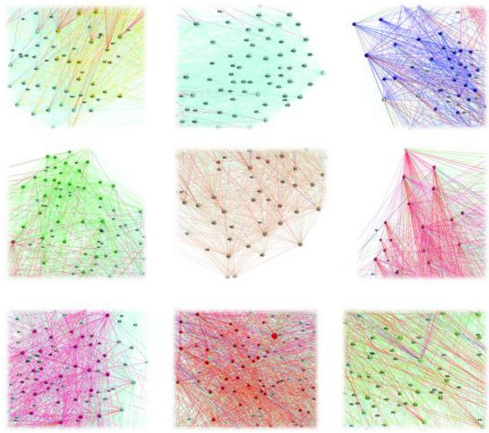


图 4 信贷异构信息网络局部可视化

Fig. 4 Local visualization of credit heterogeneous information network

4 基于特异群组的异常群体检测

通过对原始数据和图数据结构的探索性分析，可以看出在所构建的信贷异构信息网络中存在明显的信贷记录相似性集聚现象，即图网络中的社区或群组结构。本文从图网络的结构出发，首先从整个异构信息网络聚焦于存在嫌疑的局部网络（子图），提出了基于特异群组挖掘的算法 BKH-II 来实现针对局部社区的欺诈检测，目的是挖掘出相似的欺诈社区并分析其行为规律。

与聚类算法和分类模型类似，特异群组挖掘算法是通过计算数据对象间的相似性进行数据划分的挖掘任务。但与 Shen 等[17] 在上述算法的实验结果相比，经特异群组挖掘所划分出的群图相比其他个体具有明显的特殊性和异常性，同时具有明显的强相似性和紧粘合性的对象将被划分至同一群组。

4.1 特异群组挖掘

特异群组挖掘的算法 BKH-II 旨在充分挖掘存在相似欺诈行为的可疑群组，其是由经典的特异群组挖掘算法框架（AGM）改进而来的。AGM 算法的框架如图 5 所示，首先根据相似度计算的结果找出最为相似的数据对象，并采用修正策略对候选集进行二次修正计算，以排除误划入的相似对象，在修正后的候选集中再次计算特异群组的相似性，第二阶段则是将特异对象进行分组划分。

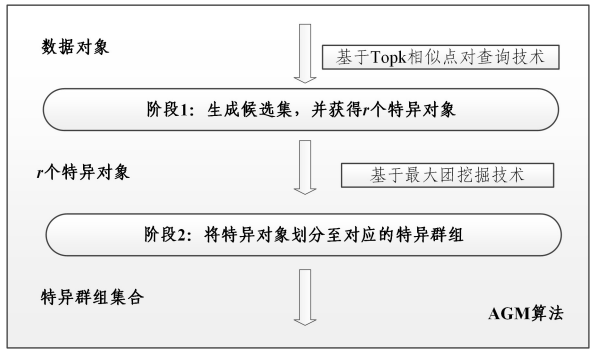


图 5 AGM 算法的思路

Fig. 5 Process of AGM algorithm

BKH-II 算法框架主要分为 4 个阶段（见图 6）：

1）根据时序信贷行为计算用户间的相似度，以此构建

人物行为相似邻接图 G；

2）对图 G 进行两阶段的基于 H 图的极大团枚举算法挖掘特异群组；

3）用局部特征工程进行二次区分；

4）根据筛选特征对特异群组进行校正，划分可疑群组和偶发群组。

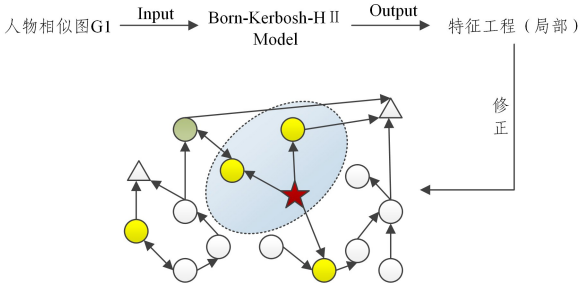


图 6　BKH-II 算法的框架

Fig. 6　Structure of BKH-II algorithm

4.1.1　图表示学习

将图结构表示为向量空间中的向量是进行后续算法挖掘的初始步骤，图表示学习也称为图嵌入（Graph Embedding），是一种将图数据映射为低维稠密向量集的过程。作为输入的图数据属于高维稀疏的抽象空间，经过映射后图数据中节点或连边等信息将在低维稠密的空间中嵌入表示。Lusseau[14] 通过图嵌入，将海量的高维、异构、动态且复杂的图数据处理为可被高效输入机器学习算法的嵌入向量或向量集。

综合考虑数据规模及网络结构特性，本文采用了经典的 node2vec 算法进行图表示学习。在 DeepWalk 中，采用深度优先采样（DFS）策略，即按照 Qian 等[18] 的方法从源节点开始以递增的距离依次采样产生节点序列。DFS 策略得到的节点序列具有同质性，即以距离为节点间相似性的度量。与 DFS 策略相反，BFS（广度优先采样）策略是从源节点开始，探索当前深度的所有邻居节点得到结构性，以节点在网络中的位置和结构表示相似性。Liu 等[19] 在实验中采用 node2vec，通过调整随机游走权重的方法使得图嵌入的结果在网络的同质性和结构性取得权衡，通过设立节点间的跳转概率进而控制对 BFS 和 DFS 的倾向性。图 7 给出了 node2vec 算法从节点 t 跳转节点 v 后下一步以节点 v 为起点继续跳转的概率。

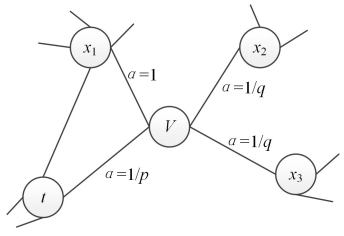


图 7　node2vec 算法节点跳转原理

Fig. 7　Node hops principle of node2vec algorithm

从节点 v 跳转到下一节点 x 的概率为 $\pi_{vx}=\alpha_{pq}(t,x)\cdot\omega_{vx}$，其中 ω_{vx} 为边 V_x 的权重，$\alpha_{pq}(t,x)$ 的定义如下：

$$\alpha_{pq}(t,x)=\begin{cases}\dfrac{1}{p}, & \text{if } d_{tx}=0\\ 1, & \text{if } d_{tx}=1\\ \dfrac{1}{q}, & if\ d_{tx}=2\end{cases} \tag{8}$$

其中，d_{tx} 指的是节点 t 到节点 x 的距离，返回参数 p 和进出参数 q 共同控制着随机游走的倾向性。其中 p 越小，随机游走回节点 t 的可能性越大即算法更注重表达网络的同质性；q 越小，则随机游走到远方节点的可能性越大，即更注重表达网络的结构性，反之，当前节点更可能在附近节点游走。

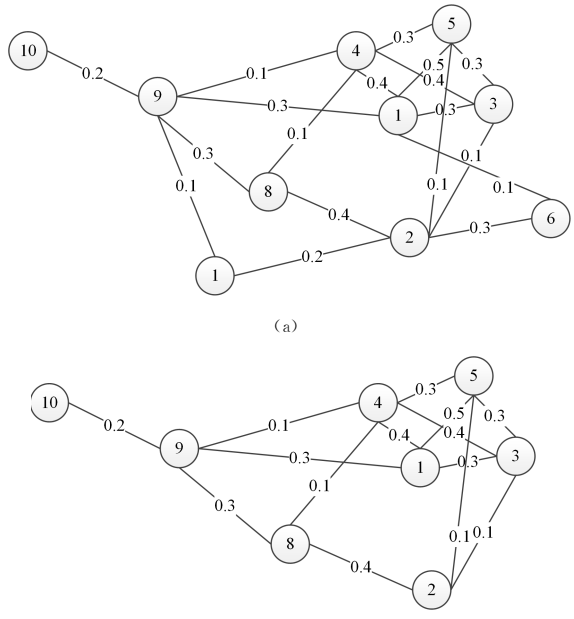
4.1.2　特异群组挖掘

通过对图结构参数及相似性问题的讨论，可以得出结论，群组内的用户在信贷时序行为上存在相似性，所以，将此现实问题嵌入至图网络中便可转换为图 Gps 中的极大团枚举问题。

定义 3（h-顶点）　对于给定图 $G=(V,E)$，h-顶点集合 H 定义为 $H=\{v:v\in V,d(v)\geqslant h\}$。其中 $d(v)$ 表示节点 v 的度，且 $\forall v\in(H\backslash V),d(v)<h$，$H\backslash V$ 表示同属于 H 和 V 的共同顶点集合。

定义 4（H 图）　对于给定图 $G=(V,E)$ 和 h-顶点集 H，图 G 的 H 图可表示为 GH，是由图 G 在 H 集合的约束下推导而得的。

如图 8(a)所示，该图拥有 10 个顶点 18 条连边，此时我们设置阈值 $h=3$，按上述定义可得此时的 h-顶点集 $H=\{1,2,3,4,5,8,9\}$，同时可以判断顶点集 H 中所包含的 7 个顶点的度都大于等于 3，而图中其他顶点的度都小于 3，故通过阈值 h 筛选顶点后的最大团如图 8(b)所示。



(a)

(b)

图 8　H 图计算示例

Fig. 8　Example of H-plot calculation

通过 H 图的转化方式可以大幅简化计算，即通过获取图 Gps 的 H 图 GpsH 进而使用特异群组挖掘算法。但与传统的 MCE 任务不同，前文所构建的信贷异构信息网络是带权网络，而 Li 等[20] 在实验中采用的 MCE 方法通常是针对 unweight 的无标度网络。因此，本文提出了一种两阶段的基于 H-Graph 的特异群组挖掘算法，其基本算法步骤为首先将所构建的人物相似图 Gps 通过 h-顶点集合推导出基于 H 图的图 $GpsH$，之后以此为基础开展基于分区的 MCE 算法。

算法首先从 h-顶点集合中依次挑出某顶点对并计算其平均相似度，其计算方法如式(9)所示，相似度将作为顶点排序的重要参考标准，其中 $adj(v')$ 表示顶点 v' 的邻接节点，

$d(v')$表示计算顶点v'的度，即计算节点v'的所有邻接节点的加权值与节点度比值的求和。

$$AS_{v'}=\frac{\sum_{\forall u\in adj(v')} w_{uv'}}{d(v')} \tag{9}$$

同时，为进一步简化后续的计算复杂性，第二阶段将采用基于分区的 MCE 算法，其主要思路借鉴了 Spark 分布式流计算的设计模式，Huang 等[21]提出将整图中的某一子图读取至内存，并在内存中完成子图本地计算最大团的任务，这将有效提升模型的计算速度并降低资源的浪费。借助种子顶点和扩展种子节点将辅助我们选取合适的子图，通常将具有最高平均相似度的前 20%的节点作为候选的种子顶点，之后应用基于分区的 MCE 算法获得极大团集合 M(C)。此时，所获得的每个极大团集合可视为特异群组，但 Yoshida 等[22]提出由于某些偶发因素及用户的特殊行为习惯可能会产生带有误差的群组划分，故此阶段获得的特异群组并非全部隶属于最终的欺诈群组，应对挖掘出的特异群组进行进一步筛选，将其区分为偶发组和可疑组。

算法 2 BKH-II

输入：Graph 阈值 h，hV=Ø
输出：极大团集合 M(C)
1. For V in (v_1, v_2, …, v_n) do
2. If d(v)>h then
3. hV=hV∪{u}
4. end if
5. end for
6. Compute H-graph GH of G
7. For i in (G_{H_1}, G_{H_2}, …, G_{H_n}) do
8. $AS_{v'}=\frac{\sum_{\forall u\in adj(v')} w_{uv'}}{d(v')}$
9. end for
10. Ranking every v in GH with AS
11. Put top 20% node in seed node set S
12. S+=Ø
13. For each v″ in S do
14. $S_+:=(S_+\cup\{v''\}\cup adj(v'')$
15. end for
16. from GH get G_{S_+}
17. Apply based partitions MCE on G_{S_+}
18. Return 极大团集合 M(C)

4.2 局部特征选择

该步骤的主要目的是选择用于区分为偶发群体和欺诈群体的特征，在输入数据中，通过筛选得到可以有效反映输入数据的变量子集，从而避免无关变量对最终结果的影响，保证模型的划分效果。

为删除输入变量中的无关变量，应定义合理的筛选标准。根据 Huang 等[23]对变量相关性对定义，若某变量与类别标签呈相关关系，此时该变量可作为独立输入数据，即可将对于最终的类别标签无影响的特征变量进行剔除。假设经 BKH-II 算法挖掘出的候选特异群组为 $temp\_G=\{g_1, g_2, \cdots, g_n\}$（此时每个特异群组为图 GpsH 中的极大团推导而得到的一组节点），每个组g_i可以用多个变量$g_i=\{f_{i1}, f_{i2}, \cdots, f_{im}\}$表示，但并非所有变量都可用于特征的筛选。

为对特征筛选选定标准，本文采用了无监督特征选择算法（分组之间的集散系数），不同群组之间关于某一特定变量特征的离散程度可由组间集散系数表示。在对观测特征f_{i1}进行筛选时，其对应的组间系数越小，移除该特征的可能性就越大，最终根据计算得到的组间系数对相关特征集合进行筛选，按得分进行排序，得分 S_j 的计算式如下：

$$s_j=\sum_{i=1}^{n}\frac{\sqrt{\frac{((f_{ij}-\mu(f_j))^2}{n}}}{\mu(f_j)} \tag{9}$$

其中，n 表示分组的数量，f_j 表示该组的第 j 个特征，f_{ij} 表示组g_i中第 j 个特征的值。通过式(9)计算得到的得分可用作特征排序的计算标准，S_j得分越高，表示该特征的预期目标类别标签的相关性越高，具体的筛选阈值可由整体得分的分布决定。

4.3 群组划分

根据 4.2 节选择的特征集合，此时每个特异群组g_i可通过 Embedding 的方式嵌入为向量 FVgi，之后可采用支持向量机的分类模型将特征向量映射至高维空间。本文参照 Xie 等[24]选用了最为广泛的高斯核函数 $K(x,y)=\exp^{-\frac{(\|x-y\|)^2}{2*(\Gamma)^2}}$ 来计算模型的最优决策面，其中 Γ 作为超参数可以通过设置交叉验证进行学习来获得。据此本文完成了对结果集的筛选，并将所挖掘出的群组划分为偶发群组和欺诈群组，此时被归于可疑群组中的节点则被看作疑似欺诈者，通过局部特征选择的二次校正，可将一些由于周期或偶发因素而被初次划分进来的正常群组筛除，这将有效提升最终欺诈识别的准确性。

5 实验

本节旨在充分评估模型的准确率、泛化性能和时间复杂度，采用了该领域内常用的 LFR benchmark 人工网络和真实 UCI 数据集构造的网络对 BaseLine 基准模型和本文提出的 BKH-II 模型进行了对比分析。首先，本节给出了详细的实验设置，证明了基准模型的有效性，同时采用了经典的 3 类评价指标评估了模型效果，并对各基准模型和目标模型在泛化性能和时间复杂度维度进行了比较，最后对算法的可解释性做了评估。

5.1 数据集及参数设置

数据集 1 LFR benchmark 人工网络

人工合成网络采用了社区发现领域经典的 LFRbenchmark 人工网络，该网络常用于检测评估社区发现算法在不同规模、结构网络中的执行效果评估，实验可以通过调整不同的参数评估模型在不同网络结构中的性能稳定性和变化幅度。

本次实验综合考虑了计算资源与效果展示的因素，参照 Xie 等[24]的实验可视化效果，选定了节点规模为 1 000、混合参数 μ 为 0 ～ 0.6 来阶梯化地对比不同模型的评估效果。所生成的人工合成网络的节点平均度控制在 20，最大节点度为 50，网络中节点度序列和社区规模序列分别服从指数为 2 和 1 的幂律分布，以此来模拟真实的网络世界。为避免实验过程中因计算资源、参数设置而导致的偶然误差，每种混合参数 μ 下随机生成 9 张人工网络图（lfr_ut_0.0_rep_0 ～ lfr_ut_0.0_rep_9）用来计算，同时每个网络上执行 30 次计算，取所有运算结果的均值作为最终的效果评分。

数据集 2 UCI 数据集所构建的信贷异构信息网络

来自 UCI 的中华大学 2016 年提供的信贷欺诈检测公开

数据集，共包含 2005 年 4 月至 9 月的 30 000 条基础时序数据，23 个变量数据和 1 个标签数据，其中变量数据分为用户属性数据、信贷及时序行为数据。

5.2 基准模型

1）Walktrap：基于随机游走思想的社区划分模型，Garza 等[25]通过计算节点与其邻居节点间的距离来定义游走的概率参数，从而实现社区的划分。

2）Surprise_communities：Hu 等[26]定义了 Surprise 函数，用于描述在某种给定的随机模型的情况下，网络的某个划分与社区节点和连边期望分布之间的距离。

3）Significance_communities：基于模块化优化参数而设定的社区发现算法，引入了 Signficance 因素，通过模块度理论的推导对节点的归并进行优化。

4）Infomap：通过双层编码方式将特异群组与信息的编码相融合，通常更短的信息编码代表了该次群组划分的效果，由此便可将社区发现问题转化为追求最短编码长度的问题。

5）Label propagation：基于图半监督学习的算法，利用数据集中打标的数据预测其他未曾标记过的数据。Zhang 等[27]认为网络结构中不同节点的标签反映了其社区属性，节点的标签数据也会随网络的迭代计算发生相应的修改，最终具有相同标签的节点将收敛于同一社区。

6）Greedy_modularity：此算法根据模块性的改变对团体进行融合，通过多次迭代记录新的聚类模式和对应的模块性得分，最终返回最高得分所对应的团体结构。

5.3 评价指标

本节将从 4 类评价指标综合评估模型的实验效果，指标定义及模型效果如下。

1）标准化互信息指标（Normalized Mutual Information，NMI）

此指标适用于已知划分标准的人工合成网络社区，可以很好地反映网络中重叠社区的划分准确性，NMI 值越大表示划分准确率越高，其定义标准如下：

$$I(A,B)=\frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B}N_{ij}\log\left(\frac{N_{ij}\times N}{N_i\times N_{\cdot j}}\right)}{\sum_{i=1}^{C_A}N_{i\cdot}\cdot\log\left(\frac{N_{i\cdot}}{N}\right)+\sum_{j=1}^{C_B}N_{\cdot j}\log\left(\frac{N_{\cdot j}}{N}\right)} \tag{10}$$

2）调整兰德系数（Adjusted Rand index，ARI）

ARI 是对兰德系数（RI）的调整与改进，反映两个数据分布的吻合情况，该系数的取值范围通常为[−1，1]，当系数的取值越接近于 1，说明该模型的聚类效果越好，其具体的定义如下：

$$ARI=\frac{\sum_{ij}\binom{n_{ij}}{2}-\frac{\left[\sum_i\binom{b_i}{2}\sum_j\binom{d_j}{2}\right]}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i\binom{b_i}{2}+\sum_j\binom{d_j}{2}\right]-\frac{\left[\sum_i\binom{b_i}{2}\sum_j\binom{d_j}{2}\right]}{\binom{n}{2}}} \tag{11}$$

3）Omega Index

Cordasco 等[28]提出该指标旨在比较不相交的聚类任务，如果算法模型都将两个节点对象聚到同一个社区中，或每个模型都将其放入不同的社区中，则称这两个模型在一对对象上达成一致。该指标取值范围为[0，1]，数值越大说明社区划分效果越好，具体的计算式如下：

$$Omega(s1,s2)=\frac{Obs(s1,s2)-Exp(s1,s2)}{1-Exp(s1,s2)} \tag{12}$$

4）综合指标 F-score

F-score 指标可以较好地描述重叠社区中节点检测的综合确定性，引入 F-score 指标可以平衡准确率和召回率的影响，可以较为全面地评价一种算法模型，其定义如下：

$$F\text{-}\ \text{score}=\frac{2\times\ \text{recall}\ \times\ \text{precision}}{\text{precision}\ +\ \text{recall}} \tag{13}$$

5.4 结果及分析

按上述实验设置，本节分别用两个数据集进行模型的训练，训练集与测试集的划分比例均为 8∶2，评估结果均采用 10 折交叉验证。针对两种数据集上的评估结果，分别使用 NMI，ARI，F-score 和 Omega 4 种评估指标在测试集上进行对比分析，具体结果如下。

1）准确性

由表 3 所列的实验结果可以看出，BKH-II 模型相比其他 6 种基准模型在 4 类评价指标上的表现均表现出了更高的准确度，分别为：$NMI=0.983$，$NRI=0.96$，$F\text{-}score=0.943$，$Omega=0.95$，其在不同指标下准确度的变化范围保持在 4.2%之内，即分数表现具有一定的稳定性，而同比经典算法 infomap 在不同指标下准确度的变化范围达到 6.7%。此外，BKH-II 模型在 4 种指标下较基准模型中的最佳表现分别提升约 9.7%，14.8%，0.5%，13.6%，最大提升幅度为 NRI 评估指标下的 14.8%，因此可以得出结论，本文提出的 BKH-II 模型在准确度上较以往模型有了较大的提升。

表 3 BKH-II 模型的实验结果

Table 3 Experimental results of BKH-II model

Algorithm	NMI	Time	NRI	Time	FI	Time	Omega	Time
BKH-II	0.983	0.0730	0.960	0.0738	0.943	0.0899	0.950	0.0778
Walktrap	0.896	0.2066	0.836	0.2089	0.938	0.2085	0.836	0.1937
Surprise_communities	0.838	0.1039	0.672	0.1030	0.869	0.1195	0.672	0.1082
Significance_communities	0.771	0.1386	0.528	0.1370	0.775	0.1227	0.513	0.1102
Infomap	0.692	0.1435	0.673	0.1302	0.718	0.1290	0.673	0.1306
Greedy_modularity	0.458	5.8245	0.218	6.0858	0.327	5.8346	0.218	5.9043
Label_propagation	0.070	0.0719	0.030	0.0748	0.143	0.0734	0.030	0.0771

2）泛化性

通过调整 LFR 网络中的混合参数 μ（设置为 0～0.6）来阶梯化地对比不同模型的评估效果，其中混合参数 μ 越大，表示网络结构越混乱，即对算法模型的泛化性能的挑战越大。Zhang 等[29]的实验结果证明，随着参数 μ 的增加，模型的准确性将逐步下降，通过对 4 个评估指标下各模型的表现可以看出在（0～0.6）范围内 0.3，0.4，0.5 的参数设置对部分模型的影响较大，其中 Label propagation 和 Greedy_modularity

两种模型随参数的上升，准确度的下降幅度越大，下降幅度分别达(20.4%，15%，16.7%)和(5.2%，17.6%，73.3%)，与 $\mu=0.5$ 处 Walktrap，Surprise_communities，Significance_communities，Infomap 和 BKH-II 均发生了较大幅度的准确率下降，但通过图 9—图 12 的曲线对比可以看出，模型 BKH-II 在随混合参数 μ 不断增大的同时，其准确率损失比例最小，且始终保持了相比其他模型更高的准确度得分，这说明该模型的泛化性能明显强于其他基准模型。

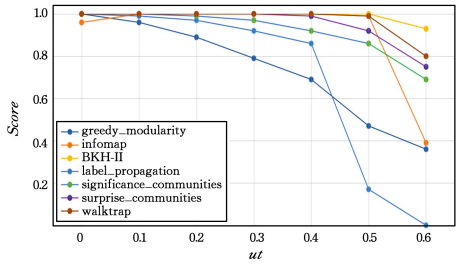


图 9 NMI 指标结果

Fig. 9 Results of NMI indicator

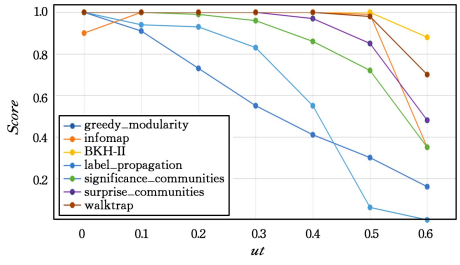


图 10 ARI 指标结果

Fig. 10 Results of ARI indicator

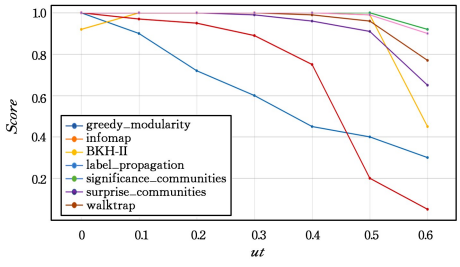


图 11 F-score 指标结果

Fig. 11 Results of F-score indicator

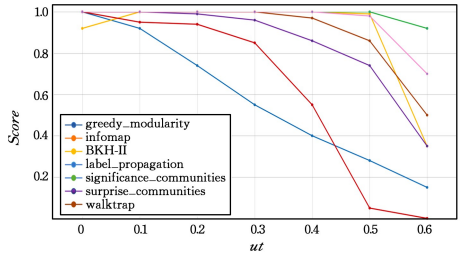


图 12 Omega 指标结果

Fig. 12 Results of Omega indicator

3)算法运行时间

本节对各算法的计算时间做了对比，以此来体现算法的时间复杂度和对计算资源的利用程度，同样采用了 LFR 人工合成网络作为测试数据集，设置节点数 N 为 1000，size=big，混合参数 $\mu=0.3$，生成的图示例为 n_1k_sz_big，同样对每个图示例随机改变其结构生成 lfr_ut_0.0_rep_0 ～ lfr_ut_0.0_rep_9 不同的图并对其做计算，最后取平均值作为模型的最终时间。

从表 4 可以看出，本文所提出的 BKH-II 模型具有较低的运行时间，Total time=17 s，综合准确度得分发现，label_propagation 模型虽具有最短的运行时间，但其准确度较低。同时 BKH-II 模型与 surprise_communities，infomap 模型具有相近的运行时间，但准确度明显高于基准模型，因此可以得出结论：本文提出的 BKH-II 模型的运行时间较短，即具有更低的时间复杂度。

表 4 BKH-II 模型的运行时间

Tabel 4 Running time of BKH-II model

Algorithm	N	SZ	Graph	Total time
BKH-II	1000	Big	110	0 min 17 sec
Walktrap	1000	Big	110	0 min 26 sec
Surprise_communities	1000	Big	110	0 min 17 sec
Significance_communities	1000	Big	110	0 min 21 sec
Infomap	1000	Big	110	0 min 19 sec
Greedy_modularity	1000	Big	110	3 min 55 sec
Label_propagation	1000	Big	110	0 min 14 sec

结束语 信贷行业的欺诈问题正随着数字化场景的升级，显示出动态化、隐蔽性和专业化的新特点，信贷数据也正向多模态、海量和异构的方向发展，这些改变将对已有的反欺诈方法提出了新的挑战。本文针对欺诈行为发生主体及其行为特点，构建了信贷异构信息网络，将群体欺诈问题表示为网络结构中特异群组问题，提出了基于特异群组的异常群体检测算法 BKH-II，通过 node2vec 算法获得图嵌入向量表示，并采用两阶段基于 H 图的 MCE 算法对可疑群组进行挖掘，进而通过局部特征工程所选择出的特征对候选群组进行修正划分，最终将特异群组区分为偶发群组和可疑群组，解决了对具有相似结构的可疑特异群组的挖掘，以及对因偶发或周期因素而碰巧具有高相似性的人群难以区分的问题。经实验结果证明，该模型在准确性、泛化性和运行时间上均明显优于基准模型。此外，异构信息网络还具备良好的可视化特性，本文在图数据结构分析和模型效果展示等环节进行了详细的网络可视化，有助于了解数据和模型的训练过程，增强训练结果的可解释性。

通过对图结构和信贷欺诈问题的进一步深挖，未来该领域内的研究具有以下未来方向：

1)深度学习算法可利用海量的数据搭建隐层的机器学习模型，通过多层网络对特征进行逐层变换，进而可以自动对相关数据的内部信息进行学习和构造，从而提升预测的准确性。

2)网络结构的演变将对相关网络特性产生较大的影响，如在重叠社区的划分中，动态演变的网络将会产生相关社区的分裂和融合，故针对动态网络中的社区划分研究将对真实世界的建模分析具有重要意义。

3)图神经网络的内存计算通常涉及对内存的频繁访问与巨大的资源开销，且由于密度和容量的增加，将导致内存组件通常会随着时间的推移而变得容易出错，因此诸如 Gan 等[30]的工作，开发高可用性内存访问的解决方案，如为各个级别设置不同的内存区域粒度至关重要。

参 考 文 献

[1] SHEN H, CHENG X, CAI K, et al. Detect overlapping and hierarchical community structure in networks[J]. Physica A: Statistical Mechanics and its Applications, 2009, 388(8): 1706-1712.

[2] LU Q, JU C. Research on Credit Card Fraud Detection Model Based on Class Weighted Support Vector Machine[J]. Journal of Convergence Information Technology, 2011, 6(1): 62-68.

[3] NIAN K, ZHANG H, TAYAL A, et al. Auto insurance fraud detection using unsupervised spectralranking for anomaly[J]. The Journal of Finance and Data Science, 2016, 2(1): 58-75.

[4] KELLER F, MULLER E, BOHM K. HiCS: High Contrast Subspaces for Density-Based Outlier Ranking[C]// 2012 IEEE 28th International Conference on Data Engineering. Arlington, VA, USA, 2012: 1037-1048.

[5] NGUYEN H V, GOPALKRISHNAN V, ASSENT I. An unbiased distance-based outlier detection approach for high-dimensional data[C]// International Conference on Database Systems for Advanced Applications. Springer, Berlin, Heidelberg, 2011: 138-152.

[6] BHATTACHARYYA S, JHA S, THARAKUNNEL K, et al. Data mining for credit card fraud: A comparative study[J]. Decision Support Systems, 2011, 50(3): 602-613.

[7] YAN H, JIANG Y, LIU G. Telecomm Fraud Detection via Attributed Bipartite Network[C]// 2018 15th International Conference on Service Systems and Service Management(ICSSSM). IEEE, 2018: 1-6.

[8] HE Y, SONG Y, LI J, et al. Hetespaceywalk: A heterogeneous spacey random walk for heterogeneous information network embedding[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 639-648.

[9] XIE Y, WANG Q, LI H H, et al. A Spatial-Temporal Graph Mining Algorithm based on Spatial-Temporal Sparse Attention Network[J/OL]. Computer Engineering: 1-8. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7ioT0BO4yQ4m_mOgeS2ml3UBXRVwfAzJsIi3Bsco8ZcxE_qZ4dUCMa-vfFpV0QH1EQ&uniplatform=NZKPT.

[10] ROY A, SUN J, MAHONEY R, et al. Deep learning detecting fraud in credit card transactions[C]// 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE, 2018: 129-134.

[11] CAI H Y, YUAN S L, WEN Y, et al. Shilling Attacks Detection Based on CNN and Hesitant Fuzzy Sets[J]. Engineering Science and Technology, 2022, 54(3): 80-90.

[12] WANG W M, ZHI L P. Fraud detection model generalization performance improvement and interpretability study based on ADASYN-SFS-RF[J/OL]. Computer application research: 1-11. https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAiTRKibYlV5Vjs7ioT0BO4yQ4m_mOgeS2ml3UH_TYrVERVm5vryPl24sxfOcsIMT5f6OT61zKGh0xRSr&uniplatform=NZKPT.

[13] SHI C, LI Y, ZHANG J, et al. A survey of heterogeneous information network analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 29(1): 17-37.

[14] LUSSEAU D, NEWMAN M E J. Identifying the role that animals play in their social networks[J]. Proceedings of the Royal Society of London. Series B: Biological Sciences, 2004, 271(suppl_6): S477-S481.

[15] ZHAO Y Q, WU Y, CHEN X. An Algorithm for Large-scale Social Network Community Detection and Visualization[J]. Journal of Computer-Aided Design and Graphics, 2017, 29(2): 328-336.

[16] GREGORY S. An algorithm to find overlapping community structure in networks[C]// European conference on principles of data mining and knowledge discovery. Springer, Berlin, Heidelberg, 2007: 91-102.

[17] SHEN H W, CHENG X Q, GUO J F. Quantifying and identifying the overlapping community structure in networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2009, 2009(7): P07042.

[18] QIAN Y, LI Y, ZHANG M, et al. Quantifying edge significance on maintaining global connectivity[J]. Scientific Reports, 2017, 7(1): 1-13.

[19] LIU H, FEN L, JIAN J, et al. Overlapping community discovery algorithm based on hierarchical agglomerative clustering[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2018, 32(3): 1850008.

[20] LI J, LI X, GAO Y, et al. Dynamic trustworthiness overlapping community discovery in mobile internet of things[J]. IEEE Access, 2018, 6: 74579-74597.

[21] HUANG F, LI X, ZHANG S, et al. Overlapping community detection for multimedia social networks[J]. IEEE Transactions on multimedia, 2017, 19(8): 1881-1893.

[22] YOSHIDA T. Weighted line graphs for overlapping community discovery[J]. Social Network Analysis and Mining, 2013, 3(4): 1001-1013.

[23] HUANG F L, ZHANG S C, ZHU X F. Discovering Network Community Based on Multi-Objective Optimization[J]. Software Journal, 2013, 24(9): 2062-2077.

[24] XIE J, SZYMANSKI B K. Community detection using a neighborhood strength driven label propagation algorithm[C]// 2011 IEEE Network Science Workshop. IEEE, 2011: 188-195.

[25] GARZA S E, SCHAEFFER S E. Community detection with the Label Propagation Algorithm: A survey[J]. Physica A: Statistical Mechanics and its Applications, 2019, 534: 122058.

[26] HU J, DONG Y H, YANG B R. Community structure discovery algorithms in large complex networks [J]. Computer Engineering, 2008 (19): 92-93, 100.

[27] ZHANG X K, REN J, SONG C, et al. Label propagation algorithm for community detection based on node importance and label influence[J]. Physics Letters A, 2017, 381(33): 2691-2698.

[28] CORDASCO G, GARGANO L. Label propagation algorithm: a semi-synchronous approach[J]. International Journal of Social Network Mining, 2012, 1(1): 3-26.

[29] ZHANG Y L, XIA X W, XU X, et al. Review on Label Propagation Algorithms for Community Detection[J]. Small Microcomputer System, 2021, 42(5): 1093-1102.

[30] GAN C, WANG B, WANG Z J, et al. A Solution for High Availability Memory Access[C]// 19th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2019). 2019.

LIU Hualing, born in 1964, Ph.D, professor. Her main research interests include financial risk control, data mining and intelligent decision-making.