计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210400096

传播路径树核学习的微博谣言检测方法

徐建民[1]　孙　朋[1]　吴树芳[2]
1 河北大学网络空间安全与计算机学院　河北 保定　071002
2 河北大学管理学院　河北 保定 071002
(hbuxjm@hbu.cn)

摘　要　微博等在线社交平台的迅猛发展,促进了各种谣言信息的广泛传播,进而给社会秩序带来了潜在的威胁。微博谣言检测能够有效遏制谣言的传播,对净化网络环境、维护社会安定具有重要意义。针对传统谣言检测模型仅考虑用户、内容、传播统计等特征,忽略了谣言传播过程中用户的影响力、情感反馈等特征随转发和评论关系变化而变化的结构问题,提出了一种基于微博信息传播树的路径树核谣言检测模型。所提模型将用户的影响力、情感反馈和内容等特征嵌入传播树的节点中,通过计算传播树中从根节点到叶子节点的路径相似度,得到微博信息传播树结构之间的相似度,进而使用基于传播路径树核的支持向量机实现对微博谣言的检测。实验结果显示,所提模型的准确率达到93.5%,其效果优于未考虑传播路径结构特征的谣言检测模型。

关键词:传播路径;传播树;微博谣言检测;核方法
中图法分类号　TP391

Microblog Rumor Detection Method Based on Propagation Path Tree Kernel Learning

XU Jian-min[1],SUN Peng[1] and WU Shu-fang[2]
1 School of Cyberspace Security and Computer,Hebei University,Baoding,Hebei 071002,China
2 School of Management,Hebei University,Baoding,Hebei 071002,China

Abstract　The rapid development of online social platforms such as microblog promotes the widespread propagation of various rumors information,thereby posing potential threats to social order. Rumor detection on microblog can effectively curb the spread of rumors and is of great significance for purifying the network environment and maintaining social stability. In view of the fact that the traditional rumor detection model only considers the characteristics of users,contents and communication statistics,and ignores the structural problem that the characteristics of users' influence and emotional feedback increase with the forwarding and comment relationship in the process of rumor communication,a path tree kernel rumor automatic detection model based on the microblog information propagation tree is proposed in this paper. It embeds users' influence,emotional feedback,contents into the nodes ofpropagation tree. By calculating the path similarity from the root node to the leaf node in propagation tree,the similarity between the microblog information propagation tree structure is obtained. Furthermore,the model uses the support vector machine classifier based on the propagation path tree kernel todetect microblog rumors. Experimental results show that the accuracy of the proposed model reaches 93.5%,which is better than that of the rumor detection models without considering the structure of propagation path.

Keywords　Propagation path,Propagation tree,Microblog rumor detection,Kernel method

1　引言

随着互联网技术的不断发展,在线社交网络已经成为人们交流与互动的重要载体,其中微博就是典型代表之一。在微博平台上用户不仅可以发布自己的内容,而且可以转发他人的内容,这种转发行为促使一条微博在极短的时间内被传播给数百万用户,从而导致了各种谣言信息的快速传播。例如,2020年2月新冠疫情期间,一则"支援湖北医疗队物资被盗,其中包括队员们的个人生活用品,导致医疗队员无法生活,还呼吁武汉当地的群众帮忙筹措物资,并附上医疗队员的联系方式"的谣言消息,在微博等社交平台上被疯狂转发,误导大众舆论,造成了一系列的不良影响。

关于谣言的研究可以追溯到20世纪40年代,以Knapp等[1]为代表的谣言心理学奠基人将谣言定义为"未经官方

到稿日期:2021-04-11　返修日期:2021-10-23
基金项目:国家社会科学基金(17BTQ068)
This work was supported by the National Social Science Foundation of China(17BTQ068).
通信作者:吴树芳(shufang_44@126.com)

证实却广泛流传的信息，与当前实事相关，且试图让他人相信它是真的”。谣言的危害性极大，会误导公众舆论、破坏政府的公信力，甚至会威胁地区或国家安全[2-3]，因此，如何及时、自动检测谣言信息，已经成为一个不可忽视的社会安全问题。当前，谣言检测仍处于初期阶段，由于社交数据数量巨大、关系繁杂、信息传播突发涌现等特点，使得谣言自动检测面临很大的挑战。

现有的谣言检测研究多是提取谣言信息中的内容、用户和时间等特征[4-8]，但有些谣言是为了模仿真实信息而故意编造的，仅靠上述特征很难将其检测出来。随着时间的推移，微博谣言信息经过评论和转发形成了某种传播结构，为了有效检测出模仿真实信息的谣言信息，相关研究利用传播结构特征[9-10]来检测谣言，但这些研究忽略了谣言传播过程中用户的影响力以及情感反馈特征。基于此，本文依据微博信息的传播过程构建微博信息传播树，提出了一种传播路径树核（Propagation Path Tree Kernel，PPTK）函数，该函数通过比较不同微博信息传播树的传播路径相似度，来计算微博信息传播树结构之间的相似度，进而使用基于 PPTK 的支持向量机（Support Vector Machines，SVM）实现微博谣言检测。

2 相关工作

目前，已有的谣言检测研究多将谣言检测视为有监督学习的二分类问题，即提取谣言特征后，使用基于机器学习的分类算法对信息进行分类。

在线社交网络谣言检测的早期研究者 Castillo 等[11]收集、分析 Twitter 上热门话题的帖子，并提取内容和用户特征，对这些帖子进行了可信度评估。Qazvinian 等[12]提取 Twitter 中推文的内容、用户行为和推文标签特征，实验结果证明组合不同特征可有效检测 Twitter 中的谣言信息。Yang 等[13]通过提取微博内容、用户、主题、传播统计、位置和客户端等特征，使用 SVM 分类器对微博信息进行谣言检测。Liang 等[14]提出了质疑率特征，并通过实验证明了该特征在谣言检测中的有效性。Kwon 等[15]提出了推文数量随时间变化的时间序列拟合模型，并利用随机森林分类算法实现谣言检测。以上谣言检测方法主要基于传统机器学习[16]实现，需人工设定特征，难以获得复杂、抽象的特征数据。

为解决人工选择特征存在的不足，相关研究使用基于深度学习的方法来自动提取各种特征。文献[17]使用正则化层次的 CNN 模型实现文本分类。Ma 等[18]基于循环神经网络模型对转发微博进行建模，学习转发帖子的语义特征，一定程度上解决了特征稀疏问题。Kaliyar 等[19]提出了一种基于内容的深度学习卷积网络来检测社交媒体上的虚假新闻信息。Bhatt 等[20]使用多层感知机将网络特征、传播统计特征以及人工设计特征结合，以识别社交网络中的谣言信息。Zhang 等[21]利用多头注意机制融合文本和视觉特征，提出了一种基于深度学习的多模态融合谣言检测网络模型。但是，上述方法不能很好地检测出基于某种动机模仿真实信息而编造的谣言信息。

为解决上述问题，Meel 等[22]提出了基于传播结构的恶意谣言信息分类算法；Vosoughi 等[23]通过分析大量 Twitter 数据得出，谣言信息和非谣言信息在传播过程中有着明显的差异，谣言信息较非谣言信息的传播范围更广、更快也更深；Wu 等[9]通过研究发现，谣言信息的传播结构特征是一个检测谣言的高阶特征，其最能突出谣言信息和正常信息的不同，将随机游走图核和 RBF 核结合，提出了一种基于混合核的谣言检测模型；Ma 等[10]提出了一种基于传播结构的树核模型，该模型通过计算传播树中子树结构之间的相似性，来区分谣言的高阶传播特征，用于识别不同类型的谣言信息；Khoo 等[24]将 Twitter 传播结构信息融入 Transform 网络中，用于捕获用户发布推文之间长距离的交互特征；Wu 等[25]基于谣言的转发过程，提出了一种基于门控图神经网络的算法，用于检测推特上的谣言信息；Bian 等[26]使用图结构模拟微博的传播过程，并提出了一种双向图卷积网络模型，用于学习谣言的传播模式。

上述基于传播结构的谣言检测研究忽略了谣言传播过程中，用户的影响力和情感反馈特征随转发和评论关系变化而变化的结构问题，基于此，本文将用户影响力、情感反馈和内容等信息嵌入到微博信息传播树的节点中，通过计算节点的相似度来得到传播路径的相似度，进而得到传播树结构之间的相似度，实现对微博等社交网络平台上的谣言的检测。

3 微博信息传播树的构造

3.1 微博信息传播树的结构

本文将源微博 $blog_k$ 的传播过程建模为树结构 $T(r)=\langle V,E\rangle$，V 表示树 T 中所有节点组成的集合，E 表示树 T 中有向边组成的集合，传播树结构如图 1 所示。

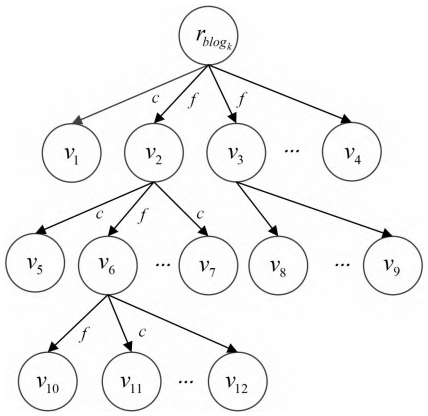


图 1 $blog_k$ 的传播树结构

Fig. 1 Propagation tree structure of $blog_k$

图 1 所示的树结构中节点可以分为两类：根节点 r_{blog_k} 和一般节点 v_i。根节点 r_{blog_k} 的内容为从源微博 $blog_k$ 中提取的用户和微博内容信息，一般节点 v_i 的内容为从响应微博中提取的用户和微博内容信息。其中，响应微博表示与源微博有直接响应（转发或评论源微博）或间接响应（转发或评论非源微博）关系的微博。树结构中的有向边分为转发（f）和评论（c）两类，例如 v_6 是对 v_2 的转发，则 v_2 到 v_6 有一条转发有向边（f），v_2 是 v_6 的父节点。

3.2 微博信息传播树中的节点数据

微博信息传播树中的节点数据使用二元组 $v=(u,c)$

表示，u 表示节点的用户信息，c 表示节点的微博内容信息。用户信息共涉及 14 个特征，详细如表 1 所列。

表 1 用户的 14 个特征描述

Table 1 14 features description of users

用户信息类型	特征	描述
背景信息	性别	用户性别
	微博龄	用户注册时间到发布该微博的时间差(h)
	V 用户	用户是否 V 认证
	认证度	微博官方认证类型
	位置	用户注册地理位置，所在的省和市
社交信息	关注数	用户关注他人微博数量
	粉丝数	用户的粉丝数量
	发布微博数	用户发布微博的数量
	互粉数	用户和粉丝互粉的数量
	收藏数	用户收藏微博的数量
	微博评论数	用户当前发布微博被评论的数量
	微博转发数	用户当前发布微博被转发的数量
影响力信息	范围影响力	用户节点在三步内引发转发和评论的数量
情感信息	情感反馈	用户当前评论微博内容的情感得分

表 1 中用户的背景信息和社交信息特征为基础特征。此外，考虑到意见领袖及用户态度对谣言传播的影响，本文新增了范围影响力特征和情感反馈特征[27]，两个新增特征的定义和计算方法如下。

(1)范围影响力

准确识别微博信息传播树中的意见领袖，可更精确地计算传播树中节点的相似度，从而有效计算出传播树结构之间的相似度。文献[28]通过研究发现，在信息传播中存在影响力逐步消散的现象，即一个节点用户的影响力范围集中在距离其三步范围的区域之内，对三步之外的节点用户不会产生明显影响，图 2 给出了用户 v_u 的三步影响范围。

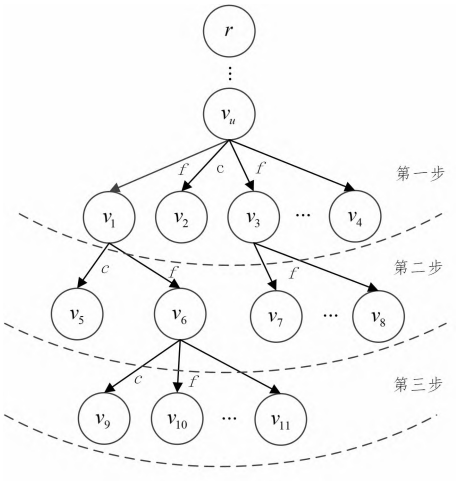


图 2 用户 v_u 三步内的影响范围示意图

Fig. 2 User's range of influence within three steps

基于上述结论，我们使用节点传播三步内引发评论和转发的有向边数量，来计算节点用户的范围影响力 $Inf(v_u)$，其计算方法如式(1)所示：

$$Inf(v_u)=\sum_{i=1}^{3}\frac{1}{i}\eta\,\|\{e_{v_jv_k}\mid v_j,v_k\in V,e_{v_jv_k}\in E_i^f(v_u)\}\|+\sum_{i=1}^{3}\frac{1}{i}(1-\eta)\,\|\{e_{v_mv_n}\mid v_m,v_n\in V,e_{v_mv_n}\in E_i^c(v_u)\}\| \tag{1}$$

其中，$E_i^f(v_u)$ 表示节点 v_u 的第 i 步转发有向边的集合，$E_i^c(v_u)$ 表示节点 v_u 的第 i 步评论有向边集合，η 表示转发行为的权重。

(2)情感反馈

本文在 Chen 等[29]提出的基于情感词典的短文本情感分析的基础上，提取了 HowNet[30]情感词典中表示支持、反对、质疑和惊讶四大类更加细粒度的情感词，并依据情感极性赋予它们不同的权重，用于识别在谣言信息传播过程中存在的“假的”“不可能”“吃惊”等语言特征，并据此计算微博内容的情感得分，具体计算过程如下。

首先，进行微博文本预处理，以标点符号为分割标志，将微博分割为若干个分句，并对每个分句进行分词，去停用词；其次，提取情感词并计算每个分句情感的得分；依据 HowNet 情感词典，提取每个分句子中的情感词，以每个情感词为基准，向前依次寻找程度副词和否定词，依据程度副词和否定词的权值来计算每个情感词的加权得分；然后将分句中的每个情感词得分求和，计算出每个分句的情感词得分；最后，将每个分句的情感词得分累加，得出微博文本的情感得分。

依据上述的情感反馈计算方法，我们计算了数据集中共涉及 3644732 个用户 2313 个谣言事件和 2315 个非谣言事件的平均情感反馈得分，部分事件的结果如图 3 所示。

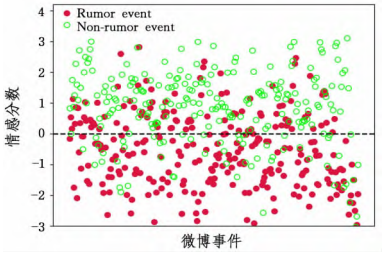


图 3 微博事件用户情感反馈散点图

Fig. 3 Scatter plot of users' emotional feedback of microblog events

图 3 中，每个点表示一个微博事件，纵坐标对应的值是该微博事件传播过程中的所有评论微博的平均情感得分，正数代表积极反馈，负数代表消极反馈，数值的大小表示情感反馈的程度。实心点代表谣言事件，空心点代表非谣言事件。通过分析图 3 可以发现，谣言信息传播过程中，用户评论的情感反馈较为消极。

4 基于传播路径树核的谣言检测

4.1 PPTK

PPTK 函数通过计算从根节点到叶子节点的传播路径相似度来得到微博信息传播树之间的相似度，其中传播路径的相似度依据路径中节点的相似度来计算，因此节点相似度的计算是路径相似度计算的基础。本文定义了一个函数 f，用于计算微博信息传播树中两个节点 v_i 和 v_j 之间的相似度，如式(2)所示：

$$f(v_i,v_j)=\beta\mathrm{sim}(\boldsymbol{u}_i,\boldsymbol{u}_j)+(1-\beta)\mathrm{sim}(c_i,c_j) \tag{2}$$

其中，$\boldsymbol{u}_i$ 和 $\boldsymbol{u}_j$ 表示包含 14 个特征的两个用户向量，c_i 和 c_j 分别是用户 $\boldsymbol{u}_i$ 和 $\boldsymbol{u}_j$ 发布的微博文本。$\mathrm{sim}(\boldsymbol{u}_i,\boldsymbol{u}_j)$ 用于计算两个节点用户之间的相似度，如式(3)所示，$\mathrm{sim}(c_i,c_j)$ 用于

计算两个节点中微博内容的相似度，如式(4)所示。β 用于度量用户相似度和内容相似度对节点相似度的贡献程度，该参数将通过实验获取。

$$(\boldsymbol{u}_i,\boldsymbol{u}_j)=\frac{\boldsymbol{u}_i\cdot\boldsymbol{u}_j}{\|\boldsymbol{u}_i\|\ \|\boldsymbol{u}_j\|} \tag{3}$$

$$\mathrm{sim}(c_i,c_j)=\frac{|n\text{-}gram(c_i)\bigcap n\text{-}gram(c_j)|}{|n\text{-}gram(c_i)\bigcup n\text{-}gram(c_j)|} \tag{4}$$

其中，$n\text{-}gram(c_i)$表示采用 n 元模型对微博文本 c_i 进行分词，本文采用 Uni-gram 和 Bi-gram 实现。

为了更好地描述 PPTK 的计算过程，本节定义 P 是从微博信息传播树 $T(r)$中提取的从根节点到每一个叶子节点的所有路径集合。$P=p_1\cup p_2\cup\cdots\cup p_{d(T)}$，其中 p_i 表示路径长度为 i 的路径集合，$p_1=\{r\}$，$p_i\in P$，$d(T)$表示微博信息传播树 T 的深度，$1\leqslant i\leqslant d(T)$。$p_{(i,j)}$表示路径集合 p_i 中第 j 个路径，$p_{(i,j)}\in p_i$，$1\leqslant j\leqslant|p_i|$，$|p_i|$表示路径集合 p_i 中的路径数。

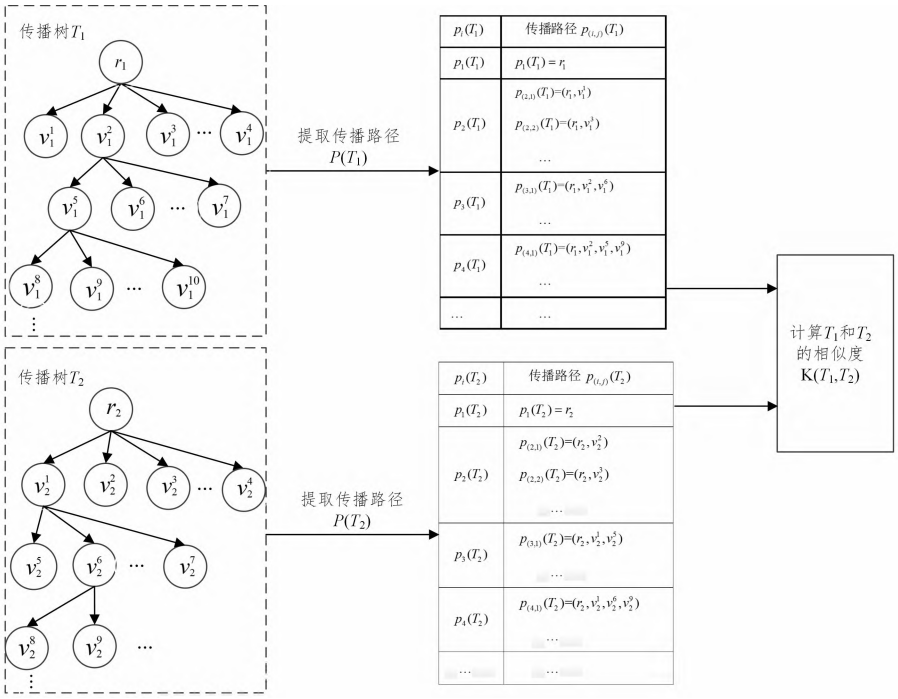


图 4 传播路径的提取及传播树相似度的计算

Fig. 4 Propagation path extraction and propagation tree similarity calculation

图 4 给出了对于给定的两个微博信息传播树 $T_1(r_1)$和 $T_2(r_2)$，分别提取从根节点到每一个叶子节点的传播路径的过程，最终利用 PPTK 函数计算得到 T_1 和 T_2 的相似度，其计算方法如式(5)所示：

$$K(T_1,T_2)=\sum_{i=1}^{\min(d(T_1),d(T_2))}\lambda^i\Delta(p_i(T_1),p_i(T_2)) \tag{5}$$

其中，λ^i 表示路径长度为 i 的传播路径的动态指数衰减权重，$\lambda(0<\lambda<1)$，长度不同的路径拥有不同的权重。$\Delta(p_i(T_1),p_i(T_2))$是计算 T_1 和 T_2 中长度为 i 的路径之间的相似度，我们将分 4 种情况对其进行计算。

(1)如果 T_1 或 T_2 中只有一个节点，则：

$$\Delta(p_1(T_1),p_1(T_2))=f(r_1,r_2) \tag{6}$$

(2)如果 $p_i(T_1)=\varnothing$ 或 $p_i(T_2)=\varnothing$，则：

$$\Delta(p_i(T_1),p_i(T_2))=0 \tag{7}$$

(3)如果$|p_i(T_1)|\leqslant|p_i(T_2)|$，则：

$$\Delta(p_i(T_1),p_i(T_2))=\sum_{j=1}^{|p_i(T_1)|}H(p_{(i,j)}(T_1),p'_{(i,j)}(T_2)) \tag{8}$$

$$H(p_{(i,j)}(T_1),p_{(i,j)}(T_2))=\frac{1}{i}\sum_{x=1}^{i}f(p_{(i,j)}(T_1)[x],p_{(i,j)}(T_2)[x]) \tag{9}$$

式(8)中，$p'_{(i,j)}(T_2)$表示与路径 $p_{(i,j)}(T_1)$最相似的 $p_i(T_2)$中的路径，$p'_{(i,j)}(T_2)\in p_i(T_2)$，其计算式如式(10)所示。式(9)用于计算两个传播路径之间的相似度，通过计算传播路径中所有节点相似度的平均值得到，$p_{(i,j)}(T_1)[x]$表示 T_1 中路径长度为 i 的路径集合中第 j 个路径的第 x 个节点。

$$p'_{(i,j)}(T_2)=\arg\max_m H(p_{(i,j)}(T_1),p_{(i,m)}(T_2)) \tag{10}$$

其中，$1\leqslant m\leqslant|p_i(T_2)|$。

(4)如果$|p_i(T_1)|>|p_i(T_2)|$，则：

$$\Delta(p_i(T_1),p_i(T_2))=\sum_{j=1}^{|p_i(T_2)|}H(p_{(i,j)}(T_2),p'_{(i,j)}(T_1)) \tag{11}$$

其中，$p'_{(i,j)}(T_1)$表示与 $p_{(i,j)}(T_2)$最相似 $p_i(T_1)$中的路径，$p'_{(i,j)}(T_1)\in p_i(T_1)$，其计算式如式(12)所示：

$$p'_{(i,j)}(T_1)=\arg\max_m H(p_{(i,j)}(T_2),p_{(i,m)}(T_1)) \tag{12}$$

其中，$1\leqslant m\leqslant|p_i(T_1)|$。

4.2 基于 PPTK 的谣言检测

本文将式(5)所示的传播路径树核函数(PPTK)作为 SVM 分类器的核函数，得到如式(13)所示的谣言分类器。首先构建好训练数据集中的微博信息传播树，然后对其进行训练得到式(13)所示的谣言分类函数中的参数，最后使用训练好的谣言分类器对待检测微博进行谣言检测。

$$class(T)=\mathrm{sgn}(\sum_{i=1}^{m}\alpha_i y_i K(T,T_i)+b) \tag{13}$$

其中，$class(T)$表示待检测微博信息传播树 T 的分类标签，α_i 表示拉格朗日乘子，b 为常数项，T_i 表示用于训练的微博信息传播树，y_i 是 T_i 的标签。

基于 PPTK 模型对待检测微博 $blog_k$ 进行谣言检测的流程如图 5 所示，其主要包括两个步骤：1)对 $blog_k$ 进行处理，获取待检测微博的转发和评论数据，从数据中提取 14 个特征，根据转发和评论过程构建待检测微博的传播树；2)依据训练好的 PPTK 分类器对 $blog_k$ 的传播树进行分类，判断其是否为谣言。

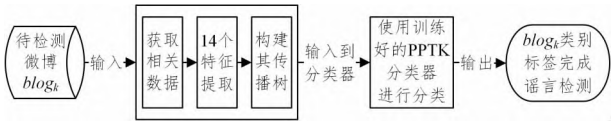


图 5 基于 PPTK 的谣言检测流程图

Fig. 5 Flow chart of rumor detection based on PPTK

5 实验

为验证本文提出的基于传播路径树核谣言检测模型的有效性，本节进行了实证研究。实验主要包括：1)参数敏感性实验，即对 PPTK 谣言检测模型中的参数进行分析；2)新增特征的有效性实验，即检验范围影响力、情感反馈及传播路径特征，以及对谣言信息检测的性能影响；3)对比实验，即采用 Ma 等[16]研究中公开的来自新浪微博的谣言数据集，比较 PPTK 模型和已有模型的谣言检测性能；4)早期谣言检测对比实验，即使用谣言事件早期的数据，将各种模型的谣言检测性能进行对比。由于实验数据有限，为使模型具有更好的泛化能力，实验采用五折交叉验证，采取评价指标的平均值作为最终的实验结果。

5.1 数据集

本文使用公开的微博数据集来评估模型的有效性。该数据集包含 2313 个谣言事件和 2351 个非谣言事件。表 2 列出了该数据集的详细统计数据。

表 2 数据集统计信息

Table 2 Dataset statistics information

统计项	数据集的数据	
	谣言	非谣言
源微博	2313	2351
用户数	2025513	1619219
微博数	2090743	1661503
最小转帖数量	3	5
最大转帖数量	59317	52156
平均转帖数量	903	706
树的最小深度	2	2
树的最大深度	36	73
树的平均深度	7.58	4.83

5.2 评价指标

本文采用谣言检测问题中常见的评价指标，即准确率(Accuracy)、精准率(Precision)、召回率(Recall)和 F_1 值(*F-Score*)，来实现对谣言检测性能的评价[31]。其对应的计算式如下：

$$Accuracy=\frac{TP+TN}{P+N} \tag{14}$$

$$Precision=\frac{TP}{TP+FP} \tag{15}$$

$$Recall=\frac{TP}{TP+FN} \tag{16}$$

$$F_1=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{17}$$

其中，P 表示谣言的样本个数，N 表示为非谣言的样本个数，TP 表示被正确检测为谣言的样本个数，TN 为被正确检测为非谣言的个数，FN 表示将非谣言检测为谣言的个数，F_1 值是查准率和查全率的综合值。精准率、召回率和 F_1 值数越大，表示谣言检测模型的性能就越好。

5.3 相关参数分析

本文中的相关参数有 3 个，分别是转发行为权重参数 η、传播路径的动态指数衰减权重中的参数 λ 和用户相似度权重 β。文献[32]在研究转发和评论行为对用户传播影响力的贡献时得出，转发行为的权重值和评论行为的权重值比为 2∶1，据此，本文中转发行为权重 $\eta=2/3$。动态指数衰减权重中参数 λ 的取值参考文献[10]，即 $\lambda=1/e$。图 6 给出了在 $\eta=2/3$，$\lambda=1/e$ 时，不同 β 值对 PPTK 模型谣言检测准确率的影响。

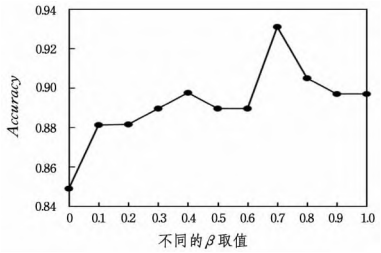


图 6 不同 β 取值下模型的准确率

Fig. 6 Accuracy of model with different β values

从图 6 中可以看出，当 β 取 0 时，表示模型计算节点相似度时只考虑用户发布微博内容的特征，模型准确率为 0.85；当 β 取 1 时，表示模型计算节点相似度时只考虑用户的特征，模型准确率为 0.90；当 β 取 0.7 时，表示模型计算节点相似度时，既考虑到了用户特征又考虑到了内容特征，模型准确率达到了相对最优，因此将参数 β 的值设为 0.7。

5.4 新增特征的有效性

为了观察范围影响力和情感反馈特征与其他特征和谣言之间的线性相关联程度，本节通过皮尔逊相关系数，从数据中提取并计算源微博的各个特征和谣言之间的线性相关联程度，正号代表随着一个特征数量的增加，另一个特征数量也在增加，负号则相反，结果如图 7 所示。

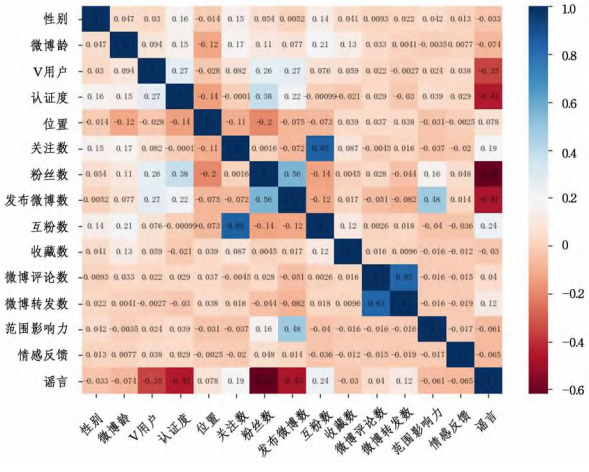


图 7 特征相关系数热力图

Fig. 7 Thermodynamic diagram of feature correlation coefficient

从图 7 可以看出，范围影响力和情感反馈与谣言呈负相关，即谣言信息的范围影响力较非谣言信息的范围影响力小，情感反馈较为消极。

为了验证新增特征的有效性，分别使用不同特征的组合来验证新增特征对谣言检测性能的影响。实验结果如图 8 所示，横坐标分别代表传统特征 BF（表 1 中用户的前 12 个特征加上文本特征）、传统特征加范围影响力和情感反馈特征 BIEF 以及 PPTK（BIEF 加上传播路径结构特征）。BF 和 BIEF 均使用基于线性核函数的 SVM 分类器。

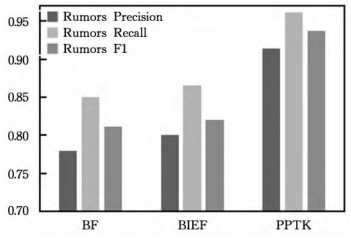


图 8 不同特征组合谣言检测性能对比图

Fig. 8 Comparison of rumor detection performance with different features combination

从图 8 可得，本文提出的特征对谣言检测的结果都有提升。BIEF 相比 BF 在谣言精准率上提升了 2.1%，尤其是加上传播路径结构特征的 PPTK 比传统特征在谣言精准率上提升了 13.5%。

5.5 对比实验

本文将提出的 SVM-PPTK 模型与已有模型进行对比。

BOW[33] 模型使用词袋 BOW（bag-of-words）表示文本，然后基于线性 SVM 分类器进行谣言信息检测。

DT-Rank[34] 模型是基于决策树排名的模型，通过一系列谣言信号特征的正则表达式匹配进行谣言事件的识别。

SVM-RBF[35] 模型提取文本内容特征，输入到基于 RBF 核的 SVM 的谣言分类器。

TF-IDF 模型将所有文本处理为 TF-IDF 向量表示，使用线性支持向量机分类器训练谣言分类。

RFC[36] 模型是基于时间序列上的用户、语言特征，使用随机森林分类器进行谣言检测。

GRU-RNN[18] 是使用带有 GRU 单元的循环神经网络，来学习转发帖子的谣言检测模型。

HSA-BLSTM[37] 结合用户的社会文本信息，使用注意力机制的分层神经网络对谣言进行检测的模型。

SVM-HK[9] 模型使用随机游走图核和 RBF 核形成的混合核函数进行谣言检测。

SVM-TK[10] 是基于树核的 SVM 谣言分类器进行谣言检测。

RvNN[38] 是通过 GRU 单元学习谣言传播结构的谣言检测方法。

表 3 列出了各种评价指标下的不同模型的谣言检测结果。本节将所有模型分为 3 组，第一组是基于传统特征的方法，包括 BOW，DT-Rank，SVM-RBF，TF-IDF 和 RFC 模型；第二组是基于最新的神经网络学习的方法，包括 GRU-RNN 和 HAS-RNN 模型；第三组是基于传播结构特征的方法，包括 SVM-HK，SVM-TK 和 SVM-PPTK 模型。从表 3 所列的实验结果可以看出：1）SVM-PPTK 模型的检测效果在准确率、召回率和 F_1 值上均优于其他模型；2）第一组基于传统特征谣言检测方法的准确率大幅度低于其他组的准确率，其中 BOW 模型仅考虑了语言特征，表现效果最差，DT-Rank 模型使用一组正则表达式去匹配特定的语言特征进行谣言检测，实际匹配到的微博数据很少，效果较考虑整体文本信息的 SVM-RBF 和 TF-IDF 模型都差，RFC 模型综合考虑了时间序列上的内容、用户特征，性能达到了相对最优；3）第二组基于神经网络的方法比第一组基于传统特征的方法在准确率上至少提升了 5%，HSA-BLSTM 模型结合各种社会信息并加入了注意力机制，性能优于使用循环神经网络仅考虑了时序特征的 GRU-RNN 模型；4）第三组基于传播结构特征核学习的谣言检测模型中，SVM-HK 仅将基于平面特征的两个核函数（随机游走图核和 RBF 核）结合，不能很好地体现出谣言信息的传播结构特征，其性能相对于 SVM-TK 模型较差，但明显优于基于传统特征的模型。SVM-TK 是基于子树的模型，在确定子树时只考虑到了子树的根节点，导致计算结果不准确。SVM-PPTK 只找两棵微博信息传播树，它们具有相同深度的传播路径，且考虑到了谣言信息传播路径中的整体节点信息，相比 SVM-TK 模型使用了更少的节点数据，但在准确率上比 SVM-TK 模型提高了 3.3%。RvNN 模型对传播树中的节点赋予相同的权重，导致模型受到新的异常节点的影响较大，谣言检测的准确率低于 SVM-PPTK 模型。

表 3 不同模型下的谣言检测性能

Table 3 Rumor detection performanceof different models

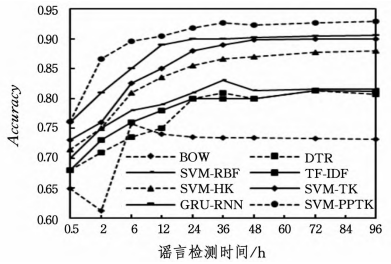
方法	准确率	谣言			非谣言		
		精准率	召回率	F_1	精准率	召回率	F_1
BOW	0.726	0.687	0.879	0.767	0.795	0.581	0.661
DT-Rank	0.732	0.738	0.715	0.726	0.726	0.749	0.737
SVM-RBF	0.815	0.793	0.848	0.819	0.840	0.782	0.810
TF-IDF	0.827	0.801	0.866	0.832	0.857	0.784	0.815
RFC	0.849	0.786	0.959	0.864	0.947	0.739	0.830
GRU-RNN	0.899	0.865	0.946	0.904	0.940	0.852	0.894
HSA-BLSTM	0.905	0.875	0.957	0.910	0.939	0.862	0.899
SVM-HK	0.880	0.866	0.930	0.896	0.911	0.861	0.885
SVM-TK	0.902	0.855	**0.968**	0.908	**0.963**	0.836	0.895
RvNN	0.908	0.912	0.897	0.905	0.904	**0.918**	0.911
SVM-PPTK	**0.935**	**0.914**	0.961	**0.937**	0.959	0.910	**0.933**

5.6 早期谣言检测

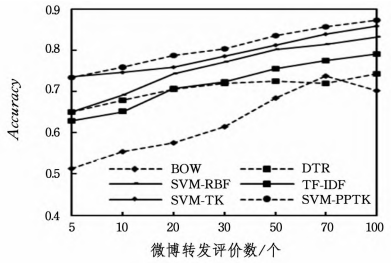
在谣言信息传播的早期阶段，对谣言信息进行实时自动检测尤为重要，如此可以方便采取预防措施，减少不必要的损失。为了评估各个模型在早期谣言检测中的效果，本节设置了一系列的时间点和微博帖子转发评论数量。在某个时间点进行检测时，微博信息传播树只包含在该时间前所有转发、评论微博的节点。检测的时间越早，设置的转发、评论的数量越少，传播树中的节点数量就越少。

图 9(a)给出了谣言信息传播过程中在一系列时间节点下，各个模型的谣言检测准确率。可以看出，SVM-PPTK 模型的谣言检测准确率优于其他所有模型。除 BOW 模型外，其他模型随着谣言传播时间在 0.5h 到 2h 内增加，准确率迅速增长，尤其 SVM-PPTK 模型在谣言信息传播 2h 后就达到了 86.7%的准确率。分析 BOW 模型发现，BOW 虽然能够很好地捕捉到一些在谣言传播初期出现的词语(如“不可能”“疑惑”“假的”等词语)，但这些词在非谣言信息中很少出现，随着时间的增加，谣言信息传播的语言特征越来越像正常信息，因此 BOW 模型表现出性能下降的趋势。几乎所有的模型性能都在谣言传播 12h 后达到稳定，此时 SVM-PPTK 准确率已达到 90.5%。

图 9(b)给出了在一定微博转发评论数量下，各个模型的谣言检测的准确率。可以看出，微博评论转发数由 5 到 100 时，各个模型性能都处于稳定增长阶段，在微博转发评论数为 100 时，SVM-PPTK 模型的谣言信息检测准确率达到 87.2%。



(a)



(b)

图 9 不同模型早期谣言检测性能比较

Fig. 9 Comparison of early rumor detection performance of different models

综上所述，无论是在较短的时间内还是当微博评论转发数较少时，SVM-PPTK 模型都表现出了良好的性能。

结束语 微博等社交网络中的谣言信息不仅严重扰乱了人们的正常生活，还会破坏政府公信力，甚至危害国家安全，因此及时检测网络中的谣言信息尤为重要。本文从谣言信息的传播过程出发，考虑到谣言信息传播过程中用户范围影响力、情感反馈以及内容等特征随转发和评论关系变化而变化的结构变化问题，提出了一种基于传播路径树核的谣言检测模型。实验结果表明，该模型的性能优于其他模型，并可以有效实现谣言信息的早期检测。本文为判断检测网络舆情中的热点话题[39]是否为谣言信息提供了一种思路，该研究虽然取得了较好的效果，但存在以下不足：1)在用户情感反馈特征方面，需要手动构建不同的情感词和程度副词，耗费人力和时间；2)PPTK 核函数仅考虑了同等深度的传播路径，未考虑子路径对谣言信息检测性能的影响。未来我们将针对以上不足展开深入研究。

参 考 文 献

[1] KNAPP R H. A psychology of rumor[J]. Public Opinion Quarterly, 1944, 8(1): 22-37.

[2] CHEN YF, LI Z Y, LIANG X, et al. Review on Rumor Detection of Online Social Networks [J]. Chinese Journal of Computers, 2018, 41(7): 1648-1677.

[3] BRATU S. The fake news sociology of COVID-19 pandemic fear: Dangerously inaccurate beliefs, emotional contagion, and conspiracy ideation[J]. Linguistic and Philosophical Investigations, 2020 (19): 128-134.

[4] MA B, LIN D, CAO D. Content representation for microblog rumor detection[M] // Advances in Computational Intelligence Systems. Cham: Springer, 2017: 245-251.

[5] ZUBIAGA A, LIAKATA M, PROCTER R. Learning reporting dynamics during breaking news for rumour detection in social media[J]. arXiv: 1610.07363, 2016.

[6] CHANG C, ZHANG Y, SZABO C, et al. Extreme user and political rumor detection on twitter[C] // International Conference on Advanced Data Mining and Applications. Cham: Springer, 2016: 751-763.

[7] LIANG G, HE W, XU C, et al. Rumor identification in microblogging systems based on users' behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.

[8] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites[C] // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015: 1751-1754.

[9] 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. IEEE, 2015: 651-662.

[10] MA J, GAO W, WONG K F. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 708-717.

[11] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C] // Proceedings of the 20th International Conference on World Wide Web. 2011: 675-684.

[12] QAZVINIAN V, ROSENGREN E, RADEV D, et al. Rumor has it: Identifying misinformation in microblogs[C] // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011: 1589-1599.

[13] YANG F, LIU Y, YU X, et al. Automatic detection of rumor on

Sina Weibo[C]// Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012:1-7.

[14] LIANG G,YANG J,XU C. Automatic rumors identification on Sina Weibo[C]// 2016 12th International Conference on Natural Computation,Fuzzy Systems and Knowledge Discovery(ICNC-FSKD). IEEE,2016:1523-1531.

[15] KWON S,CHA M,JUNG K. Rumor Detection over Varying Time Windows[J/OL]. https://www. researchgate. net/publication/311413402 _Rumor _detection _over _varying _time _windows.

[16] WANG T,LI M. Research on Comment Text Mining Based on LDA Model and Semantic Network[J]. Journal of Chongqing Technology and Business University(Natural Science Edition),2019,36(4):9-16.

[17] WANG Y,HE Y M,CHEN H X,et al. RHS-CNN:A CNN Text Classification Model Based on Regularized Hierarchical Softmax[J]. Journal of Chongqing University of Technology (Natural Science),2020,34(5):187-195.

[18] MA J,GAO W,MITRA P,et al. Detecting rumors from microblogs with recurrent neural networks[C]// International Joint Conference on Artificial Intelligence. 2016.

[19] ROHIT K K,ANURAG G,PRATIK N,et al. FNDNet-A deep convolutional neural network for fake news detection[J]. Cognitive Systems Research,2020,61:32-44.

[20] BHATT G,SHARMA A,SHARMA S,et al. Combining neural,statistical and external features forfake news stance identification[C]// Companion Proceedings of the The Web Conference. 2018:1353-1357.

[21] ZHANG S Q,DU S D,ZHANG X B,et al. Social rumor detection method basedon multimodalfusion[J]. Computer Science,2021,48(5):117-123.

[22] PRIYANKA M,DINESH K V. Fake news,rumor,information pollution in social media and web:A contemporary survey of state-of-the-arts,challenges and opportunities[J/OL]. Expert Systems With Applications. https://www. researchgate. net/publication/336271325 _Fake _News _Rumor _Information _Pollution _in _Social _Media _and _Web _A _Contemporary _Survey _of _State-of-the-arts _Challenges _and _Opportunities.

[23] VOSOUGHI S,ROY D,ARAL S. The spread of true and false news online[J]. Science,2018,359(6380):1146-1151.

[24] KHOO L M S,CHIEU H L,QIAN Z,et al. Interpretable rumor detection in microblogs by attending to user interactions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020,34(5):8783-8790.

[25] ZW A,DP A,JC A,et al. Rumor detection based on propagation graph neural network with attention mechanism[J/OL]. Expert Systems with Applications. https://www. researchgate. net/publication/341951109 _Rumor _Detection _Based _On _Propagation _Graph _Neural _Network _With _Attention _Mechanism.

[26] BIAN T,XIAO X,XU T,et al. Rumor detection on social media with bi-directional graph convolutional networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020,34(1):549-556.

[27] LI Y H,ZHENG C M,WANG W L,et al. A Multi-Dimensional and Multi-Emotion Analysis Method for Mixed Language Texts [J]. Computer Engineering,2020,46(12):113-119,141.

[28] QIN Y,MA J,GAO S. Efficient influence maximization under TSCM:a suitable diffusion model in online social networks[J]. Soft Computing,2017,21(4):827-838.

[29] CHEN X D. Research on sentiment tendency analysis of Chinese Weibo based on sentiment dictionary [D]. Wuhan:Huazhong University of Science and Technology,2012.

[30] DONG Z,DONG Q. HowNet knowledge database [J/OL]. [2013-07-25]. http://www,keenage,com.

[31] XU F,SHENG V S,WANG M. Near real-time topic-driven rumor detection in source microblogs[J/OL]. Knowledge-Based Systems. https://www. researchgate. net/publication/343791225 _Near _real-time _topic-driven _rumor _detection _in _source _microblogs.

[32] QI C,CHEN H C,YU H T. Weibo user influence evaluation method based on comprehensive analysis of user behavior[J]. Application Research of Computers,2014,31(7):2004-2007.

[33] MIHALCEA R,STRAPPARAVA C. The lie detector:Explorationsin the automatic recognition of deceptive language[C]// Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. 2009:309-312.

[34] ZHAO Z,RESNICK P,MEI Q. Enquiring minds:Early detection of rumors in social media from enquiry posts[C]// Proceedings of the 24th International Conference on World Wide Web. 2015:1395-1405.

[35] YANG F,LIU Y,YU X,et al. Automatic detection of rumor on sinaweibo[C]// Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. 2012:1-7.

[36] 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. IEEE,2013:1103-1108.

[37] GUO H,CAO J,ZHANG Y,et al. Rumor detection with hierarchical social attention network[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018:943-951.

[38] MA J,GAO W,WONG K F. Rumor detection on twitter with tree-structured recursive neural networks[C]// Association for Computational Linguistics. 2018:1980-1989.

[39] WU S F,ZHU J. Research on hot topics in Internet public opinion [M]. Beijing:Science Press,2020.

XU Jian-min, born in 1966, Ph.D, professor, Ph.D supervisor. His main research interests include information retrieval, public opinion monitoring and online social network analysis.

WU Shu-fang, born in 1979, Ph.D, professor, Ph.D supervisor. Her main research interests include information processing and online social network analysis.

(责任编辑:喻藜)