



计算机科学

COMPUTER SCIENCE

STWD-DLFRD:基于序贯三支决策与深度学习的多粒度虚假评论检测方法

辜波凯, 刘盾, 孙扬

引用本文

辜波凯, 刘盾, 孙扬. [STWD-DLFRD:基于序贯三支决策与深度学习的多粒度虚假评论检测方法](#)[J]. 计算机科学, 2026, 53(4): 188-196.

GU Bokai, LIU Dun, SUN Yang. [STWD-DLFRD:Multi-granularity Fake Review Detection via Sequential Three-way Decisions and Deep Learning](#) [J]. Computer Science, 2026, 53(4): 188-196.

相似文献推荐 (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)

[跨模态融合的少样本勒索软件分类器:基于预训练模型的多模态编码](#)

Cross-modal Fusion Few-sample Ransomware Classifier:Multimodal Encoding Based on Pre-trained Models

计算机科学, 2026, 53(4): 435-444. <https://doi.org/10.11896/jsjcx.250500078>

[知识辅助和强化句法驱动的方面级情感分析](#)

Knowledge-assisted and Reinforced Syntax-driven for Aspect-based Sentiment Analysis

计算机科学, 2026, 53(4): 406-414. <https://doi.org/10.11896/jsjcx.250600117>

[基于深度学习和多特征融合的时序社交网络关键节点识别](#)

Key Node Identification in Temporal Social Networks Based on Deep Learning and Multi-feature Fusion

计算机科学, 2026, 53(4): 143-154. <https://doi.org/10.11896/jsjcx.250300147>

[KGMamba:基于Kolmogorov-Arnold网络优化图卷积网络和Mamba的基因调控网络预测模型](#)

KGMamba:Gene Regulatory Network Prediction Model Based on Kolmogorov-Arnold Network Optimizing Graph Convolutional Network and Mamba

计算机科学, 2026, 53(4): 101-111. <https://doi.org/10.11896/jsjcx.250500097>

[基于深度学习的GIFT-128与ASCON算法神经差分区分器研究](#)

Deep Learning-based Neural Differential Distinguishers for GIFT-128 and ASCON

计算机科学, 2026, 53(3): 453-458. <https://doi.org/10.11896/jsjcx.250600176>

STWD-DLFRD: 基于序贯三支决策与深度学习的多粒度虚假评论检测方法

辜波凯 刘盾 孙扬

西南交通大学经济管理学院 成都 610031

摘要 随着在线评论对消费者决策的影响日益增强,虚假评论的检测成为保障电商平台生态健康的重要任务。现有方法多采用静态单步检测,忽视了动态特征与决策成本,导致检测效率不佳。为此,提出一种基于序贯三支决策(Sequential Three-Way Decisions)与深度学习(Deep Learning)的多粒度虚假评论检测(Fake Review Detection)方法(STWD-DLFRD)。该框架通过深度学习技术提取评论的文本、行为及社交关系特征,构建多粒度特征空间,并利用序贯三支的分层决策机制实现对不同复杂度虚假评论的动态检测。实验结果表明,与基线模型相比,STWD-DLFRD在F1值和准确率上表现最优,总分类代价显著降低。所提方法为动态环境下高成本敏感的虚假评论检测提供了一种有效的解决方案。

关键词: 虚假评论检测; 序贯三支决策; 深度学习; 多粒度特征; 决策成本

中图分类号 TP183

STWD-DLFRD: Multi-granularity Fake Review Detection via Sequential Three-way Decisions and Deep Learning

GU Bokai, LIU Dun and SUN Yang

School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China

Abstract With the increasing influence of online reviews on consumer decision-making, fake review detection has become a critical task for maintaining the integrity of e-commerce ecosystems. Existing methods predominantly rely on static single-step detection, which overlooks dynamic feature evolution and cost-sensitive decision-making, leading to suboptimal efficiency. To address these limitations, this paper proposes a sequential three-way decisions and deep learning-based multi-granularity fake review detection framework(STWD-DLFRD). The framework constructs a multi-granularity feature space by extracting textual, behavioral and social relationship features from reviews. Leveraging a hierarchical decision mechanism of sequential three-way decisions, it dynamically identifies fake reviews with varying complexity levels. Experimental results demonstrate that STWD-DLFRD outperforms baseline models in both F1-score and accuracy, while significantly reducing total classification costs. This study provides an effective solution for high-cost-sensitive fake review detection in dynamic environments, balancing detection precision and decision efficiency through adaptive granularity refinement.

Keywords Fake review detection, Sequential three-way decisions, Deep learning, Multi-granularity features, Decision cost

1 引言

随着数字经济的快速发展,在线评论已成为消费者进行决策的重要参考依据。然而,近年来大量的虚假评论开始涌入电商平台,这些评论通过传递不实的信息来误导消费者,不仅严重损害了消费者的利益,而且对电商平台的信誉造成了极大的影响。为了遏制这一现象的持续扩大,不少针对虚假评论的检测方法被提出,但随着虚假评论伪装手段的不断演变升级,检测虚假评论这一任务变得更加复杂和具有挑战性^[1]。

现有针对虚假评论的检测方法可大致分为3类:基于文本特征的检测、基于用户行为特征的检测,以及基于社交网络

特征的检测^[2]。其中,基于文本特征的检测是虚假评论检测中最常见的方法。作为虚假评论的主要载体,虚假评论在文本长度、文本传递的情感(正面与负面观点的占比)等方面与真实评论区别较大,且其文本质量总体更差,常包含过度的大写字母及数字^[3]。此外,虚假评论的文本语法、文本词汇、文本主题等特征也与真实评论存在一定差异^[4-6]。除了文本特征以外,虚假评论用户的行为特征也在虚假评论检测机制中起着重要作用。例如,如果一个新注册用户短期内发布大量虚假评论,那么它就有较大概率是一个虚假账户^[7-8]。其次,真实用户与虚假用户在评论评分、历史发布内容一致性以及踩赞比例等方面都有显著差异^[9-10]。此外,亦有研究通过群组行为特征识别虚假评论^[11-12]。相关研究表明:利用群体

到稿日期:2025-05-22 返修日期:2025-07-21

基金项目:国家自然科学基金(62276217,62402424);四川省自然科学基金(2026NSFSC0445);中央高校基本科研业务费项目(2682024KJ005,2682024ZTPY021)

This work was supported by the National Natural Science Foundation of China(62276217,62402424), Natural Science Foundation of Sichuan Province(2026NSFSC0445) and Fundamental Research Funds for the Central Universities(2682024KJ005,2682024ZTPY021).

通信作者:刘盾(newton83@163.com)

进行协同作案的虚假评论社区,其社区成员间的信息传播模式、交互关系、社交网络等特征都与真实用户有较大差异^[13-14]。近年来,国内外学者利用深度学习技术对这些特征进行了有效检测。Shang等^[15]在BERT模型的基础上对文本附加额外情感信息来识别虚假评论。Bhuvaneshwari等^[16]利用CNN和Bi-LSTM学习评论的文档级表征来进行虚假评论检测。Mukherjee等^[17]在Yelp数据集上分析语言与行为特征的有效性,发现语言特征效果有限,并推测Yelp的过滤机制可能依赖行为特征。Tang等^[18]挑选了6个用于表征用户行为的特征,搭建了一个同时包含判别器和生成器的GAN框架来识别虚假评论。Manaskasemsak等^[19]提出了一种构建评论者行为图的方法,通过识别用户的可疑行为模式判断其评论的真伪。Cheng等^[20]构建了用户-用户图以及用户-评论图,并提出了基于图神经网络的虚假评论者检测框架。He等^[21]提出了一种图注意力深度融合网络,同时从元数据和关系数据中提取信息来进行虚假评论的识别。Li等^[22]通过同构图和异构图来分别捕获评论的局部信息和全局信息,提出了基于图卷积网络的虚假评论检测模型。

通过对已有文献的分析可知,尽管现有方法在特定场景下表现良好,但仍然存在一定的局限性。一方面,已有的检测方法大多聚焦于静态检测,主要通过单步检测识别虚假评论,忽略了检测过程中的动态特征。另一方面,现有研究大多以检测精度为目标,忽略了检测过程中的决策代价问题。然而,在面对真实决策场景时,决策代价也是需要考虑的重要因素。具体而言,在虚假评论检测中,对于一些简单的虚假评论,可直接通过文本异常进行精确识别,而对一些进行了文本伪装的复杂虚假评论,则需要依赖更多的行为及关系信息才能识别,这会导致相应的检测成本增加^[23-25]。因此,对不同粒度的评论进行动态检测,判断是否需要获取更多决策信息是十分必要的。作为一种动态的、多层次的和多粒度的决策方法,序贯三支决策(Sequential Three-way Decision, S3WD)适用于动态不确定性问题的求解^[26],其在提升决策过程的可解释性和降低决策代价方面均表现出良好的性能。Zhang等^[27]利用 n -gram语言模型构建了多粒度的序贯三支情感分类模型,提高了分类准确率。Chen等^[28]为解决欺诈识别问题中检测代价不敏感的问题,提出了基于生成式对抗网络的序贯三支异常用户检测方法。Li等^[29]提出了一种代价敏感的序贯三支决策方法进行人脸识别,减少了分类误差。因此,为解决虚假评论检测中的动态性和决策代价问题,本文拟引入序贯三支决策方法。

基于以上分析,本文提出了一种基于序贯三支决策和深度学习技术的虚假评论检测方法(STWD-DLFRD)。首先,利用多关系图与BERT,LSTM,CNN以及GAT的耦合来提取各方面的评论特征。其次,借助序贯三支决策的分层决策特性,构造了多层次、多粒度的动态决策框架。最后,利用真实评论数据在多个指标下对多个模型进行对比分析,验证了本文方法的有效性。

2 相关概念

2.1 序贯三支决策

三支决策(Three-way decision, 3WD)是Yao等提出的一

种基于认知科学的不确定决策理论^[30],其核心思想是借鉴人类思维方式和行为模式,通过分治模型和序贯策略来解决复杂决策问题。它通过Trisecting-Acting-Outcome(TAO)模型(也称“分治效”模型),将整体(论域)分为3个部分,并对不同部分采取不同的决策行为或分治策略,最后对相应的结果进行评价和反馈。其框架图如图1所示。

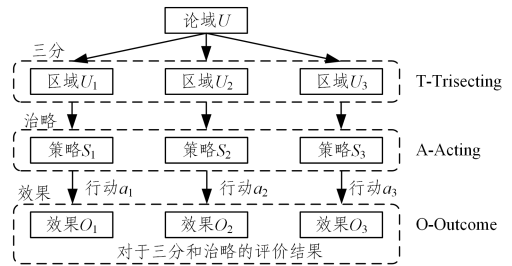


图1 三支决策框架图

Fig. 1 Framework of three-way decision

序贯三支决策理论是基于三支决策理论提出的。它利用“分而治之”的思想,针对目标对象构建由粗到细的多层次、多粒度结构,并通过一系列三支决策过程来实现对目标的动态分类。借鉴这一思想,人们在决策过程中,可以先根据少量信息对容易分类的样本做出决策,待信息充足之后,再对剩余样本进行决策。此外,该方法通过在决策过程中平衡测试成本与误分类成本,最终实现最小化总决策成本的目标^[31]。

2.2 多关系图

多关系图是一种特殊的异构图^[32]。如图2所示,不同颜色的向量表示不同评论所具有的不同特征。异构图通常包含多种类型的节点和边,而多关系图虽然只包含单一类型的节点,但节点间可以通过不同的关系进行连接,不同的关系又包含不同的语义信息,这使得多关系图具备异图图的性质。因此,大多数用于处理异图图的方法对多关系图同样适用^[33]。

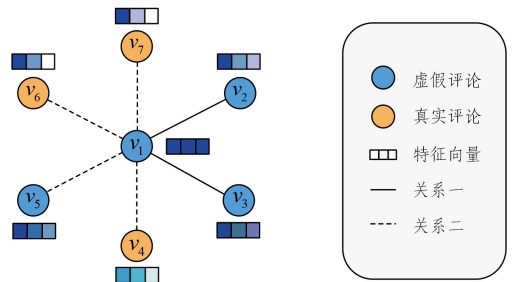


图2 多关系图基本结构框架

Fig. 2 Structure of multi-relationship diagram

为了方便起见,本文将多关系图表示为一个多元数据组:

$$G = (\mathbf{V}, \mathbf{X}, \{\mathbf{E}_r\}_{r=1}^R, \mathbf{Y}) \quad (1)$$

其中, $\mathbf{V} = \{v_1, \dots, v_n\}$ 表示节点(评论)集合; $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ 为节点特征集合; $\mathbf{x}_v \in \mathbf{R}^d$ 表示节点 v_i 的 d 维特征向量; \mathbf{E}_r 表示第 r 类关系的边集合; $e_{i,j} = (v_i, v_j) \in \mathbf{E}_r$ 表示节点间存在类型为 $r \in \{1, \dots, R\}$ 的关系; \mathbf{Y} 为节点标签集合,每个节点 v 具有二元标签 $y_v \in \{0, 1\}$ 。

2.3 深度学习技术

随着深度学习技术的兴起,BERT,LSTM,CNN和GAT等方法被广泛应用于虚假评论的检测中。BERT是一种以Transformer中Encoder结构为基础的深度双向语言表征

模型,其被证明在捕获文本特征方面是有效的^[34]。LSTM 在循环神经网络(RNN)的基础上引入“门”机制,使其能够处理长距离依赖问题,被广泛应用于虚假评论检测领域^[35]。卷积神经网络(CNN)通过局部感受野、权值共享、池化等操作,有效减少了网络的参数,缓解了模型的过拟合问题。CNN 拥有出色的捕捉邻居信息的能力,在文本处理和虚假评论检测领域展现出良好的效果^[36]。此外,图神经网络(GNN)因其在图结构数据上提取信息的出色能力,被广泛应用于虚假评论识别领域^[37]。其中,图卷积网络(GCN)通过卷积操作聚集目标节点的邻居节点特征,实现目标节点的特征更新^[38]。而图注意力网络(Graph Attention Network,GAT)在传统 GCN 的基础上引入了注意力机制,通过计算注意力系数来为每个邻居节点分配不同的权重,能有效检测出虚假评论^[39]。GAT 的基本结构框架如图 3 所示。

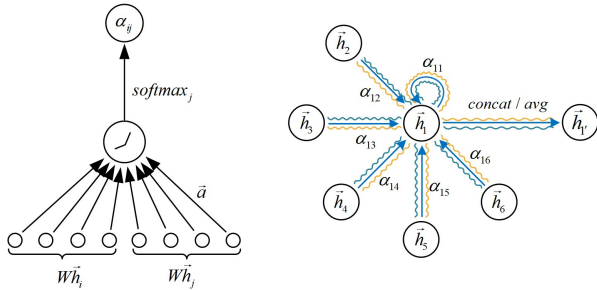


图 3 GAT 的基本结构框架

Fig. 3 Basic structural framework of GAT

在图 3 中,GAT 在多关系图中进行特征聚集的过程分为两步。首先,基于每种关系 r 对节点进行一次聚集,GAT 通过计算注意力系数和归一化权重实现关系内聚集。

注意力系数的计算式为:

$$e_{ij}^{(l)} = \text{LeakyReLU}(\mathbf{a}^T \cdot [\mathbf{W}_r^{(l)} \cdot \mathbf{x}_{v_i}^{(l-1)}, \mathbf{W}_r^l \cdot \mathbf{x}_{v_j}^{(l-1)}]) \quad (2)$$

归一化权重的计算式为:

$$\alpha_{ij}^{(l)} = \frac{\exp(e_{ij}^{(l)})}{\sum_{k \in N_i} \exp(e_{ik}^{(l)})} \quad (3)$$

关系内聚集的计算式为:

$$\mathbf{z}_{r,v_i}^{(l)} = \text{ReLU}(\sum_{v_j \in N_i^{(l)}(v_i)} \alpha_{ij}^{(l)} \cdot \mathbf{W}_r^{(l)} \cdot \mathbf{x}_{v_j}^{(l-1)}) \quad (4)$$

其次,GAT 将节点在不同关系上的特征融合起来,再与该节点本身的特征进行拼接,形成下一层的节点特征。特征关系内聚集过程如下:

$$\mathbf{x}_{v_i}^{(l)} = \text{ReLU}(\mathbf{W}_{\text{comb}}^{(l)} [\mathbf{h}_{v_i}^{(l-1)} \oplus \sum_{r=1}^R \mathbf{z}_{r,v_i}^{(l)}]) \quad (5)$$

其中, $\mathbf{x}_{v_i}^{(l-1)}$ 和 $\mathbf{x}_{v_j}^{(l-1)}$ 分别为目标节点 v_i 和邻居节点 v_j 在图神经网络第 $l-1$ 层上的特征, $\mathbf{a}^T \in \mathbf{R}^{2d^{(l-1)}}$ 为权重向量, $N_r^{(l)}(v_i)$ 为节点 v_i 在图神经网络的第 l 层上基于关系 r 的邻居节点集合, $\mathbf{z}_{r,v_i}^{(l)}$ 为图神经网络第 l 层上基于关系 r 聚集的节点 v_i 的邻居节点特征, $\mathbf{W}_r^{(l)}$ 和 $\mathbf{W}_{\text{comb}}^{(l)}$ 为权重矩阵, \oplus 表示特征拼接操作。此外,还可以引入多头注意力机制,同时计算 k 个独立的注意力权重向量。

3 基于序贯三支决策与深度学习的多粒度虚假评论检测

本章中提出了一种基于序贯三支决策和深度学习技术的虚假评论检测方法,如图 4 所示。一方面,利用多种深度学习技术提取了不同方面的评论信息,构建出多层次多粒度的特征空间。另一方面,利用序贯三支决策的分层决策机制对不同特征空间内的评论实现渐进式决策。

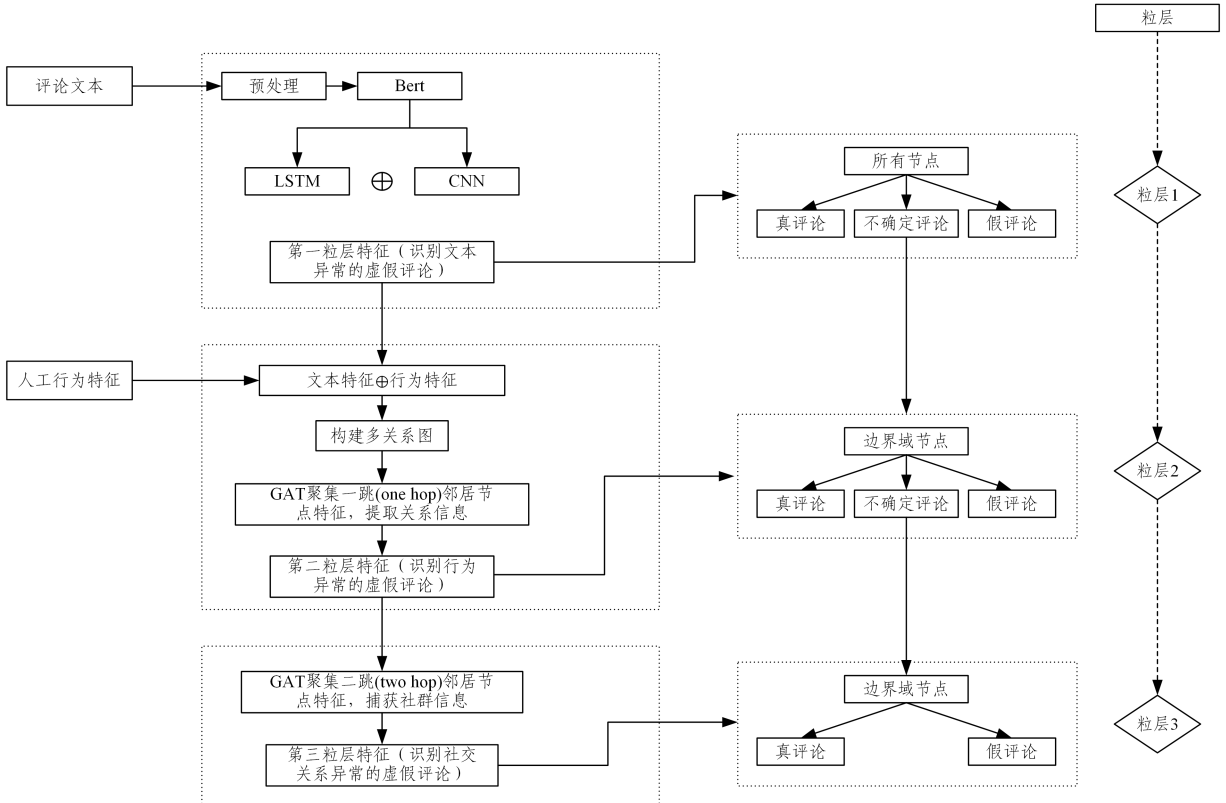


图 4 序贯三支虚假评论检测框架图

Fig. 4 Framework diagram of sequential three-branch fake review detection

3.1 基于深度学习的特征空间构造

一般而言,虚假评论的异常主要体现在3个方面:文本异常、行为异常,以及社交关系异常。本研究的粒层空间构造就是在此基础上进行的。为了方便起见,令粒层空间集合为 $\mathbf{Gr} = \{\mathbf{Gr}_1, \mathbf{Gr}_2, \mathbf{Gr}_3\}$, 评论集合为 $\mathbf{V} = \{v_1, \dots, v_n\}$ 。

3.1.1 第一粒层空间构造

由于粗粒度阶段的虚假评论仅通过文本便能有效识别,因此,第一粒层捕获评论的文本特征,用于识别文本异常的评论。为了提取到最有价值的特征空间,选择了几种最常见的深度学习技术,即BERT, LSTM和CNN,来进行文本特征的提取。具体而言,先用BERT获取文本表征,再分别利用LSTM捕获上下文的能力以及CNN捕获局部信息的能力来提取更深层次的文本信息,最后将两者进行拼接,得到第一粒层 \mathbf{Gr}_1 的评论特征:

$$h_{v_i}^{(0)} = \mathbf{W}(\text{CNN}(\text{BERT}(\text{text})) \oplus \text{LSTM}(\text{BERT}(\text{text}))) \quad (6)$$

其中, $h_{v_i}^{(0)}$ 为评论 v_i 在粒层空间 \mathbf{Gr}_0 的特征, \mathbf{W} 为权重矩阵, \oplus 为拼接操作。

3.1.2 第二粒层空间构造

考虑到经过文本伪装的评论很难仅通过文本特征检测出异常,需要进一步结合用户的异常行为来进行识别,在第一粒层文本特征的基础上加入了部分人工选定的行为特征。同时,由于大多数异常行为无法凭借单条评论进行识别(如用户历史行为异常),通常需要多条评论才能识别。因此,为了更好地检测用户的异常行为,构建了一个基于评论的多关系图来捕获各评论间的关系,用于辅助行为异常的检测。

这里,将引入的人工行为特征 f_{v_i} 与上一粒层的评论特征 $h_{v_i}^{(0)}$ 拼接起来作为多关系图的初始节点特征:

$$x_{v_i}^{(0)} = \mathbf{W}(h_{v_i}^{(0)} \oplus f_{v_i}) \quad (7)$$

其中, $x_{v_i}^{(0)}$ 为节点 v_i 在图神经网络第0层(初始)上的特征, f_{v_i} 为人工选定的节点 v_i 的行为特征, \mathbf{W} 为权重矩阵。

接着,用图注意力网络在多关系图上捕获评论间的关系信息,构造出新的特征空间,用于识别行为异常虚假评论。具体而言,初始节点特征 $x_{v_i}^{(0)}$ 在经过第一层特征聚集之后便能捕捉到一跳(One Hop)邻居节点的信息,来构造第二粒层的特征空间 \mathbf{Gr}_2 。

$$h_{v_i}^{(1)} = x_{v_i}^{(1)} = \text{GAT}(x_{v_i}^{(0)}) \quad (8)$$

3.1.3 第三粒层空间构造

为了扩大虚假评论的影响力,虚假评论者会组织社群进行协同作案,而群体作案通常涉及多位用户和多条评论。随着在多关系图上的每一次特征聚集,评论节点所包含的人际关系和交互信息也会增多。因此,为了构造能够识别社交关系异常评论的特征空间,在第3.1.2节的基础上通过图神经网络再一次进行特征聚集,用于捕捉二跳(Two Hop)邻居的特征,构造出第三粒层的特征空间 \mathbf{Gr}_3 :

$$h_{v_i}^{(2)} = x_{v_i}^{(2)} = \text{GAT}(x_{v_i}^{(1)}) \quad (9)$$

通过式(9)的二跳聚集,节点包含了来自多条评论的信息以及它们之间的交互关系,据此便可识别出社交关系异常的虚假评论。

基于上述分析,本文建立了一个具有三层次的特征空间 ($\mathbf{Gr}_1, \mathbf{Gr}_2, \mathbf{Gr}_3$)。其中,通过对评论文本特征的提取,形成最

粗粒度的特征空间 \mathbf{Gr}_1 。这一层用于检测较简单的、仅通过单一文本特征就能识别的虚假评论。接着,通过加入人工选定的行为特征,并在多关系图上进行一跳聚集,构造出更细粒度的特征空间 \mathbf{Gr}_2 。一跳聚集捕获了单个评论与其他评论间的关系,能够很好地反映出用户的异常行为模式,识别出行为异常的虚假评论。最后,通过二跳聚集继续扩大关系的覆盖范围,构造出最细粒度的特征空间 \mathbf{Gr}_3 。二跳聚集捕获了多个评论的信息,利用这些评论间的交互关系便能推断出群体协同作案的情况,识别社交关系异常的虚假评论。

3.2 序贯三支虚假评论检测方法

序贯三支虚假评论检测的核心在于通过深度学习构造的多粒度特征空间,运用“分而治之”的决策思想,实现序贯的虚假评论检测,并依据每一粒层的划分阈值,判断评论是否属于虚假评论。

3.2.1 序贯三支决策过程

根据第2.1节的思想来构建3个粒层的特征空间 $\mathbf{Gr} = \{\mathbf{Gr}_1, \mathbf{Gr}_2, \mathbf{Gr}_3\}$, 其中 \mathbf{Gr}_1 为最粗粒层, \mathbf{Gr}_3 为最细粒层。在虚假评论检测中,评论集合表示为 $\mathbf{V} = \{v_1, \dots, v_n\}$, 状态集 $\Omega = \{\mathbf{X}, \neg \mathbf{X}\}$ 表示评论的两种类型, \mathbf{X} 和 $\neg \mathbf{X}$ 分别表示真实评论(P)和虚假评论(N)。 $\Pr(\mathbf{X}_l | v_i)$ 表示在粒层 \mathbf{Gr}_l 上 v_i 属于真实评论的概率。

具体而言,序贯三支虚假评论检测的实施步骤为:在每个粒层上求出待分类评论属于真实评论的条件概率 $\Pr(\mathbf{X}_l | v_i)$, 与该粒层的阈值对 (α_l, β_l) 进行比较,据此对不同的评论采取不同的决策行为。其中 a_P, a_N 和 a_B 分别表示将评论划分到正域 $\text{POS}_l(\mathbf{X})$ (真实评论)、负域 $\text{NEG}_l(\mathbf{X})$ (虚假评论)和边界域 $\text{BND}_l(\mathbf{X})$ 3种行动。当 $\Pr(\mathbf{X}_l | v_i) \geq \alpha_l$ 时,将 v_i 划分至 $\text{POS}_l(\mathbf{X})$; 当 $\Pr(\mathbf{X}_l | v_i) \leq \beta_l$ 时,将 v_i 划分至 $\text{NEG}_l(\mathbf{X})$; 否则将 v_i 划分至 $\text{BND}_l(\mathbf{X})$ 。

值得一提的是,对于划入边界域中的评论,需要采取延迟决策的方式,即在更细粒层参考更多的特征信息再进行决策。可以看到,从最粗粒层到最细粒层的过程中,边界域中待划分的评论数量会逐渐减少,直至在最后一层进行完全划分。

3.2.2 阈值计算过程

为了方便起见,在表1中列出了三支虚假评论检测的成本代价矩阵。其中, $\lambda_{PP}^l, \lambda_{BP}^l, \lambda_{NP}^l$ 分别表示在粒层 \mathbf{Gr}_l 上,若评论为真实时采取 a_P, a_N, a_B 决策行为的代价; $\lambda_{PN}^l, \lambda_{BN}^l, \lambda_{NN}^l$ 分别表示在粒层 \mathbf{Gr}_l 上,若评论为虚假时采取 a_P, a_N, a_B 决策行为的代价。

表1 决策成本代价矩阵

Table 1 Decision cost matrix

决策行为	评论真实性	
	$\mathbf{X}(\mathbf{P})$	$\neg \mathbf{X}(\mathbf{N})$
a_P	λ_{PP}^l	λ_{PN}^l
a_B	λ_{BP}^l	λ_{BN}^l
a_N	λ_{NP}^l	λ_{NN}^l

基于此,在第 l 粒层采取不同决策行为的期望成本可表示为:

$$\text{cost}(a_P | v_i) = \lambda_{PP}^l \Pr(\mathbf{X} | v_i) + \lambda_{PN}^l \Pr(\neg \mathbf{X} | v_i) \quad (10)$$

$$\text{cost}(a_B | v_i) = \lambda_{BP}^l \Pr(\mathbf{X} | v_i) + \lambda_{BN}^l \Pr(\neg \mathbf{X} | v_i) \quad (11)$$

$$\text{cost}(a_N | v_i) = \lambda_{NP}^l \Pr(\mathbf{X} | v_i) + \lambda_{NN}^l \Pr(\neg \mathbf{X} | v_i) \quad (12)$$

根据贝叶斯决策理论的期望损失最小准则,选择期望成本最小的行动作为最佳决策方案,可以得到如下决策规则:

- 1) 如果 $\text{cost}(a_P | v_i) \leq \text{cost}(a_B | v_i)$ 且 $\text{cost}(a_P | v_i) \leq \text{cost}(a_N | v_i)$, 则采取行动 a_P 将评论划分到正域;
- 2) 如果 $\text{cost}(a_B | v_i) \leq \text{cost}(a_P | v_i)$ 且 $\text{cost}(a_B | v_i) \leq \text{cost}(a_N | v_i)$, 则采取行动 a_B 将评论划分到边界域;
- 3) 如果 $\text{cost}(a_N | v_i) \leq \text{cost}(a_P | v_i)$ 且 $\text{cost}(a_N | v_i) \leq \text{cost}(a_B | v_i)$, 则采取行动 a_N 将评论划分到负域。

此外,在虚假评论检测过程中,对评论进行错误分类的决策代价通常大于延迟判断的代价,而正确识别评论所产生的代价又小于前两者的代价。同时,将虚假评论识别为真实评论的代价又高于将真实评论识别为虚假评论的代价。因此,有 $0 \leq \lambda_{PP}^l \leq \lambda_{BP}^l < \lambda_{NP}^l$, $0 \leq \lambda_{NN}^l \leq \lambda_{BN}^l < \lambda_{PN}^l$, $\lambda_{PN}^l < \lambda_{NP}^l$ 。

基于此,可以计算在第 l 粒层上的阈值 α_l 和 β_l :

$$\alpha_l = \frac{(\lambda_{PN}^l - \lambda_{BN}^l)}{(\lambda_{PN}^l - \lambda_{BN}^l) + (\lambda_{BP}^l - \lambda_{PP}^l)} \quad (13)$$

$$\beta_l = \frac{(\lambda_{BN}^l - \lambda_{NN}^l)}{(\lambda_{BN}^l - \lambda_{NN}^l) + (\lambda_{NP}^l - \lambda_{BP}^l)} \quad (14)$$

可以看到,随着粒层的增加,边界域的评论数量逐渐减少。当达到粒层 G_{r_3} 时,令 $\alpha_l = \beta_l = 0.5$, 即在最细粒层直接进行二支决策。将 $\Pr(\mathbf{X}_l | v_i)$ 与 0.5 进行比较,当 $\Pr(\mathbf{X}_l | v_i) \geq 0.5$ 时,将 v_i 划分至 $\text{POS}_l(X)$, 评论为真;当 $\Pr(\mathbf{X}_l | v_i) < 0.5$, 将 v_i 划分至 $\text{NEG}_l(X)$, 评论为假。

3.3 算法描述

本文提出的序贯三支虚假评论检测方法算法流程如算法 1 所示。该算法通过多关系图与多种深度学习技术的耦合,逐步提取更细粒度的评论特征,构造出具有多层次动态决策框架的虚假评论检测模型。进一步地,利用序贯三支决策的思想,分别对每个粒层内的评论进行三支划分,并在最后一个粒层进行二支划分,从而完成对所有评论的分类。

算法 1 STWD-DLFRD 检测框架

输入:评论集合 \mathbf{V} , 多关系图 \mathbf{G} , 代价矩阵 COST , 粒层数 $L=3$

输出:分类结果(正域 $\text{POS}(\mathbf{X})$, 负域 $\text{NEG}(\mathbf{X})$), 总分类代价 Totalcost

1. //初始化特征提取
2. for each 评论 $v_i \in \mathbf{V}$ do
3. 通过 BERT 生成嵌入向量 $\mathbf{e}_i^{\text{BERT}}$
4. 通过 LSTM 提取上下文特征 $\mathbf{e}_i^{\text{LSTM}}$
5. 通过 CNN 提取局部特征 $\mathbf{e}_i^{\text{CNN}}$
6. 拼接特征 $\mathbf{h}_i^{(0)} \leftarrow \mathbf{W} \cdot [\mathbf{e}_i^{\text{LSTM}} \oplus \mathbf{e}_i^{\text{CNN}}]$
7. end for
8. for each 节点 $v_i \in \mathbf{G}$ do
9. 拼接行为特征 $\mathbf{x}_{v_i}^{(0)} \leftarrow \mathbf{W}(\mathbf{h}_{v_i}^{(0)} \oplus \mathbf{f}_{v_i})$
10. end for
11. //多粒度特征空间构造
12. //粒层 1: 文本特征
13. 直接使用 $\mathbf{h}_{v_i}^{(0)}$
14. //粒层 2: 行为特征
15. for each 节点 $v_i \in \mathbf{G}$ do
16. 一跳 GAT 聚集: $\mathbf{h}_{v_i}^{(1)} \leftarrow \text{GAT}(\mathbf{x}_{v_i}^{(0)}, \mathbf{G})$

17. end for
18. //粒层 3: 社交关系特征
19. for each 节点 $v_i \in \mathbf{G}$ do
20. 二跳 GAT 聚集: $\mathbf{h}_{v_i}^{(2)} \leftarrow \text{GAT}(\mathbf{x}_{v_i}^{(1)}, \mathbf{G})$
21. end for
22. //序贯三支决策
23. 初始化边界域: $\text{BND} \leftarrow \mathbf{V}$, $\text{Totalcost} \leftarrow 0$
24. for 粒层 $l=1$ to L do
25. for each $v_i \in \text{BND}$ do
26. 计算分类概率: $\Pr(\mathbf{X}_l | v_i) \leftarrow \sigma(\mathbf{W} \cdot \mathbf{h}_{v_i}^{(l)})$
27. if $\Pr(\mathbf{X}_l | v_i) \geq \alpha_l$ then
28. 将 v_i 加入 POS
29. $\text{Totalcost} \leftarrow \text{Totalcost} + \text{COST}_l^+$
30. else if $\Pr(\mathbf{X}_l | v_i) \leq \beta_l$ then
31. 将 v_i 加入 NEG
32. $\text{Totalcost} \leftarrow \text{Totalcost} + \text{COST}_l^-$
33. else
34. 保留 v_i 到 BND
35. $\text{Totalcost} \leftarrow \text{Totalcost} + \text{COST}_l^{\text{def}}$
36. end if
37. end for
38. if $l=L$ then
39. 对剩余 BND 中的评论强制二分类
40. end for
41. return $\text{POS}(\mathbf{X}) \cdot \text{NEG}(\mathbf{X})$, Total cost

4 实验结果与分析

4.1 实验数据

本文选用在线评论网站 Yelp 的公开数据集 Yelp Chi, 该数据集包含由 Yelp 自带的过滤系统筛选出的 45954 条芝加哥酒店和餐厅的过滤评论(虚假)与推荐评论(真实), 其中虚假评论占比为 14.5%。实验将数据集按 4:1 的比例划分为训练集和测试集。由于数据的类别不平衡问题较严重, 因此本研究采用小批次训练和下采样策略来进行模型训练, 即在每一次小批次训练中随机采样相同数量的正样本和负样本。

首先,对数据进行预处理,去除停用词和非英文字符,设置最大序列长度为 512,并使用分词器将文本转换为 BERT 模型能理解的形式,以便后续获取深度文本表征。此外,参考文献[40]的研究,构建 14 维人工行为特征,并与第一粒层提取的深度文本特征进行融合,组成多关系图的初始节点特征。本文将节点嵌入维度统一设置为 $\mathbf{x}_{v_i}^{(l)} \in \mathbf{R}^{64}$ 。

众所周知,虚假评论与用户、产品、时间都有关联。这里参考文献[41]的研究,构建了以评论为节点的多关系图,并设置了以下几种关系:1)R-U-R,连接由同一位用户发布的所有评论;2)R-T-R,连接同一产品下同一月份内发布的双条评论;3)R-S-R,连接同一产品下评分一致的评论。3 种关系分别形成了多关系图的 49315, 573616, 3402743 条边。针对每种关系的内聚性,设置 GAT 的注意力头数 $k=2$ 。

此外,为了验证本文方法在识别虚假内容上的有效性,还

选用了电商平台 Amazon 上乐器类目的评论数据集,并对该数据集进行虚假用户检测。该数据集一共包含了 11944 名评论用户,参照文献[42]的策略,将“helpful votes”超过 80% 的标记为良性用户,将“helpful votes”少于 20% 的标记为欺诈用户,其中欺诈用户占比为 9.5%。接着,构建了以用户为节点的多关系图,并参照文献[41]设置了以下几种关系:1)U-P-U,连接针对至少一个相同产品发表过评论的用户;2)U-S-U,连接一周内至少给出过一次相同评分的用户;3)U-V-U,通过 TF-IDF 衡量所有用户的评论文本相似度,并连接前 5% 的用户。3 种关系分别形成了多关系图的 175608, 3566479,1036737 条边。最后,构建了包括文本和行为特征 的 25 维人工特征,并经由 LSTM 和 CNN 提取更深层的特征,作为第一粒层的特征以及多关系图的初始节点特征。

4.2 序贯三支阈值设定

在真实虚假评论检测场景中,将虚假评论判断为真实评论的代价远高于将真实评论判断为虚假评论的代价,且误分类代价远高于延迟决策代价。因此,本文借鉴文献[43-44]中相关代价矩阵设置的策略,采用相对成本的方式来构造决策代价矩阵,如表 2 所列,其中 $1 \leq l \leq L$ 。

表 2 决策成本代价矩阵
Table 2 Decision cost matrix

决策行为	评论真实性	
	$X(P)$	$\neg X(N)$
a_P	$\lambda_{PP}^l = 0$	$\lambda_{PN}^l = 70$
a_B	$\lambda_{BP}^l = 4l$	$\lambda_{BN}^l = 6l$
a_N	$\lambda_{NP}^l = 30$	$\lambda_{NN}^l = 0$

4.3 评价指标

本文提出一种虚假评论检测方法的目的在于提高虚假评论的检测率,并降低检测成本。因此,本实验从检测质量和检

$$ATC = \frac{TC}{\sum_{l=1}^L count_{PP}^l + count_{NP}^l + count_{BP}^l + count_{PN}^l + count_{NN}^l + count_{BN}^l}$$

4.4 实验结果与分析

4.4.1 对比实验

为了更好地验证本文方法的有效性,针对 Yelp 数据集,本文选取了 SVM(特征选取:仅文本),BERT+LSTM+CNN(特征选取:仅文本),GAT(特征选取:选取 32 维的人工节点特征,包括文本和行为特征),BERT+LSTM+CNN+GAT(特征选取:与本实验提出模型相同)4 种基线模型进行对比

测代价两方面来构建评估指标。

4.3.1 检测质量指标

根据虚假评论检测的特性,采用 Precision, Recall, F1 值 3 个指标进行检测效果评估。

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

其中,TP, FN, FP 和 TN 分别为真正例、假负例、假正例和真负例的评论数量。

4.3.2 检测代价指标

为了评估所提出的方法在降低检测代价上的作用,引入了总分类代价和平均分类代价两个指标。总分类代价是在每一粒层上错误分类代价和延迟决策代价之和,其定义如下:

$$TC = \sum_{l=1}^L COST_l = \sum_{l=1}^L (COST_l^P + COST_l^N + COST_l^B) \quad (18)$$

其中, $COST_l^P$, $COST_l^N$ 和 $COST_l^B$ 分别表示在第 l 粒层将评论划分到正域、负域和边界域的决策代价。

在给定第 l 粒层代价损失矩阵下,令 $count_{PP}^l$, $count_{NP}^l$ 和 $count_{BP}^l$ 分别表示在第 l 粒层上将真实评论划分到正域、负域和边界域的评论数量, $count_{PN}^l$, $count_{NN}^l$ 和 $count_{BN}^l$ 分别表示在第 l 粒层上将虚假评论划分到正域、负域和边界域的评论数量。此时,代价 $COST_l$ 及总体平均分类代价 ATC 的计算式如下:

$$COST_l = COST_l^P + COST_l^N + COST_l^B \\ = \lambda_{PP}^l count_{PP}^l + \lambda_{NP}^l count_{NP}^l + \lambda_{BP}^l count_{BP}^l + \\ \lambda_{PN}^l count_{PN}^l + \lambda_{NN}^l count_{NN}^l + \lambda_{BN}^l count_{BN}^l \quad (19)$$

$$TC \quad (20)$$

实验。针对 Amazon 数据集,本文选取了 SVM(特征选取:初始人工特征),LSTM+CNN(特征选取:初始人工特征),GAT(特征选取:初始人工特征),LSTM+CNN+GAT(特征选取:初始人工特征)4 种基线模型进行对比实验。由于基线模型没有延迟决策情形,因此其分类代价仅来源于误分类代价,即设置 $\lambda_{NP}^l = 30$ 和 $\lambda_{PN}^l = 70$ 。不同模型的实验对比结果如表 3 所列。STWD-DL 在不同粒层的分类结果如表 4 所列。

表 3 不同模型的实验对比结果

Table 3 Experimental comparison results of different models

Datasets	模型	Precision	Recall	F1 值	代价	平均代价
Yelp	SVM	0.692	0.621	0.655	225 830	24.048
	BERT+LSTM+CNN	0.754	0.662	0.705	202 360	21.549
	GAT	0.725	0.658	0.689	217 590	23.171
	BERT+LSTM+CNN+GAT	0.783	0.686	0.731	197 670	21.051
	STWD-DLFRD	0.814	0.710	0.758	187 485	19.966
Amazon	SVM	0.711	0.647	0.677	51 830	21.816
	LSTM+CNN	0.772	0.689	0.728	50 330	21.201
	GAT	0.755	0.677	0.714	50 860	21.412
	LSTM+CNN+GAT	0.812	0.701	0.752	49 570	20.859
	STWD-DLFRD	0.837	0.737	0.784	48 800	20.547

表 4 STWD-DL 不同粒层的分类结果

Table 4 Classification results of different granularity layers of STWD-DL

Datasets	层数	α_l	β_l	TN	TP	FN	FP	边界域	代价	总体代价
Yelp	1	0.9251	0.1566	504	2365	806	564	5151	89415	89415
	2	0.8576	0.2528	409	1747	552	387	2056	64210	153625
	3	0.5000	0.5000	295	887	683	191	0	33860	187485
Amazon	1	0.9312	0.1728	127	665	215	144	1224	23580	23580
	2	0.8485	0.2735	112	447	121	91	453	16755	40335
	3	0.5000	0.5000	76	216	137	24	0	8465	48800

通过对表 3 进行分析可知:SVM 总体表现最差,其原因在于 LSTM,BERT,CNN 和 GAT 等均能提取深层次的评论特征,更好地进行虚假评论检测;BERT+LSTM+CNN+GAT 的效果优于 BERT+LSTM+CNN 和 GAT,这说明了虚假评论具有复杂性,提取更多粒度的特征有助于识别虚假评论;STWD-DL 在 F1 值和 Recall 值上优于所有基线模型,这表明序贯决策机制有效提升了检测能力;进一步地,STWD-DL 的总分类代价显著低于其他模型,这说明粒化和分层决策机制在降低计算资源消耗的同时,也有效降低了分类代价。从表 4 可以看出,每一粒层边界域样本呈现“逐渐减少”趋势,这说明引入序贯三支决策的延迟决策机制能够有效处理虚假评论的复杂特征。通过“分级分层”处理,实现了对不同复杂度虚假评论的动态检测,有效降低了总体代价。

4.4.2 消融实验

为了进一步验证本文方法的各组成部分的有效性,在 Yelp 数据集上进行了 6 组消融实验。首先,为验证粒层设置的合理性,设计了 STWD-DLFRD-T 和 STWD-DLFRD-TB 两个实验组。不同于 STWD-DL 方法构造的 3 层特征空间,STWD-DLFRD-T 仅使用第一粒层的特征(文本特征),且由于只有一层,决策阈值直接设为 $\alpha_1 = \beta_1 = 0.5$,而 STWD-DLFRD-TB 则使用了第一和第二粒层的特征(文本和行为特征);其次,为验证三支决策机制的有效性,设置了移除序贯三支决策机制的 Single-Step-DLFRD 模型,直接使用第三粒层的特征进行二分类决策;最后,为验证深度学习组件的有效性,分别设置了 3 组对照实验。具体而言,STWD-FRD-TF 用 TF-IDF 替代 BERT 提取文本特征,其余保持不变;STWD-FRD-BERT 移除了 LSTM 和 CNN,直接使用 BERT 提取的文本特征作为第一粒层特征;STWD-FRD-GCN 用 GCN 替代 GAT 对多关系图进行特征聚集。消融实验结果如表 5 所列。

表 5 消融实验结果

Table 5 Ablation experimental results

模型	Precision	Recall	F1 值	代价	平均代价
STWD-DLFRD-T	0.754	0.662	0.705	202351	21.549
STWD-DLFRD-TB	0.788	0.691	0.736	195578	20.828
Single-Step-DLFRD	0.783	0.686	0.731	197670	21.051
STWD-FRD-TF	0.795	0.699	0.744	192556	20.506
STWD-FRD-BERT	0.791	0.691	0.738	193289	20.585
STWD-FRD-GCN	0.801	0.702	0.748	189156	20.144
STWD-DLFRD	0.814	0.710	0.758	187485	19.966

从表 5 可以看出,单一的文本特征不足以识别复杂虚假评论,引入行为和关系特征有助于增强识别效果,这也说明了多粒度特征的协同作用。其次,三支决策机制相比于二支决策有效降低了虚假评论分类代价,凸显了序贯三支决策在

处理不确定性问题时的优势。此外,相比于传统检测手段,利用深度学习技术进行文本特征的提取可以有效提升分类准确率,且提取的特征越复杂,分类效果越好。最后,在图特征的提取过程中引入注意力机制也能增强评论的分类效果。

4.4.3 鲁棒性分析

进一步地,为了测试模型对于数据噪声的鲁棒性,针对 Yelp 数据集设置了 3 组实验。STWD-Replace 通过随机替换字符(如将“excellent”改为“exelent”),以及使用 WordNet 替换一部分关键词(如将“good”改为“great”),在测试集文本中注入噪声;STWD-Insert 通过在文本中随机插入标点或数字注入噪声;STWD-syntax 通过随机修改句子中的主谓一致性以及故意错误地使用动词时态来注入语法错误噪声。噪声前后模型的各指标变化如图 5 所示。

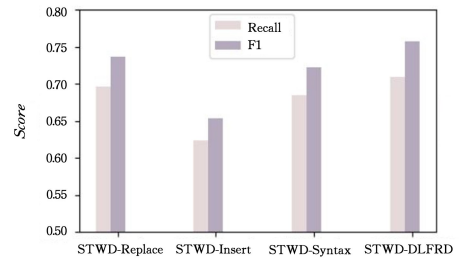


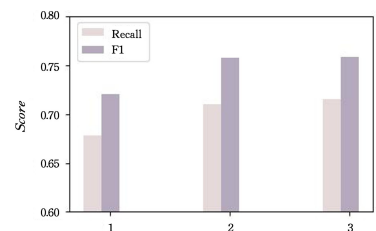
图 5 噪声测试

Fig. 5 Noise testing

由图 5 可知,模型在拼写错误、同义词替换以及语法错误下性能下降较少,表明文本特征具有鲁棒性;而无关字符的插入导致 F1 值下降了 10%,需在预处理阶段加强过滤。

4.4.4 灵敏度分析

在使用本文方法进行虚假评论检测时,为了探究数据不平衡问题以及模型的配置对检测效果的影响,针对 Yelp 数据集设置了 3 组灵敏度分析,即分别使用了不同图注意力网络的注意力头数 k ,以及不同的下采样比例和不同的多关系图节点嵌入维度,来进行检测结果的对比分析。相关灵敏度分析的结果如图 6—图 8 所示。

图 6 多头注意力 k Fig. 6 Multi-head attention k

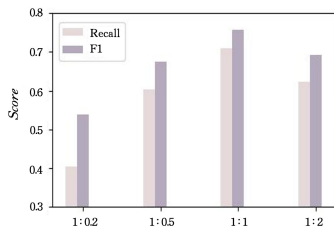


图7 数据采样比例(正:负)

Fig. 7 Data sampling ratio(positive:negative)

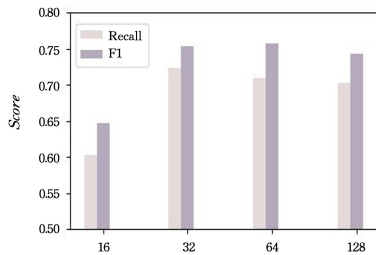


图8 节点嵌入维度

Fig. 8 Node embedding dimension

从图6可以看出,随着多头注意力的头数 k 增加,模型检测效果更好,但在 $k=2$ 和 $k=3$ 时的差别不大,说明2头注意力足以捕捉不同的关系信息。从图7可以看出,正样本和负样本的不均衡会造成模型的过拟合,正负样本一比一的采样策略有助于提高模型的表现。图8是模型在不同节点嵌入维度下的表现,当嵌入维度为32,64以及128时,模型表现接近,当嵌入维度为16时,表现最差。

结束语 考虑到虚假评论检测中静态决策导致检测成本偏高的问题,本文提出了一种基于序贯三支决策和深度学习的虚假评论检测方法,利用多关系图与深度学习技术的耦合提取了多方面的评论特征,构造出3个粒层的特征空间,并借助序贯三支决策“分而治之”的思想实现了渐进式的虚假评论检测。实验表明,本文方法提高了虚假评论的检测精度,降低了检测代价,对数据噪声和参数波动具有较强的鲁棒性。同时,通过消融实验进一步验证了多粒度特征、深度学习技术,以及序贯决策机制的必要性。未来的研究将在序贯三支决策的框架下继续探索不同的粒层构造策略,以进一步提升复杂环境下的虚假评论检测效率。

参考文献

[1] LI L Y, QIN B, LIU T. A Survey on Fake Review Detection[J]. Chinese Journal of Computers, 2018, 41(4): 946-968.

[2] INSTANTO R S H, MAHMUDY W F, BACHTIAR F A. Detection of online review spam: A literature review[C]// Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology (SIET). New York: ACM, 2020: 57-63.

[3] LAI C L, XU K Q, LAU R Y K, et al. High-order concept associations mining and inferential language modeling for online review spam detection[C]// 2010 IEEE International Conference on Data Mining Workshops. IEEE, 2010: 1120-1127.

[4] KUMAR A, GOPAL R D, SHANKAR R, et al. Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering

[J]. Decision Support Systems, 2022, 155: 113728.

[5] JABEUR S B, BALLOUK H, ARFI W B, et al. Artificial intelligence applications in fake review detection: Bibliometric analysis and future avenues for research[J]. Journal of Business Research, 2023, 158: 113631.

[6] CHIEN S Y, YANG C J, YU F. XFlag: Explainable fake news detection model on social media[J]. International Journal of Human-Computer Interaction, 2022, 38(18/19/20): 1808-1827.

[7] ZHANG J Y. Research on Feature Extraction and Detection Technology of Fake Reviews[J]. Intelligent Computer and Applications, 2023, 13(10): 52-55.

[8] WEI J R, WANG R T, WANG H. Large-scale Fake Review Group Detection Based on Network Structure Features [J]. Operations Research and Management Science, 2023, 32(1): 194-200.

[9] SCHOENMÜLLER V, NETZER O, STAHL F. The extreme distribution of online reviews: Prevalence, drivers and implications[J]. Journal of Marketing Research, 2020, 57(5): 853-877..

[10] HAJEK P, SAHUT J M. Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection [J]. Technological Forecasting and Social Change, 2022, 177: 121532.

[11] MUKHERJEE A, LIU B, GLANCE N. Spotting fake reviewer groups in consumer reviews[C]// Proceedings of the 21st International Conference on World Wide Web(WWW). New York: ACM, 2012: 191-200.

[12] XU C, ZHANG J, CHANG K, et al. Uncovering collusive spammers in Chinese review websites[C]// Proceedings of the 22nd ACM International Conference on Information and Knowledge Management(CIKM). New York: ACM, 2013: 979-988.

[13] WANG Z, WEI W, MAO X L, et al. User-based network embedding for opinion spammer detection[J]. Pattern Recognition, 2022, 125: 108512.

[14] PAUL H, NIKOLAEV A. Fake review detection on online E-commerce platforms: a systematic literature review[J]. Data Mining and Knowledge Discovery, 2021, 35(5): 1830-1881.

[15] SHANG Y, LIU M, ZHAO T, et al. T-Bert: A spam review detection model combining group intelligence and personalized sentiment information[C]// Proceedings of the International Conference on Artificial Neural Networks(ICANN). Cham: Springer, 2021: 409-421.

[16] BHUVANESHWARI P, RAO A N, ROBINSON Y H. Spam review detection using self attention based CNN and bi-directional LSTM[J]. Multimedia Tools and Applications, 2021, 80(12): 18107-18124.

[17] MUKHERJEE A, VENKATARAMAN V, LIU B, et al. What Yelp fake review filter might be doing? [C]// Proceedings of the International AAAI Conference on Web and Social Media (ICWSM). Palo Alto, CA: AAAI, 2013: 409-418.

[18] TANG X, QIAN T, YOU Z. Generating behavior features for cold-start spam review detection with adversarial learning[J]. Information Sciences, 2020, 526: 274-288.

[19] MANASKASEMSAK B, TANTISUWANKUL J, RUNGSAWANG A. Fake review and reviewer detection through beha-

- vioral graph partitioning integrating deep neural network[J]. *Neural Computing and Applications*, 2023, 35: 1169-1182.
- [20] CHENG L C, WU Y T, CHAO C T, et al. Detecting fake reviewers from the social context with a graph neural network method[J]. *Decision Support Systems*, 2024, 179: 114150.
- [21] HE L, XU G, JAMEEL S, et al. Graph-aware deep fusion networks for online spam review detection[J]. *IEEE Transactions on Computational Social Systems*, 2022, 10(5): 2557-2565.
- [22] LI A, QIN Z, LIU R, et al. Spam review detection with graph convolutional networks[C]// *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*. New York: ACM, 2019: 2703-2711.
- [23] DUMA R A, NIU Z, NYAMAWA A S, et al. A Deep Hybrid Model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings[J]. *Soft Computing*, 2023, 27(10): 6281-6296.
- [24] ZHANG D, LI W, NIU B, et al. A deep learning approach for detecting fake reviewers; Exploiting reviewing behavior and textual information [J]. *Decision Support Systems*, 2023, 166: 113911.
- [25] SAHOO S R, GUPTA B B. Classification of spammer and non-spammer content in online social network using genetic algorithm-based feature selection[J]. *Enterprise Information Systems*, 2020, 14(5): 710-736.
- [26] YAO Y, DENG X. Sequential three-way decisions with probabilistic rough sets[C]// *Proceedings of the 10th IEEE International Conference on Cognitive Informatics and Cognitive Computing*. Piscataway, NJ: IEEE, 2011: 120-125.
- [27] ZHANG G Q, LIU Q, JI L H. Multi-granularity Sentiment Classification Method Based on Sequential Three-way Decisions[J]. *Computer Science*, 2018, 45(12): 153-159.
- [28] CHEN R, YANG X, LUO J F, et al. Research on Sequential Three-way Abnormal User Detection Based on GANs-LightGBM[J]. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2021, 33(5): 816-825.
- [29] LI H, ZHANG L, HUANG B, et al. Sequential three-way decision and granulation for cost-sensitive face recognition [J]. *Knowledge-Based Systems*, 2016, 91: 241-251.
- [30] YAO Y Y, WONG S K M. A decision theoretic framework for approximating concepts [J]. *International Journal of Man-Machine Studies*, 1992, 37(6): 793-809.
- [31] YANG X, LI Y, LI T. A review of sequential three-way decision and multi-granularity learning[J]. *International Journal of Approximate Reasoning*, 2023, 152: 414-433.
- [32] LIU Y, AO X, QIN Z, et al. Pick and choose: A GNN-based imbalanced learning approach for fraud detection[C]// *Proceedings of the Web Conference 2021 (WWW)*. New York: ACM, 2021: 3168-3177.
- [33] ZHANG G, LI Z, HUANG J, et al. eFraudCom: An e-commerce fraud detection system via competitive graph neural networks [J]. *ACM Transactions on Information Systems*, 2022, 40(3): 1-29.
- [34] KHAN A, SOHAIL A, ZAHOORA U, et al. A survey of the recent architectures of deep convolutional neural networks[J]. *Artificial Intelligence Review*, 2020, 53: 5455-5516.
- [35] ALSHARIF N. Fake opinion detection in an e-commerce business based on a long-short memory algorithm[J]. *Soft Computing*, 2022, 26(16): 7847-7854.
- [36] JAVED M S, MAJEED H, MUJTABA H, et al. Fake reviews classification using deep learning ensemble of shallow convolutions[J]. *Journal of Computational Social Science*, 2021, 4(2): 883-902.
- [37] OAK R. Detecting review fraud using metaheuristic graph neural networks[J]. *International Journal of Information Technology*, 2024, 16(7): 4019-4025.
- [38] WANG J, WEN R, WU C, et al. FdGars: Fraudster detection via graph convolutional networks in online app review system[C]// *Proceedings of the Companion of the 2019 World Wide Web Conference (WWW)*. New York: ACM, 2019: 310-316.
- [39] REN Z, ZHANG X, ZHANG S, et al. Fake review detection via heterogeneous graph attention network[C]// *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. Cham: Springer, 2023: 364-376.
- [40] RAYANA S, AKOGLU L. Collective opinion spam detection: Bridging review networks and metadata[C]// *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York: ACM, 2015: 985-994.
- [41] DOU Y, LIU Z, SUN L, et al. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters[C]// *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM)*. New York: ACM, 2020: 315-324.
- [42] ZHANG S, YIN H, CHEN T, et al. GCN-based user representation learning for unifying robust recommendation and fraudster detection[C]// *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2020: 689-698.
- [43] LIU D, YE X. A matrix factorization based dynamic granularity recommendation with three-way decisions[J]. *Knowledge-Based Systems*, 2020, 191: 105243.
- [44] FAN Q, LIU D, YE X Q. Cost-sensitive Text Sentiment Analysis Based on Sequential Three-way Decision[J]. *Pattern Recognition and Artificial Intelligence*, 2020, 33(8): 732-742.



GU Bokai, born in 2001, master, is a student member of CCF (No. Z6696G). His main research interests include data mining, fake reviews detection, three-way decision and granular computing.



LIU Dun, born in 1983, Ph. D, professor, is a distinguished member of CCF (No. 17413D). His main research interests include data mining and knowledge discovery, rough set theory and granular computing, decision support systems.