文章编号：1003-0077(2019)06-0132-9

微博谣言事件自动检测研究

王志宏[1]，过 弋[1,2,3]

(1. 华东理工大学 信息科学与工程学院，上海 200237；
2. 大数据流通与交易技术国家工程实验室 商业智能与可视化研究中心，上海 200237；
3. 石河子大学 信息科学与技术学院，新疆 石河子 832003)

摘 要：互联网大数据环境下，谣言事件的散播已成为以微博为代表的在线社交网络持续健康稳定发展的主要障碍之一，因此及时有效地进行谣言事件自动检测对营造清朗的网络环境和维护社会和谐发展有着现实意义。该文以微博事件为背景，综合谣言事件特征随时间变化特性以及时间维度上谣言事件的分布特点，引入论域划分思想，基于模糊聚类算法提出了随时间动态变化的事件时序特征构建模型；同时，基于社会学中谣言的传播原理，提出将事件流行度、模糊度和流传度作为微博谣言事件检测分类器的三项新特征。实验结果表明，该文提出的动态时序特征表示方法和三项新特征使谣言事件自动检测效果得到了可观提升。

关键词：谣言事件检测；动态时序特征；SVM；在线社交网络

中图分类号：TP391 **文献标识码**：A

Automatic Rumor Event Detection in Chinese Microblogs

WANG Zhihong[1], GUO Yi[1,2,3]

(1. Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Business Intelligence and Visualization Research Center, National Engineering Laboratory for Big Data Distribution and Exchange Technologies, Shanghai 200237, China; 3. School of Information Science and Technology, Shihezi University, Shihezi, Xinjiang 832003, China)

Abstract: Online Social Networks (OSNs) rumor events detection has realistic significance to improve the quality of OSN information ecology environment and maintain social harmony. This paper, integrating the variation of rumor events features over time and the distribution of rumor events in time dimension, proposes a fuzzy algorithm based model to construct the dynamic time series features over time by introducing the idea of domain division. Meanwhile, we introduce the popularity, the ambiguity and the spread as three new features based on the communication theory of rumor events in sociology. The experimental results testify that our proposed model and new features improve the performance of automatic rumor events detection on Chinese Microblogs.

Keywords: rumor events detection; dynamic time series features; SVM; online social networks

0 引言

随着在线社交网络的迅速发展，以微博为代表的碎片化信息创造了一个原子型的世界，谎话、流言、绯闻等大量不实信息在其中高速传播。“信息污染”导致人们难以从纷繁复杂的信息中甄别出可靠信息，严重影响了人们正常的生活秩序。微博内容主要通过人与人之间的“关注-被关注”网络进行传播。人与人、人与信息之间的高度互联融合，使得人人都可以参与到信息的产生和传播中，这种病毒式的传播方式促使一条信息能够在极短时间内传播到数百万的用户。例如，2011 年 3 月 16 日突发的“抢盐事件”[1]，主要是一条关于“日本核辐射会污染海

收稿日期：2018-07-20 定稿日期：2018-09-03

基金项目：国家重点研发计划(2018YFC0807105)；国家自然科学基金(61462073)；上海科学技术委员会(17DZ1101003，18511106602，18DZ2252300)

水导致以后生产的盐都无法食用，且吃含碘的食用盐可防核辐射”的谣言信息在社交网络上疯狂传播。从而，导致在我国大部分城市和农村一夜之间出现了“抢盐潮”。很多商店纷纷打出“盐已售完”等标识，并出现了一盐难求、高价售盐等现象。因此，自动高效的识别在线社交网络中的谣言事件意义重大，尤其是微博等在线社交媒体。

日常生活中，人们大多基于自己的常识或通过新闻网站、公共社区等来分辨微博事件的真假。例如，Snopes①、微博社区管理中心②、新浪微博官方辟谣账号(“@微博辟谣”③)等。但是这类网站媒体的报道并不完整且具有一定的时滞性，因此对谣言事件进行自动识别，可以帮助我们更好地防范谣言，辅助管理机构进行谣言干预和治理。

目前，微博平台上的谣言事件自动检测研究仍处于起步阶段，大部分的研究工作都将这一问题作为分类任务来处理，即根据人工构造的特征使用传统机器学习的分类算法进行谣言事件的识别。主要包括浅层的统计特征，如谣言事件的内容[2-4]、用户属性[5]、传播方式[6-7]等；以及深层的文本内容特征，如谣言事件情感倾向性[8-9]、事件主题[10]、事件关键词[11]等。本文在上述特征的基础上，根据传播学者 Crouse [12]提出的谣言传播公式“谣言的流通量＝事件的重要性×事件的模糊性/公众批判能力”，考虑谣言事件的传播原理，提出了事件流行度、模糊度和流传度三项新特征，用于微博谣言事件的自动检测。

另外，上述研究工作在构建分类特征时，忽略了事件特征随着事件发展的时间变化特性。仅仅基于单个观察窗口或固定的观察点进行特征构建，往往难以表示谣言事件的一般发展传播模式。因此，Kwon 等[13]首次指出了谣言事件传播过程中时间属性的重要性，并提出了推文数量随时间变化的时间序列拟合模型，在 Twitter 数据集上获得了较好的检测效果。Ma 等[14]在 Kwon 等研究的基础上进一步扩展了随时间变化的特征集合，利用简单的等长时间序列划分来观察谣言事件特征随时间的变化，并在 Twitter 数据集和新浪微博数据集上获得了不错的识别结果。但他们在构建谣言事件时间序列特征的过程中，均未考虑事件时序数据的分布特点，即在时间维度上事件本身的聚合程度。为了更好地观察和表示谣言事件特征随时间的变化，本文引入模糊时间序列模型中的论域划分思想，将事件的时间跨度作为论域，提出了基于模糊聚类的事件时序数据动态划分算法，并在此基础上构建了随时间变化的事件特征集合。实验结果表明，本文提出的基于动态时间序列的事件特征表示方法，可以有效提高谣言事件检测的效果。

1　相关工作

谣言事件在社交网络环境下发展迅猛，其滋生和传播容易误导社会舆论，导致线下的“群体性恐慌”以及线上的“网络暴力”。社交网络谣言事件治理工作正变得日益重要。其中，微博谣言事件检测引起了学术界广泛的关注。现有谣言事件检测方法一般分为两大类：人工检测和基于机器学习的自动检测[15]。

在人工谣言事件检测方面，就国内而言，新浪微博提供了官方辟谣账号“@微博辟谣”和基于众包的辟谣平台“微博不实信息举报中心”。但由于微博平台谣言检测工作量大、人力资源不足等，截止到 2018 年 7 月 16 日，共发布和审核谣言事件数为 40 624 条(其中，“@微博辟谣”发布了 4 654 条辟谣信息，“微博不实信息举报中心”共审核判定 35 970 条不实事件)，难以反映微博平台上实际的谣言事件规模，覆盖率不足。Snopes 是国外一家专门核查并揭穿谣言和传闻的网站，该网站对谣言事件会使用“真/假/不确定”的可信度评定，目前 Snopes 已经公布了 11 887 条信息的判定结果。但是相对于社交网络上的谣言事件来说，该网站所能发挥的作用依然很小。所以，由于不能提供足够的人力资源进行谣言事件的判定和检测，人工谣言事件检测方法存在以下局限性：(1)对信息的覆盖率不足；(2)谣言检测周期较长，如果在谣言带来大量危害前仍无法进行谣言事件的判定，那么谣言事件检测的工作将失去意义。

在自动谣言事件检测方面，现有大部分研究工作主要将这一问题作为分类任务来处理，重点在于分类算法的选择和改进，以及构造更有效的检测特征。Yang 等[2]提出基于传统的内容、用户、传播特征以及新增的客户端类型和事件地理位置特征共五大类谣言事件检测特征，并使用 SVM 模型进行单文本的谣言事件自动检测。文献[4]从源微博评论内容角度定义了支持性、置信度、内容相关性三个特

① https://www.snopes.com/

② http://service.account.weibo.com/?type=5&status=0

③ https://weibo.com/weibopiyao

征，构建了SVM分类模型，并有效地识别出了微博虚假消息。文献[5]则从用户行为的角度出发，提出了基于用户行为的新的谣言事件检测特征，并对Logistics回归、SVM、朴素贝叶斯、决策树和K近邻五种算法做了实验对比。有学者还提出基于微博特有的转发行为形成的传播网络进行谣言事件检测，Wu等[6]通过对单文本谣言事件传播规律的分析，明确指出了谣言和非谣言在传播过程中转发模式的区别，并将信息发布、转发行为特征与内容特征相结合，利用混合SVM分类器进行谣言识别，取得了较好的结果。Kwon等[13]则从时序、结构和语言三个方面对谣言事件的传播特征进一步细分和研究，并在SVM、决策树和随机森林三种算法上进行了实验对比。Ma等[14]针对多文本谣言事件的特征会随着事件的传播不断变化的情况，建立了一种时序结构用以描述对时间敏感的谣言事件检测特征在谣言事件全生命周期的时间序列上的变化，并使用SVM、随机森林和决策树构建谣言事件自动识别模型。上述方法大都是基于谣言事件浅层的统计特征或信息传播特征，并未挖掘谣言事件传播过程中的深层语义特征。

毛二松等[8]考虑微博谣言事件的情感倾向性、意见领袖传播影响力等深层语义特征，通过训练集成分类器对微博谣言事件进行检测。祖坤琳等[9]首次提出将微博评论的情感倾向作为谣言事件检测分类器的新特征，使谣言检测的分类效果得到可观提升。杨文太等[10]从谣言事件主题角度出发，借鉴了物理学中的动力学理论对微博突发话题特征进行建模，以较小的时间窗口来捕获谣言事件语义特征，同时也解决了检测工作的及时性问题。武庆圆等[11]则针对短文谣言事件词语稀疏、语义提取困难等问题，通过在文本与标签之间引入语义层构建了一个多标签双词主题模型，用于发现社交媒体上短文本属于谣言的倾向。上述研究的核心是为谣言事件构造合适的特征，使用传统机器学习的分类算法进行谣言事件自动检测。

近年来，随着深度神经网络技术在自然语言处理、图像处理等领域取得的一系列突破性研究成果，其强大的特征学习与特征表示能力引起了广泛关注。在谣言事件检测领域，Ma等[16]首次引入神经网络模型对微博谣言事件的多文本序列数据进行深层特征表示，通过构建循环神经网络(RNN)模型对谣言事件进行检测，一定程度上克服了传统手工特征构造的复杂性问题，提高了谣言事件自动检测的准确率。但深度神经网络模型的训练需要大量的数据和计算资源，同时网络的层数、模型的架构以及模型的可解释性都是复杂且具有挑战的问题。本文的主要研究工作是针对多文本微博事件信息寻找更具有表示能力的谣言事件特征，使用传统分类算法进行谣言事件的自动检测。

综上所述，从研究方法的角度来讲，谣言事件检测的主要研究工作大多是通过构造事件特征，采用机器学习的分类算法进行谣言事件检测，主要包括浅层的统计特征[2-7]及深层的文本内容特征[8-11]。本文在上述特征的基础上，基于社会学的谣言传播原理提出了事件流行度、模糊度和流传度三项新特征用于微博谣言事件的自动检测；从研究对象的检测粒度上来说，微博谣言事件的检测对象可分为单文本事件的细粒度谣言检测[2-9,11]和多文本事件的粗粒度谣言检测[10,13-14,16]。本文研究主要面向多文本时间序列数据的谣言事件检测。为了更好地观察谣言事件特征随时间的变化，本文综合时间维度上事件数据本身的聚合程度，提出基于模糊聚类的事件时序数据动态划分算法，构建了随时间变化的事件特征集合，有效提高了谣言事件检测效果。

2 微博谣言事件自动检测模型

给定微博事件E，与该事件相关的微博消息集合为$P=\{p_1, p_2, \cdots, p_n\}$。本文首先对事件相关的微博按时间升序排列，然后采用时序数据动态划分算法在时间维度上对事件进行分割，并在此基础上构建随时间变化的事件特征集合(包括基础特征和新增特征)。最后，融合所有特征向量训练SVM模型进行微博谣言事件的自动检测。其流程如图1所示。

2.1 基于动态时间序列的特征构建

现有大部分谣言事件检测的研究工作中，在构建分类特征时忽略了特征随事件发展的时间变化特性，仅对固定时间窗口内的事件进行特征构建。Ma等[14]指出了事件特征随时间变化的特性，并利用简单的等长时间序列划分来捕捉谣言事件特征的时间变化特性，检测效果得到了提升。但等长的时间序列划分忽略了事件时序数据在时间维度上的聚合程度。为了更好地观察谣言事件特征随时间的变化，本文提出了基于模糊聚类的事件时序数据动态

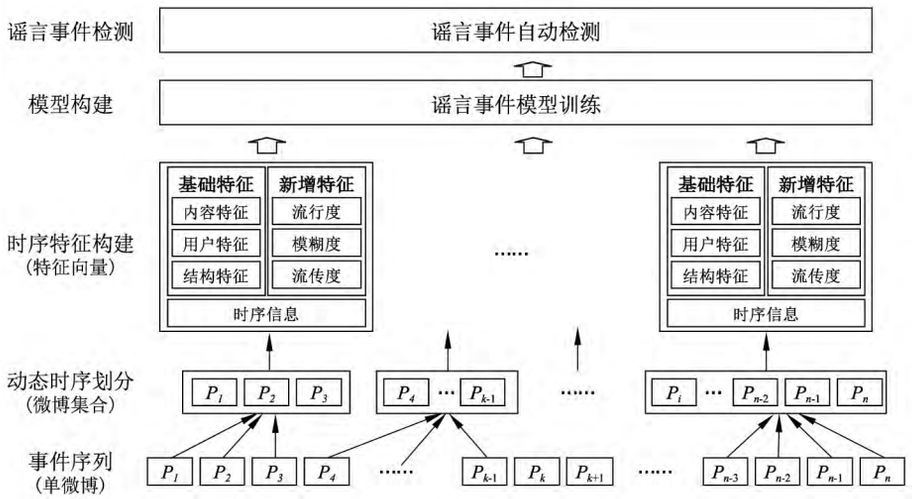

图 1　微博谣言事件自动检测流程

划分算法，并在此基础上构建了随时间变化的事件特征集合。

2.1.1　事件时序数据动态划分算法

对于事件时序数据，数据分布较密集的区域划分子区间长度应较短，而数据分布较稀疏的区域子区间长度应较长。即合理的事件时间序列划分后得到的子区间的长度应该跟数据的分布有密切关系。基于此，本文引入了模糊时间序列模型中的论域划分思想，将事件的时间跨度当作论域，提出了基于模糊聚类算法的事件时序数据动态划分算法。本文采用的模糊聚类算法是模糊 C 均值(FCM)算法，该算法是由 Bezdek[17] 在 1981 年提出，是目前应用最为广泛和成功的一种模糊聚类算法。FCM 算法将 N 个 L 维向量分为 C 个模糊组，通过迭代不断更新隶属度以及聚类中心，最小化目标函数对数据进行聚类。目标函数及约束条件如式(1)所示。

$$J(U,V)=\sum_{i=1}^{N}\sum_{c=1}^{C}(u_{ic})^m d^2(x_i,v_c)$$

$$s.t.\begin{cases}0\leqslant u_{ic}\leqslant 1,\forall i,c\\ 0\leqslant\sum_{i=1}^{N}u_{ic}\leqslant 1,\forall c\\ \sum_{c=1}^{C}u_{ic}=1,\forall i\end{cases}\tag{1}$$

其中，$m\geqslant 1$ 是模糊加权系数，$d(x_i,v_c)$ 表示第 i 个数据点与第 c 个聚类中心的距离，u_{ic} 是数据点 x_i 属于 v_c 的隶属度。

为了求含有约束条件的目标函数的极值，引入拉格朗日因子构造新的目标函数，如式(2)所示。

$$J_\lambda(U,V)=\sum_{c=1}^{C}(u_{ic})^m d^2(x_i,v_c)+\lambda\left(\sum_{c=1}^{C}u_{ic}-1\right)\tag{2}$$

对于目标函数求极值的最优化条件如下：

$$\frac{\partial J_\lambda}{\partial\lambda}=\sum_{c=1}^{C}u_{ic}-1=0$$

$$\frac{\partial J_\lambda}{\partial u_{ic}}=\sum_{c=1}^{C}m(u_{ic})^{m-1}d^2(x_i,v_c)-\lambda=0$$

$$\frac{\partial J_\lambda}{\partial v_c}=\sum_{i=1}^{N}(u_{ic})^m x_i-v_c\sum_{i=1}^{N}(u_{ic})^m=0$$

从而得到隶属度和聚类中心的计算如式(3)所示。

$$u_{ic}=\frac{1}{\sum_{k=1}^{C}\left(\frac{d(x_i,v_c)}{d(x_i,v_k)}\right)^{\frac{2}{m-1}}},\quad v_c=\frac{\sum_{i=1}^{N}(u_{ic})^m x_i}{\sum_{i=1}^{N}(u_{ic})^m}\tag{3}$$

本文中使用的 FCM 算法所涉及的参数设置如下：模糊加权系数 $m=2$，聚类中心数 $C=50$，FCM 算法停止的条件是迭代次数达到 100 次，或相邻两次迭代目标函数改进小于 $1*10^{-5}$。

根据隶属度可以获得时间序列的一个模糊分割，得不到确切的时间分割点。所以本文基于 FCM 算法所计算的聚类中心点，选取相邻两个聚类中心的中点作为本文时间跨度论域的临界点，得到区间 I_1，I_2，…，I_C，其中 C 为聚类中心个数。

本文提出的基于 FCM 的事件时序数据动态划分算法(Dynamic Time Series，DTS)描述如下：对于每一个事件 $E_i=\{(p_{ij},t_{ij})\}_{j=1}^{n}$，$p_{i,j}$ 表示事件相关的微博，$t_{i,j}$ 是其对应的时间戳。设置 FCM 的聚类中心点数为 C，即事件时间序列划分为 C 个不等长的区间。对于微博数不少于 C 的事件，采用 FCM 算法对时序数据进行不等长划分；对于微博数少于 C 的事件，本文按照单条微博进行时序数据划分，区

间数不足的事件进行补空处理，从而统一事件时序数据划分区间数。

2.1.2 事件时序特征构建

对于每一个微博事件 E_i，本文使用 D-维特征向量 F_i^D 表示。其中，特征向量包括基础特征（内容特征、用户特征和结构特征）和新增特征（流行度、模糊度和流传度），所有特征的具体定义将在 2.2 节进行详细描述。为了更好的表示谣言事件特征随时间的变化，基于 2.1.1 节提出的事件时序数据动态划分算法，所以本文的微博事件动态时序特征向量（DTS_F）定义如式（4）所示。

$$\begin{aligned} DTS\_F(E_i) &= (F_{i,1}^D, F_{i,2}^D, \cdots, F_{i,C}^D) \oplus \\ &\quad (S_{i,1}^D, S_{i,2}^D, \cdots, S_{i,C-1}^D) \\ F_{i,t}^D &= (\tilde{f}_{i,t,1}, \tilde{f}_{i,t,2}, \cdots, \tilde{f}_{i,t,D}) \\ S_{i,t}^D &= \frac{\Delta F}{\Delta t} = \frac{F_{i,t+1}^D - F_{i,t}^D}{\left|\frac{I_{t+1}}{2}\right| - \left|\frac{I_t}{2}\right|} \end{aligned} \tag{4}$$

其中，$F_{i,t}^D$ 表示事件 E_i 的第 t 个时间区间的特征集合，$S_{i,t}^D$ 表示事件 E_i 的第 $t+1$ 个时间区间与第 t 个时间区间上特征的波动程度，$\oplus$ 表示 $F_{i,t}^D$ 特征集合和 $S_{i,t}^D$ 特征集合的连接操作。$\left|\frac{I_t}{2}\right|$ 表示第 t 个时间区间的中点。$\tilde{f}_{i,t,d}$ 表示事件 E_i 的第 t 个时间区间内第 d 个特征的无量纲数值。本文采用 z-score 标准化方法将微博事件中各类特征值转化为无量纲的纯数值，便于不同量级指标能够进行比较和加权，$\tilde{f}_{i,t,d}$ 定义如式（5）所示。

$$\begin{aligned} \tilde{f}_{i,t,d} &= \frac{f_{i,t,d} - \bar{f}_{i,d}}{\sigma(f_{i,d})} \\ \bar{f}_{i,d} &= \frac{1}{C}\sum_{t=1}^{C} f_{i,t,d} \\ \sigma(f_{i,d}) &= \sqrt{\frac{1}{C-1}\sum_{t=1}^{C}(f_{i,t,d} - \bar{f}_{i,d})^2} \end{aligned} \tag{5}$$

其中，$f_{i,t,d}$ 表示事件 E_i 在第 t 个时间区间上第 d 个特征的原始值。$\bar{f}_{i,d}$ 事件 E_i 的第 d 个特征在 C 个时间段上的平均值，$\sigma(f_{i,d})$ 则是在 C 个时间段上的事件 E_i 的第 d 个特征的标准差。

2.2 特征工程

本节将重点介绍文中微博谣言事件自动检测的过程中使用的所有特征，含基础特征和新增特征，及各类特征的定义和计算方式。

2.2.1 基础特征

本文所采用的基础特征如表 1 所示，包括基于内容的特征、基于用户的特征和基于结构的特征。本文会针对微博谣言事件发展过程中划分的每一个时间区间分别使用公式（5）计算下表中的每个特征值。与之前研究不同的是，文中微博内容主题使用 LDA 模型计算了微博热点话题下的 48 个主题分布，另外，情感词的识别和情感倾向主要基于大连理工大学情感词汇本体库。

表 1 基础特征表

基于内容的特征	
Topic_Distribution	微博内容主题分布
Avg_Length	微博内容平均长度
Pos/Neg_Words	微博中正/负向情感词数量
Avg_Sentiment	微博内容平均情感得分
URL_Rate	含 URL 的微博占比
Pos/Neg_Rate	正向情感微博占比
First_Person_Rate	微博内容中第一人称占比
Hashtag_Rate	含＃的微博占比
Mention_Rate	含@的微博占比
QMark_Rate	含？的微博占比
ExclMark_Rate	含！的微博占比
Multi_Mark_Rate	含多个？或！的微博占比
基于用户的特征	
User_Desc_Rate	有简介的用户占比
Head_Img_Rate	有头像的用户占比
Verified_Rate	认证用户占比
Male/Female_Rate	男性/女性用户占比
Large_City_Rate	位于大城市的用户占比
Avg_Followers	平均粉丝数
Avg_Followees	平均关注数
Avg_Register_Time	平均注册天数
Avg_Reputation	平均声望（粉丝数/（粉丝数＋关注数））
基于结构的特征	
Avg_Reposts	平均转发数
Avg_Comments	平均评论数

2.2.2 新增特征

传统基础特征主要针对数据本身的特性，未考虑谣言事件传播的社会必要属性。美国社会学家

Allport 和 Postman 认为谣言事件得以流传的一个必要条件就是其模糊性，同时指出模糊性乘以重要性决定了谣言的流传程度。在该定义中，谣言传播是无意识主体作出的反应，对此 Crouse 在上述基础上引入人的影响因素，重新定义为“谣言的流通量＝事件的重要性×事件的模糊性/公众批判能力”。为了对微博谣言事件进行区分，本文提出了事件流行度、模糊度、流传度三个新的特征对微博事件进行表示。对于微博事件 E_i，有 C 个时间分割，即 C 个事件发展阶段，那么这三个特征在各阶段的定义和数学表示如下：

事件流行度（Posts Popularity，PPop）：是指微博事件发展过程中各阶段的重要程度。本文采用各时间段内用户对微博内容的转发、评论和点赞数来计算各阶段事件的流行程度，如式(6)所示。

$$\mathrm{PPop}_{i,t}=\begin{cases}0, & |P_t|=0\\ \dfrac{1}{|P_{i,t}|}\displaystyle\sum_{p=1}^{|P_{i,t}|}(r_{i,p}+c_{i,p}+l_{i,p}), & |P_t|\neq 0\end{cases} \tag{6}$$

其中，$P_{i,t}$ 表示第 i 个事件中第 t 个时间段的微博集合，$|P_{i,t}|$ 则是指该集合中微博的总数，$r_{i,p}$，$c_{i,p}$，$l_{i,p}$ 分别表示该集合中第 p 条微博的转发数、评论数和点赞数。

事件模糊度（Posts Ambiguity，PAmb）：是指微博事件发展过程中各阶段的模糊程度。对于每个时间段，本文使用当前时间段内微博内容与前置时间段内微博内容的不相似程度来表示该时间段的模糊程度，并采用 tf-idf 计算内容关键词对微博内容进行表示。同时，使用 Jaccard 距离计算各时间段内事件的模糊程度，如式(7)所示。

$$\mathrm{PAmb}_{i,t}=1-\frac{TW_{i,t}\cap\left(\bigcup_{m=0}^{t-1}TW_{i,m}\right)}{\bigcup_{m=0}^{t}TW_{i,m}} \tag{7}$$

其中，$TW_{i,t}$ 表示第 i 个事件中第 t 个时间段的内容关键词集合，$\bigcup_{m=0}^{t-1}TW_{i,m}$ 表示第 i 个事件中第 t 个时间段所有前置时间段的内容关键词集合。

事件流传度（Posts Spread，PSpr）：是指微博事件发展过程中各阶段的流传程度。文献[18]中指出公众批判能力从本质上看体现的是公众的态度，因此本文使用表 1 中的“微博内容平均情感得分”来计算公众的批判能力。根据 Crouse 的谣言传播公式，则本文的事件流传度＝事件流行度×事件模糊度/事件情感度，如式(8)所示。

$$\mathrm{PSpr}_{i,t}=\begin{cases}\dfrac{\mathrm{PPop}_{i,t}\times\mathrm{PAmb}_{i,t}}{min\{|\mathrm{Sentiment}_{i,t}|\}_{t=0}^{C}}, & \mathrm{Sentiment}_{i,t}=0\\ \dfrac{\mathrm{PPop}_{i,t}\times\mathrm{PAmb}_{i,t}}{|\mathrm{Sentiment}_{i,t}|}, & \mathrm{Sentiment}_{i,t}\neq 0\end{cases} \tag{8}$$

3　实验与分析

3.1　实验数据

为方便实验对比，本文采用文献[16]中公开的微博谣言事件数据集。该数据集主要来自新浪微博社区管理中心的不实信息，共包含 2 313 个谣言事件和 2 351 个非谣言事件，其中 1 表示谣言事件(R)，0 表示非谣言事件(NR)。这些数据都是通过微博开放 API 从微博社区管理中心获取。数据集的详细统计信息如表 2 所示。

表 2　数据集详细统计信息

数据集统计指标	新浪微博数据集
用户数	2 819 338
微博总数	3 752 459
微博事件总数	4 664
谣言事件总数	2 313
非谣言事件总数	2 351
事件平均微博数	804
事件最小微博数	10
事件最大微博数	59 318
事件平均时间跨度	1 808.74 Hours
事件最小时间跨度	0.02 Hours
事件最大时间跨度	34 312.00 Hours

3.2　实验对比

为保证实验的公平性，所有模型使用相同的训练集和测试集，并针对谣言(R)和非谣言(NR)两个类别分别使用准确率(Acc)、精准率(P)、召回率(R)和 $F1$ 值来评价模型的性能。

3.2.1　参数选择

根据文献[2,5,13-14]等的实验发现，在谣言事件检测领域，SVM 模型略优于决策树、随机森林等

其他分类模型。故本文选择SVM作为基础模型。本文首先对SVM模型的核函数和参数的选择进行了实验讨论。表3是分别使用四种核函数(默认参数和所有特征下)实验结果,可以看出RBF核函数更加适合本文的分类任务。

表3 四种核函数训练结果对比

Kernel	Class	Acc	*P*	*R*	*F*1
Linear	R	0.787	0.795	0.770	0.782
	NR		0.781	0.804	0.792
Poly.	R	0.809	0.820	0.787	0.803
	NR		0.799	0.830	0.814
RBF	R	**0.837**	**0.854**	**0.810**	**0.831**
	NR		**0.822**	**0.864**	**0.842**
Sigmod	R	0.836	**0.854**	0.808	0.830
	NR		0.821	**0.864**	**0.842**

SVM模型的核函数确定之后,我们需要确定误差代价参数cost以及针对RBF核函数的γ参数。首先,固定核函数参数$\gamma=\frac{1}{\text{特征数}}$(默认值),通过不断减小搜索步长,对误差代价参数cost进行实验选择,结果如图2所示,最终确定cost=2.1。

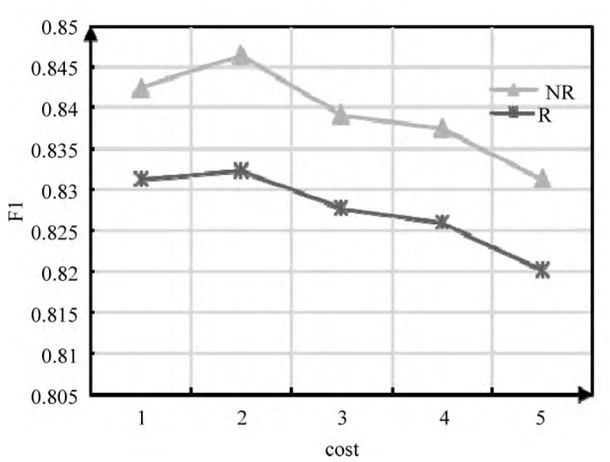

图2 参数cost选择

同样的方法可获得$\gamma=0.000\ 35$,如图3所示。

3.2.2 实验结果与分析

本文共使用6种模型进行微博谣言事件的检测,包括DT-Rank[19]、LK-RBF[20]、SVM-TS[14]、GRU-2[16]以及本文提出的SVM-DTS(共包括两个模型,SVM^{DTS}_{com}:基础特征;SVM^{DTS}_{all}:基础特征+新增特征)。实验结果比较如表4所示。

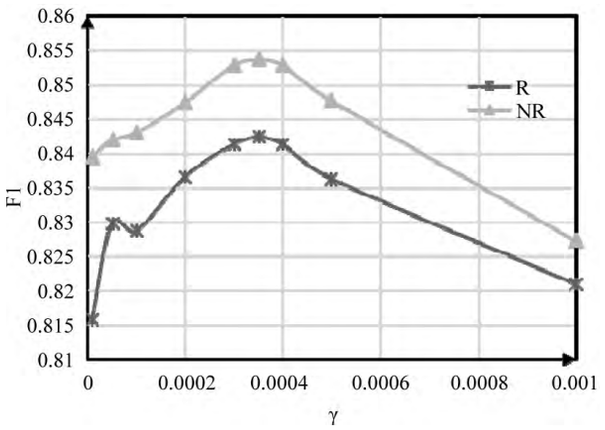

图3 参数γ选择

表4 模型试验结果对比

Method	Class	Acc	*P*	*R*	*F*1
DT-Rank	R	0.648	0.683	0.624	0.652
	NR		0.615	0.674	0.643
LK-RBF	R	0.681	0.768	0.629	0.692
	NR		0.604	0.749	0.669
SVM-TS	R	0.796	0.813	0.763	0.788
	NR		0.780	0.828	0.804
GRU-2	R	0.833	0.829	**0.842**	0.835
	NR		**0.837**	0.825	0.830
SVM^{DTS}_{com}	R	0.832	0.851	0.803	0.826
	NR		0.816	0.862	0.839
SVM^{DTS}_{all}	R	**0.849**	**0.868**	0.820	**0.843**
	NR		0.832	**0.878**	**0.854**

从表4可以看出,决策树模型DT-Rank的表现相对于其他模型来说,效果最不理想。这是由于DT-Rank模型是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识别,而这些正则表达式在本文使用的新浪微博谣言事件数据集中仅能匹配到1.63%的微博数据。相对来说,基于SVM模型的LK-RBF和SVM-TS在谣言事件识别上表现良好。尤其是SVM-TS模型,相对DT-Rank模型的准确率提高了14.8%左右,*F*1值提高了13.6%~16.1%。一方面是由于SVM模型本身良好的泛化能力,更能适应微博内容的多样性,更重要的是由于SVM-TS模型中,考虑了谣言事件特征随时间变化的特性,因此,检测效果得到了大幅提升。另外,GRU-2模型是基于RNN的深度神经网络模型。该模型通过谣言事件中,所有词之间的关系自动构

建特征，更好地捕捉了谣言事件内容的高层次特征。相对传统的机器学习模型的准确率和 $F1$ 值等都得到了有效提升。

本文提出的 SVM_{com}^{DTS} 模型考虑了事件时序数据在时间维度上的聚合程度，使用模糊聚类对时序数据动态划分，并在此基础上构建了随时间变化的事件特征集合。相对于 SVM-TS 模型在准确率(3.6%)、精准率(3.6%～3.8%)、召回率(3.4%～4.0%)和 $F1$ 值(3.5%～3.8%)上都有明显提升。一方面说明了事件特征的时间波动性对谣言事件检测的重要性，另一方面也验证了本文提出的动态时序特征构建模型可以更好地捕捉和表征谣言事件特征随时间的变化特性。SVM_{all}^{DTS} 模型是在 SVM_{com}^{DTS} 模型的基础上，考虑社会学的谣言传播原理加入了流行度、模糊度和流传度三项新特征，在准确率(1.7%)、精准率(1.6%～1.7%)、召回率(1.6%～1.7%)和 $F1$ 值(1.5%～1.7%)上都有一定的提升，验证了本文提出的新特征对谣言事件检测的有效性。此外，本文提出 SVM-DTS 模型相对于深度神经网络模型 GRU-2 在准确率和 $F1$ 值分别提高了 1.6%和 0.8%～2.4%左右，在精准率上，R 类别上提高了 3.9%，NR 类别有所下降(约 0.5%)。同样，在召回率上，R 类别上下降了 2.2%，NR 类别上上升了 5.3%。但总体来说，本文的 SVM-DTS 模型在微博谣言事件检测方面效果优于深度神经网络 GRU-2 模型。一方面是由于 GRU-2 模型仅考虑了微博事件内容特征，忽略了微博数据特有的结构特征；另一方面是因为 GRU-2 模型中使用等时划分方法对事件时序数据进行分割，未考虑事件时间维度上的分布特征。综上所述，本文提出的 SVM-DTS 模型在微博谣言事件检测方面相对于其他模型表现较好。

3.2.3　新特征影响

对本文提出的三个新特征，分别使用不同的特征组合(基础特征＋单项新特)研究了每项新特征对模型识别效果的影响。

实验结果如图 4 所示，横坐标为基础特征和各项新特的组合(即：PPop、PAmb 和 PSpr)，纵坐标为谣言事件检测的准确率 Acc。总体来看，在传统基础特征基础上，本文提出的各项新特征对谣言事件检测结果都有所提升，准确率上升约 1.0%～1.4%，进一步说明了本文提出的三个新特征对于谣言事件检测的有效性。其中，PAmb 特征提升效果最为显著。这也说明在事件传播过程中，事件的模糊程度极大地影响了人们对于事件真实性的判断，符合人们的一般认知规律。

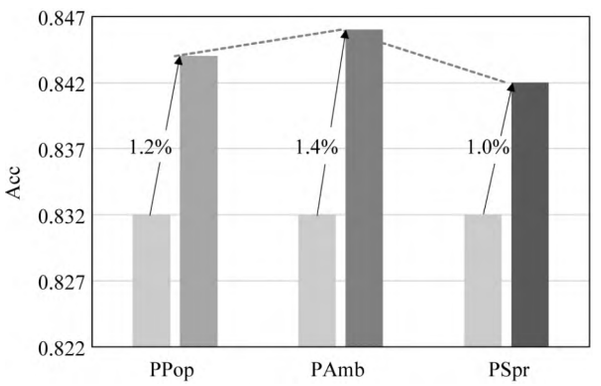

图 4　单项新特征模型影响对比

4　总结

本文提出的 SVM-DTS 谣言事件自动检测模型，不仅考虑了谣言事件特征随时间变化的特性，而且综合了时间维度上谣言事件的分布特征，有效提高了抽取出的特征对谣言事件的表征能力；另外，基于社会学中谣言的传播原理，定义了事件流行度、模糊度和流传度三项谣言事件检测的新特征。实验结果表明，本文提出的模型使谣言事件检测效果得到了可观的提升。

在未来的工作中，一方面，我们将寻找更加符合微博谣言事件传播模式的计算方法和新增特征的表示方法，同时深入考察和分析文中提到的基础特征和新增特征对谣言事件检测效果的影响，从而，选择最佳的特征组合；另外一方面，我们将考虑事件传播学原理，构建更符合事件发展传播的时序特征表示模型。同时，我们也将考虑使用深度神经网络模型来解决人工特征构建复杂和特征语义性不强等问题。

参考文献

[1] 张蕾，郭晓桐. 谣言、信任与群体性事件——基于谣“盐”和抢盐风波的调查研究[J]. 国际新闻界，2012(7)：12-18.

[2] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on Sina Weibo[C]//Proceedings of ACM, 2012：1-7.

[3] Liu X, Nourbakhsh A, Li Q, et al. Real-time rumor debunking on Twitter[J]. 2015：1867-1870.

[4] 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.
[5] 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.
[6] Wu K，Yang S，Zhu K Q. False rumors detection on Sina Weibo by propagation structures[C]// Proceedings of the 31st International Conference on Data Engineering，IEEE，2015：651-662.
[7] Ma J，Gao W，Wong K F. Detect rumors in Microblog posts using propagation structure via kernel learning [C]// Proceedings of ACL2017. 2017.
[8] 毛二松，陈刚，刘欣，等. 基于深层特征和集成分类器的微博谣言检测研究[J]. 计算机应用研究，2016，33(11)：3369-3373.
[9] 祖坤琳，赵铭伟，郭凯，等. 新浪微博谣言检测研究[J]. 中文信息学报，2017，31(3)：198-204.
[10] 杨文太，梁刚，谢凯，等. 基于突发话题和领域专家的微博谣言检测方法[J]. 计算机应用，2017，37(10)：2799-2805.
[11] 武庆圆，何凌南. 基于多标签双词主题模型的短文本谣言分析研究[J]. 情报杂志，2017，36(3)：92-97.
[12] 胡钰. 大众传播效果[M]. 北京：新华出版社，2000.
[13] Kwon S，Cha M，Jung K，et al. Prominent features of rumor propagation in online social media[C]// Proceedings of the 13th International Conference on Data Mining，IEEE. 2014：1103-1108.
[14] 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. ACM，2015：1751-1754.
[15] Liang G，Yang J，Xu C. Automatic rumors identification on Sina Weibo[C]// Proceedings of the 12th International Conference on Natural Computation，Fuzzy Systems and Knowledge Discovery，IEEE. 2016：1523-1531.
[16] Ma J，Gao W，Mitra P，et al. Detecting rumors from Microblogs with recurrent neural networks[C]// Proceedings of IJCAI2016. 2016.
[17] Bezdek，James C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York：Plenum Press，1981.
[18] 王倩，于风. 奥尔波特和波斯特曼谣言传播公式的改进及其验证：基于东北虎致游客伤亡事件的新浪微博谣言分析[J]. 国际新闻界，2017，39(11)：47-67.
[19] Zhe Zhao，Paul Resnick，Qiaozhu Mei. 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.
[20] Justin Sampson，Fred Morstatter，Liang Wu，et al. Leveraging the implicit structure within social media for emergent rumor detection[C]// Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. 2016：2377-2382.

王志宏(1990－)，博士研究生，主要研究领域为文本挖掘、机器学习。
E-mail：zhwang817@foxmail.com

过弋(1975－)，通信作者，博士，教授，主要研究领域为文本挖掘、知识发现、商业智能分析。
E-mail：guoyi@ecust.edu.cn

(上接第123页)

张茜(1993－)，硕士，主要研究领域为自然语言处理、机器学习、情感计算等。
E-mail：1048604035@qq.com

张士兵(1962－)，博士，教授，主要研究领域为无线通信、OFDM系统、认知无线电等。
E-mail：zhangshbntu@163.com

任福继(1959－)，通信作者，博士，教授，主要研究领域为自然语言处理，人工智能，语言理解与交流，情感计算等。
E-mail：ren@is.tokushima-u.ac.jp