

RT-NIG: 在邻域信息图中重构二元组用于谣言早期检测

王 标¹, 卫红权^{1,2}, 刘树新^{1,2}, 王 凯^{1,2}, 江昊聪^{1,2}, 李 燃¹

¹中国人民解放军战略支援部队信息工程大学 郑州 中国 450001

²国家数字交换系统工程技术研究中心 郑州 中国 450002

摘要 随着谣言在网络上不断扩散和传播,其危害会越来越大。在谣言尚未传播的最早期,利用源发布账号的用户信息和文本信息对其进行识别并遏止具有十分重要的意义。当前的检测方法局限于自然语言处理技术,注重从文本中提取信息来识别谣言,缺乏对用户信息的深度挖掘和有效结合,导致模型的检测性能低。为此,本文提出了一种谣言早期检测的新途径 RT-NIG,通过在交叉分布的邻域信息图中重构二元组信息来识别谣言。首先针对谣言传播最初阶段缺乏传播信息,无法形成图结构数据的情景,利用对象的潜在相关性构造虚拟邻域图,解决了数据不确定性以及不完备性等问题;通过图神经网络捕获邻域图中潜在的对象关系,关注用户之间潜在的可信度关系以及文本之间的情感极性关系,在两个邻域信息图中交叉传递用户信息和语义信息,分别重构了用户信息和语义信息;最后通过加权集成的方式,重新构造“用户-推文”二元组信息,有效地结合了这两种信息,并用于下游的谣言分类任务。分别在中文 Weibo 和英文 PHEME 两个真实数据集上进行了实验验证,本文方法在准确率、精确率、召回率、F₁ 值等指标上优于多种先进的早期检测方法,在两个数据集上准确率分别比最优的对比方法提升了 5% 和 8%;并且通过消融研究以及超参数分析,进一步证明了用户信息在早期检测中具有的重要作用以及二元组信息重构方式的有效性。针对无传播信息可用的场景,RT-NIG 也为其他一些早期检测问题提供了新的解决方法,例如假新闻、网络暴力、误导消息等问题。

关键词 谣言早期检测; RT-NIG; 邻域信息图; 重构二元组; 图神经网络

中图法分类号 TP391.1 DOI 号 10.19363/J.cnki.cn10-1380/tn.2024.11.10

RT-NIG: Reconstructing Two-tuples in Neighborhood Information Graphs for Early Rumor Detection

WANG Biao¹, WEI Hongquan^{1,2}, LIU Shuxin^{1,2}, WANG Kai^{1,2}, JIANG Haocong^{1,2}, LI Ran¹

¹PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China

²National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China

Abstract As rumors continue to diffuse and spread on the network, their influence will become more and more serious. In the earliest stage that rumors have not yet spread, it is of great significance to use the user information and text information of the source account to identify and suppress them. The current detection methods are limited to natural language processing technology and focus on extracting information from the text to identify rumors. The lack of in-depth mining and an effective combination of user information leads to low detection performance of the model. In this paper, we propose a new approach for early rumor detection, RT-NIG, which can identify rumors by reconstructing two-tuples information in the cross-distributed neighborhood information graphs. Firstly, in response to the situation where there is a lack of dissemination information in the initial stage of rumor propagation and graph-structured data cannot be formed, a virtual neighborhood graph is constructed using the potential correlation of objects to solve the problems of data uncertainty and incompleteness. Then the potential object relationships in the neighborhood graph are captured by graph neural network. The user information and semantic information are transferred in two neighborhood information graphs and reconstructed, paying attention to the potential credibility relationship between users and the emotional polarity relationship between texts. Finally, using weighted integration, the "User-Tweet" two-tuple information is reconstructed, effectively combining the two kinds of information, and is used for the downstream rumor classification task. Experiments are conducted on two real datasets, Chinese Weibo and English PHEME. The proposed method outperformed various advanced early detection methods in accuracy, accuracy, recall, F1 value, and other indicators. The accuracy was improved by 5% and 8% compared to the optimal comparison method on the two datasets. Ablation research and super parameter analysis further prove the important role of user information in early detection and the effectiveness of two-tuple information reconstruction. RT-NIG also provides new solutions for other early detection problems, such as fake news, online violence, misleading information,

通讯作者: 刘树新, 博士, 助理研究员, Email: liushuxin11@126.com。

本课题得到嵩山实验室项目(No. 221100210700-2)和河南省重大科技专项(No. 221100210100)的资助。

收稿日期: 2022-12-08; 修改日期: 2023-05-10; 定稿日期: 2024-09-23

etc., in scenarios where no disseminated structural information is available.

Key words early rumor detection; RT-NIG; neighborhood information graphs; reconstruct two-tuples; GNN

1 引言

谣言被定义为充满了个人世界观的舆论交流^[1]。造谣者带着个人目的编撰谣言,一般是为了博取眼球,或者是为了某种利益,其带来的影响通常是恶劣的。随着社交网络平台的发展,在线用户制造谣言的成本变得低廉,谣言变得随处可见。而且谣言在传播过程中,会影响到一部分人的判断,形成固化的确认偏见,所以尽早地发现并遏制谣言对网络空间的健康有序发展具有重要的意义。

谣言识别可利用的信息主要有四类:用户信息、语言信息、结构信息以及时间信息。其中,利用结构和时间信息可以在更长的时间窗口内将谣言和非谣言区分开来,但这种场景建立在事件开始传播的基础上。但是,当越来越多的人接触到谣言时,人们就越可能相信它,而在信任之后,便很难再修正这种偏见,即塞梅尔韦斯反射^[2]、确认偏见^[3]、锚定偏见^[4]等。在信息的传播过程中,相同兴趣的人往往会忽略其他不同的意见,导致“回音室”现象^[5-6]。因此,在传播的最初阶段进行谣言检测,不仅能及时发现谣言,还可以避免确认偏见的产生。但在传播的最初阶段,即推文在社交媒体上发布但尚未开始有效传播时,结构和时间信息一般是不可用的,或者是不存在的^[7-8]。相反,用户和语言信息更加容易获得,并且在传播的最初阶段可以用作良好的指示符^[8]。

本文旨在谣言一经发布的最早期即谣言尚未开始传播的阶段进行检测,也即本文早期检测的研究不依赖于传播结构,仅仅依靠源发布账户的用户信息以及文本信息进行谣言检测。当前的检测模型大多从自然语言处理技术出发,通过在文本中提取不同表现形式的信息,实现了有效的谣言检测并取得了一定的效果。模型一般从两个方面研究:一种方法是利用传统机器学习方法人工提取文本中的模式特征,如词频信息、词袋信息等;另一种方法是采用 CNN、LSTM、GRU、BERT 等模型,深度提取文本中的语义信息。这些模型的性能取决于提取语言信息的质量,而用户信息仅仅是作为语言信息的增益,其作用没有得到深度挖掘。

基于图的方法具有强大的表达能力,能够关注到数据对象间潜在的关系。但是在谣言传播的最初阶段缺乏传播信息,无法形成图结构数据。在一些无图场景中,利用邻域信息系统从数据对象间潜在的

相关性出发,构造虚拟的邻域信息图(在后文中简称邻域图),可以系统解决数据的不确定性、不完备性等问题。在粗糙集理论中,邻域信息系统是基于邻域关系划分知识的一种信息系统,可以处理不完备的混合数据,而传统的信息系统基于严格的不可分辨关系,只能处理类别型数据。特别地,本文所指的邻域图是指通过邻域信息系统划分的邻域关系构造的图数据。为了处理具有任意形状簇和不同密度的数据集,Zhu 等人^[9]提出了加权自然邻域图,用于异常值检测。Wang 等人^[10-11]提出了一种新的异常值检测模型,在一系列预定义的邻域图上采用马尔可夫随机游走过来定义异常值分数。为了处理具有混合值属性的数据集,Wang 等人^[12]提出了一种基于加权邻域信息网络模型的混合值数据集离群点检测方法。

基于以上考虑,本文通过构造虚拟的邻域图来解决无图可用的早期检测场景。首先构造“用户-推文”二元组,遵循“不可信用户更容易发表谣言”^[13]以及“负面情感导向明显的用户更可能是造谣者”^[14]两个理论驱动的假设,基于邻域信息系统,分别构造了用户邻域图以及语义邻域图,然后在两个邻域图中分别传递对象的语义信息与用户信息,重构为新的二元组,用来进行谣言分类。本文的主要贡献有:

- 提出了一种谣言早期检测的新方法 RT-NIG,通过在构造的邻域图中交叉传递用户信息和语义信息重构“用户-推文”二元组信息表示,用于谣言分类。
- 在谣言传播最初阶段无传播信息可用的情况下,基于邻域信息系统以及两个理论驱动的假设,从数据对象间潜在的相关性以及经验知识出发,构造了虚拟的用户邻域图以及语义邻域图,解决了早期检测时无法形成图结构数据的问题。
- 分别在中文 Weibo 和英文 PHEME 两个真实谣言数据集上实验验证,结果表明本文方法优于多种已知的早期检测基线方法,在两个数据集上准确率分别比最优的对比方法提升了 5% 和 8%。

2 相关工作

本文的早期检测旨在利用源发布账户的用户信息和文本信息在谣言一经发布的最初阶段进行甄别,而不是在谣言已经开始传播并造成影响后再进行检

测。因此, 早期检测的相关工作主要围绕可利用的特征信息以及用于分类的算法模型分别进行介绍。

(1) 早期检测的特征工程

当推文刚发布但尚未传播时, 可利用的信息仅有源发布账户的用户信息及文本信息。

理论驱动的一些文献指出, “谣言传播者是希望获得关注和流行的人”^[15]。造谣者跟正常用户的账号在追随者数量、朋友或关注者以及发表的推文数量等方面可能存在较大不同^[16], 可以用来识别谣言, 还可以通过用户可信度来判断发布者的意图, 是无意发布错误信息还是故意发布。Vicaria 等人^[5]考虑了在线社交媒体上与用户行为相关的一系列特征, 用于平滑两极分化和减少错误信息。Gupta 等人^[17]基于用户的特征, 如用户粉丝数以及用户名长度等, 对推特的可信度进行评分。Ruchansky 等人^[18]提出的模型结合了推特源用户的信息, 用于检测推特中的虚假信息。但一些用户账号是被刻意培养的, 跟正常用户差异很小, 这就需要使用推文的语言信息来做进一步的判定。

对文本信息的探索一般是从知识核查、模式识别以及潜在语义等几个方面进行的。

知识是从语句中提取的三元组信息, 即主语、谓语和宾语。知识核查时, 首先从开源的网页和数据库中抽取需要的原始事实, 通过抽取、链接、消歧、融合等技术构建知识图(Knowledge Graph, KG)。然后将从发布推文中提取的知识与 KG 中的知识进行比较, 来识别推文中信息的真实性^[19]。但是这种方法工作量大^[20], 实时性要求高, 只能判断真实性, 不能判断发布者的意图^[21]。

模式识别是通过数据分析以及理论驱动等找到区分谣言与真实的一组特征^[21], 这些特征可以通过字词统计的方法获得, 如 TF-IDF^[22], n-grams, 语言探究与字数统计工具(Linguistic Inquiry and Word Count, LIWC)^[23]等。Riedel 等人^[24]以及 Takahashi 等人^[25]利用 TF-IDF 提取假新闻或者谣言的加权关键词列表, 提取语言内容特征, 用来识别假新闻或谣言。Vogel 等人^[26]使用 n-grams 提取字词表示, 结合 SVM 模型分析假新闻的传播。Kwon 等人^[13]则通过 LIWC 获得的语言特征来识别谣言。但有效特征的选取一般是基于专家经验或者是大量数据分析, 费时费力, 而且不能提取深层次的特征, 难以检查模仿正常推文写作的谣言。

提取推文中潜在的语义信息, 用来发现谣言的深层特征, 越来越多地被用在谣言识别中。Chen 等人^[27]基于一种结合 CNN 的残差网络提取文本中的词

语级语义信息。Liu 等人^[28]基于内核图注意力网络提取文本的句子级语义信息对事实进行联合推理, 但需要使用其他的声明和证据。张明书等人^[29]利用 Bi-LSTM 提取文本的情感语义信息, 并结合谣言的种类信息来识别谣言。虽然这种潜在语义的提取可解释性不强, 但在谣言的自动检测中效果明显高于传统的模式识别方法。

(2) 早期检测的算法模型

早期检测的算法根据输入特征的不同可以分为两类, 一类是采用人工提取特征的传统机器学习方法, 如决策树^[13, 30]、支持向量机^[13, 30-33]、随机森林^[13, 34]、逻辑回归^[34]、朴素贝叶斯^[34]、传统自然语言处理^[25]等方法。传统方法易于实现, 2 计算速度快, 可解释性强, 鲁棒性强, 但要依赖于人工提取特征, 这些特征的好坏将直接影响模型的性能。另外一类是采用神经网络提取特征的深度学习方法, 如 CNN^[27, 35], RNN^[36], BERT 预训练模型^[37-38]以及这几种模型的结合^[39-40]。将深度学习方法用于自然语言处理技术, 可以提取推文内容的深层信息。而且这是一种端到端模型, 能够自动提取特征, 进行深度拟合, 识别谣言的效果更好。

可以看出, 当前早期检测方法重视从源推文文本中提取有效的语言特征用于谣言分类, 从自然语言处理技术上提高识别效果, 对于用户信息的结合也只是粗略地进行了简单的向量拼接, 忽略了从用户信息和语言信息等可用信息的深度结合上获取更好的特征表示。

3 准备工作

3.1 邻域信息系统

令 $NIS = (X, A, V, f)$ 是一个邻域信息系统, 其中 $X = \{x_1, x_2, \dots, x_n\}$ 为对象的非空有限集合, 称为论域(universe);

$A = \{a_1, a_2, \dots, a_m\}$ 为属性的非空有限集合, 其中包含 k 个数值属性和 d 个类别属性, 即 $k + d = m$;

$V = \bigcup_{a \in A} V_a$, V_a 表示属性 a 的值域;

$f: X \times A \rightarrow V$ 是一个函数, 即对 $\forall x \in U, a \in A$, 有 $f(x, a) \in V_a$ 。

用 $u_i^{h'}$ ($i = 1, 2, \dots, n; h' = 1, 2, \dots, k$,

$k + 1, \dots, k + d - 1, m$) 表示对象 x_i 在属性 $a_{h'}$ 上的值。

对象 x_i 在属性集合 A 上的邻居集合为:

$$N_i = \{x_j \mid x_i, x_j \in X, S_{ij} \geq r\} (i, j = 1, 2, \dots, n), \quad (1)$$

其中 S_{ij} 是对象 x_i 和 x_j 之间的相似度; r 是邻域半径, 用于判断对象 x_i 和 x_j 是否是邻域关系。

则 X 中的邻域关系为:

$$R = \left\{ (x_i, x_j) \in X \times X \mid x_j \in N_i \right\} (i, j = 1, 2, \dots, n), \quad (2)$$

则商集 X/R 构成了 X 上的邻域知识。

3.2 问题描述

谣言早期检测本质上是一个分类问题。可以描述为:

用 $\mathcal{D} = \{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_i, \dots, \mathcal{E}_n\}$ 表示谣言检测任务的事件集合, 其中 $\mathcal{E}_i = \{\tilde{h}_i^{user}, \tilde{h}_i^{sem}\}$ 表示第 i 个事件的二元组信息, n 表示数据集的事件总数, \tilde{h}_i^{user} 表示事件 \mathcal{E}_i 的用户信息, \tilde{h}_i^{sem} 表示事件 \mathcal{E}_i 的文本信息或语义信息。谣言检测的目的是训练一个分类器做到: $\mathcal{F}: \mathcal{D} \rightarrow \mathcal{Y}$ 。其中, \mathcal{Y} 是数据集的标签 y_i 的集合, 其中数据集的 $y_i \in \{F, T\}$ (分别代表谣言和非谣言)。相关符号说明见表 1。

表 1 符号表示

Table 1 Glossary of notations

符号	定义
\mathcal{R}	JSON 格式原始数据;
\mathcal{D}	预处理后数据;
\mathcal{D}_i^{ext}	第 i 个事件的文本内容;
\mathcal{E}_i	第 i 个事件的二元组信息;
\tilde{h}_i^{user}	第 i 个二元组的初始用户信息;
\tilde{h}_i^{sem}	第 i 个二元组的初始语义信息;
\mathcal{Y}	数据集的标签集;
NIS	邻域信息系统;
X	对象的非空有限集合;
A	属性的非空有限集合;
a	A 中的属性;
V	属性的值域;
f	邻域信息系统的映射函数;
$u_i^{h'}$	对象 x_i 在属性 $a_{h'}$ 上的值;
N_i	在 A 上对象 x_i 的邻居集合;
R	X 上的邻域关系;
S_{ij}	x_i 和 x_j 的相似度;
h_i^l	第 i 个对象在第 l 层的向量表示
$l; L$	第 l 层神经网络; 神经网络总层数
$e; E$	第 e 轮训练; 训练总轮数

4 基于二元组重构的谣言早期检测方法

在无谣言传播信息情况下, 当前研究注重从文本中提取有效信息来优化模型, 为了突破自然语言

处理技术的局限, 本文提出了一种早期谣言检测的新方法 RT-NIG。如图 1 技术框架, RT-NIG 方法主要包含五个内容:

- (1) 初始数据处理
- (2) 初始二元组表示
- (3) 邻域信息图
- (4) 信息重构模型
- (5) 重构二元组表示及分类

核心为构造邻域信息图及信息重构模型。具体流程为:

首先, 对原始的 json 数据进行预处理, 提取跟用户和推文有关的数据, 构造初始二元组来刻画早期事件信息; 然后, 基于邻域信息系统分别构造用户邻域图以及语义邻域图, 通过加权集成的图神经网络捕捉两个邻域图中对象潜在的相关性, 并分别在其中传递和重构语义信息和用户信息, 得到新的二元组表示, 用于下游的谣言分类任务。

4.1 初始数据处理

数据的质量将直接影响模型的性能, 低质量的数据必然会限制分析并降低挖掘结果的质量^[41], 因此, 进行数据预处理至关重要。本文使用了中文 Weibo 以及英文 PHEME 两个真实的谣言数据集, 均为平台 API 接口下的真实数据, 包括用户、推文、社交、时间等信息。

在推文尚未传播的早期仅有源发布用户的用户信息和文本信息可用, 因此, 筛选数据集中的用户信息以及文本内容用于本证研究。第 i 个事件的用户信息可表示为:

$$\tilde{h}_i^{user} = \{u_i^1, u_i^2, \dots, u_i^{h'}, \dots, u_i^m\}, \quad (3)$$

在用户信息组成的邻域信息系统中, $u_i^{h'}$ 表示第 i 个对象在属性 $a_{h'}$ 上的值。用户信息为字典格式, 需要进行预处理的内容较少, 本文将时间转化为时间戳格式, 并进行了数值的标准化处理。如公式 4:

$$u_i^h = \left(u_i^{h'} - \min_{i=1,2,\dots,n} \{u_i^{h'}\} \right) / \left(\max_{i=1,2,\dots,n} \{u_i^{h'}\} - \min_{i=1,2,\dots,n} \{u_i^{h'}\} \right) \in (0,1). \quad (4)$$

则第 i 个事件的初始用户信息可表示为:

$$\tilde{h}_i^{user} = \{u_i^1, u_i^2, \dots, u_i^h, \dots, u_i^m\}. \quad (5)$$

但是文本数据含有的信息一般比较混杂, 不能直接提取关键特征, 要进一步处理停用词、乱码字符、提及符号、URL 链接、特殊字符、无用空格等, PHEME 数据集中的文本还需要进行字母大小写转化、拼写校对等处理。

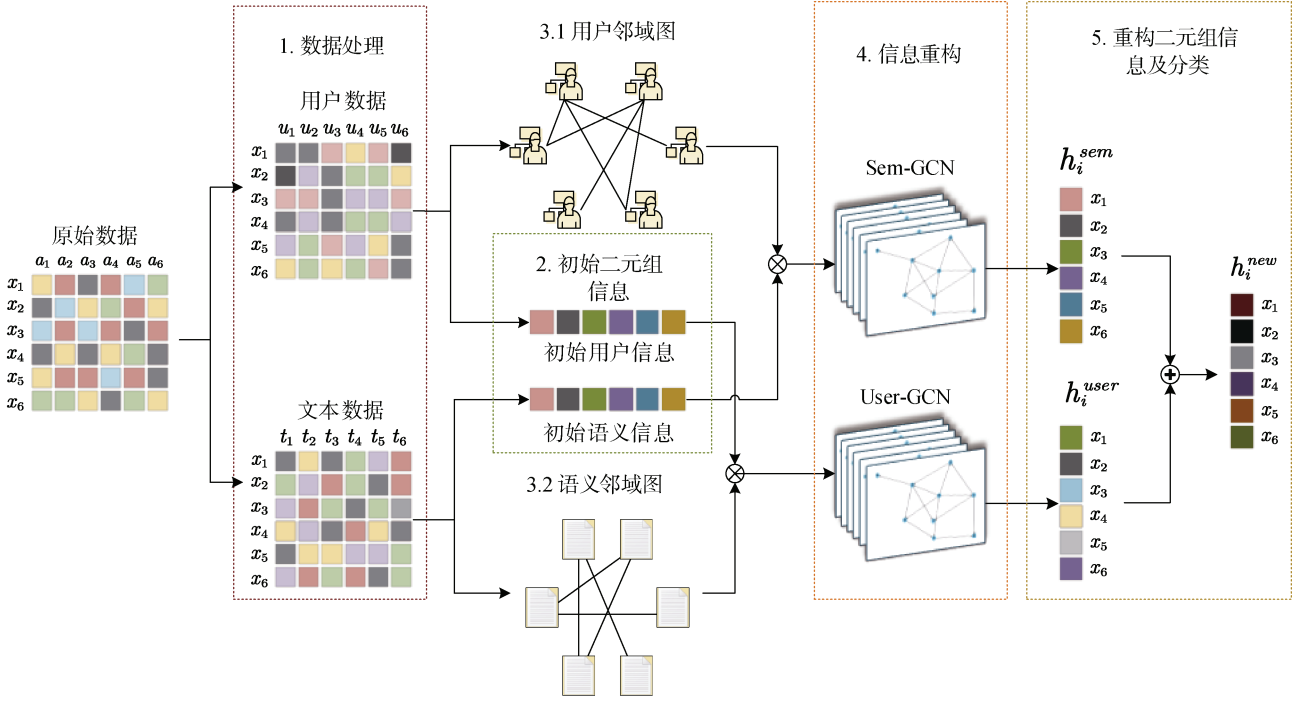


图 1 早期检测技术 RT-NIG 流程框架

Figure 1 Early detection technology RT-NIG process framework

4.2 初始二元组表示

如图 1 所示, 数据预处理后, 为了得到初始二元组信息, 需要进一步对文本数据进行模式特征或者深层语义特征的提取。

模式特征是使用字词统计的方法对文本数据进行特征提取。如实验设置中, 传统机器学习的对比算法采用的就是人工选取特定的文本特征。在使用 Weibo 数据集时, 基于知网情感词典对文本内容进行情感分析, 获得了积极词个数、积极评分、消极词个数、消极评分以及文本长度等 5 个关键特征。在使用 PHEME 数据集时, 采用 LIWC 提取了词性、情感、社会、心理、健康等不同类别的词统计, 共 73 种, 用作文本的关键特征。

深层语义特征是采用神经网络提取文本内容的潜在语义信息。本文采用的方法是基于 BERT 预训练模型的语义提取, 获得句子级的语义表示。则第 i 个事件的语义信息可表示为:

$$\tilde{h}_i^{sem} = Bert(\mathcal{D}_i^{text}), \quad (6)$$

其中 \mathcal{D}_i^{text} 为数据预处理后第 i 个事件的文本内容, $Bert(\cdot)$ 为 BERT 模型, 输出为提取的 768 维的特征向量, \tilde{h}_i^{sem} 表示事件的初始语义信息。此外, 还有 CNN、LSTM、GRU 等模型可用来提取文本的语义信息。

最后得到第 i 个事件的初始二元组信息表示:

$$h_i^{ini} = concat(\tilde{h}_i^{user}, \tilde{h}_i^{sem}). \quad (7)$$

现有的模型都是利用初始二元组信息进行谣言分类的, 即重点在从文本中提取语言信息, 然后与用户信息进行简单的拼接, 得到用于分类的信息表征。文本特征提取的好坏将直接影响到算法结果的质量, 而用户信息只能作为语言信息的增益, 缺乏更深度的挖掘。而 RT-NIG 算法的贡献是在初始二元组表示的基础上, 通过关注对象的邻域信息来重构二元组信息, 更好地结合了用户信息和语言信息。

4.3 邻域信息图

两个理论假设是在研究文献的基础上得出的一些经验, 不一定适用于所有场景, 但也符合一定的认知。本文遵循理论假设基础, 是为了依据经验构造更适合的邻域信息图。

邻域图构造实质上是将相似的对象节点之间建立联系, 从而构造图结构数据, 过程类似于聚类。在后续设计中, “不可信用户更容易发表谣言”, 实质上是将用户按照可信程度聚类成不同的簇, 在不同簇构成的网络中传递语义信息, 同时结合了用户信息以及语义信息, 即使用户是刻意培养的水军账户, 通过语义信息的结合也可以发现谣言。同理, 对于“负面情感导向明显的用户更可能是造谣者”, 实质上是将文本按照抽象出的不同语义情感聚类成不同的簇, 并在其网络结构中传递用户信息, 同样结合了用户信息以及语义信息。

(1) 用户邻域图

用户的可信度可以用来辅助发现意图, 不可信用户更容易发表谣言。谣言的可信度低可能会损害个人的声誉, 具有高度影响力的个人可能会试图避免参与谣言的传播^[13]。遵循这一理论驱动的假设, 构造用户的邻域信息图, 来关注用户之间潜在的可信度关系, 如图 1 所示。

在用户信息构成的信息系统中, 用户信息一般为混合数据类型, 包括数值型、符号型以及空值等。不精确的数据测量造成的误差、不同的数据理解以及数据获取的严格限制等都可能产生空值的产生。考虑到需要处理不完备混合数据, 本文借鉴异质距离函数^[42], 将 \tilde{h}_i^{user} 和 \tilde{h}_j^{user} 的相似度定义为:

$$S_{ij} = \sum_{h=1}^m (s_{ij}^h)^2, \quad (8)$$

$$s_{ij}^h = \begin{cases} 0, & \text{if } u_i^h \text{ or } u_j^h \text{ is None;} \\ 1 - |u_i^h - u_j^h|, & \text{when } h=1, 2, \dots, k; \\ 1, & \text{if } u_i^h = u_j^h (h=k+1, k+2, \dots, m); \\ 0, & \text{if } u_i^h \neq u_j^h (h=k+1, k+2, \dots, m), \end{cases} \quad (9)$$

用 \mathbf{S} 表示 $n \times n$ 的相似度矩阵, 将 \mathbf{S} 展开成 $1 \times n^2$ 的矩阵, 并从小到大排序得到 \mathbf{S}^r 。

设置一个邻域半径阈值参数 $\theta \in [0, 1]$, 定义邻域半径为:

$$r = \mathbf{S}^r \left[\text{int}(\theta * n^2) \right], \quad (10)$$

式(10)表示取 \mathbf{S}^r 特定位置的相似度数作为邻域半径, $\text{int}(\cdot)$ 表示取整。根据式(1)和式(2)可以确定每个对象的邻居 N_i , 最终得到图的邻接矩阵

$\mathbf{U}^{user} \in \{0, 1\}^{n \times n}$, \mathbf{U}^{user} 中的元素 u_{ij} 满足:

$$u_{ij} = \begin{cases} 0, & \text{if } i = j; \\ 1, & \text{if } x_j \in N_i; \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

显然, 当 θ 值逐渐变大时, 构造的邻域信息图由密变疏。如图 2a 和 2c 所示, 分别为 Weibo 和 PHEME 数据集中构造的用户邻域图, 选择其中的 2000 条边进行展示。

(2) 语义邻域图

虚假的故事在传播中激发了恐惧、厌恶和惊讶, 而真实的故事激发了期待、悲伤、喜悦和信任, 负面情感导向明显的用户更可能是造谣者^[14]。遵循这一理论驱动的假设, 构造语义邻域图, 来关注对象间潜在的情感极性关系, 如图 1 的 3.2 部分所示。

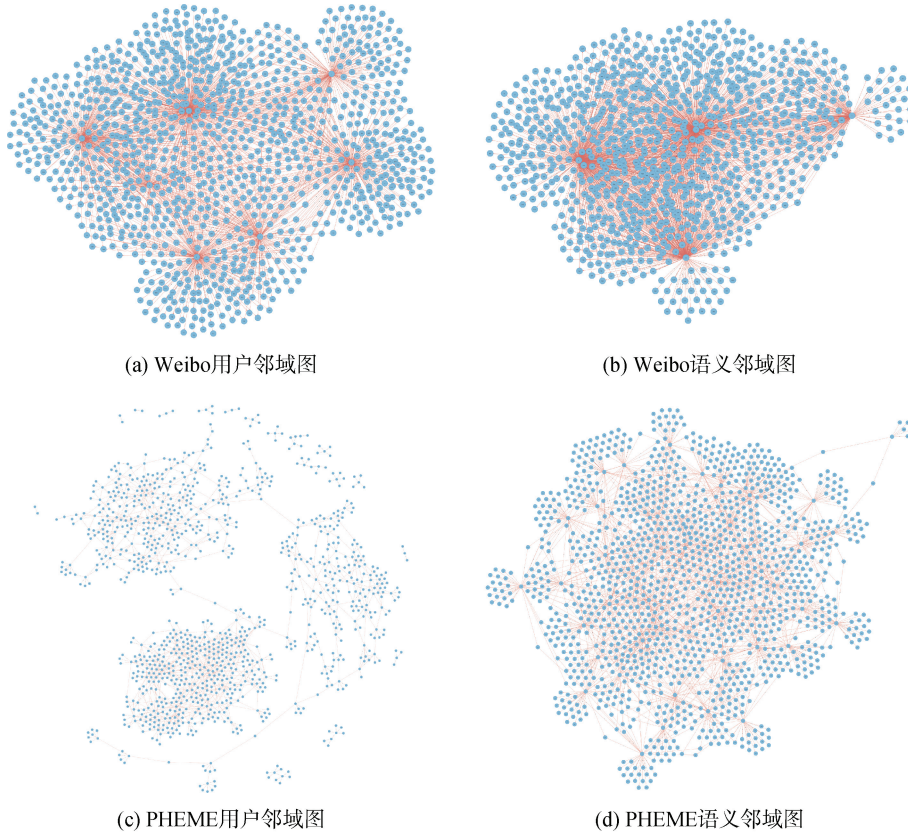


图 2 Weibo 和 PHEME 数据集上的邻域信息图

Figure 2 Neighborhood information graphs on Weibo and PHEME datasets

在语义信息构成的信息系统中, 第 i 个事件的语义信息表示为:

$$\tilde{h}_i^{sem} = (e_i^1, e_i^2, \dots, e_i^h, \dots, e_i^{m'}), \quad (12)$$

其中 e_i^j 为第 i 个事件的第 j 维语义特征, 总共有 m' 维。将 \tilde{h}_i^{sem} 转换到低维空间, 获得更加稠密的推文表示。定义如下:

$$\tilde{h}_i^{sem} = FC(\tilde{h}_i^{sem}), \quad (13)$$

其中, $FC(\cdot)$ 为全连接层, 可以将 \tilde{h}_i^{sem} 进行降维表示。降维后的 \tilde{h}_i^{sem} 可表示为:

$$\tilde{h}_i^{sem} = (e_i^1, e_i^2, \dots, e_i^h, \dots, e_i^m), \quad (14)$$

降维后维度为 m 维。

采用余弦相似度^[43-45]指标衡量两个文本之间的相似度。则 \tilde{h}_i^{sem} 和 \tilde{h}_j^{sem} 的相似度定义为:

$$S_{ij} = \frac{\sum_{h=k}^m (e_i^h \times e_j^h)}{\sqrt{\sum_{h=k}^m (e_i^h)^2} \times \sqrt{\sum_{h=k}^m (e_j^h)^2}}, \quad (15)$$

S_{ij} 值域为 $[-1, 1]$, 越接近 -1 , 两个对象越不相似, 反之越相似。

设置一个邻域半径阈值参数 $\gamma \in [0, 1]$, 根据式(1)和式(2)可以确定每个对象的邻域关系 R , 最终得到对象间的邻接矩阵 $U^{sem} \in \{0, 1\}^{n \times n}$, U^{sem} 中的元素 u_{ij} 满足:

$$u_{ij} = \begin{cases} 0, & \text{if } i = j; \\ 1, & \text{if } x_j \in N_i; \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

显然, 当 γ 值逐渐变大时, 构造的邻域信息图由密变疏。如图 2b 和 2d 所示, 展示了语义邻域图的 2000 条边。

4.4 信息重构模型

在构建的邻域信息图基础上, 通过构建交叉集成的网络模型, 捕获邻域信息图中对象间的语义关系与可信关系, 重新构造事件的语义信息与可信信息。注意, 这里的交叉是指在邻域图中信息的交叉传递, 即构造用户邻域图, 在图中传递的是语义信息; 构造语义邻域图, 但传递的是用户信息, 通过交叉集成的方式可以更好地结合这两种异质信息。

具体如图 1 所示, 首先构造两个子模型: 通过捕捉用户邻域图中潜在可信关系来重构语义信息的 RT-Sem 模型以及通过捕捉语义邻域图中潜在语义关系来重构用户信息的 RT-User 模型, 这两个模型都是基于图卷积神经网络构建的。然后通过加权集成这

两个交叉构建的子模型, 得到了本文的模型 RT-NIG 模型。

一般的, 图卷积层(Graph Convolutional Layer, GCL)的前向传播公式为:

$$h_i^{l+1} = \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} h_j^l W^{l+1} \right), \quad (17)$$

其中, h_i^{l+1} 代表第 i 个事件在 $l+1$ 层时的信息表示, 特别地, $h_i^0 = \tilde{h}_i^{user}$ 或者 $h_i^0 = \tilde{h}_i^{sem}$, 表示初始的信息表示。 $c_{ij} = \sqrt{d_i d_j}$, d_i, d_j 表示对象 i, j 的邻居数。

W^{l+1} 为第 $l+1$ 层的权重矩阵, $\sigma(\cdot)$ 为激活函数。为了防止出现过平滑^[46]问题, 加入残差连接, 将前一层节点属性跳跃连接到下一层, 用来增强初始的用户信息。如图 3 所示, 前向传播之后进行跳跃连接:

$$h_i^{l+1} = \text{concat}(h_i^{l+1}, h_i^l). \quad (18)$$

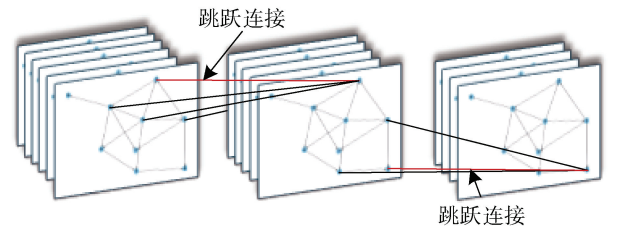


图 3 GCN 前向传播

Figure 3 GCN forward propagation

4.5 重构二元组信息的刻画与分类

由式 17 和式 18, 重构后的用户信息和语义信息分别为 h_i^{user} 和 h_i^{sem} , 通过对两个子模型的性能进行分析来确定权重系数。Sem-GCN 子模型与 User-GCN 子模型的准确率分别为 α 和 β , 则集成后的“用户-推文”二元组信息为:

$$h_i^{new} = \text{concat}(\eta \cdot h_i^{user}, (1-\eta) h_i^{sem}) \quad (19)$$

其中, $\eta = \alpha / (\alpha + \beta)$ 为权重系数。最后是将新二元组信息通过全连接层以及激活函数进行分类:

$$\hat{y} = \sigma(FC(h_i^{new})) \quad (20)$$

$\hat{y} \in R^{1 \times C}$ 为预测所有类标签的概率向量, C 为标签类别数。此外, 使用交叉熵作为损失函数来训练模型的参数, 并应用 L_2 正则化。

整个过程算法如算法 1 所示。

算法 1. RT-NIG 算法

输入: 原始数据 \mathcal{R} ; 标签集合 \mathcal{Y} ; 神经网络训练层

数, 轮数 L , E ; 邻域半径阈值参数 θ , γ ;

输出: 类概率向量 \hat{y} ;

//数据预处理和构造初始二元组信息;

1 $\mathcal{D} \leftarrow \mathcal{R}$;

2 **for** $i=1:n$ **do**

3 **for** $h=1:k$ **do**

4 $u_i^h \leftarrow \left(u_i^{h'} - \min_{i=1,2,\dots,n} \{u_i^{h'}\} \right) / \left(\max_{i=1,2,\dots,n} \{u_i^{h'}\} - \min_{i=1,2,\dots,n} \{u_i^{h'}\} \right) \in (0,1)$ Eq. (4);

5 **end**

6 **end**

7 **for** $i=1:n$ **do**

8 $\tilde{h}_i^{user} \leftarrow \{u_i^1, u_i^2, \dots, u_i^h, \dots, u_i^m\}$ Eq. (5);

9 $\tilde{h}_i^{sem} \leftarrow Bert(\mathcal{D}_i^{text})$ Eq. (6);

10 $\tilde{h}_i^{sem} \leftarrow FC(\tilde{h}_i^{sem})$ Eq. (13);

11 $\tilde{h}_i^{sem} \leftarrow (e_i^1, e_i^2, \dots, e_i^h, \dots, e_i^m)$ Eq. (14);

12 **end**

//构造邻域信息图;

13 **for** $i=1:n$ **do**

14 **for** $j=1:n$ **do**

15 $S_{ij} \leftarrow \sum_{h=1}^m (s_{ij}^h)^2$ Eq. (8);

16 $N_i \leftarrow \{x_j | x_i, x_j \in X, S_{ij} \geq \theta\} (i, j=1, 2, \dots, n)$

Eq. (1);

17 $U^{user} \leftarrow \begin{cases} 0, & \text{if } i=j; \\ 1, & \text{if } x_j \in N_i; \\ 0, & \text{otherwise,} \end{cases}$ Eq. (11);

18 $S_{ij} \leftarrow \frac{\sum_{h=k}^m (e_i^h \times e_j^h)}{\sqrt{\sum_{h=k}^m (e_i^h)^2} \times \sqrt{\sum_{h=k}^m (e_j^h)^2}}$ Eq. (15);

19 $N_i \leftarrow \{x_j | x_i, x_j \in X, S_{ij} \geq \gamma\} (i, j=1, 2, \dots, n)$

Eq. (1);

20 $U^{sem} \leftarrow \begin{cases} 0, & \text{if } i=j; \\ 1, & \text{if } x_j \in N_i; \\ 0, & \text{otherwise,} \end{cases}$ Eq. (16);

21 **end**

22 **end**

//构造信息重构模型;

23 $h_i^{user}, h_i^{sem} \leftarrow \begin{cases} h_i^{l+1} = \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} h_j^l W^{l+1} \right) \\ h_i^{l+1} = \text{concat}(h_i^{l+1}, h_i^l) \end{cases}$ Eq.

(17) and Eq. (18);

//重构二元组信息的刻画及分类;

24 $h_i^{new} \leftarrow \text{concat}(\eta \cdot h_i^{user}, (1-\eta) h_i^{sem})$ Eq. (19);

25 **for** $e=1, 2, \dots, E$ **do**

26 **for** $l=1, 2, \dots, L$ **do**

27 $\hat{y} \leftarrow \sigma(FC(h_i^{new}))$ Eq. (20);

28 **end**

29 **end**

5 实验和结果分析

在本节, 对比了 RT-NIG 和其他基线模型在不同数据集上的性能, 并进行了 RT-NIG、RT-User 与 RT-Sem 的消融研究以及超参数的设置。

5.1 数据集

本文使用了 Ma 等人^[47]提供的中文 Weibo 数据集以及 Zubiaga 等人^[48]提供的英文 PHEME 数据集。Weibo 数据集是谣言检测领域常见的数据集, 文献[49-53]等都使用了该数据集。PHEME 数据集是 Zubiaga 等人采集的 5 个有价值的媒体事件中的转发评论活动, 包括 Ferguson unrest(August 9, 2014)、Ottawa shooting(October 22, 2014)、Sydney siege(December 15, 2014)、Charlie Hebdo shooting(January 7, 2015)、Germanwings plane crash(March 24, 2015)等 5 个媒体事件。简单统计信息如表 2。

5.2 实验设置

(1) 基线方法

为了验证本文所提方法的有效性, 将 RT-NIG 模型与两类基线模型进行了对比实验。一类是经典机器学习算法, 包括最大熵^[54], 逻辑回归^[55], 岭回归^[56], 支持向量机^[57], 决策树^[58], GBDT^[59], 随机森林^[60]等; 一类是先进的深度学习方法, 包括 CNN^[35], CNN-LSTM^[39], CNN-GRU^[39], CNN-SimpleRNN^[39], BERT^[38], BERT-LSTM^[40], TextGCN^[61]等。特别地, 本文的早期检测研究的背景是在谣言在网络上刚刚发布但尚未开始传播时, 所有的算法都只使用了源账户的用户信息和文本信息, 没有任何传播信息。而当前的图神经网络模型都是在谣言开始传播后利用其传播结构进行检测, 所以在与其他图神经网络模型的对比上只选择了适用于本文场景的文本分类模型 TextGCN。

在对比算法实验设置中, 其他算法采用手工方法或者神经网络从文本中提取有效信息, 然后将用户信息与提取的文本信息拼接, 获得事件的信息表征, 最后送入到下游的谣言分类器中。这些算法有:

表 2 数据集统计信息
Table 2 Dataset Statistics

数据集	样本数	谣言数	非谣言数	用户特征数	数值型属性数	类别型属性数
Weibo 数据集	4664	2313	2351	12	6	6
Charlie Hebdo	2079	458	1621	8	6	2
Ferguson	1143	284	859	8	6	2
PHEME 数据集	Germanwings Crash	469	238	8	6	2
	Ottawa Shooting	890	470	8	6	2
	Sydney Siege	1221	522	8	6	2
	Total	5802	1972	3830	6	2

- 1) ME(IIS): 最大熵是一种在满足已知条件之后, 选择不确定性最大(不确定的部分是等可能的)的模型。
- 2) LR: 逻辑回归是一种广义的线性回归分析模型, 常用于二分类问题。
- 3) DT: 决策树递归地选择最优特征, 并根据该特征对训练数据进行分割, 使得对各个子数据集有一个最好的分类。
- 4) SVM: 支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器。
- 5) RR: 岭回归是一种专用于共线性数据分析的有偏估计回归方法。
- 6) GBDT: 梯度提升决策树是一种集成回归树的 Boosting 算法。
- 7) RF: 随机森林一种由决策树构成的 Bagging 算法。
- 8) CNN: 一种前馈神经网络, 由若干卷积层和池化层组成, 可挖掘文本深层的特征。
- 9) CNN-LSTM: 一种结合 CNN 和 LSTM 的新型混合深度学习模型, 可用于谣言分类。
- 10) CNN-GRU: 一种结合 CNN 和 GRU 的新型混合深度学习模型。
- 11) CNN-SimpleRNN: 一种结合 CNN 和 SimpleRNN 的新型混合深度学习模型。
- 12) BERT: 一种使用 BERT 预训练模型提取文本语义信息的假新闻检测模型。
- 13) BERT-LSTM: 一种结合 BERT 和 LSTM 的混合深度学习模型, 通过获得更高级的语义表示来进行文本的情感分析与分类。
- 14) TextGCN: 一种能够构建基于文本和词的异构图, 并应用 GCN 对文本进行分类的模型。

(2) 实验环境

本文采用 sklearn 机器学习库实现 LR、DT、SVM、RR、GBDT、RF 等传统机器学习的模型; 采用 NLTK 自然语言处理工具包实现 ME。在 TensorFlow 环境中实现 CNN、CNN-LSTM、CNN-GRU、

CNN-SimpleRNN、TextGCN 等深度学习模型; 在 PyTorch 环境中实现 BERT、BERT-LSTM 以及本文 RT-NIG 模型。所有模型都运行在 1 GeForce RTX 3090 GPU、64 GB RAM、16 coresIntel(R)Xeon(R)Gold 5218 CPU@2.30 GHz Linux 服务器上。

(3) 模型参数

所有算法包括本文算法均采用 5 折交叉验证, 并运行 10 次取平均值, 并采用早期训练停止。本文算法还配置了卷积层数(1)、学习率(0.01)、权重衰减(0.00005)、丢失率(0.5)、训练周期(50epoch)和优化器(Adam)。训练迭代周期超过 25 轮时, 进入早期停止判断阶段, 当损失值连续升高 10 轮, 训练停止。

(4) 评价指标

在两个真实数据集上, 采用以下指标进行评价: 整体准确度(Accuracy), 每个类别各自的精确度(Precision)、召回率(Recall)以及 F1 值。

5.3 总体性能

(1) 在 Weibo 和 PHEME 数据集上的对比实验

表 3 和表 4 展示了 RT-NIG 与基线方法在 Weibo 和 PHEME 两个真实数据集上的性能。具体分析如下:

1) 在 PHEME 数据集上, 深度学习全面优于传统机器学习方法, 说明了深度提取文本语义信息的有效性及其重要性, 但深度学习时间消耗普遍高于传统机器学习方法, 这是因为大部分时间消耗在了深度学习模型的训练部分。在 Weibo 数据集上, BERT、BERT-LSTM 等模型的效果依然优于传统的机器学习方法, 然而 CNN、CNN-LSTM、CNN-GRU 以及 CNN-SimpleRNN 等模型的效果反而没有传统机器学习好。为此, 对 Weibo 用户特征在两类标签下的概率密度分布进行了分析, 如图 4, 分别是互粉数、用户定位、关注朋友数、粉丝数、认证类型、发布微博数、推文跨越时间(从用户建立到发布推文)、收藏微博数等。可以看出这些都是强特征, 说明 Weibo 数据集统计特征更加明显, 适宜采用传统机器学习算法进行建模分析。此外, CNN 以及 CNN

表 3 Weibo 数据集上的谣言早期检测结果(T: 非谣言; F: 谣言)

Table 3 Early detection results of rumors on Weibo dataset (T: non-rumor; F: rumor)							
模型类别	基线方法	标签类	准确率	精确率	召回率	F1 值	消耗时间(min)
传统机器学习方法	ME(IIS)	T	0.8600	0.8868	0.8282	0.8564	6.15
		F		0.8364	0.8923	0.8634	
	LR	T	0.8707	0.8805	0.8605	0.8702	0.55
		F		0.8617	0.8811	0.8711	
	DT	T	0.8844	0.8854	0.8856	0.8853	0.04
		F		0.8839	0.8832	0.8834	
	SVM	T	0.8902	0.9630	0.8137	0.8819	0.29
		F		0.8365	0.9680	0.8974	
	RR	T	0.8904	0.9416	0.8346	0.8848	0.11
		F		0.8493	0.9472	0.8956	
	GBDT	T	0.9204	0.9517	0.8877	0.9183	0.62
		F		0.8935	0.9537	0.9224	
深度学习 方法	RF	T	0.9204	0.9547	0.8843	0.9180	0.43
		F		0.8908	0.9572	0.9227	
	CNN-GRU	T	0.8574	0.8708	0.8548	0.8627	1.24
		F		0.8433	0.8604	0.8518	
	CNN-SimpleRNN	T	0.8585	0.8324	0.9141	0.8713	0.68
		F		0.8939	0.7973	0.8428	
	CNN-LSTM	T	0.8714	0.8951	0.8548	0.8745	1.20
		F		0.8476	0.8896	0.8681	
	CNN	T	0.8875	0.9085	0.8732	0.8905	1.31
		F		0.8661	0.9032	0.8843	
	BERT	T	0.9273	0.9329	0.9219	0.9260	29.82
		F		0.9241	0.9320	0.9268	
	BERT-LSTM	T	0.9302	0.9310	0.9325	0.9309	46.0
		F		0.9302	0.9259	0.9271	
	TextGCN	T	0.8944	0.9288	0.8562	0.8910	3.99
		F		0.8647	0.9333	0.8976	
	RT-NIG ¹	T	0.9799	0.9674	0.9937	0.9803	34.25
		F		0.9934	0.9659	0.9794	+9.85

(注:¹RT-NIG 算法的前一个时间为构造邻域信息图的时间, 后一个时间为模型运行时间)

表 4 PHEME 数据集上的谣言早期检测结果(T: 非谣言; F: 谣言)

Table 4 Early detection results of rumors on PHEME dataset (T: non-rumor; F: rumor)							
模型类别	基线方法	标签类	准确率	精确率	召回率	F1 值	消耗时间(min)
传统机器学习方法	ME(IIS)	T	0.7044	0.7345	0.8650	0.7944	28.15
		F		0.6002	0.3925	0.4743	
	LR	T	0.6272	0.7303	0.6898	0.7094	1.01
		F		0.4570	0.5056	0.4800	
	DT	T	0.6682	0.7498	0.7467	0.7482	0.10
		F		0.5118	0.5157	0.5134	
	SVM	T	0.6579	0.7454	0.7316	0.7384	0.10
		F		0.4969	0.5147	0.5055	
	RR	T	0.7504	0.7629	0.9026	0.8268	0.26
		F		0.7065	0.4548	0.5530	
	GBDT	T	0.7528	0.7635	0.9065	0.8288	1.02
		F		0.7142	0.4544	0.5552	
	RF	T	0.7623	0.7686	0.9157	0.8357	0.57
		F		0.7397	0.4645	0.5705	

续表

模型类别	基线方法	标签类	准确率	精确率	召回率	F1 值	消耗时间(min)
深度学习方法	CNN-GRU	T	0.8475	0.8606	0.9179	0.8883	1.46
		F		0.8163	0.7107	0.7598	
	CNN-SimpleRNN	T	0.8424	0.8605	0.9087	0.8839	0.87
		F		0.8006	0.7132	0.7544	
	CNN-LSTM	T	0.8398	0.8591	0.9061	0.8820	1.48
		F		0.7955	0.7107	0.7507	
	CNN	T	0.8432	0.8580	0.9140	0.8851	0.76
		F		0.8081	0.7056	0.7534	
	BERT	T	0.8832	0.8820	0.9507	0.9148	33.82
		F		0.8892	0.7524	0.8140	
	BERT-LSTM	T	0.8992	0.8982	0.9556	0.9260	45.70
		F		0.9017	0.7899	0.8421	
	TextGCN	T	0.8289	0.8498	0.9042	0.8760	4.32
		F		0.7804	0.6771	0.7238	
	RT-NIG	T	0.9788	0.9771	0.9912	0.9841	54.22
		F		0.9831	0.9547	0.9684	+5.95

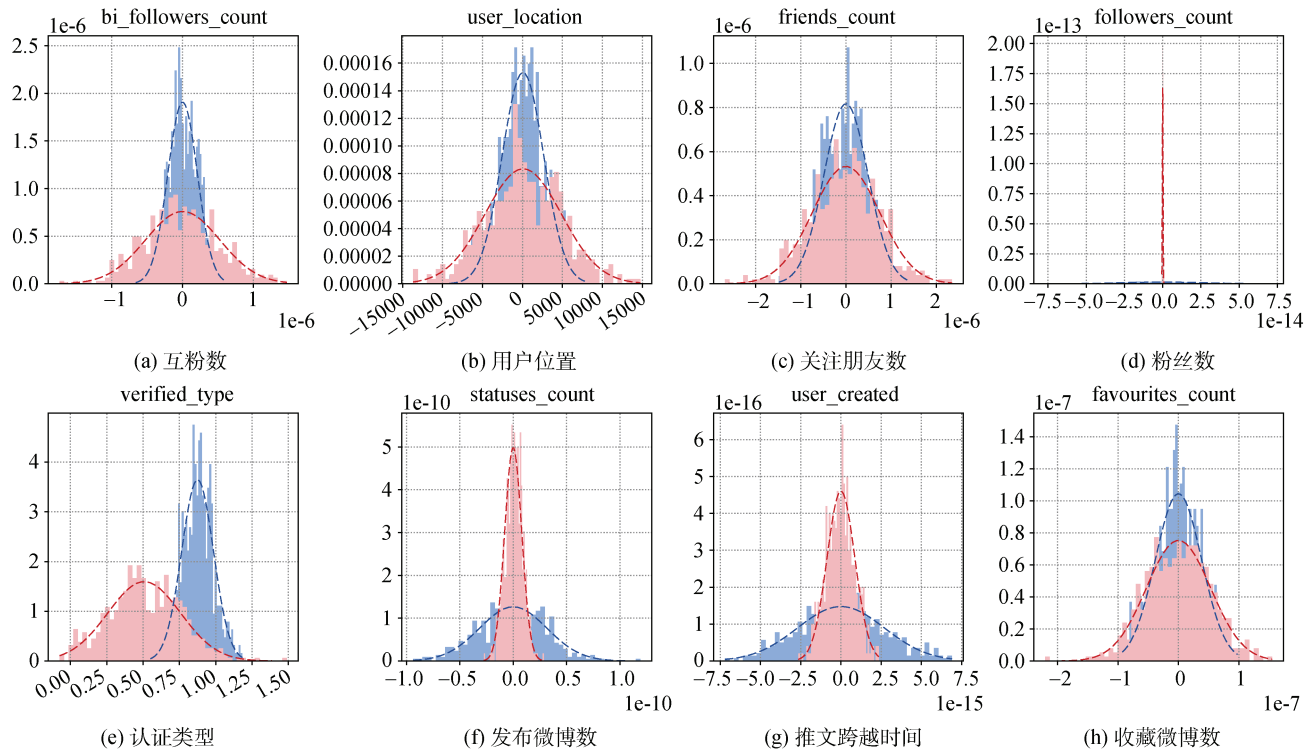


图 4 用户特征的概率密度分布(红色为谣言, 蓝色为非谣言)

Figure 4 Probability density distribution of user features (Red is rumor, blue is non-rumor)

和 RNN 的混合模型在提取文本语义信息上不如 BERT 预训练模型, 这可以理解, 毕竟 CNN 模型更适合图像处理, RNN 模型更适合音频处理, 而 BERT 模型更适合自然语言处理。

2) 从表 3 和表 4 可以看出, RT-NIG 模型性能优于所有的基线模型。尤其是在 PHEME 数据集上表现更加突出。RT-NIG 性能明显优于 TextGCN, 说明

RT-NIG 模型的构图模式以及交叉集成更加有效。BERT 模型使用初始二元组信息进行模型的训练与谣言分类, 而 RT-NIG 使用邻域信息图进行二元组的信息重构, 并在此基础上进行训练和预测。证明了图数据能够捕获对象之间潜在的相关性, 有着强大的表达能力。也证明了本文模型在早期谣言检测任务中的有效性和优越性。但是 RT-NIG 模型的时间消耗

比其他模型更长, 这是因为构建邻域图消耗了更多的时间, 在模型的训练和预测方面, 消耗时间反而不多。

(2) 消融研究

为了验证构建邻域信息图以重构信息的有效性, 首先对子模型 RT-User 与 RT-Sem 进行了消融研究。

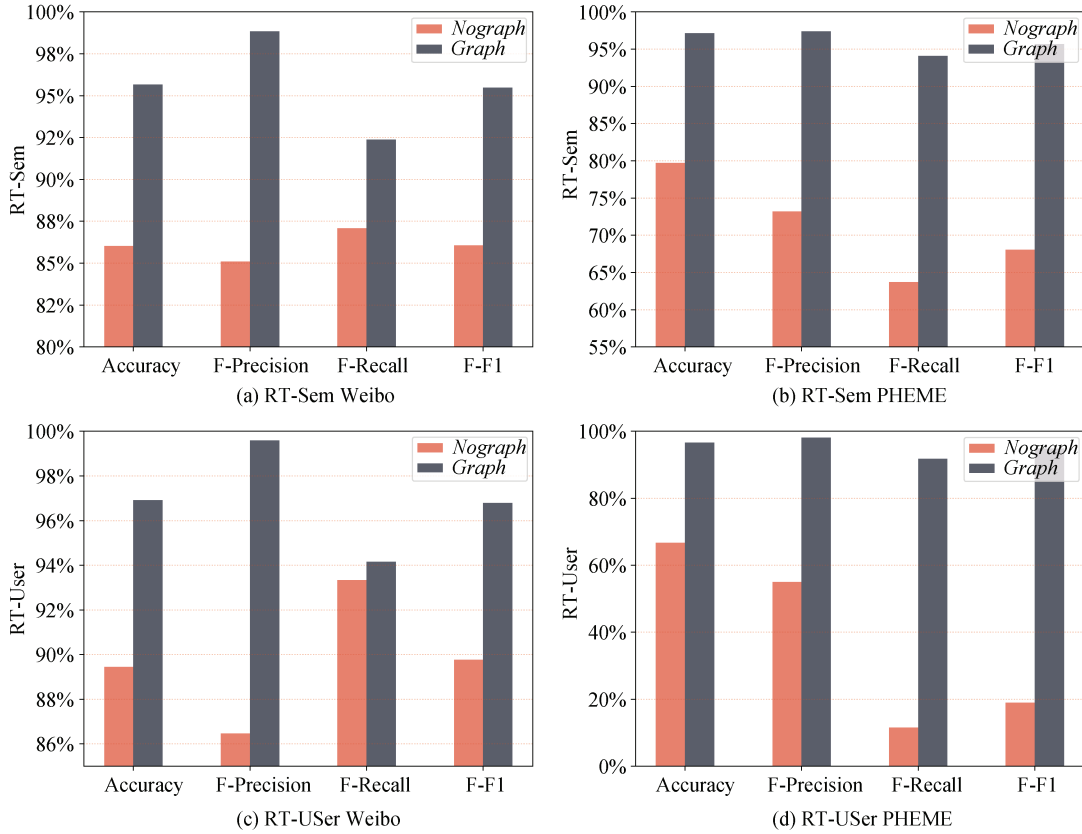


图 5 RT-User 与 RT-Sem 在构建邻域信息图与无图两种情况下的性能对比

Figure 5 The performance comparison between RT-User and RT-Sem in using neighborhood information graph and no graph

由图 5 可以看出, 采用邻域信息图重构用户信息和语义信息后, 模型的性能得到了极大的提升。说明在邻域信息图中传递和重构信息表示, 能够充分表达事件的模式特征, 有利于下游的分类任务。

如表 3、表 4 和图 5, 单一的 RT-User 与单一的 RT-Sem 模型性能都优于 TextGCN, 说明了遵循理论假设构造邻域图的有效性。同时, 构图和无图两类情景的对比也呼应了这两个假设。以 RT-User 子模型为例, 在 4.3 节中提到, 构图模式是按照抽象出的不同语义情感聚类成不同的簇, 并在其网络结构中传递用户信息, 结合了语义信息与用户信息, 而无图模式只使用了用户信息, 缺乏语义信息的利用, 两者之间的差异说明了通过语义情感的聚类可以更好地挖掘用户信息。

其次, 为了验证通过交叉集成来构造网络的有

对比了 RT-User 与 RT-Sem 在构建邻域信息图与无图两种情况下的性能。如图 5 所示。特别地, 为了突出对于谣言事件的检测性能, 本节之后的性能及参数分析采用 Accuracy、F-Precision、F-Recall、F-F1 等 4 个指标, 它们分别为整体准确率、以及代表谣言事件单类的精确率、召回率、F1 值。

效性, 对比了子模型 RT-User 与 RT-Sem 与集成模型 RT-NIG 的性能。如图 6 所示。

如图 6, RT-User 提取的是重构的用户信息, RT-Sem 提取的是重构的语义信息, RT-NIG 模型是集成两个子模型得到的。经过交叉集成后, RT-NIG 模型的性能普遍优于 RT-User 和 RT-Sem, 说明交叉集成的方式是有效的, 即重构后的用户信息和文本信息具有互相增益的效果, 两者缺一不可。而且在一些指标上, RT-User 性能优于 RT-Sem, 即重构的用户信息分类效果优于重构的语义信息。即使在部分指标上 RT-User 的效果不理想, 但从用户信息对语义信息的增益以及在部分指标上的优势也可以说明用户信息在早期检测中具有重要的作用, 不应只专注于通过自然语言技术提取语义信息, 而忽视了对用户信息的结合。从图 7 中可以看出, 经过集成后, RT-NIG

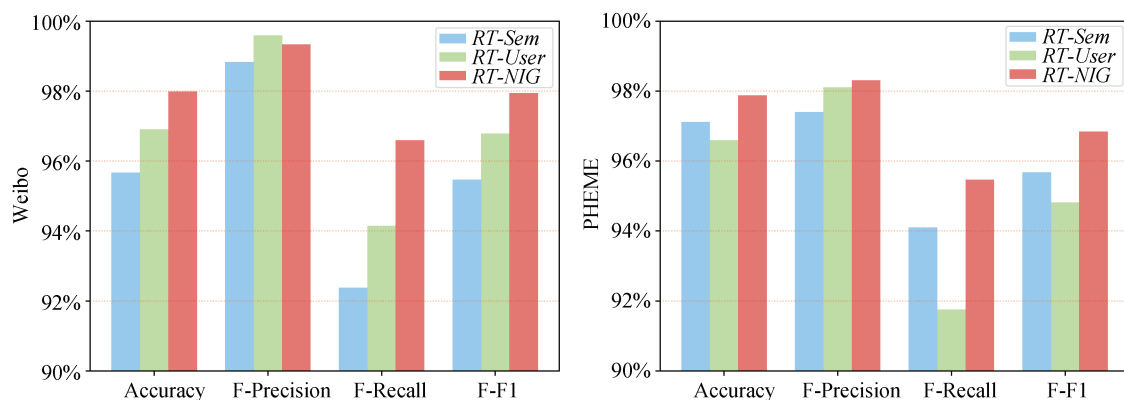


图 6 RT-NIG 及其子模型在两个数据集上的性能对比

Figure 6 Performance comparison of RT-NIG and its sub models on two datasets

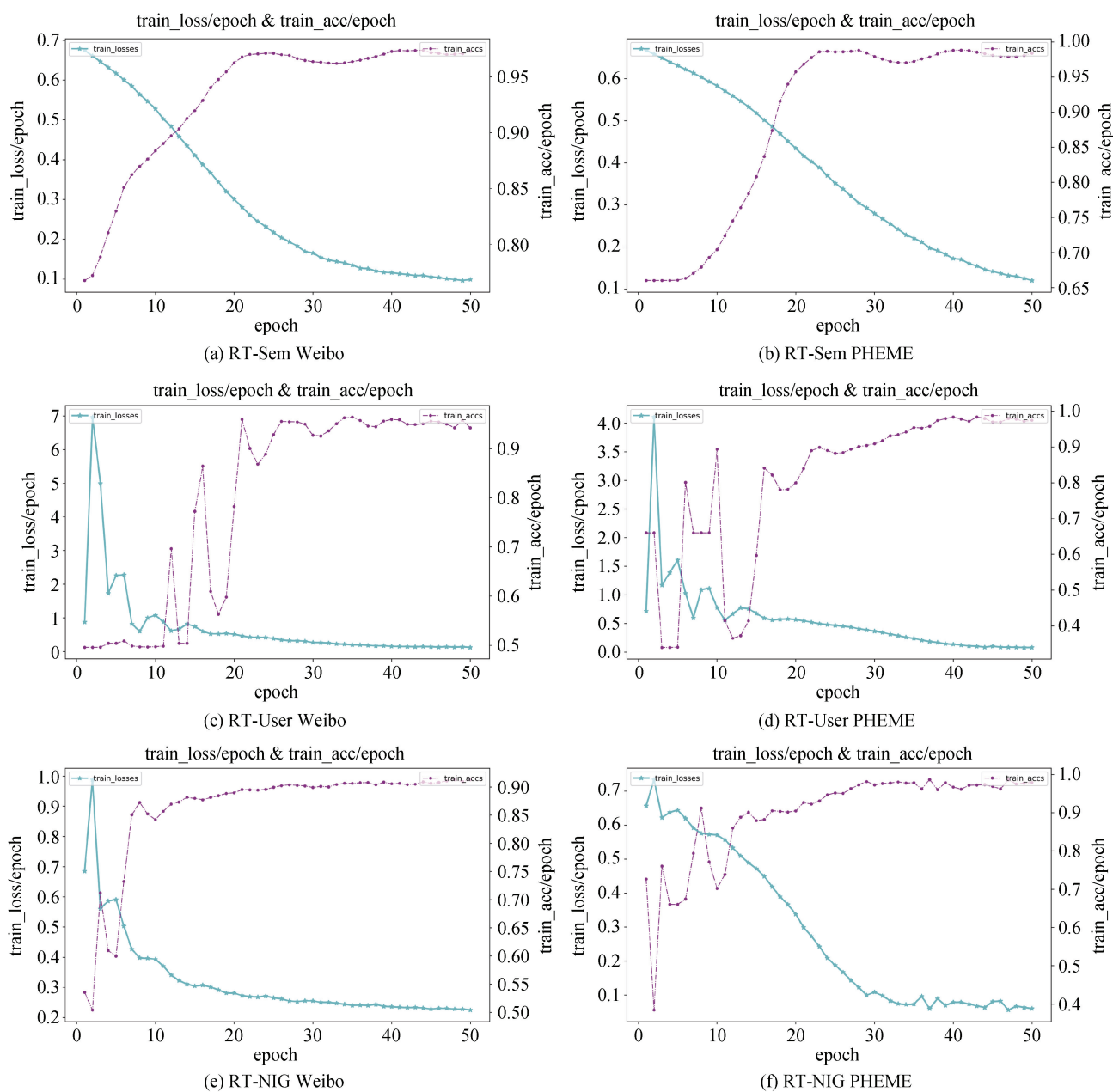


图 7 RT-NIG 及其子模型在两个数据集上训练时的损失以及准确率的变化情况

Figure 7 The changes of the loss and accuracy rate of RT-NIG and its sub models when training on two datasets

收敛速度比 RT-Sem 更快, 且训练比 RT-User 更平稳, 兼容了两个子模型训练时的优点。

5.4 超参数设置

本节分别就邻域半径阈值参数 θ 、 γ 以及训练集比例 μ 对性能的影响进行了分析。由于选择的阈值越大, 构造的邻域信息图越稀疏, 训练时间越少, 本文在[0.7,1]之间讨论阈值参数对模型性能的影响。需要注意的是, 图并不是越稀疏越好, 如果对象连边

过少的话, 部分对象将成为孤立节点, 无法完成信息传递和重构任务, 所以应该保证每个对象都能形成一定的邻居数。

(1) 邻域半径阈值参数 θ

本小节分析了 θ 在[0.7,1]上对子模型 RT-Sem 的影响, 如表 5 和表 6 所示, “连边数”表示构建的邻域信息图中的总边数, “最小连边数”表示所有对象中最少的连边数, 而“时间轮”表示每轮训练所需要的时间。

表 5 邻域半径阈值参数 θ 对 RT-Sem 性能的影响(Weibo dataset)

Table 5 The impact of neighborhood radius threshold parameter θ on the performance of RT-Sem (Weibo dataset)

邻域半径 阈值参数 θ	标签类	准确率	精确率	召回率	F1 值	连边数	最小连边数	时间/轮 (s)
0.99	T	0.9134	0.9117	0.9174	0.9144	106386	1	0.08
	F		0.9156	0.9093	0.9123			
0.95	T	0.9264	0.9080	0.9512	0.9288	541258	1	0.17
	F		0.9483	0.9013	0.9239			
0.90	T	0.9468	0.9219	0.9779	0.9489	1084847	3	0.23
	F		0.9763	0.9151	0.9445			
0.85	T	0.9514	0.9209	0.9892	0.9537	1628436	13	0.30
	F		0.9881	0.9130	0.9489			
0.80	T	0.9567	0.9301	0.9890	0.9584	2172025	27	0.34
	F		0.9883	0.9238	0.9547			
0.75	T	0.9528	0.9253	0.9866	0.9547	2715614	60	0.39
	F		0.9861	0.9184	0.9507			
0.70	T	0.9551	0.9247	0.9924	0.9571	3259204	91	0.41
	F		0.9920	0.9173	0.9529			

表 6 邻域半径阈值参数 θ 对 RT-Sem 性能的影响(PHEME dataset)

Table 6 The impact of neighborhood radius threshold parameter θ on the performance of RT-Sem (PHEME dataset)

邻域半径 阈值参数 θ	标签类	准确率	精确率	召回率	F1 值	连边数	最小连边数	时间/轮 (s)
0.99	T	0.8843	0.8952	0.9345	0.9143	165415	1	0.06
	F		0.8609	0.7871	0.8219			
0.95	T	0.9433	0.9468	0.9689	0.9576	838679	1	0.19
	F		0.9372	0.8934	0.9142			
0.90	T	0.9582	0.9627	0.9749	0.9685	1680259	1	0.33
	F		0.9514	0.9258	0.9377			
0.85	T	0.9711	0.9727	0.9841	0.9783	2521839	1	0.39
	F		0.9687	0.9461	0.9569			
0.80	T	0.9712	0.9704	0.9868	0.9784	3363419	1	0.45
	F		0.9740	0.9410	0.9568			
0.75	T	0.9668	0.9553	0.9809	0.9675	4204999	1	0.51
	F		0.9810	0.9525	0.9660			
0.70	T	0.9648	0.9672	0.9803	0.9735	5046579	1	0.58
	F		0.9625	0.9347	0.9476			

由表 5、表 6 以及图 8, 可以看出 θ 取值在 [0.75,0.8]之间时, 效果较好。为了能够兼顾训练时间

的要求, 在本文中设置 $\theta = 0.8$ 。

(2) 邻域半径阈值参数 γ

由表 7, 表 8 以及图 9, 可以看出 γ 取值在 $[0.70, 0.75]$ 之间时, 效果较好。为了能够兼顾训练时

间的要求, 在本文中设置 $\gamma = 0.75$ 。同时, 从表 5、6、7、8 以及图 8、9 可以看出, 不同邻域信息图的构造对模型的性能产生了很大的影响, 从另一个方面证明了构造邻域信息图以重构信息的有效性。

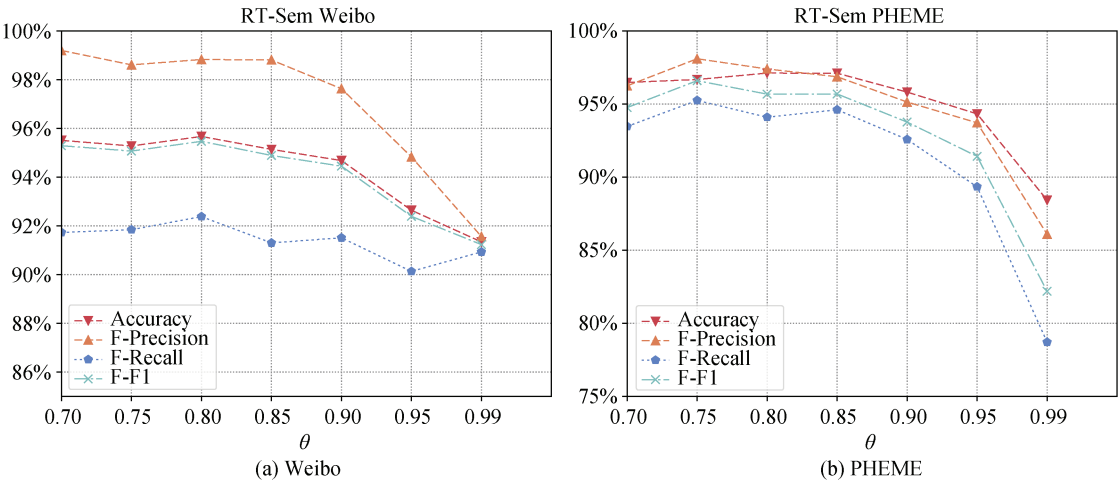


图 8 邻域半径阈值参数 θ 在两个数据集上对模型 RT-Sem 的影响

Figure 8 The impact of neighborhood radius threshold parameter θ on the model RT-Sem on two data sets

表 7 邻域半径阈值参数 γ 对 RT-User 性能的影响(Weibo dataset)

Table 7 The impact of neighborhood radius threshold parameter γ on the performance of RT-User (Weibo dataset)

邻域半径 阈值参数 γ	标签类	准确率	精确率	召回率	F1 值	连边数	最小连边数	时间/轮 (s)
0.99	T	0.8905	0.9302	0.8465	0.8863	382	1	0.02
	F		0.8570	0.9352	0.8944			
0.95	T	0.8927	0.9378	0.8433	0.8880	977	1	0.03
	F		0.8927	0.8556	0.9430			
0.90	T	0.8757	0.9057	0.8412	0.8721	6004	1	0.03
	F		0.8495	0.9108	0.8790			
0.85	T	0.8412	0.8268	0.8669	0.8463	97055	1	0.03
	F		0.8578	0.8151	0.8358			
0.80	T	0.9320	0.9032	0.9694	0.9350	679619	1	0.04
	F		0.9667	0.8939	0.9287			
0.75	T	0.9624	0.9358	0.9938	0.9639	2276541	1	0.05
	F		0.9933	0.9304	0.9608			
0.70	T	0.9691	0.9455	0.9962	0.9702	4716647	1	0.10
	F		0.9959	0.9415	0.9679			

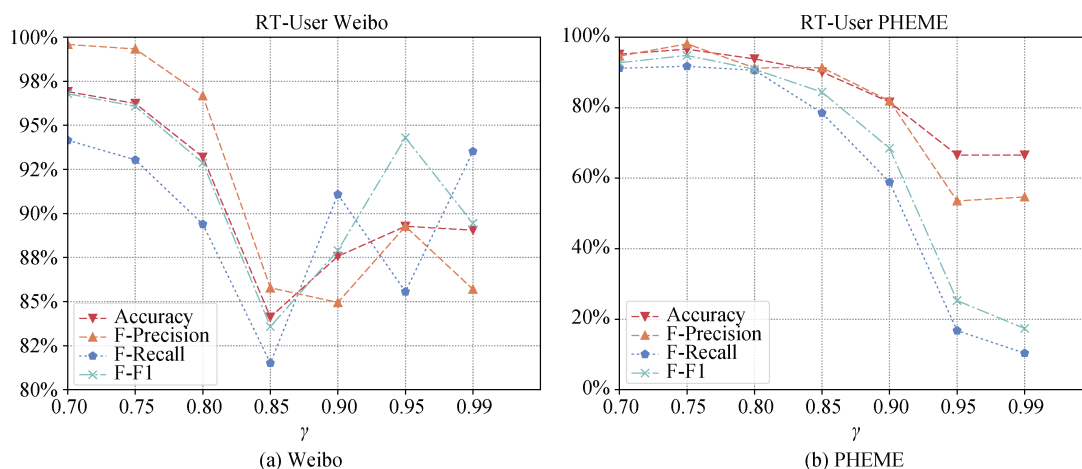
表 8 邻域半径阈值参数 γ 对 RT-User 性能的影响(PHEME dataset)

Table 8 The impact of neighborhood radius threshold parameter γ on the performance of RT-User (PHEME dataset)

邻域半径 阈值参数 γ	标签类	准确率	精确率	召回率	F1 值	连边数	最小连边数	时间/轮 (s)
0.99	T	0.6654	0.6739	0.9551	0.7902	248	1	0.01
	F		0.5467	0.1036	0.1738			
0.95	T	0.6658	0.6825	0.9227	0.7845	23070	1	0.02
	F		0.5354	0.1676	0.2528			

续表

邻域半径 阈值参数 γ	标签类	准确率	精确率	召回率	F1 值	连边数	最小连边数	时间/轮 (s)
0.90	T	0.8157	0.8148	0.9328	0.8697	1114386	1	0.03
	F		0.8190	0.5886	0.6847			
0.85	T	0.9014	0.8966	0.9615	0.9279	5817716	1	0.08
	F		0.9135	0.7848	0.8441			
0.80	T	0.9383	0.9521	0.9548	0.9533	11131545	1	0.15
	F		0.9129	0.9061	0.9088			
0.75	T	0.9659	0.9588	0.9909	0.9746	14366835	1	0.18
	F		0.9811	0.9175	0.9482			
0.70	T	0.9518	0.9558	0.9725	0.9638	15848378	1	0.20
	F		0.9462	0.9116	0.9275			

图9 邻域半径阈值参数 γ 在两个数据集上对模型 RT-User 的影响Figure 9 The impact of neighborhood radius threshold parameter γ on the model RT-User on two data sets

(3) 训练集比例 μ

一般来讲, 训练集比例越大, 模型拟合得越好, 测试效果越好。分别在两个数据集上测试了训练集比例 μ 对 RT-NIG 模型的影响。如图 10, 其中测试集比例为 $1-\mu$ 。

从图 10 可以看出, 随着训练集比例增大, 测试结果整体上是越来越好, 但是在一些指标上变化不是很明显, 说明 RT-NIG 模型在处理小比例训练数据时也能很好地拟合维度信息, 为下游分类任务提供良好的特征表示。也证明了基于邻域信息图的二元组信息重构技术的有效性。

6 结论

本文提出了一种谣言早期检测的新途径, 通过在交叉分布的邻域信息图中重新构造“用户-推文”二元组信息, 用于谣言分类。该方法首先构造初始的二元组信息; 然后遵循两个理论驱动的假设, 基于邻域信息系统, 分别构造了用户邻域图以及语义邻域

图; 利用 GCN 捕获邻域图中对象间潜在的可信度关系以及语义关系, 并传递和重构语义信息和用户信息; 最后通过交叉集成构建了 RT-NIG 模型, 用于二元组信息重构。该方法能够充分利用用户信息与语言信息, 并有效地结合, 在 Weibo 和 PHEME 两个真实谣言数据集上的表现优于先进的早期检测方法。通过消融以及超参数分析等进一步的研究, 证明了用户信息在早期谣言检测中的重要作用以及基于邻域信息图的二元组重构技术的有效性。

本文方法针对尚未开始传播的谣言进行检测, 在无传播信息可用的情况下, 通过构造邻域信息图重构二元组, 充分结合了源发布账户的用户信息以及文本信息, 提高了谣言的识别效果。这也为一些早期检测问题提供了新的解决方法, 例如假新闻、网络暴力、虚假信息等问题。但也存在一些不足, 比如输入数据图结构化会导致模型整体的时间消耗增加, 下一步还需要提出时间复杂度更小的构图算法, 另外, 还可以结合图片、视频等更多早期检测能利用的信息。

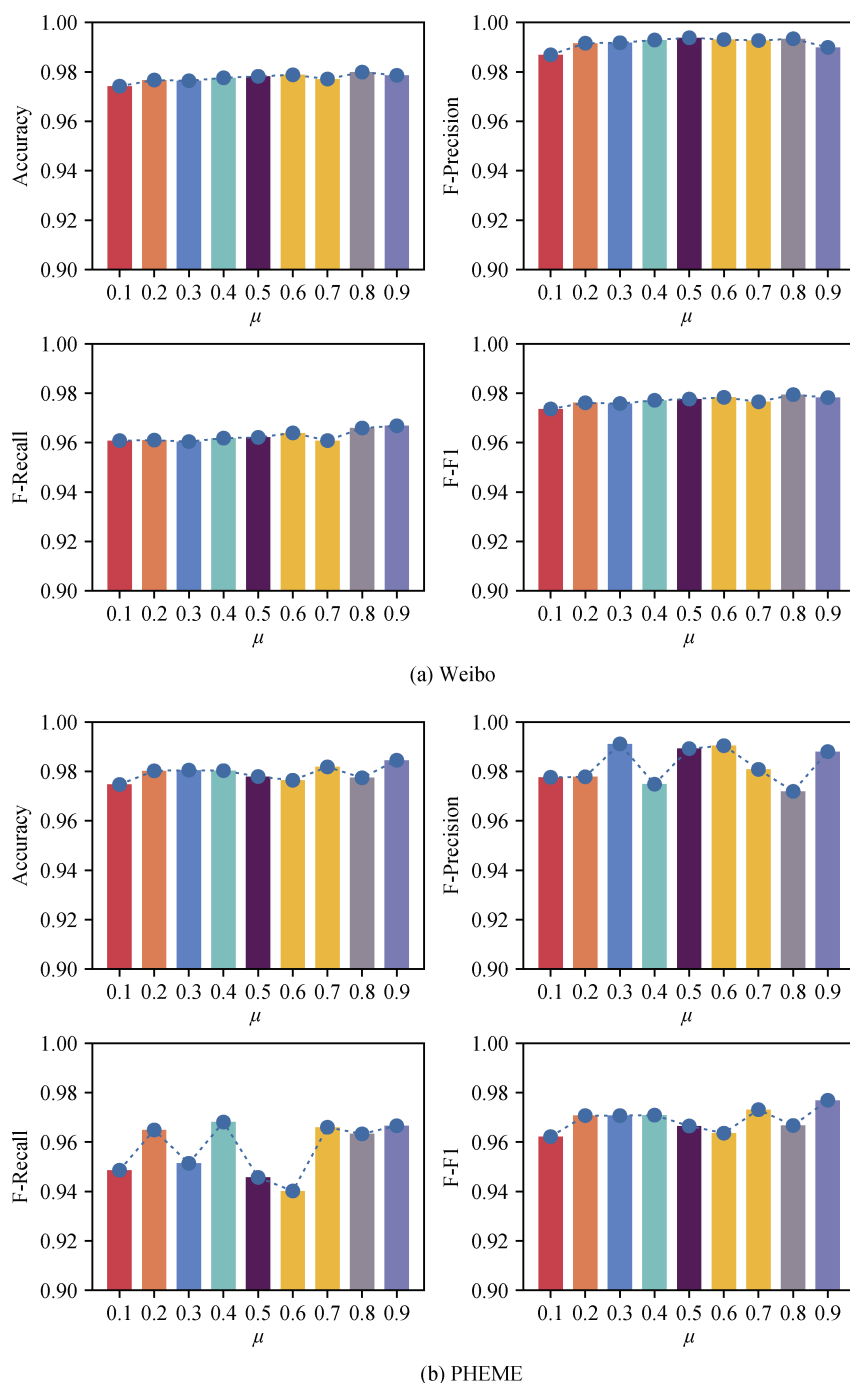


图 10 训练集比例 μ 在两个数据集上对模型 RT-NIG 的影响

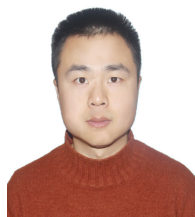
Figure 10 The impact of training set proportion μ on the model RT-NIG on two data sets

参考文献

- [1] Rosnow R L. Inside Rumor: A Personal Journey[J]. *American Psychologist*, 1991, 46(5): 484-496.
- [2] Bálint P, Bálint G. The Semmelweis-Reflex[J]. *Orvosi Hetilap*, 2009, 150(30): 1430.
- [3] Nickerson R S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises[J]. *Review of General Psychology*, 1998, 2(2): 175-220.
- [4] Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases[M]. *Judgment under Uncertainty*. Cambridge: Cambridge University Press, 1982: 3-20.
- [5] Del Vicario M, Quattrociocchi W, Scala A, et al. Polarization and Fake News[J]. *ACM Transactions on the Web*, 2019, 13(2): 1-22.
- [6] Pierri F, Piccardi C, Ceri S. Topology Comparison of Twitter Diffusion Networks Effectively Reveals Misleading Information[J]. *Scientific Reports*, 2020, 10(1): 1372.
- [7] Zhou X Y, Jain A, Phoha V V, et al. Fake News Early Detection[J]. *Digital Threats: Research and Practice*, 2020, 1(2): 1-25.
- [8] Kwon S, Cha M, Jung K. Rumor Detection over Varying Time

- Windows[J]. *PLoS ONE*, 2017, 12(1): e0168344.
- [9] Zhu Q S, Feng J, Huang J L. Weighted Natural Neighborhood Graph: An Adaptive Structure for Clustering and Outlier Detection with no Neighborhood Parameter[J]. *Cluster Computing*, 2016, 19(3): 1385-1397.
- [10] Wang C, Liu Z, Gao H, et al. VOS: A New Outlier Detection Model Using Virtual Graph[J]. *Knowledge-Based Systems*, 2019, 185: 104907.
- [11] Wang C, Gao H, Liu Z, et al. Outlier Detection Using Diverse Neighborhood Graphs[C]. *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 2018: 58-62.
- [12] Wang Y, Li Y P. Outlier Detection Based on Weighted Neighbourhood Information Network for Mixed-Valued Datasets[J]. *Information Sciences*, 2021, 564: 396-415.
- [13] Kwon S, Cha M, Jung K, et al. Prominent Features of Rumor Propagation in Online Social Media[C]. *2013 IEEE 13th International Conference on Data Mining*, 2013: 1103-1108.
- [14] Vosoughi S, Roy D, Aral S. The Spread of True and False News Online[J]. *Science*, 2018, 359(6380): 1146-1151.
- [15] Sunstein C R. On Rumors: How Falsehoods Spread, Why We Believe Them, and What Can Be Done[M]. Princeton, NJ: Princeton University Press, 2014.
- [16] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]. *The 20th international conference on World wide web*, 2011: 675-684.
- [17] Gupta A, Kumaraguru P. Credibility Ranking of Tweets during High Impact Events[C]. *The 1st Workshop on Privacy and Security in Online Social Media*, 2012: 2-8.
- [18] Ruchansky N, Seo S, Liu Y. CSI: A Hybrid Deep Model for Fake News Detection[C]. *The 2017 ACM on Conference on Information and Knowledge Management*, 2017: 797-806.
- [19] Dong X, Gabrilovich E, Heitz G, et al. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion[C]. *The 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014: 601-610.
- [20] Nickel M, Murphy K, Tresp V, et al. A Review of Relational Machine Learning for Knowledge Graphs[J]. *Proceedings of the IEEE*, 2016, 104(1): 11-33.
- [21] Zhou X, Zafarani R. Fake news: A survey of research, detection methods, and opportunities [J]. *arXiv preprint arXiv:181200315*, 2018.
- [22] Potthast M, Kiesel J, Reinartz K, et al. A Stylometric Inquiry into Hyperpartisan and Fake News[J]. *ArXiv e-Prints*, 2017: arXiv: 1702.05638.
- [23] Castelo S, Almeida T, Elghafari A, et al. A Topic-Agnostic Approach for Identifying Fake News Pages[C]. *Companion Proceedings of The 2019 World Wide Web Conference*, 2019: 975-980.
- [24] Riedel B, Augenstein I, Spithourakis G P, et al. A Simple but Tough-to-Beat Baseline for the Fake News Challenge Stance Detection Task[J]. *ArXiv e-Prints*, 2017: arXiv: 1707.03264.
- [25] Takahashi T, Igata N. Rumor Detection on Twitter[C]. *The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems*, 2012: 452-457.
- [26] Vogel I, Meghana M. Fake News Spreader Detection on Twitter Using Character N-Grams. Notebook for PAN at CLEF 2020 [EB/OL]. 2020: 2009.13859. <https://arxiv.org/abs/2009.13859v1>.
- [27] Chen Y X, Sui J, Hu L, et al. Attention-Residual Network with CNN for Rumor Detection[C]. *The 28th ACM International Conference on Information and Knowledge Management*, 2019: 1121-1130.
- [28] Liu Z H, Xiong C Y, Sun M S, et al. Fine-Grained Fact Verification with Kernel Graph Attention Network[C]. *The 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 7342-7351.
- [29] Zhang M S, Ge X Y, Wei B. Rumor Detection Based on Emotional Features and Types of Rumors[J]. *Computer Simulation*, 2023, 40(6): 21-25, 60.
(张明书, 葛晓义, 魏彬. 基于情感特征和谣言种类的谣言检测[J]. *计算机仿真*, 2023, 40(6): 21-25, 60.)
- [30] Zhao Z, Resnick P, Mei Q Z. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts[C]. *The 24th International Conference on World Wide Web*, 2015: 1395-1405.
- [31] Cai G Y, Wu H, Lv R. Rumors Detection in Chinese via Crowd Responses[C]. *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014: 912-917.
- [32] Yang F, Liu Y, Yu X H, et al. Automatic Detection of Rumor on Sina Weibo[C]. *The ACM SIGKDD Workshop on Mining Data Semantics*, 2012: 1-7.
- [33] Wu K, Yang S, Zhu K Q. False Rumors Detection on Sina Weibo by Propagation Structures[C]. *2015 IEEE 31st International Conference on Data Engineering*, 2015: 651-662.
- [34] Yang Z F, Wang C, Zhang F, et al. Emerging Rumor Identification for Social Media with Hot Topic Detection[C]. *2015 12th Web Information System and Application Conference*, 2015: 53-58.
- [35] Liu Z, Wei Z H, Zhang R X. Rumor Detection Based on Convolutional Neural Network[J]. *Journal of Computer Applications*, 2017, 37(11): 3053-3056, 3100.
(刘政, 卫志华, 张韧弦. 基于卷积神经网络的谣言检测[J]. *计算机应用*, 2017, 37(11): 3053-3056, 3100.)
- [36] Chen T, Li X, Yin H Z, et al. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection[M]. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2018: 40-52.
- [37] Miao X, Rao D N, Jiang Z H. Syntax and Sentiment Enhanced BERT for Earliest Rumor Detection[M]. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2021: 570-582.
- [38] Heidari M, Zad S, Hajibabae P, et al. BERT Model for Fake News Detection Based on Social Bot Activities in the COVID-19 Pandemic[C]. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, 2021: 103-109.
- [39] Nasir J A, Khan O S, Varlamis I. Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach[J]. *International Journal of Information Management Data Insights*, 2021, 1(1): 100007.
- [40] Shen J, Xu M. Sentiment Analysis of Chinese Comments on OTA Website Using \square BERT and LSTM[C]. *International Conference on Cloud Computing, Internet of Things, and Computer Applications*, 2022: 388-392.
- [41] Al-Taie M Z, Kadry S, Lucas J P. Online Data Preprocessing: A

- Case Study Approach[J]. *International Journal of Electrical and Computer Engineering*, 2019, 9(4): 2620.
- [42] Wilson D R, Martinez T R. Improved Heterogeneous Distance Functions[J]. *Journal of Artificial Intelligence Research*, 1997, 6: 1-34.
- [43] Rahutomo F, Kitasuka T, Aritsugi M. Semantic cosine similarity [C]. *Proceedings of the The 7th international student conference on advanced science and technology ICAST*, 2012.
- [44] Xia P P, Zhang L, Li F Z. Learning Similarity with Cosine Similarity Ensemble[J]. *Information Sciences*, 2015, 307: 39-52.
- [45] Nguyen H V, Bai L. Cosine Similarity Metric Learning for Face Verification[M]. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 709-720.
- [46] Chen D L, Lin Y K, Li W, et al. Measuring and Relieving the Over-Smoothing Problem for Graph Neural Networks from the Topological View[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(4): 3438-3445.
- [47] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [J]. *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2016: 3818-3824.
- [48] Zubiaga A, Liakata M, Procter R. Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media[J]. *ArXiv e-Prints*, 2016: arXiv: 1610.07363.
- [49] Bian T, Xiao X, Xu T Y, et al. Rumor Detection on Social Media with Bi-Directional Graph Convolutional Networks[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 549-556.
- [50] Ye A S, Wang L N, Wang R, et al. An End-to-End Rumor Detection Model Based on Feature Aggregation[J]. *Complexity*, 2021, 2021(1): 1-16.
- [51] Song C G, Shu K, Wu B. Temporally Evolving Graph Neural Network for Fake News Detection[J]. *Information Processing & Management*, 2021, 58(6): 102712.
- [52] Rao D N, Miao X, Jiang Z H, et al. STANKER: Stacking Network Based on Level-Grained Attention-Masked BERT for Rumor Detection on Social Media[C]. *The 2021 Conference on Empirical Methods in Natural Language Processing*, 2021: 3347-3363.
- [53] Ma T H, Zhou H H, Tian Y, et al. A Novel Rumor Detection Algorithm Based on Entity Recognition, Sentence Reconfiguration, and Ordinary Differential Equation Network[J]. *Neurocomputing*, 2021, 447: 224-234.
- [54] Wu N. The maximum entropy method [M]. *Springer Science & Business Media*, 2012.
- [55] LaValley M P. Logistic Regression[J]. *Circulation*, 2008, 117(18): 2395-2399.
- [56] Mcdonald G C. Ridge regression [J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2009, 1(1): 93-100.
- [57] Noble W S. What Is a Support Vector Machine?[J]. *Nature Biotechnology*, 2006, 24(12): 1565-1567.
- [58] Quinlan J R. Learning Decision Tree Classifiers[J]. *ACM Computing Surveys*, 1996, 28(1): 71-72.
- [59] Friedman J, Tibshirani R, Hastie T. Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors)[J]. *The Annals of Statistics*, 2000, 28(2): 337-407.
- [60] Breiman L. Random forests [J]. *Machine learning*, 2001, 45(1): 5-32.
- [61] Yin C, Yuan M. GCN-Based Text Classification Research[M]. *Communications in Computer and Information Science*. Cham: Springer International Publishing, 2022: 55-66.



王标 于 2012 年在解放军信息工程大学电子工程专业获得学士学位。现在战略支援部队信息工程大学网络空间安全专业攻读硕士学位。研究兴趣包括: 虚假舆情检测、图神经网络、粗糙集理论。Email: wangbiao9911@163.com



卫红权 于 2014 年在解放军信息工程大学通信与信息系统专业获得博士学位。现任国家数字交换系统工程技术研究中心研究员。研究领域为融合网络安全、可重构网络理论与技术等。Email: whq@ndsc.com.cn



刘树新 于 2016 年在解放军信息工程大学获得博士学位。现为国家数字交换系统工程技术研究中心助理研究员。研究领域为复杂网络、链路预测、通信网络安全。Email: liushuxin11@126.com



王凯 于 2019 年在战略支援部队信息工程大学军事信息学专业获得博士学位。现为国家数字交换系统工程技术研究中心副研究员。研究领域为网络安全治理、通信网络安全等。Email: wangkai0508@126.com



江昊聪 于 2016 年在英国布里斯托大学无线通信与信号处理专业获得硕士学位。现为国家数字交换系统工程技术研究中心助理研究员。研究领域为通信网络安全等。Email: nancilia@163.com



李燃 于 2020 年在郑州大学运筹学与控制论专业获得硕士学位。现在战略支援部队信息工程大学网络空间安全专业攻读博士学位。研究领域为粗糙集理论, 信息论, 数据挖掘等。Email: liran9955@163.com