Journal of Computer Applications ISSN 1001-9081 2018-02-10
计算机应用,2018,38(2):410-414,420 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2018)02-0410-05 DOI:10.11772/j.issn.1001-9081.2017082368

基于梯度提升决策树的微博虚假消息检测

段大高[1,2],盖新新[1],韩忠明[1,2*],刘冰心[3]

(1.北京工商大学 计算机与信息工程学院,北京 100048; 2.北京工商大学 食品安全大数据技术北京市重点实验室,北京 100048;
3. University of Liverpool, Department of mathematical Sciences, Liverpool, GB L69 7ZX)

(* 通信作者电子邮箱 hanzhongming@btbu.edu.cn)

摘 要:微博是信息共享的重要平台,同时,也成为虚假消息产生和推广的重要平台,虚假消息的传播严重扰乱了社会秩序。为了快速、有效地识别微博虚假消息,提出一种基于梯度提升决策树(GBDT)的虚假消息检测方法。首先,从评论的角度分析微博虚假消息和真实消息之间存在的差异,在此基础上提取评论中的文本内容、用户属性,信息传播和时间特性的分类特征;然后,基于分类特征,采用GBDT算法实现微博虚假消息识别模型;最后,在两个真实的微博数据集上进行验证。实验结果表明,基于GBDT的识别模型能有效提高微博虚假消息检测的准确率。

关键词:微博;社交网络;虚假消息;梯度提升决策树;评论

中图分类号:TP391.4;TP181 **文献标志码**:A

Micro-blog misinformation detection based on gradient boost decision tree

DUAN Dagao[1,2], GAI Xinxin[1], HAN Zhongming[1,2*], LIU Bingxin[3]

(1. *School of Computer and Information Engineering, Beijing Technology and Business University, Beijing* 100048, *China*;
2. *Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing* 100048, *China*;
3. *Department of Mathematical Sciences, University of Liverpool, Liverpool, GB L69 7ZX*)

Abstract: Micro-blog has become an important platform for information sharing. Meanwhile, it is also one of the main ways for spreading of different misinformation. In order to detect the micro-blog misinformation quickly and effectively, a method based on Gradient Boost Decision Tree (GBDT) was proposed. Firstly, classification features of content, user properties, information dissemination and time characteristic were extracted from the comments of micro-blog. Then an identification model based on GBDT algorithm was proposed to detect misinformation. Finally, two real micro-blog datasets were used to verify the efficiency and effectiveness of the model. The experimental results show that the proposed model can effectively improve the accuracy of micro-blog misinformation detection.

Key words: micro-blog; social network; misinformation; gradient boost decision tree; comment

0 引言

微博是如今网民发布信息和获取信息的主要渠道之一。根据中国互联网信息中心(China Internet Network Information Center, CNNIC)2017年1月发布的全国互联网发展统计报告[1],我国网民规模达7.31亿,其中微博用户超过2.67亿,占整体网民的36.5%。微博的低门槛特性使得用户可以不受时间地域的限制,自由表达自己的观点,使用户之间分享信息更加迅速、便捷。微博已经逐步渗透进人们的生活,影响人们的生活方式。

微博平均每天会增加数亿条博文,这些博文中既有真实的信息,也有大量的虚假信息,而虚假信息的泛滥对群众的影响非常大。例如:2017年3月初,一篇文章在网络上引起轩然大波,该文称疫苗会损害人体健康,危害无穷,家长们应该让孩子远离疫苗,甚至声称孩子自然感染疾病比打疫苗强。该虚假消息在传播的过程中,误导了网民的思想,很多家长选择了不再给孩子注射疫苗。4月,微博上纷纷在转一条如何鉴别草莓变色催熟的文章。该文称如果草莓籽是红色的,便是用了染色剂的缘故。浙江宁波、江苏徐州等地市民也的确发现,市面上很多草莓的籽是红色的,顿时心生不安。很多市民不再食用草莓,给社会造成了巨大的经济损失。然而,真实的情况是,草莓自然成熟后,有部分草莓籽是会变成红色的。有效地识别虚假信息对营造诚信、公平、健康的网络环境以及维持正常的社会秩序是十分必要的。

现有的研究主要是通过选取文本内容、用户属性和传播特性等方面的特征,然后构建合适的分类模型,以达到识别微博虚假消息的目的。但是,这些研究中往往只选取局部、片面的特征(如选取文本内容特征的统计特征、浅层传播特征或者简单的用户属性特征),没有全面、深入地分析并挖掘影响虚假消息识别的主要因素。另外,以往的研究中只是选用单一的分类器对微博虚假消息进行检测,如朴素贝叶斯(Naive Bayes, NB)、支持向量机(Support Vector Machine, SVM)、决

收稿日期:2017-08-28;**修回日期**:2017-10-10。 **基金项目**:教育部人文社会科学研究基金资助项目(13YJC860006);北京市自然科学基金资助项目(4172016);北京市科技计划项目(Z161100001616004)。

作者简介:段大高(1976—),男,湖南邵阳人,副教授,博士,CCF会员,主要研究方向:多媒体信息处理、现代网络通信、嵌入式系统、智能数据分析; 盖新新(1990—),女,河北邢台人,硕士研究生,主要研究方向:数据挖掘; 韩忠明(1972—),男,山西文水人,副教授,博士,CCF会员,主要研究方向:海量数据分析与挖掘、互联网挖掘、生物信息学; 刘冰心(1996—),女,北京人,主要研究方向:数据挖掘。

策树(Decision Tree,DT) 等,没有考虑使用组合多个弱分类器构建强分类器来识别虚假消息,故而识别的精度不高。因此,基于微博的短文本特性,本文提取微博评论的文本内容、用户属性、信息传播和时间特性四个方面的特征,构建基于梯度提升决策树(Gradient Boost Decision Tree, GBDT) 算法的微博虚假消息识别模型。实验结果表明,本文提出的识别方法能够有效提高虚假消息检测的准确率。

1　相关工作

近几年国内外关于微博虚假消息的研究逐渐增多。在国外方面,2011 年,Castillo 等[2] 提出了对 Twitter 话题可信度进行评估的方法,通过提取消息特征、用户特征、话题特征以及传播特征,采用 J48 决策树分类方法来预测热门话题是否可信。2012 年,Yang 等[3] 提出客户端类型和微博事件发生的地理位置两种新特征,采用 SVM 分类方法对谣言进行检测。实验结果表明,当微博所涉及的事件发生在国外而且使用非移动客户端时,此微博被判断为谣言微博的概率较高。2015 年,Dayani 等[4] 通过提取用户特征和内容特征,并采用 K 最近邻(K-Nearest Neighbors, KNN) 分类器以及 NB 分类器在 Twitter 中检测谣言中支持、反对、质疑、中性的评论。实验结果表明: 对于用户特征,KNN 分类器的效果并不理想; 而对于内容特征,朴素贝叶斯能有效检测出谣言话题下的评论数量。2015 年,Ma 等[5] 提出基于谣言生命周期的时间序列的社交上下文特征,包括微博内容特征、用户特征和传播特征,并采用线性 SVM 分类器分别在 Twitter 数据集与 DT、随机森林(Random Forest, RF) 以及 SVM-RBF 方法作比较。实验结果表明: 该文中提出方法的精确性比 DT、RF 以及 SVM-RBF 方法高,且达到与 DT、RF 以及 SVM-RBF 相同的精确性的用时最少。2015 年,Liu 等[6] 提出在 Twitter 上的实时谣言揭露,通过使用“群众智慧”和系统性方法来挖掘语言特征,并采用 DT 分类器、RF 分类器以及 SVM 分类器进行实验。实验结果表明: 该文中提出的方法在事件只有最初的 5 条 Tweets 以及最初的一小时内的预测结果都要高于其他方法; 而选取两个实时谣言跟踪网站 snopes. com 和 emergent. info 与人工验证方法相比,结果显示该方法能将检测延迟减少 25% 和 50% 。

与国外相比,国内关于虚假消息检测的研究相对较少。2013 年,蒋盛益等[7] 对现有成果进行了梳理,总结了这些研究的不足,指出了微博信息可信度分析的关键问题和核心方法,并对未来进行了展望。2013 年,贺刚等[8] 提出利用符号特征、链接特征、关键词分布特征和时间差等新特征,将微博谣言识别形式化为分类问题,利用 SVM 分类算法对微博进行分类,识别结果可以辅助识别谣言。2016 年,路同强等[9] 在分析微博谣言传播特点的基础上,结合微博文本内容、微博用户等方面的特征构建特征集合,将半监督学习算法应用到谣言检测中,以解决人工标注语料代价高昂的问题。2016 年,吴树芳等[10] 在 HITS(Hyperlink-Induced Topic Search) 算法的基础上,提出了融合用户交互行为和博文内容的微博用户可信度评估算法,分别构建基于交互行为和基于博文内容的微博用户有向链接图,通过反复训练法获得可信度阈值,绘制不同可信度算法的用户可信度曲线,验证了算法的可行性和有效性。2016 年,谢柏林等[11] 提出一种基于把关人行为的微博虚假信息及早检测方法。该方法利用模型状态持续时间概率为 Gamma 分布的隐半马尔可夫模型来刻画信息转发者和评论者对流行的真实信息的把关行为,基于此来及早识别微博上流行的虚假信息。实验结果表明该方法具有较好的性能和较高的在线检测速度。

2　特征选取

微博虚假消息与真实消息的评论存在着很大的差异。在文本内容方面,虚假消息的评论具有语气不确定程度强、消极词汇多、内容与源消息相关程度弱的特点; 在用户属性方面,虚假消息的发布者一般是非认证用户,其注册日期比较短、注册地信息不够详细,朋友数量远高于粉丝数量,并且不使用顶级域名; 在传播特性方面,网络大 V 用户对源消息的转发和 URL、@ 、hashtag 等符号信息将会影响用户对源消息的信任程度,进而影响微博的转发量; 在时间特性方面,距离源微博发布时间越久的微博,其是虚假消息的概率会越小。

基于以上分析,本文中总共选取了 11 个特征,并将这些特征分为四类: 基于文本内容的特征、基于用户的特征、基于信息传播的特征和基于时间的特征。其中, 基于文本内容的特征已在文献[12]中详细介绍,故在此不再多作阐述。表 1 列出了文中所使用的全部特征,并对特征作了简单的介绍。

表 1　特征及其描述

Tab. 1　Features and their description

分类	特征	描述
文本内容特征	支持性(SUP)	用户对消息倾向性的态度
	内容相关性(COR)	评论是否是对消息的补充
	置信度(CON)	用户发表评论语气的确定程度
用户特征	是否认证(VER)	用户身份是否经新浪微博认证
	关注度(ATT)	账户被关注程度
	注册日期(RED)	用户注册的实际天数
	注册地(POR)	用户注册位置信息的详细程度
	顶级域名(TLD)	个人介绍中是否有顶级域名
信息传播特征	URL/@ /hashtag	评论包含 URL/@ /hashtag 的数量
	意见领袖(OPL)	意见领袖传播过程中的影响力
时间特征	时间差(TID)	当前微博发布与原微博发布的时间跨度(单位: d)

2.1　关注度特征

微博用户之间存在的关系有两种: 关注与被关注。关注其他账户,则此账户为所关注账户的粉丝,可以看到其关注账户发表的博文。两个账户互相关注,两个账户即为朋友关系,都可以看到彼此发表的博文。郭浩等[13] 指出, 积极关注别人,保持较高的发文数量,就可以吸引更多的粉丝,获得更高的关注度,使社会化网络媒体营销更加有效。这说明一些在微博上传播虚假消息的账户,可能会关注多个其他账户,以希望这些账户能够关注自己,看到自己发表的博文并传播这些消息,结果表现为朋友数量远远多于粉丝数量。正常用户的朋友和粉丝的数量一般相差不多,其微博上的关注关系一般是现实中朋友关系的映射。因此,将关注度特征计算公式表示如下:

$$ATT_u = FOL_u/(FOL_u + FRI_u) \tag{1}$$

其中: FOL_u 表示用户 u 的粉丝数量, FRI_u 表示用户 u 的朋友

数量。正常用户的关注度值要高于虚假消息传播用户的关注度值。

2.2 顶级域名特征

顶级域名是付费服务,它具有易查找、可信度高、独立性等优点,一般来说,只有一些有需要的个人或者是公司才会使用这项服务。而虚假消息传播用户本身是为了盈利,故而只会注册一些免费的账户来传播信息,所以此特征具有明显的区分性。顶级域名特征(TLD)的取值是{0,1},0 表示个人介绍中有顶级域名的用户的特征值,1 表示个人介绍中无顶级域名的用户的特征值。

2.3 意见领袖特征

王永强[14]指出,所谓意见领袖,指的是人际传播网络中经常为他人提供信息、意见、评论并对他人施加影响的“活跃分子”,是大众传播效果形成过程的中介或过滤环节。意见领袖在信息传播过程中的影响是巨大的。例如,2010 年 12 月 6 日,微博上爆出金庸先生“去世”的消息,当晚《中国新闻周刊》在官方微博上转发了这则微博,这则消息事后被证实为谣言。但网络大 V 的转发加速了消息的传播,导致此谣言在数分钟内即被转发近千条。为了衡量意见领袖在传播过程中的影响,本文中将用户分为两类:认证用户和普通用户,主要获取认证用户在传播过程中的影响。由此,将意见领袖特征的计算公式表示为:

$$OPL_u = \begin{cases} REP_{\text{ver}}/REP_{\text{ori}}, & \text{认证用户} \\ 0, & \text{普通用户} \end{cases} \tag{2}$$

其中:REP_{ver} 表示通过认证用户微博被转发的数量,REP_{ori} 表示源消息的转发数量。如果是普通用户,则意见领袖特征为 0。

2.4 时间差特征

谣言的传播有四个阶段:潜伏期、变异期、爆发期和消亡期。谣言的爆发期通常时间比较短暂。在谣言微博发布后,随即会出现一系列辟谣的微博,并且其传播要比谣言微博快很多,所以,距离谣言源微博时间越久的微博,它是谣言的概率会越小。根据以上分析,用时间差特征来表示当前评论发布时间距微博源消息发布时间的间隔,其计算公式表示如下:

$$TID_w = TIM_w - TIM_m \tag{3}$$

其中:TIM_w 表示当前评论 w 的发布时间,TIM_m 表示源消息 m 的发布时间。时间差特征以天为单位。

2.5 其他特征

是否认证特征(VER)、注册日期特征(RED)、注册地特征(POR)在一定程度上反映了用户的可信度。本文中通过是否认证特征将用户分为两类:认证用户和普通用户。是否认证特征的取值是{0,1},0 表示普通用户的特征值,1 表示认证用户的特征值。注册日期特征是指用户注册的实际天数,通过计算用户当前评论的发表时间与用户的注册日期的差值来实现。注册地特征衡量用户注册位置信息的详细程度,其取值是{0, 0.5, 1},0 表示注册位置信息为空的用户的特征值,0.5 表示注册位置信息中只有省份的用户的特征值,1 表示注册位置信息中既有省份又有城市的用户的特征值。

3 特征选取

本文在微博消息的评论中提取四个方面的特征,从不同的角度衡量微博虚假消息与真实消息之间的区别。与真实消息相比,在文本内容方面,虚假消息中 SUP 特征值为负、COR 特征值较低、CON 特征值较低的评论更多;在用户属性方面,虚假消息的发布者一般是 VER 特征值为 0,且 ATT 特征值较低、RED 特征值较低、POR 特征值较低、TLD 特征值为 0;在传播特性方面,虚假消息的评论中 URL、@、hashtag 特征值较低,OPL 特征值较低;在时间特性方面,虚假消息中 TID 特征值较小的评论更多。特征提取的目的是为了分析影响类别之间差异的主要因素。

微博虚假消息识别问题,可以看作一个分类问题。在数据量较大的情况下,需要选择一个分类速度高且准确率也高的模型。因此本文中选用 GBDT 算法,它是由 Friedman[15]提出的组合决策树模型,是一种由多个弱分类器经过多次迭代形成的强分类器。与传统 Boosting 算法(如 Adaboost)不同的是,GBDT 算法的基分类器是回归树,其迭代的目的是通过计算上一次模型的负梯度来改进模型,然后在残差减少的梯度方向上建立新的决策树;Adaboost 算法通过简单地调整正确、错误样本的权重来改进模型,二者有本质区别。

现给定微博数据样本$\{(x_i, y_i)\}(i = 1,2,\cdots,n)$。由于虚假消息识别是一个分类过程,故采用对数损失函数,即:

$$L(y, F(x)) = 2\sum_{i=1}^{q}\log(1 + \exp(-2y_i p_i)) \tag{4}$$

其中:$x_i = (x_{1i}, x_{2i}, \cdots, x_{qi})$,$n$ 为样本的数量,q 为虚假消息识别中特征的数量,y_i 为样本的实际标签,p_i 为样本的预测标签。GBDT 算法的详细步骤如下:

1) 初始化模型,估计使损失函数最小化的常数值 β:

$$F_0(x) = \arg\min_{\beta}\sum_{i=1}^{n} L(y_i, \beta) \tag{5}$$

2) 在上一次模型损失函数的梯度下降方向上建立模型,从 $m = 1$ 到 M(M 为迭代次数):

① 计算损失函数的负梯度在当前模型的值,将它作为残差 r_{im} 的估计值:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F_m(x) = F_{m-1}(x)}; \quad i = 1,2,\cdots,n \tag{6}$$

② 将 ① 中得到的估计残差作为输入,拟合一棵回归树,求得回归树的叶节点区域 $R_{j,m}(j = 1,2,\cdots,J)$。

③ 为使损失函数极小化,对于 $j = 1,2,\cdots,J$,求得沿梯度下降方向的最优步长 β_{jm}:

$$\beta_{jm} = \arg\min_{\beta}\sum_{x_i \in R_{j,m}} L(y_i, F_{m-1}(x) + \beta) \tag{7}$$

④ 更新模型 $F_m(x)$:

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J}\beta_{jm} I;\ x \in R_{jm} \tag{8}$$

3) 迭代结束,得到模型 $F_M(x)$:

$$F_M(x) = \sum_{m=1}^{M}\sum_{j=1}^{J}\beta_{jm} I;\ x \in R_{jm} \tag{9}$$

4) 根据得到的模型,估算样本预测为正类的概率 $p_+(x)$ 和预测为负类的概率 $p_-(x)$:

$$\begin{cases} p_+(x) = p_r(y = 1 \mid x) = \dfrac{1}{1 + e^{-2F_M(x)}} \\ p_-(x) = p_r(y = -1 \mid x) = \dfrac{1}{1 + e^{2F_M(x)}} \end{cases} \tag{10}$$

5) 据以下准则预测样本标签 $y(x)$，其中 $c(-1,1)$ 是代价函数，表示当真实类别为 1，预测类别为 -1 时的代价:

$$y(x) = 2^* \ l\{c(-1,1)p_+(x) > c(1,-1)p_-(x)\} - 1 \tag{11}$$

其中: l{} 是将布尔值转换为{0,1} 函数。

4　实验结果与分析

4.1　实验数据

本文实验数据集有两个: 数据集 1 选自文献[16]，其数据采集自新浪微博社区管理中心和新浪微博 API 接口，总共包含 2313 个谣言和 2351 个非谣言，内容包括旅游、球赛、娱乐、生活、常识等话题。数据集 2 是在文献[17]中数据集的基础上，采集新浪微博社区管理中心中的不实信息作为谣言数据，然后在新浪微博上爬取与谣言微博具有相同时间跨度的微博作为非谣言数据，保留原微博字数超过 10，评论数超过 200 条的微博。处理后的数据集 2 总共包含 447 个谣言和 455 个非谣言，内容主要是 2013 年和 2014 年的热点新闻。两个数据集的统计情况见表 2。相比数据集 2，数据集 1 包含的特征的相关信息更多，本文在数据集 1 中提取了表 1 中介绍的所有特征; 而数据集 2 则缺少表 1 中某些特征的相关信息，最终在数据集 2 中提取了 SUP、COR、CON、URL、@、hashtag 和 TID 特征。本文中提出的虚假消息识别模型是一个综合模型，如果需要针对具体某个事件进行识别，可以结合本文中的模型，并使用和事件本身相关的特征进行识别。实验按照 8: 2的比例随机划分数据集，即数据集的 80% 作训练集，余下 20% 作测试集，均采用十折交叉验证。

表 2　数据集的统计情况

Tab. 2　Statistics of the data set

数据集	用户数量	微博数量	事件数量	谣言数量	非谣言数量
1	2746818	3805656	4664	2313	2351
2	205391	286714	902	447	455

4.2　特征归一化

从评论中提取的特征如果直接用于分类，其相差过大的权重范围将会影响分类器的准确性。为此，对特征进行归一化处理是十分有必要的。本文使用式(12) 对特征进行归一化处理，归一化后特征权重限定在[0,1]区间，可以消除离群数据对分类的影响，也可以使计算过程收敛得更快。

$$y_{i,j} = \frac{x_{i,j} - \min(x_{\cdot j})}{\max(x_{\cdot j}) - \min(x_{\cdot j})} \tag{12}$$

其中: $\min(x.j)$ 表示第 j 列特征权重的最小值，$\max(x.j)$ 表示第 j 列特征权重的最大值。

4.3　评价指标

为了评测微博虚假消息检测的结果，本文选用查准率(P)、查全率(R) 以及 $F1$ 值作为评价标准。

$$P = TP/(TP + FP) \tag{13}$$

$$R = TP/(TP + FN) \tag{14}$$

$$F1 = 2PR/(P + R) \tag{15}$$

其中: TP 是被正确判别为谣言的微博数，FP 是被错误判别为谣言的微博数，FN 是被错误判别为非谣言的微博数。另外，为了衡量总体的分类效果，采用下面的公式计算总体分类正确率:

$$Acc = \text{识别正确的微博数} / \text{总微博数} \tag{16}$$

4.4　结果分析

微博虚假消息的评论存在着语气不确定程度强、消极词汇多、重复源消息等的特点。基于此，通过统计微博消息中被模型判定为虚假消息评论的比例，可以得到一个阈值，当微博消息中的虚假评论达到这个阈值的时候，则此微博被判定为虚假消息。

为了比较不同分类器分类的结果，本文选择 Castillo 等[2]使用的 J48 决策树分类器、Yang 等[3]使用的 SVM 分类器以及 Kwon 等[18]使用的 RF 分类器。其中，SVM 核函数选择径向基核函数(Radial Basis Function, RBF)，使用 LIBSVM[19] 中的 grid 来寻找最优的参数 c 和 γ。

4.4.1　实验阈值

实验以正确率 Acc 为基准，使用不同分类器获得使正确率 Acc 最高的阈值，称为最佳阈值，它可以最好地将虚假消息与真实消息区分开。两个数据集的最佳阈值统计结果如表 3 所示。

表 3　数据集的最佳阈值

Tab. 3　The best threshold of the data sets

数据集	分类器	最佳阈值	数据集	分类器	最佳阈值
1	GBDT	0.16	2	GBDT	0.17
	RF	0.27		RF	0.28
	SVM	0.40		SVM	0.45
	J48	0.45		J48	0.48

4.4.2　特征重要性

为了验证特征在分类过程中的影响，以正确率 Acc 为基准，用 GBDT 分类器的默认参数来对不同的特征进行训练，数据集 1 使用表 1 中的全部特征，数据集 2 使用 SUP、COR、CON、URL、@、hashtag 和 TID 特征，两个数据集的训练结果如表 4 所示。其中，特征前面的“-”符号表示不包括该特征的特征集，Acc 中的“—”表示实验没有使用该特征集。

表 4　不同特征对分类的影响

Tab. 4　Influence on classification with different features

特征集	Acc		特征集	Acc	
	数据集 1	数据集 2		数据集 1	数据集 2
(-) SUP	0.891	0.872	(-) TLD	0.880	—
(-) COR	0.883	0.773	(-) URL	0.881	0.881
(-) CON	0.888	0.886	(-) @	0.882	0.889
(-) VER	0.887	—	(-) hashtag	0.891	0.890
(-) ATT	0.884	—	(-) OPL	0.887	—
(-) RED	0.870	—	(-) TID	0.831	0.887
(-) POR	0.886	—	ALL	0.894	0.892

从表 4 中可以明显看出，实验中用到的所有特征都有助于提升微博虚假消息的检测效果。其中，数据集 1 使用所有特征(ALL) 的正确率 Acc 是 0.894，高于数据集 2(0.892)。这是因为数据集 1 使用了表 1 中的全部特征，数据集 2 只使用表 1 中的部分特征。在数据集 1 中，时间差特征(TID) 和注册日期特征(RED) 对总体分类结果影响是最大的; 在数据集 2 中，内容相关性特征(COR) 和支持性特征(SUP) 对总体分类结果影响是最大的。这是因为数据集 1 中的话题，例如生活、常识等，其讨论的时间会比较长，所以在数据集 1 中，关

于时间特征的重要性会比较高;数据集 2 的话题是热点新闻,其评论内容比数据集 1 更加规范,所以在数据集 2 中,起重要作用的主要是基于文本内容的特征,而新闻的时效一般都比较短,故时间差特征(TID)在数据集 2 中体现的重要性没有在数据集 1 中的重要性高。

4.4.3 分类结果

为了便于比较,实验将 GBDT、RF、J48 中决策树的最大深度统一设定为 15,SVM 核函数选择 RBF,使用 LIBSVM 寻找最优的参数 c 和 γ。两个数据集的实验结果如表 5 所示。其中,F 表示虚假消息,T 表示真实消息。

从表 5 中可以看出,GBDT 分类器的正确率 *Acc* 要明显高于 SVM 和 J48。这是因为 GBDT 是一种由多个弱分类器形成的强分类器,其效果要好于单一的分类器;GBDT 分类器的分类效果要好于 RF,这是因为 GBDT 的输出是所有结果的累积,RF 采用多数投票原则决定最终结果,且 RF 训练调参时依赖于决策树的最大深度,而 GBDT 只需很小的深度就可以达到很高的精度,实验中为了提高分类速度,没有给 RF 增大深度。数据集 1 中 GBDT 分类器的正确率 *Acc* 要高于数据集 2 中 GBDT 分类器的 *Acc*,因为数据集 1 中使用了表 1 中的全部特征,数据集 2 只使用表 1 中的部分特征,且数据集 1 比数据集 2 数据量大,故分类模型加精确。

表 5 不同分类器的分类结果

Tab. 5 Classification results of different classifiers

数据集	分类器	*Acc*	*P*(*F*)	*R*(*F*)	*F*1(*F*)	*P*(*T*)	*R*(*T*)	*F*1(*T*)
1	GBDT	0.940	0.932	0.949	0.940	0.946	0.930	0.938
	RF	0.908	0.869	0.948	0.907	0.956	0.877	0.915
	SVM	0.881	0.873	0.897	0.885	0.889	0.864	0.876
	J48	0.858	0.853	0.893	0.873	0.836	0.863	0.849
2	GBDT	0.929	0.920	0.936	0.928	0.929	0.910	0.919
	RF	0.910	0.900	0.930	0.915	0.925	0.887	0.906
	SVM	0.900	0.880	0.911	0.895	0.907	0.868	0.887
	J48	0.883	0.869	0.876	0.872	0.874	0.869	0.871

5 结语

本文从微博评论的角度在文本内容、用户属性、信息传播和时间特性四个方面分析影响分类的因素并提取分类特征,并基于 GBDT 算法设计微博虚假消息识别模型。通过在两个微博数据集上的对比实验分析可以看到,模型在数据集 1 上的实验结果要好于在数据集 2 上的实验结果;在数据集 1 中,起主要作用的是基于时间的特征,在数据集 2 中,起主要作用的是基于文本内容的特征。两个数据集上的实验均表明,本文提出的基于 GBDT 的方法能够有效提高微博虚假消息检测的准确率。

但是,微博虚假消息检测的价值体现在能够及早地发现并处理,以减少对社会的危害。因此,下一步的工作重点是通过借助传播模型以及消息传播过程中用户的认知与识别能力,综合更复杂的特征来构建合适的模型,实现实时检测微博虚假消息的目的。

参考文献:

[1] 中国互联网络信息中心. 中国互联网络发展状况统计报告[R]. 北京: 中国互联网信息中心,2017. (China Internet Network Information Center (CNNIC). Statistical report on Internet development in China [R]. Beijing: China Internet Network Information Center, 2017.)

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

[3] YANG F, LIU Y, YU X, et al. Automatic detection of rumor on Sina Weibo [C]// MDS '12: Proceedings of the 2012 ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: Article No. 13.

[4] DAYANI R, CHHABRA N, KADIAN T, et al. Rumor detection in Twitter: an analysis in retrospect [C]// ANTS 2015: Proceedings of the 2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems. Piscataway, NJ: IEEE, 2015: 1 - 3.

[5] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites [C]// CIKM '15: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015: 1751 - 1754.

[6] LIU X, NOURBAKHSH A, LI Q, et al. Real-time rumor debunking on twitter [C]// CIKM '15: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015: 1867 - 1870.

[7] 蒋盛益,陈东沂,庞观松,等. 微博信息可信度分析研究综述[J]. 图书情报工作,2013,57(12):136 - 142. (JIANG S Y, CHEN D Y, PANG G S, et al. Research review of information credibility analysis on microblog [J]. Library and Information Service, 2013, 57(12): 136 - 142.)

[8] 贺刚,吕学强,李卓,等. 微博谣言识别研究[J]. 图书情报工作,2013,57(23):114 - 120. (HE G, LYU X Q, LI Z, et al. Automatic rumor identification on microblog [J]. Library and Information Service, 2013, 57(23): 114 - 120.)

[9] 路同强,石冰,闫中敏,等. 一种用于微博谣言检测的半监督学习算法[J]. 计算机应用研究,2016,33(3):744 - 748. (LU T Q, SHI B, YAN Z M, et al. Semi-supervised learning algorithm applied to microblog rumors detection [J]. Application Research of Computers, 2016, 33(3): 744 - 748.)

[10] 吴树芳,徐建民. 基于 HITS 算法的微博用户可信度评估[J]. 山东大学学报(工学版),2016,46(2):1 - 7. (WU S F, XU J M. Evaluation of microblog users' credibility based on HITS algorithm [J]. Journal of Shandong University (Engineering Science), 2016, 46(2): 1 - 7.)

[11] 谢柏林,蒋盛益,周咏梅,等. 基于把关人行为的微博虚假信息及早检测方法[J]. 计算机学报,2016,39(4):730 - 744. (XIE B L, JIANG S Y, ZHOU Y M, et al. Misinformation detection based on gatekeepers' behaviors in microblog [J]. Chinese Journal of Computers, 2016, 39(4): 730 - 744.)

[12] 段大高,王长生,韩忠明,等. 基于微博评论的虚假消息检测模型[J]. 计算机仿真,2016,33(1):386 - 390. (DUAN D G, WANG C S, HAN Z M, et al. A rumor detection model based on Weibo' reviews [J]. Computer Simulation, 2016, 33(1): 386 - 390.)

(下转第 420 页)

-590.
[6] LONG C, WONG R C-W, YU P S, et al. On optimal worst-case matching [C]// SIGMOD '13: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2013: 845 - 856.
[7] TONG Y, SHE J, DING B, et al. Online mobile micro-task allocation in spatial crowdsourcing [C]// ICDE 2016: Proceedings of the 2016 IEEE 32nd International Conference on Data Engineering. Piscataway, NJ: IEEE, 2016: 49 - 60.
[8] LEONG HOU U, MOURATIDIS K, MAMOULIS N. Continuous spatial assignment of moving users [J]. The VLDB Journal — The International Journal on Very Large Data Bases, 2010, 19(2): 141 - 160.
[9] GAREY M R, JOHNSON D S. Computers and Intractability: A Guide to the Theory of NP-completeness [M]. New York: W. H. Freeman, 1979: 90 - 91.
[10] MEHTA A. Online matching and ad allocation [J]. Foundations & Trends in Theoretical Computer Science, 2013, 8(4): 265 - 368.
[11] WANG Y, WONG S C-W. Two-sided online bipartite matching and vertex cover: beating the greedy algorithm [C]// ICALP 2015: Proceedings of the 2015 International Colloquium on Automata, Languages, and Programming, LNCS 9134. Berlin: Springer, 2015: 1070 - 1081.
[12] TING H F, XIANG X. Near optimal algorithms for online maximum edge-weighted b-matching and two-sided vertex-weighted b-matching [J]. Theoretical Computer Science, 2015, 607(P2): 247 - 256.
[13] HASSAN U U, CURRY E. A multi-armed bandit approach to online spatial task assignment [C]// UIC-ATC-ScalCom '14: Proceedings of the 2014 IEEE 11th International Conference on Ubiquitous Intelligence and Computing, and 2014 IEEE 11th International Conference on Autonomic and Trusted Computing, and 2014 IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops. Washington, DC: IEEE Computer Society, 2014: 212 - 219.
[14] 宋天舒,童咏昕. 空间众包环境下的三类对象在线任务分配[J]. 软件学报,2017,28(3):611 - 630. (SONG T S, TONG Y X. Three types of objects online task allocation space crowdsourcing environment. [J] Journal of Software, 2017, 28(3): 611 - 630.)
[15] REN J, ZHANG Y, ZHANG K, et al. SACRM: Social Aware Crowdsourcing with Reputation Management in mobile sensing [J]. Computer Communications, 2014, 65: 55 - 65.
[16] DAI W, WANG Y, JIN Q, et al. An integrated incentive framework for mobile crowdsourced sensing [J]. Tsinghua Science and Technology, 2016, 21(2): 146 - 156.
[17] 张晓航,李国良,冯建华. 大数据群体计算中用户主题感知的任务分配[J]. 计算机研究与发展,2015,52(2):309 - 317. (ZHANG X H, LI G L, FENG J H. User topic aware task assignment in large data group computing [J]. Journal of Computer Research and Development, 2015, 52 (2): 309 - 317).
[18] CHEN Z, FU R, ZHAO Z, et al. gMission : a general spatial crowdsourcing platform [J]. Proceedings of the Very Large Data Base Endowment, 2014, 7(13): 1629 - 1632.

This work is partially supported by the National Natural Science Foundation of China (61170052), the Jinan City University Institute of independent innovation project (201401211).

LIU Hui, born in 1992, M. S. candidate. His research interests include crowdsourcing, deep learning.

LI Sheng´en, borin in 1963, Ph. D., professor. His research interests include data warehouse, online analysis, data mining.

(上接第 414 页)

[13] 郭浩,陆余良,王宇,等. 多特征微博垃圾互粉检测方法[J]. 中国科技论文,2012,7(7):548 - 551. (GUO H, LU Y L, WANG Y, et al. Detection of spam mutual concerns in micro-blogs based on multi-features [J]. China Sciencepaper, 2012, 7(7): 548 - 551.)
[14] 王永强. 微博“意见领袖”少数派的权利[N]. 中国经营报, 2011-09-19 (C05). (WANG Y Q. Micro-blog “opinion leaders” the minority´ rights [N]. China Business Journal, 2011-09-19 (C05).)
[15] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine [J]. The Annals of Statistics, 2001, 29(5): 1189 - 1232.
[16] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]// IJCAI 2016: Proceedings of the 25th International Joint Conference on Artificial Intelligence. London: dblp Computer Science Bibliography, 2016: 3818 - 3824.
[17] JIN Z, CAO J, JIANG Y-G, et al. News credibility evaluation on microblog with a hierarchical propagation model [C]// ICDM '14: Proceedings of the 2014 IEEE International Conference on Data Mining. Washington, DC: IEEE Computer Society, 2014: 230 - 239.
[18] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media [C]// ICDM 2013: Proceedings of the 2013 IEEE 13th International Conference on Data Mining. Piscataway, NJ: IEEE, 2013: 1103 - 1108.
[19] CHANG C-C, LIN C-J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): Article No. 27.

This work is partially supported by the Humanities and Social Sciences Foundation of Ministry of Education (13YJC860006), the Beijing Municipal Natural Science Foundation (4172016), the Beijing Science and Technology Project (Z161100001616004).

DUAN Dagao, born in 1976, Ph. D., associate professor. His research interests include multi-media information processing, modern network communication, embedded system, intelligent data analysis.

GAI Xinxin, born in 1990, M. S. candidate. Her research interests include data mining.

HAN Zhongming, born in 1972, Ph. D., associate professor. His research interests include mass data analysis and mining, Web mining, bioinformatics.

LIU Bingxin, born in 1996. Her research interests include data mining.