

一种加权图卷积神经网络的新浪微博谣言检测方法

王昕岩¹, 宋玉蓉¹, 宋波²

¹(南京邮电大学 自动化学院、人工智能学院, 南京 210023)

²(南京邮电大学 现代邮政学院, 南京 210003)

E-mail: wangxy11@foxmail.com

摘要: 信息化时代下, 广泛传播的谣言极大地影响了人们的日常生活, 甚至威胁了社会稳定, 因此对谣言的检测任务具有现实意义. 目前基于深度学习模型的谣言检测方法忽略了事件之间的联系或事件之间联系的紧密程度, 对检测效果造成了一定影响. 本文考虑事件之间联系的异质性, 将事件之间联系的紧密程度描述为连边权重, 提出了一种基于加权图卷积神经网络 (Weighted-Graph Convolutional Network, W-GCN) 模型的新浪微博谣言检测方法. 该方法通过 W-GCN 模型, 学习得到节点的隐层表示, 进而对节点进行分类, 最终完成谣言检测任务. 实验结果表明, 与现有的谣言检测方法相比, 本文提出的基于 W-GCN 模型的谣言检测方法, 可以提高谣言检测的正确率、精确率、召回率和 F1 值, 即能更有效地识别谣言.

关键词: 图卷积神经网络; 连边权重; 新浪微博; 谣言检测

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2021)08-1780-07

Sina Microblog Rumor Detection Method Based on Weighted-graph Convolutional Network

WANG Xin-yan¹, SONG Yu-rong¹, SONG Bo²

¹(College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

²(School of Modern Posts & Institute of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In the era of information, widespread rumors greatly affect the daily life of people and even threaten social stability. So it has practical significance for the task of rumor detection. At present, rumor detection methods based on the deep learning model ignore the connection between events or the closeness of the connection, which affects the detection results in a certain degree. Considering the heterogeneity of the connection between events, describing the closeness of the connection as the weight of the edge, we propose a new Sina Microblog Rumor Detection method based on Weighted-Graph Convolutional Network (W-GCN). In this method, we use W-GCN model to study the hidden representation of the nodes, and then classifies the nodes to finally complete the rumor detection task. Experiment results show that, compared with the existing rumor detection methods, the rumor detection method based on W-GCN model we proposed can improve the accuracy, precision, recall and F1-measure of rumor detection, that is to say, it can be more effective for rumor identification.

Key words: graph convolutional network; edge weight; sina microblog; rumor detection

1 引言

谣言是在人和人之间传播的, 含有公众关心信息的一种特殊陈述, 而其真实性不能及时得到证明或是无法证明^[1], 往往会带来恶劣的影响. 例如新冠肺炎期间, 美国总统特朗普在得知消毒剂能够在一分钟之内杀死病毒后, 呼吁给新冠病毒患者直接注射消毒剂, 或者用强光照, 以起到杀死病毒的作用. 据美国疾控中心报告显示, 家用清洁剂产品中毒的数量在几周内都在上升. 在人们的日常生活中, 这类误导性的谣言比比皆是, 因此为了避免谣言可能造成的伤害, 采取谣言检测的措施是十分必要的.

谣言能够在社会中快速而广泛地传播, 与社交媒体的快速发展是密切相关的. 社交媒体在提高人们信息获取效率的

同时, 也产生了许多与真实信息相背的谣言信息^[2]. 由于每个普通人都可以成为信息的发布源, 就会造成信息质量的良莠不齐, 不可避免地会有不法分子为了达到目的恶意地发布谣言信息误导民众, 使其做出符合谣言散布者期望的举动, 影响了社会的稳定. 近年来, 社交媒体的谣言治理已经成为热门研究课题, 引起了国内外学者的广泛关注. 对社交媒体的谣言检测任务, 通过收集与事件相关的各种信息, 自动地完成检测任务来判断被检测事件的真伪. 对谣言事件的自动检测方法, 一方面减轻了人工辟谣的资源消耗, 另一方面能更快地检测出流散的谣言, 从而能够营造积极健康的网络环境, 维护社会的稳定^[3].

目前针对社交媒体的谣言检测方法, 如传统的机器学习方法, 以及深度学习方法包括卷积神经网络 (Convolutional

收稿日期: 2020-07-14 收修改稿日期: 2020-08-28 基金项目: 国家自然科学基金项目 (61672298) 资助; 江苏高校哲学社会科学研究重点项目 (2018SJDZ1142) 资助; 教育部人文社会科学研究规划基金项目 (17YJAZH071) 资助. 作者简介: 王昕岩, 男, 1996 年生, 硕士, 研究方向为复杂网络、谣言检测; 宋玉蓉, 女, 1971 年生, 博士, 教授, 博士生导师, CCF 会员, 研究方向为复杂网络中信息传播及其控制策略、自适应网络建模、仿真与智能优化; 宋波, 男, 1987 年生, 博士, 讲师, 研究方向为复杂网络传播动力学、自适应网络建模与分析、信息安全.

Neural Network, CNN)^[4]、循环神经网络(Recurrent Neural Network, RNN)^[5]等,对事件的学习只是将这些事件看作成独立的个体,平等地学习事件的特征。然而在社交媒体中,事件之间的关系并不是独立的,而是具有拓扑关系的。社交媒体中的每个事件和与该事件有关的用户都存在着联系,包括发布、评论、转发等,并且这些用户之间的交互关系,也是与这些用户有关的事件之间关系的体现。只考虑每个事件本身的特征而忽略事件之间的联系,会对检测效果造成影响。因此有学者考虑了事件之间的联系,提出使用图网络模型来进行谣言检测,但是他们忽略了事件之间联系的紧密程度,也会对检测效果造成一定的影响。在社交网络中,事件之间的联系具有异质性,其紧密程度存在着差异。谣言事件往往具有鼓动性,这是由于大部分的谣言事件的发布者在发布谣言时,擅长使用一些煽动性的语句来激发人们的情绪。往往对这些事件进行评论和转发的用户,更加容易受到这些刺激性语句的蛊惑,也会受到其他谣言事件中这类语句的鼓动。与非谣言相比,谣言的发布者更加希望所发布的谣言信息能够有更多转发和评论,来引起人们的关注。于是许多谣言发布者为了达到这一目的,会有意识地组织用户进行转发、评论来提高事件的热度,而这些用户也更易受到其他谣言发布者的利用。因此本文考虑事件之间相同数量的评论者或转发者,来确定事件之间联系的紧密程度。

为了充分考虑事件之间的联系及其异质性,本文提出了一种基于加权图卷积神经网络(Weighted-Graph Convolutional Network, W-GCN)模型的谣言检测方法。该方法以事件作为节点,以事件之间是否有联系构建连边并根据事件之间联系的紧密程度,赋予连边不同的权重。接着通过 BERT(Bidirectional Encoder Representation from Transformers)模型和 CNN 模型对事件的文本提取特征,并将该特征作为 W-GCN 模型中节点的特征向量,来完成模型的训练,从而完成对社交媒体中事件的谣言检测任务。考虑到新浪微博是一种基于用户关系信息分享、传播以及获取的实时社交媒体,是中国最大的社交媒体平台,因此本文选择新浪微博作为研究对象来完成谣言检测的任务。实验结果表明,与传统的机器学习模型以及 GRU^[5]、PPC-RNN + CNN^[6]等深度学习模型相比,本文提出的 W-GCN 模型能更有效地完成新浪微博中谣言的检测任务。并且与这些方法相比,本文方法同样能更加有效地完成新浪微博中谣言的早期检测任务。

本文剩余章节安排如下:第 2 节介绍社交媒体谣言检测的相关研究工作;第 3 节介绍本文提出的新的谣言检测模型;第 4 节通过实验及分析,验证本文提出的新的谣言检测方法的有效性;第 5 节为结束语。

2 相关工作

目前国内外对于谣言检测的工作主要分为两个方面,一个是使用传统的机器学习方法,另一个是使用深度学习的方法。使用传统的机器学习方法大多是通过选取人工特征,主要包括文本特征、用户特征以及传播特征,并使用包括贝叶斯^[7]、支持向量机(Support Vector Machine, SVM)、决策树^[8]等分类器来对事件进行学习分类。Qazvinian 等^[9]使用了贝叶

斯分类器,提取了用户行为特征及文本特征等特征。Yang 等^[10]使用了 SVM 分类器,提取了发布微博的客户端及其所在的地理位置等特征。Liang 等^[11]提取了质疑文本特征以及用户行为等特征。Yang 等^[12]提取了热门话题等特征。Wu 等^[13]提出了基于图核的混合 SVM 分类器,提取了语义特征以及高阶传播特征。Ma 等^[14]提出了传播树模型来获取事件的特征表示。

近年来随着深度学习的快速发展,深度学习模型包括卷积神经网络(CNN)、循环神经网络(RNN)等在计算机视觉^[15]等领域上的应用得到了快速发展,使用深度学习模型来解决自然语言处理(Natural Language Processing, NLP)任务^[16],逐步取得了显著的成果。并且与传统的机器学习模型相比,深度学习模型对人工特征的依赖较小。因此在 NLP 任务中,使用深度学习模型可以减少设计特征带来的人力、物力的消耗,并能提高模型的鲁棒性。Ma 等^[5]首次提出将深度学习模型 RNN 应用于社交媒体的谣言检测,首先通过 TF-IDF(Term Frequency-Inverse Document Frequency)算法来获取时间段内的微博文本向量,接着通过 RNN 模型的学习,最后获取文本向量的特征表示。Yu 等^[17]使用生成句向量的方法来获取微博文本向量,再用 CNN 模型来学习获取文本向量的特征表示。Liao 等^[18]提取文本特征和用户局部特征,使用 GRU(Gated Recurrent Unit)网络和注意力机制来获取微博事件的特征表示。Hu 等^[19]使用图卷积神经网络(Graph Convolutional Network, GCN)模型,根据新闻的相似性构建连边,并提取历史信用特征对新闻进行分类。Tian^[20]等人使用 GCN 模型,根据谣言传播的方向性构建连边对事件进行分类。虽然这些方法在分类效果上比传统的机器学习方法更加出色,但是它们没有考虑事件之间的联系或事件之间联系的异质性,对谣言检测的效果造成了一定的影响。

3 模型方法

3.1 问题描述

对新浪微博平台的谣言检测任务,实际上是一个对事件进行二分类的问题。对于给定的事件组 $E = \{e_1; e_2; e_3; e_4; \dots; e_i; \dots; e_n\}$ 和标签组 $L = \{l_1; l_2; l_3; l_4; \dots; l_i; \dots; l_n\}$ 。在事件组 E 中, e_i 表示第 i 个事件,包括所有与这个事件相关的微博原文以及转发、评论等信息, n 表示事件总数;在标签组 L 中, l_i 表示第 i 个事件的标签,分为谣言标签和非谣言标签。在新浪微博平台的谣言检测任务中,首先将事件组 E 分为事件训练集 E_{train} 和事件测试集 E_{test} ,将标签组分为训练集标签 L_{train} 和测试集标签 L_{test} 。让模型对事件训练集 E_{train} 以及它们对应的训练集标签 L_{train} 进行学习,构建出一个以事件信息 e_i 作为输入,分类标签 l_i 作为输出的映射模型。接着将事件测试集 E_{test} 输入该模型,输出对应的预测标签组 L_{pred} 。最后根据预测标签组 L_{pred} 与实际标签组 L_{test} 的相符程度,来衡量该谣言检测模型的检测效果。

3.2 模型介绍

1) BERT 预训练模型

自然语言处理(NLP)任务一般分为上游任务和下游任务。上游任务主要对数据进行预处理,将文本数据转化为向量

表示.下游任务主要对转化后的向量进行操作.上游任务使用的模型主要包括 Word2Vec^[21]、ELMo^[22]等. Word2Vec 模型将单词通过高维空间映射成词向量,可以表现现实生活中词语之间的关系.但是通过 Word2Vec 模型,每个单词与一个固定的词向量相对应,使得每个单词无法理解上下文的语义. ELMo 模型通过双向语言模型,并根据具体的输入来得到上下文依赖的词向量表示,将 NLP 的一部分下游任务转移到了上游任务中,增强了向量的特征表示能力.但是该模型的工作效率低,并且只能将数据用词向量来表示. BERT 模型^[23]作为一个语言表征模型,由谷歌团队使用双向编码操作,通过大量数据和多层模型训练而成.该模型的并行性好,效率高.并且该模型可以将数据转化为维度具有一致性的句向量,因此本文采用 BERT 模型对新浪微博中的文本数据进行预处理操作.

2) 卷积神经网络(CNN)模型

卷积神经网络(CNN)模型^[24]可以使用通过 BERT 模型转化后的向量,来作为模型的输入,进而完成模型的训练,如图 1 所示. CNN 模型由 4 部分构成:输入层(Input Layer)、卷积层(Convolution Layer)、池化层(Pooling Layer)和全连接层(Fully Connected Layer).

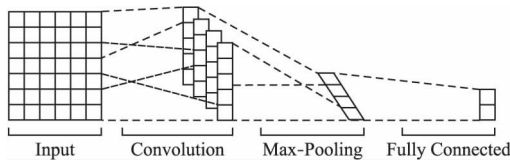


图 1 卷积神经网络(CNN)结构图

Fig. 1 Structure of convolutional neural network(CNN)

CNN 模型将事件 e_i 转化后的向量 v_i 作为输入,通过卷积层(Convolution Layer)选取宽度与向量 v_i 维度等长的卷积核,使该卷积核始终保持在竖直方向滑动,来完成对向量的卷积操作,继而得到与卷积核数量相同的特征向量 c_i .接着通过池化层(Pooling Layer)对特征向量 c_i 进行最大池化(Max-Pooling)操作,提取局部接受域中向量的最大值,得到特征向量 p_i .最后通过全连接层,将最大池化操作后提取的特征向量 p_i ,在全连接层中与分类标签进行全连接(Fully Connected)操作,得到该事件 e_i 的预测标签 l_i ,最终完成谣言的检测任务.本文提出的模型将最大池化操作后的特征向量 p_i 作为事件 e_i 的特征向量.

3) 图卷积神经网络(GCN)

图卷积神经网络^[25](GCN)是一种多层神经网络,可以通过图中的节点和连边信息直接对图中的节点进行端对端的学习并完成分类任务.给定一个图 $G=(N,E)$,其中 N 表示节点, E 表示连边. GCN 模型的输入包括特征矩阵 $X \in \mathbf{R}^{n \times m}$,邻接矩阵 $A \in \mathbf{R}^{n \times n}$ 以及标签组 $L \in \mathbf{R}^{n \times 2}$.其中特征矩阵 X 包括图中的所有节点 n 以及每个节点特征向量的维数 m .邻接矩阵 A 表示图中各节点之间的连边关系.如果节点 n_i 和节点 n_j 之间有连边,则 $A_{ij} = \alpha$,其中 α 为该连边的权重,否则 $A_{ij} = 0$. GCN 模型通过节点的相邻节点信息来更新该节点的隐层信息,每一层只捕获该节点的最近邻节点的信息.当 GCN 模型有多个层数时,模型会整合更多的邻居节点信息.因此 GCN 模型中的第 i 个隐藏层的特征矩阵 H_i ,如式(1)所示.

$$H_i = f(H_{i-1}, A) \quad (1)$$

其中, $H_0 = X$, f 是传播函数.

GCN 模型的第 i 个隐藏层的特征矩阵 H_i ,与特征矩阵 $H_i \in \mathbf{R}^{n \times m}$ 相对应,代表这个节点在该层的特征表示.第 i 个隐藏层的特征矩阵 H_i 中节点的特征被聚合后再通过传播函数 f ,得到该节点在下一个隐藏层中的特征表示.对于只含有一个隐藏层的 GCN 模型,节点的 k 维特征矩阵 $H_1 \in \mathbf{R}^{n \times k}$ 如式(2)所示.

$$H_1 = \rho(\tilde{A} X W_0) \quad (2)$$

其中 ρ 为激活函数; $W_0 \in \mathbf{R}^{m \times k}$ 是 GCN 模型第 0 层的权重矩阵; \tilde{A} 是 GCN 模型中标准化对称邻接矩阵,如式(3)所示.

$$\tilde{A} = D^{-1/2} A D^{-1/2} \quad (3)$$

对于多层 GCN 模型,下一个隐藏层的特征矩阵 H_{i+1} 可以通过上一个隐藏层的特征矩阵 H_i 及权重矩阵 W_i ,如式(4)所示.

$$H_{i+1} = \rho(\tilde{A} H_i W_i) \quad (4)$$

3.3 加权图卷积神经网络(W-GCN)模型

1) W-GCN 模型介绍

本文在考虑事件之间联系的基础上,充分考虑了事件之间联系的异质性,即事件之间联系的紧密程度,赋予了 GCN 模型中连边的权重,提出了一种新型社交网络的谣言检测方法.该方法对微博事件中的用户信息以及文本信息进行提取处理.对于用户信息,通过考虑用户与事件之间的评论与转发行为来确定事件与事件之间是否有联系关系,并根据事件之间相同的评论者或转发者数量来确定事件之间联系的紧密程度.将新浪微博中的事件作为节点,事件之间的联系作为连边,联系的紧密程度作为连边的权重,从而搭建了一个加权图卷积神经网络(W-GCN)模型,如图 2 所示.对于微博事件中的文本信息,首先通过 BERT 模型将文本信息进行预处理操作,将文本信息转化成维数为 768 的向量表示,紧接着使用 CNN 模型对转化后向量进行特征提取.最后将 CNN 模型中最大池化(Max-Pooling)操作后得到的特征输入到 W-GCN 模型中,最后模型学习得到事件的隐层表达,进而实现对事件的分类操作,最终完成新浪微博中谣言的检测任务.

本文提出的加权图卷积神经网络模型 W-GCN 的输入包括两个部分:由所有事件的特征向量组成的节点的特征矩阵 X ,以及根据事件之间的联系及其紧密程度构建的 W-GCN 中的邻接矩阵 A .

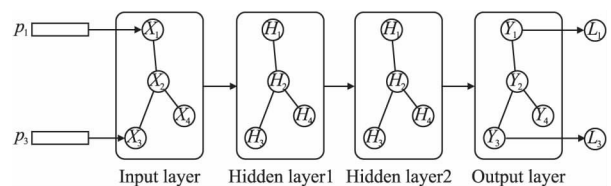


图 2 加权图卷积神经网络(W-GCN)结构图

Fig. 2 Structure of Weighted-Graph Convolutional Network(W-GCN)

2) W-GCN 模型的节点特征矩阵

考虑到 GCN 中,每个节点的特征向量需为固定的长度.

仅使用预训练模型对一个微博事件的所有文本信息进行预处理转化为向量,无疑会破坏文本的结构特征,会对检测效果造成巨大影响. 经过实验,仅通过预训练模型对一个微博事件的所有文本信息进行预处理,不具有很好的分类特征. 考虑到现有的深度学习模型中, CNN 模型在经过最大池化(Max-Pooling) 操作后得到的特征向量, 维度相同并且维数较少, 可以最大程度地保留文本的结构特征, 因此本文使用卷积神经网络模型(CNN) 对事件文本进行特征提取. 对于图网络模型中作为节点 n_i 的事件 e_i , 首先将事件 e_i 的文本信息, 逐条通过 BERT 模型转化为向量表示, 并作为 CNN 模型的输入. 经过 CNN 的卷积和池化操作后得到特征 p_i , 将其作为 W-GCN 模型中节点的特征向量 x_i .

3) W-GCN 模型的邻接矩阵

W-GCN 模型中节点 n_i 和节点 n_j 之间的连边, 由对应的事件 e_i 和事件 e_j 之间的联系来确定. 对于事件 e_i 和事件 e_j , 根据这两个事件之间是否有相同的评论者或转发者来确定两个事件之间是否存在联系, 继而决定图网络模型中节点 n_i 和节点 n_j 之间是否存在连边. 节点之间的连边通过邻接矩阵 A 来表示, 为了简便起见, 本文不考虑事件传播的方向, 即不考虑连边的有向性. 如果事件 e_i 和事件 e_j 之间没有相同的评论者或转发者, 那么这两个事件之间不构成传播关系, 即事件 e_i 和事件 e_j 之间不存在联系, 定义节点 n_i 和节点 n_j 之间不存在连边, 令 $A_{ij} = 0$; 如果事件 e_i 和事件 e_j 之间含有相同的评论者或转发者, 那么事件 e_i 和事件 e_j 之间就可能存在传播关系, 对此定义节点 n_i 和节点 n_j 之间存在连边, 那么 A_{ij} 的值至少为 1. 再通过考虑事件 e_i 和事件 e_j 之间联系的紧密程度, 即事件 e_i 和事件 e_j 之间相同的评论者或转发者数量 α , 给 A_{ij} 赋予不同的值来定义节点 n_i 和节点 n_j 之间连边的权重, 即:

$$A_{ij} = \begin{cases} \alpha, & \text{相同的评论或转发者数量} \\ 0, & \text{无相同的评论或转发者} \end{cases}$$

4 实验结果及分析

4.1 参数及评价指标

1) 实验数据集及模型参数设置

本文选用的数据集是 Ma 等^[5] 使用的用于新浪微博平台的谣言检测数据集. 该数据集包括 4664 个事件, 每个事件对

表 1 数据集参数表

| Statistic | Number |
|-------------------------------|---------|
| Events | 4664 |
| Rumors | 2313 |
| Non-rumors | 2351 |
| Microblogs | 3805656 |
| Users | 2746818 |
| Average time length/event(h) | 2460.7 |
| Average of posts/event | 816 |
| Max of posts/event | 59318 |
| Min of posts/event | 10 |

应的标签, 事件信息包括微博用户信息、文本信息、时间等信息, 如表 1 所示.

在模型的参数设置上, 本文将该数据集 4664 个事件中的 85% 作为训练集, 15% 作为测试集, 并完成了 5 次 5 折交叉验证. 对 BERT 模型, 设置向量的维度为 768; 对卷积神经网络(CNN) 模型, 设置卷积核的宽度设为 3, 长度为 768, 卷积核数量为 300, dropout 为 0.5. 对加权图卷积神经网络(W-GCN) 模型, 沿用 Kipf 等^[25] 的设置, 第一个隐藏层单元数为 16, 第二个隐藏层单元数为 7, dropout 设置为 0.5. 模型在学习时, 在每个 Epoch 中迭代训练所有的训练集事件, 直到模型收敛或者 Epoch 数量达到最大值为止. 此外, 本文选取了以小时作为单位的 9 个时间点, 分别为 1, 3, 6, 12, 24, 36, 48, 72, 96. 从事件中微博原文发出的时间算起, 在以上的时间节点内, 分别通过 W-GCN 模型来完成新浪微博的谣言检测任务, 来检验该模型在谣言早期检测方面的有效性.

2) 实验评估参数

谣言检测任务实际上是一个文本分类的子任务, 是一种二分类的文本分类问题, 输出的结果为谣言及非谣言的概率, 选取概率较大的一类作为模型预测的类别. 模型的预测类别与实际类别的分类依据, 如表 2 所示.

表 2 分类依据

| | 预测为谣言 | 预测为非谣言 |
|--------|-------|--------|
| 实际为谣言 | TP | FN |
| 实际为非谣言 | FP | TN |

本文选取了评价分类器优良最常见的 4 个指标: 正确率(Accuracy, Acc)、精确率(Precision, Pre)、召回率(Recall, Rec)、F1 值(F1-Measure, F1), 来评价本文提出的 W-GCN 模型的有效性. 正确率表示被模型正确分类的事件数量与事件总数的比值, 如式(5) 所示; 精确率表示被模型正确分类为谣言(或非谣言) 的事件数量与被分类为谣言(或非谣言) 的事件总数的比值, 如式(6) 所示; 召回率表示被模型正确分类为谣言(或非谣言) 的事件数量与实际为谣言(或非谣言) 的事件总数的比值, 如式(7) 所示; F1 值表示准确率和召回率加权调和平均, 是对准确率和召回率的综合考虑, 如式(8) 所示.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Pre = \frac{TP}{TP + FP} \quad (6)$$

$$Rec = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (8)$$

4.2 实验对比模型

本文将本文方法与同样使用该数据集的基准方法进行对比, 所选取的基准方法如下:

1) DT-Rank 模型^[26]. 该模型通过搜索微博文本中有争议的事实性声明, 将它们按相似度进行聚类并根据统计特征, 使用决策树的排名方法对聚类结果进行排名, 来完成微博事件的分类任务.

2) SVM-TS 模型^[27]. 该模型通过使用事件的时间序列来得到随时间变化的事件特征, 使用线性 SVM 分类模型来完成微博事件的分类任务.

3) GRU-2 模型^[5]. 该模型基于树结构递归神经网络, 捕捉相关帖子随时间变化的上下文信息, 使用 GRU 模型通过传播结构学习事件特征, 来完成微博事件的分类任务.

4) PPC_RNN + CNN 模型^[6]. 该模型根据事件的传播路径, 将循环神经网络(RNN)模型与卷积神经网络(CNN)模型进行模型融合, 使用融合后的模型学习事件特征, 来完成微博事件的分类任务.

4.3 实验结果

1) 谣言检测效果对比

由于 GCN 的特性, 本文在具体实验时, 对事件的训练集 E_{train} 只将其中的 150 个微博事件的标签设置为已知标签, 实验结果如表 3 所示. 表格中, 第 1 列为所选取的基准方法的模型及本文的模型, 从上至下分别为: DT-Rank, SVM-TS, GRU-2, PPC_RNN + CNN 和 W-GCN 模型; 第 2 列为事件的类别, 其中 R 表示谣言(Rumor), N 表示非谣言(Non-rumor); 第 3 列为正确率; 第 4 列为精确率; 第 5 列为召回率; 第 6 列为 F1 值.

表 3 谣言检测结果

Table 3 Rumor detection results

| Method | Class | Acc | Pre | Rec | F1 |
|---------------|-------|-------|-------|-------|-------|
| DT-Rank | R | 0.732 | 0.738 | 0.715 | 0.726 |
| | N | | 0.726 | 0.749 | 0.737 |
| SVM-TS | R | 0.857 | 0.839 | 0.885 | 0.861 |
| | N | | 0.878 | 0.83 | 0.857 |
| GRU-2 | R | 0.910 | 0.876 | 0.956 | 0.914 |
| | N | | 0.952 | 0.864 | 0.906 |
| PPC_RNN + CNN | R | 0.916 | 0.884 | 0.957 | 0.919 |
| | N | | 0.955 | 0.876 | 0.913 |
| W-GCN | R | 0.956 | 0.937 | 0.977 | 0.957 |
| | N | | 0.976 | 0.934 | 0.955 |

从结果中可以看到在基准方法中, 传统的机器学习方法中 SVM-TS 模型的检测效果最好, 可以达到 0.857 的正确率; 深度学习方法中 PPC_RNN + CNN 模型的检测效果最好, 可以达到 0.916 的正确率, 可见在谣言检测任务中, 使用深度学习的模型可以取得比传统的机器学习方法更加显著的效果. 而本文提出的 W-GCN 模型在检测结果上, 正确率达到了 0.956, 且各项评估指标均优于上述基准方法. 而本文在谣言检测的评估指标中, 召回率可以达到 0.977, 表明本文方法对谣言事件有更高的敏感性. 因此本文模型在谣言检测的任务上, 降低了谣言事件的误判率, 能够尽可能多地识别出可能的谣言事件. 模型的谣言检测评估指标中的召回率数值更高, 意味着更多的谣言事件被筛选出来, 能够更大程度地识别出社交网络中潜藏的谣言事件, 来减少潜藏谣言事件带来的不利影响, 这对社交媒体是十分必要的. 综上所述, 本文方法在对新浪微博中的谣言检测任务中表现出了更好的检测效果.

与此同时, 由于社交媒体信息量的快速增长, 对社交媒体中的事件予以谣言或非谣言的标签, 往往需要花费大量的时间与精力. 因此社交媒体中的事件, 拥有标签的事件大多数是已经过时的事件, 只有极少部分的事件是较新的事件. 因此在对新浪微博的谣言检测任务中, 传统的机器学习方法和深度学习方法包括 CNN、RNN 等在事件标签数量不足的情况下, 检测效果会大幅下降, 泛化能力较弱. 本文提出的模型能够在

标签数量有限的情况下, 达到同样的检测效果. 综上所述, 与现有的谣言检测方法相比, 本文提出的模型在新浪微博的谣言检测任务中具有更好的实用性.

2) 5 折交叉验证结果

本文使用 5 次 5 折交叉验证, 进一步验证本文模型的谣言检测效果. 在对数据集的划分中, 本文将数据集中的 85% 划分为训练集, 15% 划分为测试集. 在交叉验证环节中, 本文将训练集的数据平均分成 5 份, 每次验证实验只取其中的 1 份, 并将其作为测试集, 其余 4 份作为训练集, 通过 W-GCN 模型来得到这份数据的谣言检测结果, 并记录各项评估指标. 将该过程重复进行 5 次, 最终得到所有作为本文训练集数据的检测结果, 并记录各项评估指标, 结果如表 4 所示.

表 4 5 折交叉验证结果

Table 4 5-fold cross validation results

| Number | Class | Acc | Pre | Rec | F1 |
|---------|-------|-------|-------|-------|-------|
| 1 | R | 0.942 | 0.912 | 0.976 | 0.943 |
| | N | | 0.975 | 0.909 | 0.941 |
| 2 | R | 0.952 | 0.922 | 0.988 | 0.954 |
| | N | | 0.987 | 0.917 | 0.951 |
| 3 | R | 0.966 | 0.942 | 0.991 | 0.966 |
| | N | | 0.991 | 0.942 | 0.966 |
| 4 | R | 0.948 | 0.923 | 0.976 | 0.948 |
| | N | | 0.975 | 0.921 | 0.947 |
| 5 | R | 0.950 | 0.915 | 0.993 | 0.952 |
| | N | | 0.992 | 0.908 | 0.948 |
| average | R | 0.952 | 0.922 | 0.985 | 0.952 |
| | N | | 0.984 | 0.919 | 0.951 |

结果显示, 5 次 5 折的交叉验证结果中, 正确率最大值为 0.966, 最小值为 0.942, 相差 2.4%, 正确率的平均值为 0.952. 本文测试集实验结果的正确率高于这 5 次交叉验证结果的最小值, 低于这 5 次交叉验证结果的最大值, 略高于这 5 次交叉验证结果的平均值. 由此可见本文模型在新浪微博的谣言检测任务中具有较强的泛化能力.

3) 早期谣言检测效果对比

本文在谣言的早期检测方面与其他基准方法进行了对比, 结果如图 3 所示.

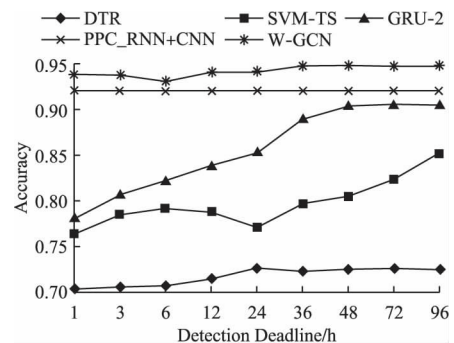


图 3 谣言的早期检测结果

Fig. 3 Result of rumor early detection

在谣言的早期检测任务中, 只能使用从微博原文发布的时间算起之后的若干小时内的微博信息. 为了评估本文模型

与其他基准方法在谣言检测早期性的表现,本文选取了以小时作为单位的9个时间点,分别为1,3,6,12,24,36,48,72,96小时.可以看出与现有的谣言检测方法相比,本文提出的模型在谣言的早期检测任务上同样表现出了更好的效果.

4) 与无权图卷积神经网络模型的效果对比

为了分析连边的权重对检测效果的影响,我们将本文模型与无权图卷积神经网络(Unweighted-Graph Convolutional Network, UW-GCN)模型进行对比,实验结果如图4所示.通过比较W-GCN模型与UW-GCN模型在选取的9个时间点上谣言检测的正确率,不难看出考虑事件之间联系的紧密程度,能够对谣言检测的效果有较大提升.

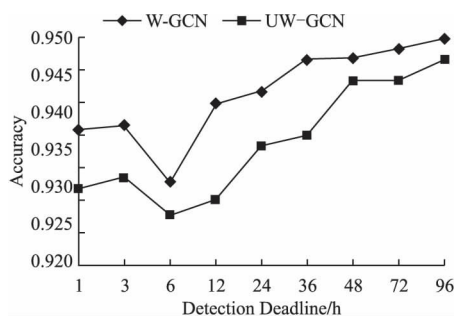


图4 加权图卷积神经网络(W-GCN)模型与无权图卷积神经网络(UW-GCN)模型的谣言检测结果
Fig. 4 Rumor detection results of Weighted-Graph Convolutional Network(W-GCN) and Unweighted-Graph Convolutional Network(UW-GCN)

5) 模型训练的Epoch数量对谣言检测效果的影响

对W-GCN模型,不同的Epoch数量下训练集事件与测试集事件的谣言检测效果进行实验,设置了25个不同的Epoch数量,结果如图5所示.

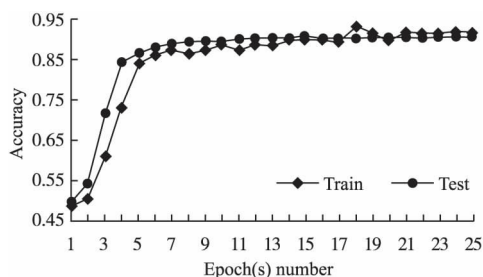


图5 Epoch(s)数量对谣言检测效果的影响
Fig. 5 Effects of Epoch(s) number on rumor detection

可以看出,当模型训练1个Epoch时,训练集和测试集数据的检测正确率都比较低,此时的模型无法学习到微博事件的隐层特征,模型的谣言检测效果较低.当模型训练到了第5个Epochs时,训练集和测试集数据的检测正确率均快速上升,此时的模型已经初步完成了对微博事件隐层特征的学习,模型的谣言检测效果一般.随着模型训练到第25个Epochs时,训练集和测试集数据的检测正确率在一定范围内波动,总体呈上升趋势,此时的模型已经具有比较良好的谣言检测效果.而随着Epoch(s)数量的进一步增加,模型最终会收敛,对谣言事件的检测效果会达到最优值.

5 结束语

目前传统的机器学习方法以及深度学习方法被广泛地应用于社交媒体的谣言检测中,但是这些方法忽略了事件之间的联系或者事件之间联系的紧密程度,会影响检测的效果.因此本文充分考虑了事件之间联系的异质性,基于图卷积神经网络(GCN)提出了一种新型的新浪微博平台谣言检测的方法.该方法通过将事件作为节点,事件之间的联系作为连边,并且考虑了连边的权重,构建出了一个加权图卷积神经网络模型W-GCN.接着使用BERT模型和卷积神经网络(CNN)模型得到微博事件的特征表示,并将该特征输入到W-GCN模型中,对W-GCN中的节点完成分类,最终完成新浪微博的谣言检测任务.与所选的基准方法相比,本文的方法在谣言事件的检测上取得了最好的效果,体现了本文模型在新浪微博谣言检测问题中对谣言事件检测的有效性.同时在谣言的早期检测方面,本文的方法也高于基准方法,体现了本文提出的模型在新浪微博谣言的早期检测方面的有效性.为了提高模型检测的性能,在之后的研究中,可以考虑使用其他事件之间的联系作为GCN模型中节点之间的连边,或考虑使用不同的图网络模型,在节点的特征向量中,融入一些代表性的微博事件特征等方法,来进一步增强模型的谣言检测性能.

References:

- [1] Liu Zhi-yuan, Zhang Le, Tu Cun-chao, et al. Statistical and semantic analysis of rumors in Chinese social media [J]. Scientia Sinica Informationis, 2015, 45(12): 1536-1546.
- [2] Vosoughi, Soroush, Deb Roy, et al. The spread of true and false news online [J]. Science, 2018, 359(6380): 1146-1151.
- [3] Liu Ya-hui, Jin Xiao-long, Shen Hua-wei, et al. A survey on rumor identification over social media [J]. Chinese Journal of Computers, 2018, 41(7): 1536-1558.
- [4] Lecun Y, Boser B, Denker J, et al. Backpropagation applied to handwritten zip code recognition [J]. Neural Computation, 1989, 1(4): 541-551.
- [5] Ma Jing, Gao Wei, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks [C] // Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI), 2016: 3818-3824.
- [6] Liu Yang, Wu Yi-fang. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks [C] // Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI), 2018, 354-361.
- [7] He Hui. Ranking based bayesian sequence recommendation algorithm [J]. Journal of Chinese Computer Systems, 2018, 39(7): 1560-1564.
- [8] Chen Jia-jun, Miao Duo-qian. Adaptive decision tree algorithm based on fusion of decision costs and attribute preference [J]. Journal of Chinese Computer Systems, 2018, 39(6): 1208-1212.
- [9] Qazvinian V, Rosengren E, Radev D R, et al. Rumor has it: identifying misinformation in microblogs [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2011: 1589-1599.
- [10] Yang Fan, Liu Yang, Yu Xiao-hui, et al. Automatic detection of rumor on Sina Weibo [C] // Proceedings of the ACM SIGKDD

- Workshop on Mining Data Semantics(MDS) ,2012: 13-20.
- [11] Liang Gang, He Wen-bo, Chun Xu, et al. Rumor identification in microblogging systems based on users' behavior [J]. Proceedings of IEEE Transactions on Computational Social Systems, 2015, 2 (3) : 99-108.
- [12] Yang Zhi-fan, Wang Chao, Zhang Fan, et al. Emerging rumor identification for social media with hot topic detection [C] //Proceedings of the 12th Web Information System and Application Conference(WISA) ,2015: 53-58.
- [13] Wu Ke, Song Yang, Zhu Kenny-Q. False rumors detection on Sina Weibo by propagation structures [C] //Proceedings of IEEE 31th International Conference on Data Engineering(ICDE) ,2015: 651-662.
- [14] Ma Jing, Gao Wei, Wong Kam-fai. Detect rumors in microblog posts using propagation structure via kernel learning [C] //Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(ACL) ,2017: 708-717.
- [15] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3(4) : 212-223.
- [16] Xi Xue-feng, Zhou Guo-dong. A survey on deep learning for natural language processing [J]. Acta Automatica Sinica, 2016, 42 (10) : 1445-1465.
- [17] Yu Feng, Liu Qiang, Wu Shu, et al. A convolutional approach for misinformation identification [C] //Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) ,2017: 3901-3907.
- [18] Liao Xiang-wen, Huang Zhi, Yang Ding-da, et al. Rumor detection in social media based on a hierarchical attention network [J]. Scientia Sinica Informationis, 2018, 48(11) : 1558-1574.
- [19] Hu Guo-yong, Ding Ye, Qi Shu-han, et al. Multi-depth graph convolutional networks for fake news detection [M]. Natural Language Processing and Chinese Computing, Cham: Springer International Publishing, 2019: 698-710.
- [20] Tian Bian, Xu Ting-ang, Xiao Xi, et al. Rumor detection on social media with Bi-Directional graph convolutional networks [J]. Proceedings of the Association for the Advancement of Artificial Intelligence(AAAI) ,2020, 34(1) : 549-556.
- [21] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [C] //Proceedings of the International Conference on Learning Representations(ICLR) ,2013: 1-12.
- [22] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations [C] //Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL) ,2018: 2227-2237.
- [23] Devlin J, Chang M, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C] //Proceedings of North American Chapter of the Association for Computational Linguistics(NAACL) ,2019: 4171-4186.
- [24] Wang Gen-sheng, Huang Xue-jian. Convolution neural network text classification model based on word2vec and improved TF-IDF [J]. Journal of Chinese Computer Systems, 2019, 40(5) : 1120-1126.
- [25] Kipf T, Welling M. Semi-supervised classification with graph convolutional networks [C] //Proceedings of the International Conference on Learning Representations(ICLR) ,2017: 1-14.
- [26] Zhao Z, Resnick P, Mei Q, et al. Enquiring minds: early detection of rumors in social media from enquiry posts [C] //Proceedings of the 24th International Conference on World Wide Web (ICWW) ,2015: 1395-1405.
- [27] Ma J, Gao W, Wei Z, et al. Detect rumors using time series of social context information on microblogging websites [C] //Proceedings of the Conference on Information and Knowledge Management (CIKM) ,2015: 1751-1754.

附中文参考文献:

- [1] 刘知远, 张乐, 涂存超, 等. 中文社交媒体谣言统计语义分析 [J]. 中国科学: 信息科学, 2015, 45(12) : 1536-1546.
- [3] 刘雅辉, 靳小龙, 沈华伟, 等. 社交媒体中的谣言识别研究综述 [J]. 计算机学报, 2018, 41(7) : 1536-1558.
- [7] 何慧. 基于 Ranking 的贝叶斯序列推荐算法 [J]. 小型微型计算机系统, 2018, 39(7) : 1560-1564.
- [8] 陈家俊, 苗夺谦. 一种决策风险代价与属性偏好融合的适应性决策树算法 [J]. 小型微型计算机系统, 2018, 39(6) : 1208-1212.
- [16] 奚雪峰, 周国栋. 面向自然语言处理的深度学习研究 [J]. 自动化学报, 2016, 42(10) : 1445-1465.
- [18] 廖祥文, 黄知, 杨定达. 基于分层注意力网络的社交媒体谣言检测 [J]. 中国科学: 信息科学, 2018, 48(11) : 1558-1574.
- [24] 王根生, 黄学坚. 基于 Word2vec 和改进型 TF-IDF 的卷积神经网络文本分类模型 [J]. 小型微型计算机系统, 2019, 40(5) : 1120-1126.