

毕业设计说明书

|  |  |  |  |
| --- | --- | --- | --- |
| **作 者:** | 李庆贺 | **学 号：** | 915106840425 |
| **学 院:** | 计算机科学与工程学院 | | |
| **专业(方向):** | 智能科学与技术 | | |
| **题 目:** | 基于机器学习的微博突发事件分析与 | | |
|  | 谣言检测 | | |

**指导者：** 夏 睿 教授

(姓 名) (专业技术职务)

(姓 名) (专业技术职务)

**评阅者：** \*\*\* 副教授

(姓 名) (专业技术职务)

2019 年 5 月

声 明

我声明，本毕业设计说明书及其研究工作和所取得的成果是本人在导师的指导下独立完成的。研究过程中利用的所有资料均已在参考文献中列出，其他人员或机构对本毕业设计工作做出的贡献也已在致谢部分说明。

本毕业设计说明书不涉及任何秘密，南京理工大学有权保存其电子和纸质文档，可以借阅或网上公布其部分或全部内容，可以向有关部门或机构送交并授权保存、借阅或网上公布其部分或全部内容。

学生签名：

年 月 日

指导教师签名：

年 月 日

毕业设计说明书中文摘要

|  |
| --- |
| 随着微博平台的流行，微博平台上产生了大量难于验证的谣言信息，并在用户之间快速传播，对平台的发展和社会稳定带来巨大危害。如何自动检测并识别谣言对于制止谣言传播、减小谣言危害具有极其重要的意义。基于以上背景，开展了谣言检测的相关研究。  本文对来源于新浪微博的谣言数据集的4664条标注数据进行研究，利用谣言数据中的文本信息以及时间信息进行建模，构建出门控循环单元网络(GRU)模型实现了基于序列模型的微博谣言检测算法，在测试集上达到了89.36%的准确率。在此基础上，利用卷积神经网络(CNN)来提取文本的深层特征，构建出CNN-GRU复合网络模型并获得了测试集上92.43%的准确率。基于上述实验结果，本文进一步对两种谣言检测模型进行早期谣言检测实验，取得了不错的效果。  关键词 微博谣言 早期检测 机器学习 深度神经网络 |

毕业设计说明书外文摘要

|  |
| --- |
| **Title**  Microblog Burst Event Analysis and Rumors  Detection Based on Machine Learning  **Abstract**  With the popularity of the Weibo platform, a large number of rumor messages that are difficult to verify are generated on the Weibo, and spread rapidly among users, which brings great harm to the development of the Weibo platform and social stability. How to automatically detect and identify rumors is of great significance for stopping the spread of rumors and reducing the harm of rumors. Based on the above background, relevant research on rumor detection was carried out.  This paper studies the 4664 annotation data from the rumor dataset of Sina Weibo, and uses the text information and time information in the rumor data to model and construct a Gated Recurrent Unit(GRU) model to detect rumors. This rumor detection algorithm achieved 89.36% accuracy on the test set. On this basis, the Convolutional Neural Network(CNN) is used to extract the deep features of the text, and the CNN-GRU composite network model is constructed and the accuracy of 92.43% on the test set is obtained. Based on the above experimental results, this paper further conducts early rumor detection experiments on two rumor detection models, and has achieved great effects.  Keywords Rumors Early Detection Machine Learning Deep Neural Network |

目 次

1 绪论…………………………………………………………………………………….……..1

1.1 研究背景及意义……………………………………………………………….…....... 1

1.2 相关技术的现状…………………………………………………………………….... 2

1.3 总体技术方案及其社会影响………………………………………………………… 4

1.4 论文章节安排………………………………………………………………………… 4

2 问题描述及方法基础………………………………………………………………..……… 6

2.1 问题描述……………………………………………………………………………… 6

2.2 卷积神经网络........…………………………………………………………………… 6

2.3 门控循环单元………………………………………………………………………… 7

3 微博谣言检测模型…………………………………………………………………..…...… 10

3.1 数据预处理……………………………………………………………………………10

3.2 基于门控循环单元的网络模型…………………………………………………...… 12

3.3 基于卷积-门控循环单元的网络模型………..………………………………………14

3.4 目标函数及优化方法…………………………………………………………………15

4 模型实验与评估..…………………………………………………………………....………16

4.1 数据集…………………………………………………………………………………16

4.2 早期谣言检测…………………………………………………………………………18

4.3 评估方法及实验参数设置……………………………………………………………18

4.4 实验结果分析与改进………………………………………………….……………...19

结论..…………………………………….………………………………………………...…..... 24

致谢.………………………………………………………….……………………………….….25

参考文献……………………………………………………………………………………....…26

# 1 绪论

本章首先对谣言检测的背景进行介绍，并阐述了研究谣言检测技术的重要意义，然后对于谣言检测技术的发展历程以及国内外谣言检测相关的研究情况进行了介绍。最后介绍本文的研究内容以及章节结构安排。

## 1.1 研究背景及意义

近几十年来，互联网的发展已经掀起了一个科技浪潮，同时，我国互联网在用户规模及普及率、互联网基础设施和应用领域等方面也保持快速发展，从规模与用户数来看，中国互联网已成为国际互联网中不可忽略的部分。从中国互联网络信息中心(CNNIC)最新发布的第43次《中国互联网络发展状况统计报告》的数据可以看出，截止2018年12月，我国互联网用户规模已达到8.29亿，全年新增互联网用户5653万，互联网普及率达59.6%；其中移动互联网用户规模为8.17亿，全年新增6433亿；移动互联网用户比例由2017年底的97.5%提升至2018年底的98.6%[1]。

互联网的快速发展，用户规模的扩大和网络应用普及率的提高，使得社会媒体获得迅速发展，信息呈现出前所未有的多样性，人们获取信息、意见表达的方式得到了空前拓展。以微博为代表的社会化网络新媒体正在成为人们生活中获取最新信息、表达观点的新平台。随着互联网技术的应用，网络新媒体的功能更加先进和多元，内容更加自由和开放，同时信息发布费用更加低廉、使用更加便捷；人们获取信息的方式也由线下传统方式转变为线上网络获取，为网络媒体带来了海量用户。

随着微博等平台的流行，微博上产生了大量难于验证的谣言信息，给用户体验、平台发展和国家稳定带来巨大危害[2]。微博谣言的危害性主要表现在其突发性与不确定性。首先微博谣言常伴随着突发事件产生，在某个事件发生后，与之相关的谣言也随之产生并进行传播。其次，人们希望在突发事件发生后第一时间获取到相关信息，而事件的突发性导致官方新闻媒体无法短时间内公布相关信息，这从一定程度上促进了谣言的产生，也为事件相关的信息增加了不确定性。这些谣言泛滥的危害显而易见，它们不仅能影响个人的判断，误导社会舆论导向，更严重者会引起公众的恐慌和社会的动荡[3]。近年来，国内外先后爆发了由谣言传播引发的群体性突发事件，严重扰乱了人们的生活秩序，影响了社会的稳定，造成的损失超出了社会资源所能承受的能力[4]。例如2015年８月一则关于“韦拉克鲁斯附近出现吸毒团伙的枪击和孩童绑架事件”的谣言在美国社交网络中传播，其引发的骚动导致部分民众将车辆留在马路中间，冲向学校去接自己的孩子，从而造成了26起车祸[5]。又如2018年10月28日，重庆万州公交车坠江事件相关微博引起了人们的关注，事件发生后由于官方权威媒体没有及时发布调查结果通报，导致一时间网络上出现了许多有关公交车司机不规范驾驶导致事故的相关谣言，而处在消息不对称一端的用户在网络谣言的误导下纷纷针对公交车司机进行舆论攻击，直到权威媒体公布事故调查真相时，才将谣言止住。这给同为受害者的公交车司机及其家属带来了莫大的伤害。

据新浪微博官方发布的《2018，2018年，信息7.48万条，辟除新增谣言352例，标记不实信息1811条，全年单条不实信息从举报到处理平均用时16.97小时，每条谣言澄清用时较2017年减少近1个小时[6]。由此可见，未知谣言鉴别己成为社交网络内容安全中不可忽视的课题，谣言传播具有突发性且传播速度极快，对正常的社会秩序造成不良影响，公众的从众心理又在客观上促进了谣言的传播，因此在传播早期揭露谣言对于减少其有危害性影响尤为重要。为了区分谣言和事实，个人和组织往往依靠经验常识和调查性新闻，著名的谣言报道网站snopes.com、fatcheck.org等发挥了一定的作用，但是在这些途径都涉及到人工验证步骤，因此调查性新闻难以全面准确，并且存在长时间的判别延迟[5]。因此考虑如何高效率鉴别谣言信息、如何测量、量化相关传播信息的可信度，已成为目前学术界、政府职能机构等相关部门共同关注的重要问题，对构建和谐的社会环境具有重要意义。

## 1.2 相关技术的现状

随着微博等网络社交平台的快速发展，关于微博谣言检测的研究工作越来越多。在研究早期，国外的一些学者主要基于机器学习的经典思路，将其作为分类问题进行研究。Castillo等[7]Twitter，在Twitter上与信息、用户特征和传播特征，采用J48，实现了一定的效果。Qazvinian 等[8]分析了Twitter，斯分类器和。Mendoza等[9]分析了突发时间下Twitter用户的行为特征，，提出了微博转发形成的。Takahashi等[10]针对相关事件在Twitter上，和的谣言检测系统。相对于国外研究，国内针对谣言检测的相关研究起步较晚，早期研究较少。其中，程亮等[11]，利用BP神。许晓东等[12]，现微博谣言传播。

上述方法在一定程度上解决了问题，但是仍然具有较大缺点，就是其未能较好的利用微博本身具有的文本信息特征。Yang等[13]标注了微博数据，然后对标注数据利用支持向量机的方法建模抽取特征，但其未能挖掘出文本信息的深层特征，导致模型的识别率并不高，因此贺刚等[14]在其研究基础上，引入识别谣言的符号特征、链接特征、关键词分布特征和时间差等新特征来预测谣言，获得了改善的效果。

近年来，随着自然语言处理以及深度学习的快速发展，国内外对于谣言检测的相关研究逐渐谣言事件本身转向更细粒度的文本属性特征等方面，模型方法也开始采用能够获取更高层次特征水平的深度网络学习模型。

Friggeri等[15]和Hannak等[16]利用谣言传播。Kwon等[17]根据推文的数量推出了。Ma等[18]在模型中加入了时间序列的社会背景特征。这些方法通常需要进行大量的数据预处理和特征工程。Zhao等[19]通过使用一组正则表达式来挖掘用户对论文的态度。Ma等[20]提出使用循环神经网络模型来自动学习相关推文帖子上下文信息随时间的变化；最近他们[21]又提出了两种基于自下而上和自上而下的用于谣言的表示学习和分类，通过整合结构特征和语义特征来生成更好的表示；同时Ma等[22]又首先提出和立场分类结合在一起的工作，该方法成功地学习为两个核心任务共同表示和分类数据。

此外，Wu等[23]利用一些，提出了一种混合SVM分类器，它，以捕获新浪微博上的传闻和传播模式。Ma等[24]使用，以便在Twitter上识别不同类型的谣言。最近，He等[25]认为，并首次尝试探索用于，通过使用时间平滑框架将任意神经网络扩展到进化学习的简单方法，提出了一个历时传播框架，通过历时连接将历史影响纳入当前学习的特征。Yavary等[26]将用户反馈视为谣言检测的重要特征，提出了于用户会话树的模式和基于子树的模式，并将提取的模式用作极限学习机的特征，构建谣言检测模型。

值得注意的是，Guo等[27]将谣言事件建模为包含不同语义信息水平的层次时间序列，提出了一种包含社会注意力机制的谣言检测的新型层次网络(HSA-BLSTM)；它将一个事件分为几个包含几个帖子的子事件。每个帖子进一步分为几个单词。然后将结构化事件输入分层Bi-LSTM网络，社会特征被用作识别谣言的突出部分的另一条线索并通过Bi-LSTM中的注意机制实现这些特征，获得谣言检测的准确表示。该模型不仅考虑了源微博及其评论转发的特征信息，也将他们的层次结构信息、社会特征等利用注意力机制来实现其表示。

## 1.3 总体技术方案及其社会影响

本课题主要研究基于机器学习方法的微博谣言检测问题，本文主要利用了近年来发展迅速的深度神经网络模型进行实现。首先利用微博谣言数据集中的文本信息与时间信息进行建模分析，并依据谣言事件的发布与转发时间对文本信息进行划分构建模型的输入序列，并将预处理好的数据送进构建的门控循环单元网络(GRU)进行处理从而实现谣言分类检测任务。在此基础上，本文通过对谣言数据的转发数量特征和时间分布特征进行研究分析，发现了谣言事件与非谣言事件在传播过程中的差异，从而在序列划分方式上进行改进，对源微博文本与转发评论信息分别利用卷积神经网络(CNN)提取文本深层特征，并将提取的深层特征送入门控循环单元网络实现了准确率更高、性能更好的微博谣言检测模型。

此外，本文分别对两种模型进行早期谣言检测实验，分别从时间早期和传播早期测试两种模型的早期检测性能。

随着互联网革命的展开，相关的社交平台，如微博、推特等不断涌现并开始改变人们的生活。媒介不仅便利了人们的生活，也使得人们交流的范式从过去的电话、手机变为了今天的微博平台上的留言、交流。然而在微博平台上，一个事件的发生常常伴随着谣言的产生，使得人们无法分辨信息的正确性。而微博谣言借助微博平台的便捷性进行广泛传播，对网络舆论、社会秩序甚至国家稳定造成有害的影响。针对微博谣言的检测识别问题，现在主要依靠用户举报、人工审查方式来进行识别信息是否为谣言，这种做法需要依靠大量的人力资源，而且识别具有滞后性，无法在谣言传播的早期进行识别，因此不能很好的控制谣言信息造成的危害。本课题提出的微博谣言检测模型不仅很好地解决了自动识别微博信息是否是谣言的问题，更在早期检测方面取得进展，使得在谣言产生并进行传播的早期对信息进行甄别，帮助人们及早发现可能造成不良影响的谣言信息，及时采取措施制止其传播，从而降低其危害性。

## 1.4 论文章节安排

1. 进行了介绍，并对本论文提出的模型方法进行了简单介绍。
2. ，并详细介绍构建微博谣言检测模型使用的卷积神经网络和门控循环单元的原理。
3. 详细介绍了基于门控循环单元构建的谣言检测模型的结构及优化方法。
4. 详细介绍了基于卷积神经网络和门控循环单元构建的复合深度神经网络模型的结构及优化方法。
5. 首先介绍研究所使用的微博数据集的结构和规模。然后介绍了本文进行早期谣言检测的实现原理。接着对本文所实现的谣言检测模型性能的评估方法进行了说明。最后对实验结果进行了分析，并提出了本文实验中存在的一些不足以及未来模型改进的方法。

# 2 问题描述及方法基础

本章主要对课题研究所涉及的机器学习、自然语言处理的原理和方法进行介绍，主要分为四部分，第一部分是将本课谣言检测任务的符号化描述；第二部分是微博数据的预处理，包括语言模型、文本分词等技术；第三部分与第四部分分别是本文搭建的微博谣言检测模型所使用的机器学习方法——卷积神经网络和门控循环单元网络。

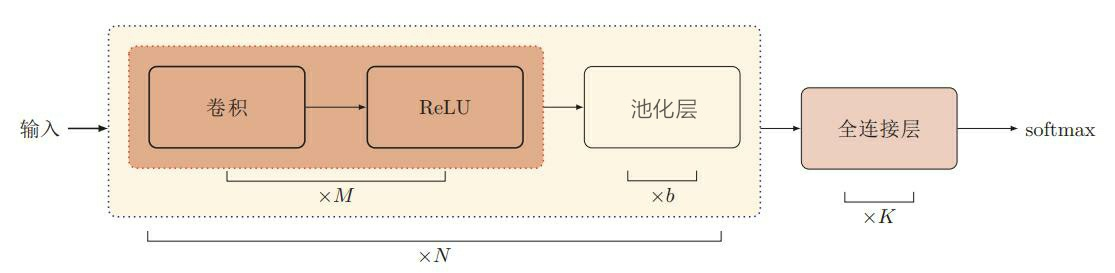
## 2.1 问题描述

社交媒体平台上的微博通常是有一定字数限制的短文本信息，因此源微博包含有限的谣言文本特征。因此，本文参照Ma等人在[20]中对数据的处理方式，利用源微博的转发、评论等序列信息进行建模。

假设有一微博数据集M = {*M*}，每个微博事件都有一系列与之相关的转发和评论，记为*M* = {()}，其中表示一条转发或评论的文本信息； 表示该转发或评论对应的时间戳。因此谣言检测的目标是，对于每条微博*M*，根据其转发序列 {()} 来预测其类别标签 。在本文中，规定 为谣言信息， 为非谣言信息。

## 2.2 卷积神经网络

卷积神经网络(Convolutional Neural Network，CNN或ConvNet)是、权。起初卷积神经网络主要应用在图像处理、计算机视觉等领域，目前已经成功应用于自然语言处理(NLP)的诸多任务中，如句子语义分析[29]、文本分类[30]和强化学习任务[31]等。卷积神经网络架构能，从而改进模型性能。一个典型的卷积网络是由卷积层、网络，。一个卷积块为连续*M*个卷积层和*b*个池化层，一个卷积网络中可以堆叠*N*个连续的卷积块，然后再接着*K*个全连接层。目前常用的卷积神经网络结构如图2.1所示。



**图2.1 典型的卷积神经网络结构**

利用卷积神经网络，可以获得划分后的转发序列*F* = 的特征向量。首先，用词嵌入 来表示每个单词 ，其中 为词嵌入的维度。然后每个长度为*n* 的转发序列由以下计算：

(2.1)

卷积层是通过加权矩阵 以逐行的形式进行运算，其中 *h* 是卷积运算中窗口的长度，而卷积，则是一组固定的权重和不同窗口内数据做内积。之后卷积的结果应用 *ReLU* 函数激活获得元素的特征映射 ：

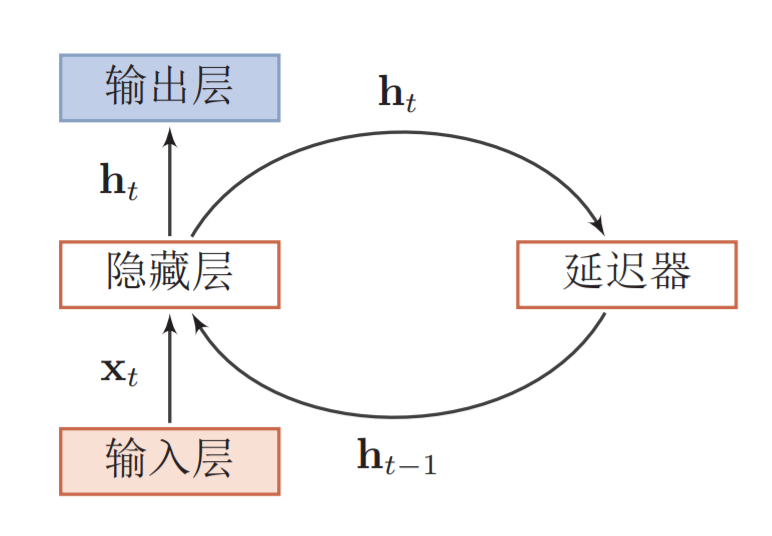
(2.2)

其中，表示从第*i*个词到第 (*i+h-1*) 个词的词嵌入，*b*为偏置项， 表示卷积运算。之后卷积层的输出将送入池化层进行最大池化操作，并将最大值 作为对应滤波器的提取特征。

CNN模型使用多个滤波器(窗口大小不同、权重不同的矩阵)来获取多个特征。假设经CNN处理获得的特征矩阵长度为E，那么对于每个转发序列，可以得到可变长度的矩阵 。通常可以重复卷积层与池化层来获得更深层次的交互特征。这些特征可以连接到全连接层，或者作为输入数据送入到循环神经网络(RNN)中进行下一步的处理。

## 2.3 门控循环单元

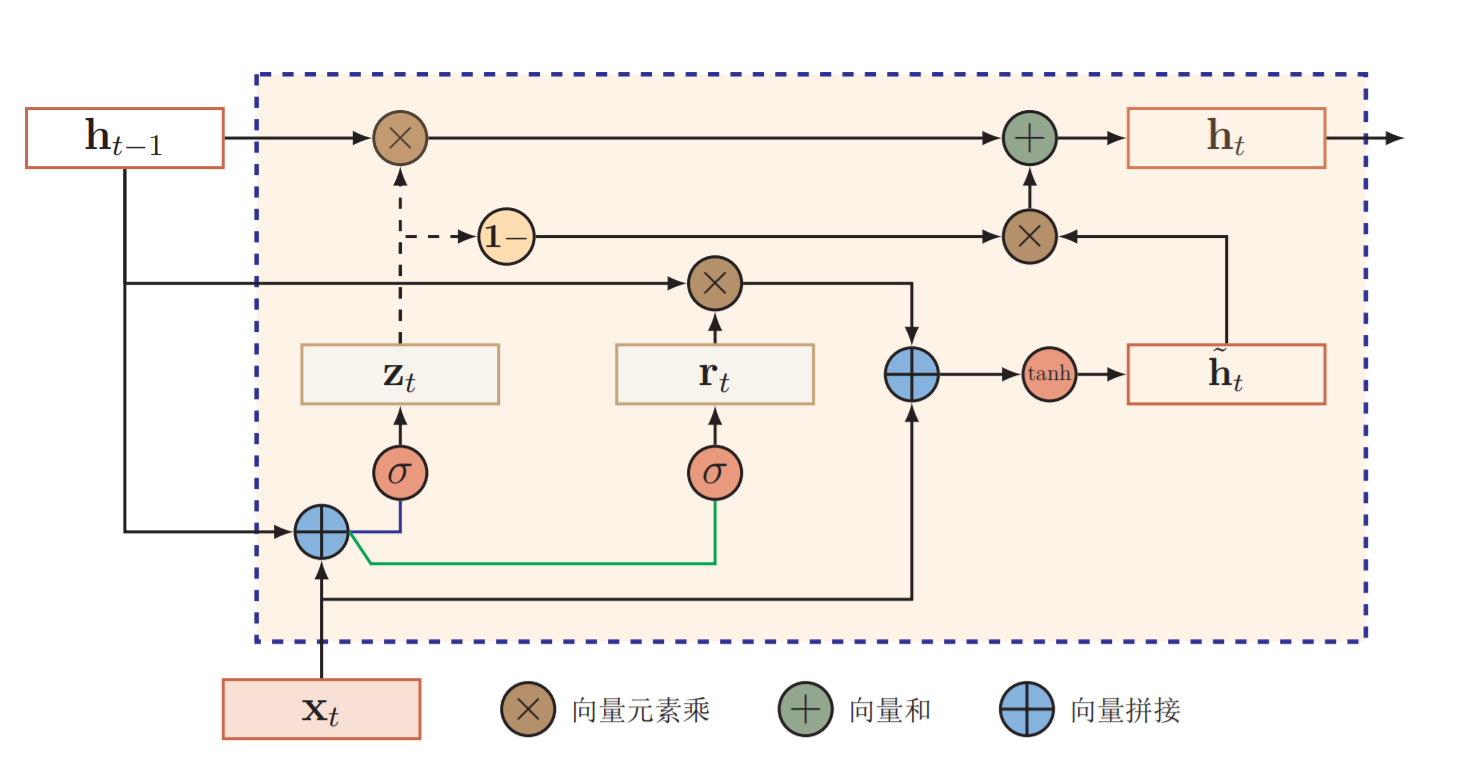
循环神经网络(Recurrent Neural Network，RNN)。在循环神经网络中，，也可以接受相邻单元传递的隐层信息，，如图2.2所示。和前馈神经网络相比，循环神经网络根据其内部的机制，针对任意时序的输入序列进行处理，因此已经被广泛应用在语音识别、语言模型以及自然语言生成等任务上。循环神经网络的来学习。



**图2.2 循环神经网络结构**

虽然从理论上说，RNN，但是当输入序列比较长时，，实际上只能够学习到关系，也称为长期依赖问题。为了解决这个问题，人们对循环神经网络进行了很多的改进，其中最有效的改进方式引入门控机制。

门控循环单元(Gated Recurrent Unit，GRU)网络，是一种更新方式的循环神经网络。基于LSTM，GRU是其经过实践检验后证明得出的一种效果很好的变体，它较LSTM网络的结构更加简单，且针对RNN网络中的长依赖难以得到的问题具有较好的效果。在LSTM网络中，输入门和遗忘门是互补关系，因此相关的发明者由此入手，对于这较为冗余的两个门(输入门与遗忘门)，GRU将其直接合并成一个门：更新门。同时，GRU也不引入额外的记忆单元，直接在当前状态前一状态之间引入线性依赖关系，其结构如图2.3所示。



**图2.3 GRU循环单元结构**

在GRU网络中，当前时刻的候选状态

(2.3)

其中 为重置门，用来控制候选状态的计算是否依赖上一时刻的状态。

(2.4)

当时，候选状态只和当前输入相关，和历史状态无关。当时，候选状态和当前输入和历史状态相关，与简单循环网络一致。

GRU网络的隐层状态的更新方式为

(2.5)

其中为更新门，，以及需要。

(2.6)

最后，该单元的输出为

(2.7)

上述各式中，分别为权重参数、隐层参数和偏置项，均可以通过网络进行学习。

3 微博谣言检测模型

本章将从数据预处理、模型设计原理、模型结构以及优化目标及方法等方面详细介绍本文构建的基于门控循环单元的微博谣言检测模型，以及在此基础上引入积神经网络来分别提取源微博和转发评论序列的深层特征，从而利用更高层次的信息构建出基于卷积—门控循环单元的网络模型。

## 3.1 数据预处理

微博数据集中事件的数据不仅包含源微博及其转发序列的文本、时间信息，还包含了相关用户的属性特征等信息，因此首先需要在微博数据集中提取出模型需要的文本特征和时间特征。此外，不同事件微博的转发序列长度不同，每条转发附带的文本信息长度也并不相同，所以需要将数据进行预处理，对文本信息进行聚合、划分，利用合适的语言模型进行建模并最终处理为包含丰富特征信息的且适合作为模型输入的向量数据。

### 3.1.1 序列划分

为了解决不同事件微博序列长度不一致问题，同时更好的利用序列的文本信息，本文引入聚合度*N*来表示微博序列的聚合程度。因此微博事件数据*M* = {()} 转换为*F* = ，其中 表示文本聚合后的新序列，而*N*的取值将与采用的模型方法进行适当设置。

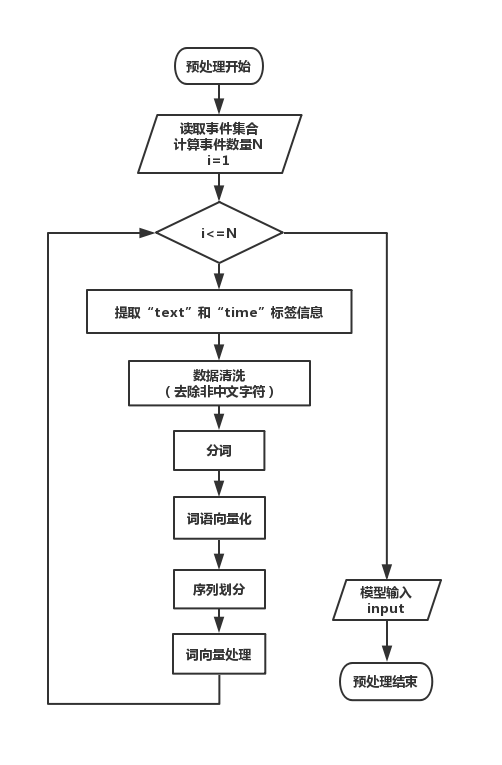
### 3.1.2 文本分词

基于划分的序列，需要将每个序列中的文本符号信息进行分词，将每个句子切分成单词。本文使用结巴(jieba)分词工具对中文文本进行分词，在分词之前，本文已经将文本信息中的标点符号以及非法字符等去除，以保证分词后获得正确的单词。

### 3.1.3 语言模型

分词后需要对文本单词进行语言建模，将文本符号数据转换为模型输入需要的词向量形式的数据。本文采用Mikolov等人在[28]中提出的Skip-gram模型来学习词语的分布表示。Skip-gram模型对目标词到上下文的后验概率建模，即模型已知目标词的条件下预测上下文词语出现的概率。对于窗口中的文本序列，模型通过目标词 ，分别预测其上下文词语中任意一个词语出现的概率。

### 3.1.4 数据预处理流程



**图3.1 数据预处理流程图**

第一部分，首先需要在微博数据集中提取出每一个事件对应微博序列的“text”和“time”标签信息，并将同一事件的“posts\_num”个“text-time”对构成集合。

第二部分，分别对每个事件的“text-time”对集合进行处理。将“text”去除标点符号、英文字符以及网址链接等干扰项，完成数据清洗工作。

第三部分，对清洗后的文本进行分词，利用“结巴分词”工具包，将文本句子切分为单词序列。

第四部分，利用预训练好的词嵌入模型，将单词序列转换为对应的词向量序列。

第五部分，根据模型选取合适的序列划分算法，将不同长度的微博事件序列转换为模型输入对应长度。

第六部分，将划分好的新序列进行聚合，将同一序列内的词嵌入进行加和取平均，得到最终的模型输入数据。

基于以上设计，得到相应的数据预处理流程图如图3.1所示。

## 3.2 基于门控循环单元的网络模型

微博是一系列内容相关的时间序列，而由前文介绍可知，门控循环单元网络是一种优化的循环神经网络，能够获取输入序列之间的依赖关系，从而对输入进行建模。本章将以门控循环单元为基本结构，构建微博谣言检测模型。

微博数据集中数据的形式为同一事件相关的源微博、转发及评论，是一系列以发布时间为序的中文字符序列。不同事件的序列长度不同，而GRU网络需要的输入为长度大致相同的序列数据，因此，首先需要设计算法将每个事件的序列进行划分和聚合。

### 3.2.1 平均划分(Average division，AD)

假设事件*M* = {()}的序列长度为，模型GRU网络的时间步(Time Steps)为，那么平均划分算法将会把事件*M*的所有序列划分为相同长度的*N*份，其中每一份长度为 。

对于划分后的新序列，将每个新序列内的文本信息分别转换为词向量后进行取平均，以最终的平均值作为该序列的特征向量。

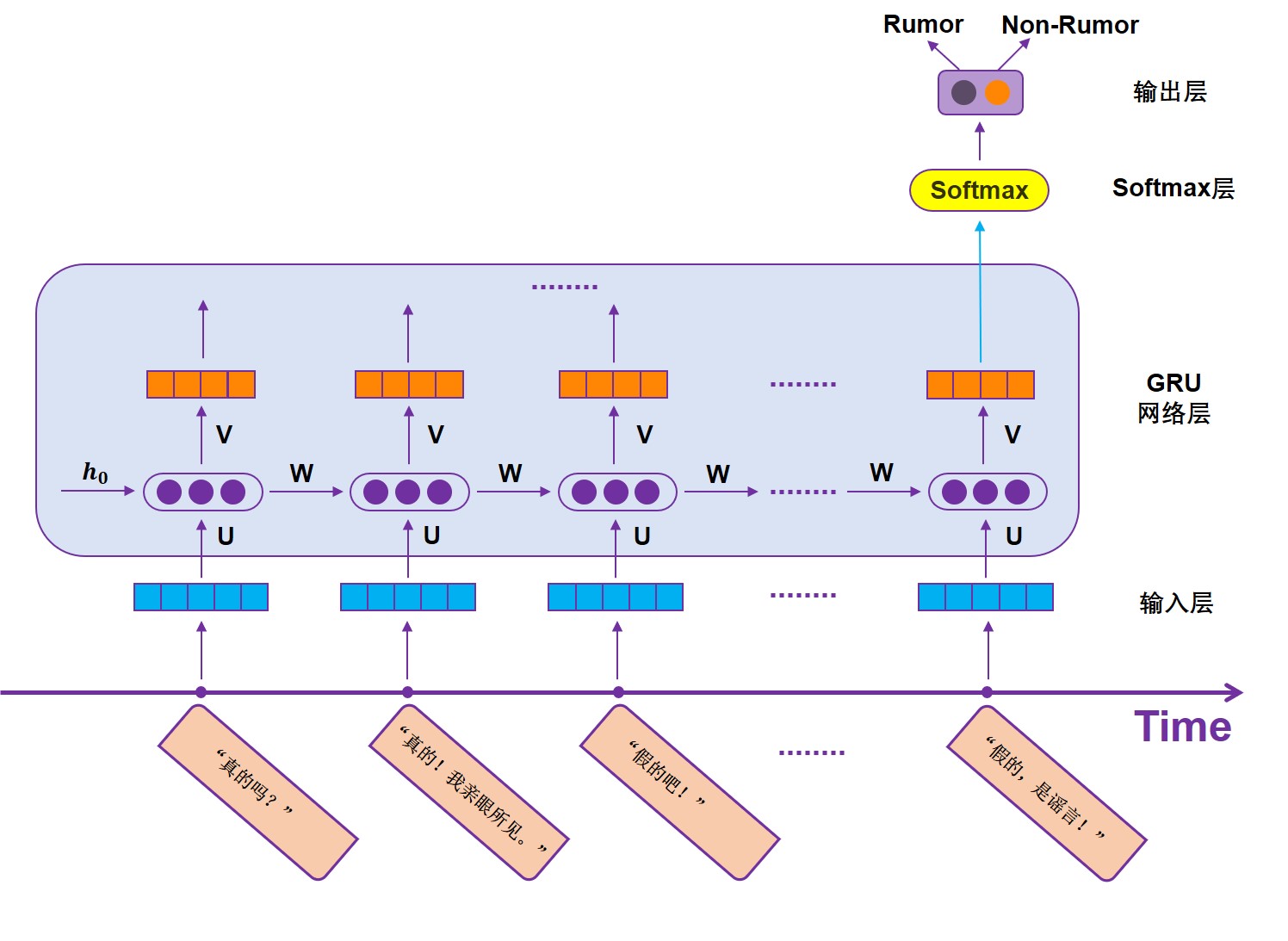
### 3.2.2 以时间间隔划分(Divided by time interval，DT)

本算法将根据每个事件的时间跨度 将序列划分为近似等于GRU网络的时间步长度[20]。首先，将整个时间序列均等的划分为*N*个时间间隔(*N*为时间步长度)，此时时间间隔长度为 每条转发或评论依据发布时间分布对应的时间间隔内，这*N*个时间间隔构成集合。然后，去除中的空间隔(不含有转发或评论序列的时间间隔)得到非空间隔集合(即每个时间间隔内至少包含一条转发或评论)。在集合中选择连续的、时间跨度最长的时间间隔加入到集合中。如果集合中时间间隔数小于*N*且大于前一轮得到的集合内的时间间隔数，那么将时间间长度减半 ，然后继续开始新一轮的划分，否则，返回连续时间间隔的集合。需要注意的是，每个事件划分后的时间间隔数量不一定等于*N*，只是接近于*N*，而同一事件中各个时间间隔的长度相等。DT算法的伪代码如下：

|  |  |
| --- | --- |
| 输入：相关事件序列，GRU网络的时间步数N； | |
| 输出：时间间隔的集合 | |
| 1: |  |
| 2: | **while** true **do** |
| 3: |  |
| 4: | ; |
| 5: |  |
| 6: |  |
| 7: | Find such that has the longest time span continues intervals; |
| 8: | **if** **then** |
| 9: |  |
| 10: | **else** |
| 11: | **return** |
| 12: | **end if;** |
| 13: | **end** |
| 14: | **return** |

**算法3.1 构造可变长度时间序列的算法（DT）**

对于划分后的时间间隔集合，将其中的每个时间间隔内的所有文本序列信息分别转换为词向量后进行取平均，以最终的平均值作为该时间间隔的特征向量。



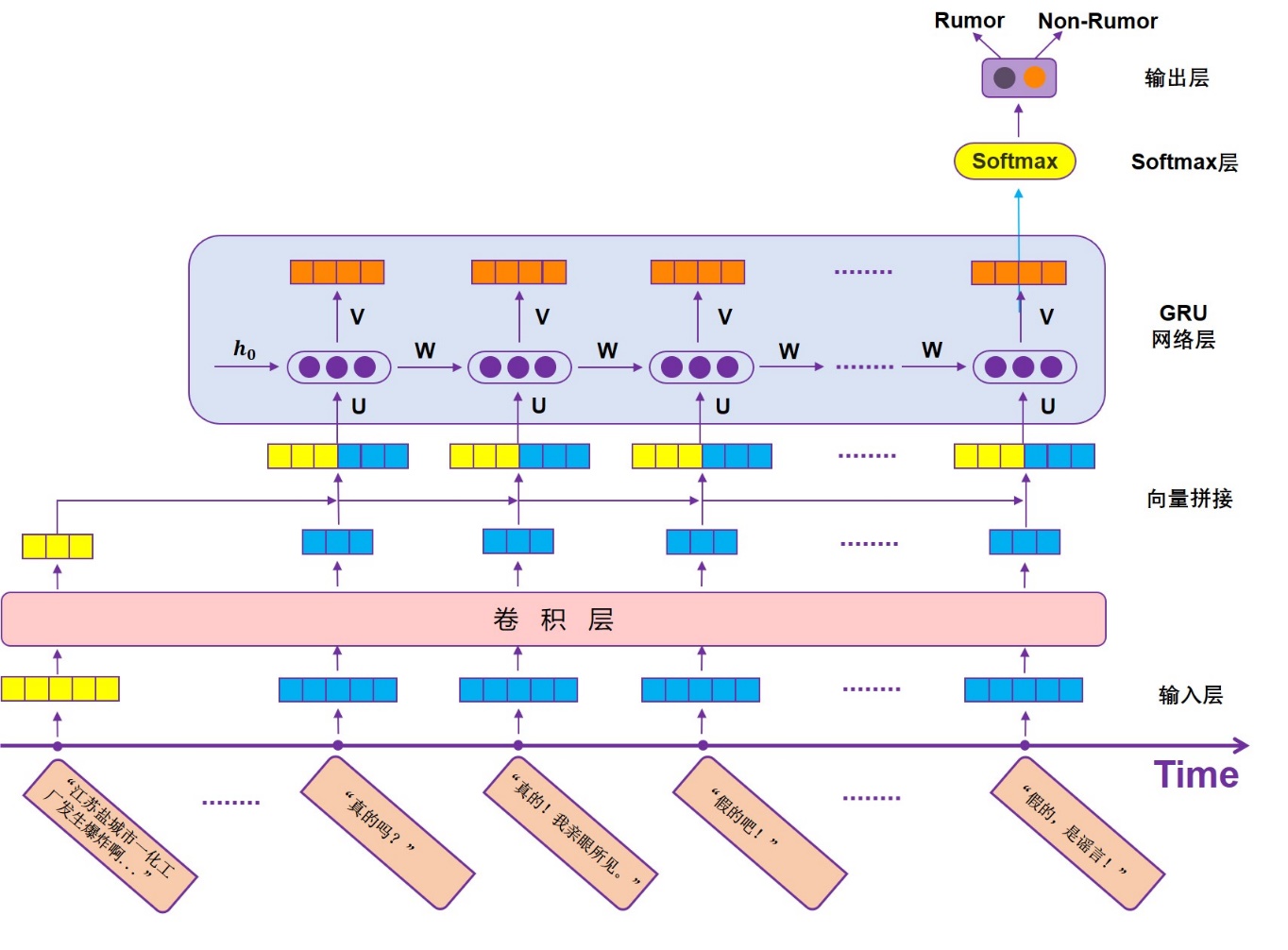
**图3.2 基于GRU网络的谣言检测模型结构**

微博数据经过预处理后即可作为模型输入。基于门控循环单元网络的谣言检测模型主要由输入层、GRU网络层和Softmax层组成。其中可通过增加GRU网络层来设计更为复杂的网络模型获取更深层次信息。该模型的整体结构如图3.2所示。

## **3.3 基于卷积—门控循环单元的网络模型**

通过对数据集的进一步研究，一个相关事件的微博数据是由一系列内容相关的时间序列信息，其中由一条发布时间最早的源微博(Original Microblog，OM)和若干转发/评论(Repost)组成，通过比较源微博与转发的信息特征可以发现，一个事件的主要内容特征信息主要分布在源微博上，而转发序列上主要是一些内容较短的文本特征。

基于以上分析，本模型将每个事件的源微博和转发序列分别独立处理，经过数据预处理过程后，分别送入卷积神经网络模型处理，提取深层特征信息。然后将处理后的源微博特征与转发序列特征拼接成新向量，送入GRU网络模型进行处理，将GRU网络的输出送入全连接层并通过Softmax层进行计算两类的概率。改进后的模型结构如图3.3所示。



**图3.3 CNN-GRU微博谣言检测模型结构**

需要注意的是，在CNN-GRU模型中，其数据预处理过程中序列划分阶段并没有采用AD和DT算法，而是统一设置序列长度为10000(序列长度不足10000的需要进行补全操作)，然后顺序的将每10个序列进行划分为新序列，同一个新序列内的转发信息转换为词向量后进行加和取平均来获得该新序列的向量表示。因此经过数据预处理后，每个事件的输入形式为1000个相同维度的向量。

## **3.4 目标函数及优化方法**

本文所构建的GRU、CNN-GRU微博谣言检测模型，本质上是一个利用深度神经网络模型进行的二分类(谣言、非谣言)任务，而对于分类任务一般都采用交叉熵函数来作为模型的优化目标，因此本模型的损失函数为

(3.1)

其中，为样本数量，为样本的真实标签， 为经过网络模型计算所得的预测标签。这样 表示了模型整体的分类误差，在训练过程中随着模型不断学习，其分类越来越准确，而 也将越来越小。因此，该模型的优化目标为

(3.2)

模型训练过程选择梯度下降(Gradient Descent，GD)算法进行参数学习与迭代优化。需要注意的是，相比GRU网络模型，CNN-GRU模型除需要对GRU网络层的参数进行迭代学习外，还要根据模型的输出对卷积层中的相关参数进行反向迭代更新，因此CNN-GRU模型比GRU网络模型复杂的多，但能够获取文本信息中更高水平的特征信息，从而改进谣言检测的性能。

基于以上模型原理及结构设计，本文下一章将对构建的基于门控循环单元的网络模型以及基于卷积—门控循环单元的网络模型分别进行微博谣言的识别实验和早期检测实验以测试模型的性能。

# 4 模型实验与评估

本章将介绍实验所使用的微博数据集信息，并对微博数据的特征进行统计分析。然后将介绍微博谣言检测实验的一些参数设置及对实验结果进行评估的一些方法。同时本章也会利用构建的两种谣言检测模型进行早期谣言检测实验，并对实验结果进行对比分析，验证模型的性能。

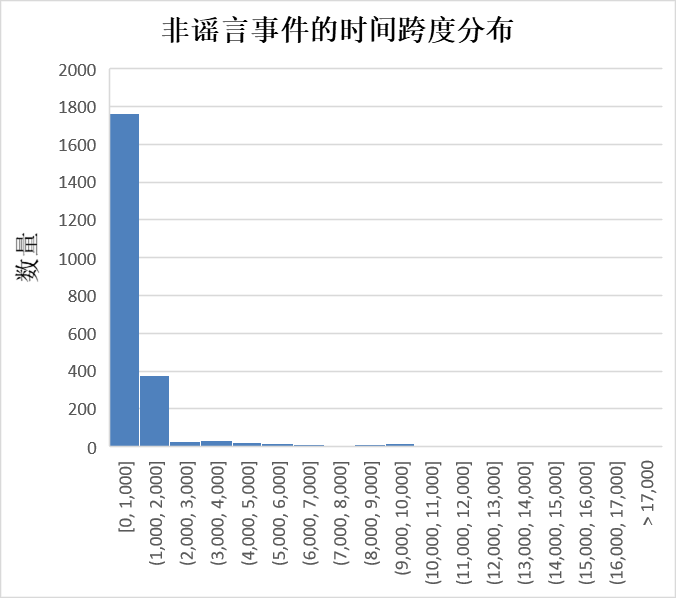
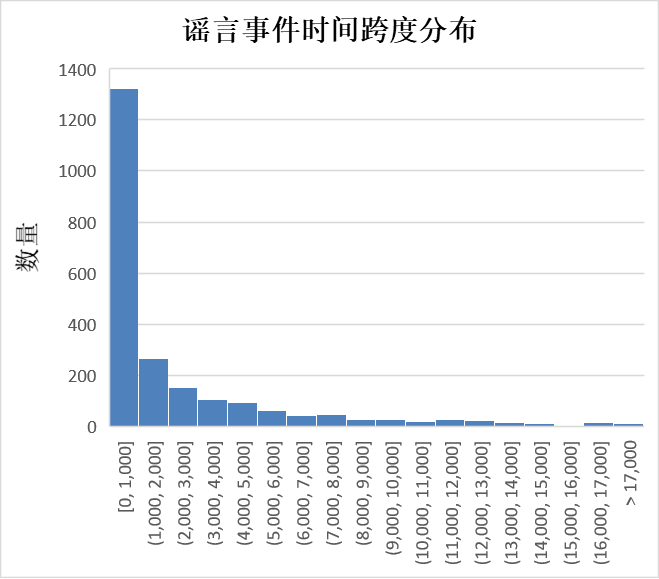
## 4.1 数据集

本文所有的实验数据来源于新浪微博平台，所有谣言事件与非谣言事件数据皆为中文信息。该数据集共包含4664个事件，其中有2313个被标记为谣言事件，2351个为非谣言事件。该数据集的一些统计数据信息如表4.1所示。

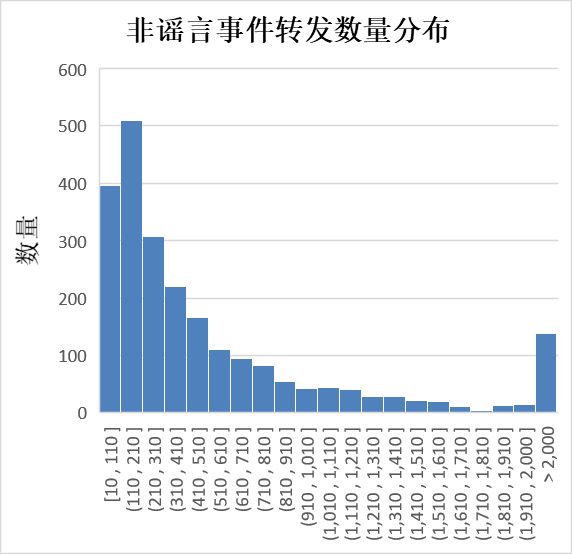
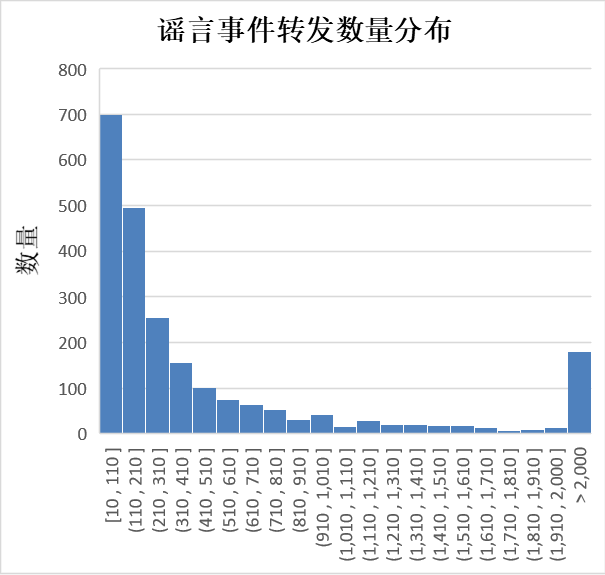
**表5.1 微博数据集**

|  |  |  |  |
| --- | --- | --- | --- |
| **中文微博数据集** | **谣言事件** | **非谣言事件** | **合计** |
| 数量 | 2,313 | 2,351 | 4,664 |
| 转发数 | 2,090,743 | 1,661,716 | 3,752,459 |
| 最小转发数 | 10 | 10 | 10 |
| 最大转发数 | 59,318 | 52,157 | 59,318 |
| 平均转发数 | 805 | 708 | 805 |
| 最小时间跨度(min) | 3 | 1 | 1 |
| 最大时间跨度(h) | 28,095 | 27,682 | 28,095 |
| 平均时间跨度(h) | 2,344 | 1,028 | 1,680 |

由微博数据集的统计数据分析，谣言事件与非谣言事件的数量近似相等；但从转发数量与时间跨度角度分析，无论是谣言事件还是非谣言事件，其最大值与最小值差距非常大，从一定程度上表明两类事件在转发数与时间跨度方面分布不平衡。为进一步具体分析两类事件在转发数与时间跨度上的分布特征，我又提取出每个事件的转发数量及每条微博的发布时间进行数据分析，并将统计分析结果绘制成直方图以便更清晰地观察其特征。



1. (b)



(c) (d)

**图4.1 微博数据集时间跨度分布特征及转发数分布特征**

由图4.1 的(a)、(b)可见，在时间跨度方面，虽然两类事件的时间跨度都很长，但其整体大都分布在2000 mins内，两者的区别在于，非谣言事件分布更为集中，而谣言事件较为分散。因此，在设计模型时，可选择2000~3000 mins作为时间的截断点便可利用大部分序列信息。在转发数量方面分析，两类事件的转发数量分布也较为相似，大部分事件序列长度在17000以内，这对于设置模型的时间步长是一个重要的参考，从事件的传播特征角度也可以发现谣言事件与非谣言事件的一些差异，例如谣言事件的“爆发式”传播，其在传播初期会引发大量转发与评论，随后热度衰退，转发与评论也逐渐减少，其在谣言检测中的影响也会随时间的增加而减小。相比较的，非谣言事件在传播初期的热度并没有谣言事件高，其热度随时间的衰减也没有谣言事件快，非谣言事件的传播，是一个递进的、持续的过程。

## 4.2 早期谣言检测

微博谣言常伴随着突发事件产生，因此会在事件爆发的初期引起大量的讨论，并随着时间的推移和相关事件的真实信息增多，微博谣言的影响及危害也会逐渐降低。因此，构建早期谣言检测模型，在谣言产生的初期识别出谣言信息从而加以制止其传播具有十分重要的现实意义。

本文构建的GRU网络模型及CNN-GRU模型除可以对谣言事件进行准确识别外，在谣言的早期检测方面，也具有不错的效果。

对于GRU网络模型，其序列划分方式以发布时间为参考，从时间的角度对整个序列进行了切分处理。因此，在对GRU网络模型进行早期谣言检测实验时，将从时间角度来表达“早期”。本实验设置了1、6、12、24、48、72、96、120(单位：小时)8个时间跨度作为微博的截止发布时间，利用时间跨度以内的数据来进行早期谣言检测实验，验证模型的早期谣言检测性能。

对于CNN-GRU模型，其序列划分方式是将顺序的序列每10个为一组进行划分，可视为从传播的角度对整个序列进行了切分处理。因此，对CNN-GRU模型的早期谣言检测实验将从传播角度来表达“早期”。本实验设置了5%、15%、25%、35%、45%、55%、65%、75%、85%和95%(数据占比)10个数据的利用率，利用每个事件的部分数据来对谣言进行识别，完成模型的早期谣言检测实验。

## 4.3 评估方法及实验参数设置

本文各项实验的评估将采用经典分类任务的评估指标准确率(Accuracy)、精确率(Precision)、召回率(Recall)和值(-Score)来评估模型。在介绍四个评估指标的含义之前，首先将介绍以下几个概念。

假设对微博事件进行分类，其中谣言为正类(Positive)，非谣言为负类(Negative)。那么把谣言正确分类为谣言记为TP，非谣言正确分类为非谣言记为TN，谣言被错分为非谣言记为FN，非谣言被错分为谣言记为FP。

准确率，

(4.1)

精确率，又叫查准率，

(4.2)

召回率，又叫查全率，

(4.3)

值，是精确率与召回率的加权调和平均，是对模型的综合评估

(4.4)

针对本文所构建的微博谣言检测模型，本文共设置了4个实验以验证模型的效果。其中实验1和实验2用来验证GRU网络模型与CNN-GRU模型对谣言事件与非谣言事件的检测识别能力；实验3和实验4分别用来验证两类模型的早期谣言检测能力。

各实验进行时相关的模型参数设置表4.2所示。

**表4.2 实验相关参数设置**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型参数 | 实验1  (GRU) | 实验2  (CNN-GRU) | 实验3  (GRU) | 实验4  (CNN-GRU) |
| embedding\_size | 300 | 300 | 300 | 300 |
| time\_steps | 50 | 1000 | 50 | 1000 |
| filter\_size | \ | 50 | \ | 50 |
| filter\_width | \ | 4/5 | \ | 4/5 |
| hidden\_num | 128 | 128 | 128 | 128 |

其中embedding\_size、time\_steps、filter\_size、filter\_width和hidden\_num分别表示词向量维度、GRU网络的时间步数、卷积层的滤波器尺寸、滤波器宽度和隐层单元数。基于以上参数设置，开展实验。

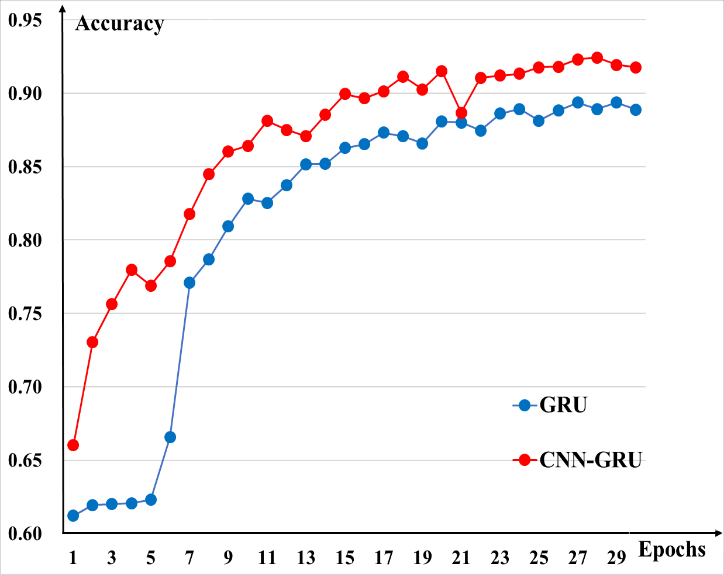
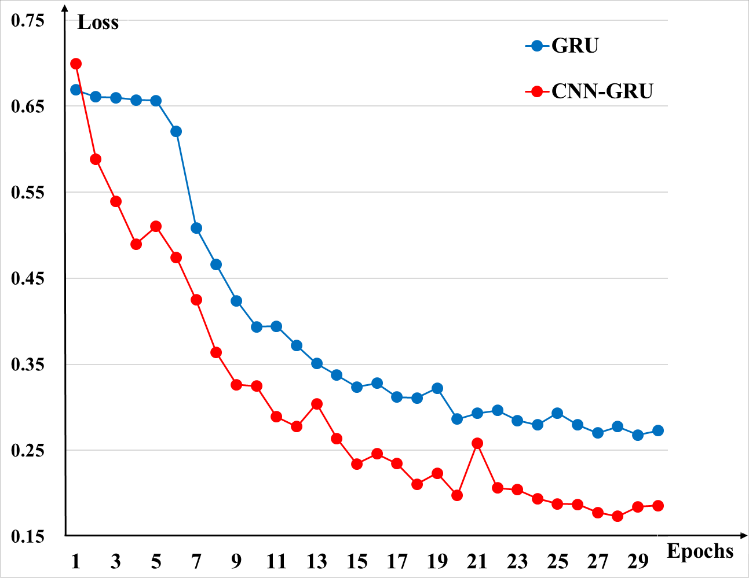
## 4.4 实验结果分析与改进

在谣言检测模型的识别率方面，根据实验1与实验2的测试结果，得到表4.3。

**表4.3 GRU模型与CNN-GRU模型实验结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | Accuracy | Precision | Recall |  |
| GRU | 0.8936 | 0.8763 | 0.9011 | 0.8885 |
| CNN-GRU | 0.9243 | 0.9314 | 0.9345 | 0.9329 |

同时，绘制出实验1与实验2的损失和准确率在训练过程中的曲线如图4.2。



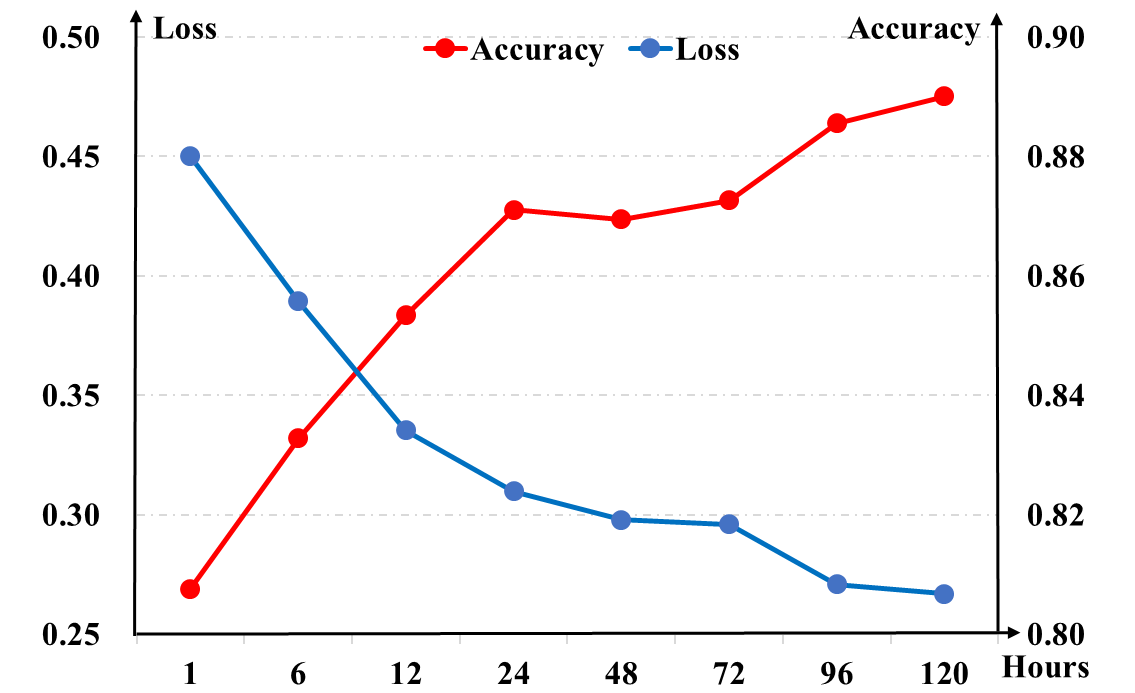
**图4.2 GRU网络模型与CNN-GRU模型损失及准确率的训练过程曲线**

根据实验结果分析可知：(1)GRU网络模型与CNN-GRU模型都可以进行微博谣言的识别，两种模型都可以较少的迭代训练中收敛，达到较高的准确率；(2)两种模型的区别在于CNN-GRU模型利用卷积神经网络来提取序列的深层特征，并且将源微博单独处理并运用到之后的输入数据中，因此在训练过程中CNN-GRU模型收敛更快，准确率更高。

在早期谣言检测实验中，实验3对GRU网络模型设置了数据的时间截点，从微博事件产生与传播的时间角度来控制微博事件的早期数据量来进行早期谣言检测实验；实验4对CNN-GRU模型设置了模型输入数据占全部数据的比例，从微博事件的传播过程角度来进行早期谣言检测实验。根据实验3与实验4的实验结果，分别整理得到以下数据表并绘制曲线。其中，Hours列数据表示输入数据的时间截点，Percents列数据表示采用的输入数据占全部数据的百分比。

**表4.4 GRU网络模型早期检测实验结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hours | Accuracy | Precision | Recall |  |
| 1 | 0.8076 | 0.8125 | 0.8561 | 0.8337 |
| 6 | 0.8328 | 0.8564 | 0.8746 | 0.8654 |
| 12 | 0.8534 | 0.8496 | 0.8841 | 0.8665 |
| 24 | 0.8711 | 0.8689 | 0.8763 | 0.8726 |
| 48 | 0.8695 | 0.8705 | 0.8951 | 0.8826 |
| 72 | 0.8727 | 0.8876 | 0.8943 | 0.8909 |
| 96 | 0.8856 | 0.8903 | 0.8721 | 0.8811 |
| 120 | 0.8901 | 0.8832 | 0.9014 | 0.8922 |

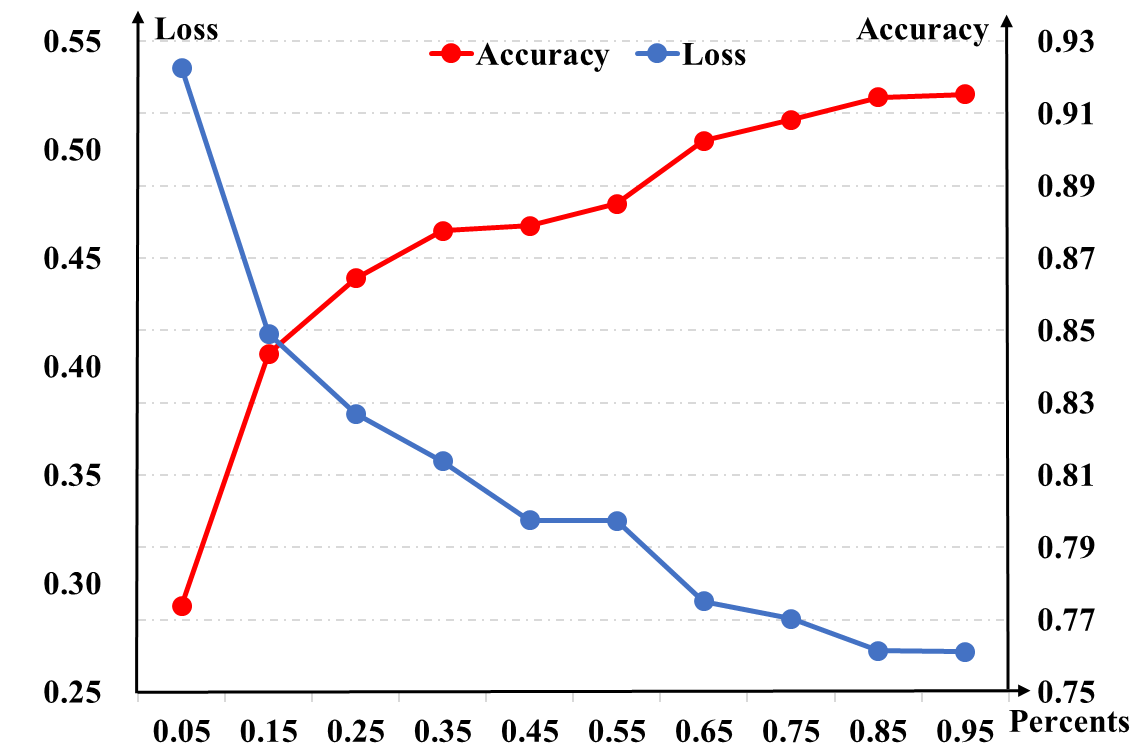
****

**图4.3 GRU网络模型早期检测实验Loss与Accuracy曲线**

以上结果表明，GRU网络模型与CNN-GRU模型分别从传播时间和传播过程两个角度进行了实验验证，在利用有限的输入数据时仍可以提取深层特征信息进行识别，识别的准确率随着输入数据的增多而迅速提高，具有不错的早期检测性能：GRU网络模型能够在微博事件发生后的1小时内识别谣言的准确率达到80%以上；CNN-GRU模型可以只利用微博整个事件传播数据的15%而达到84%的识别准确率，表明了本文构建的两种模型能够在事件突发后对相关微博谣言快速检测与识别。

**表4.5 CNN-GRU模型早期检测实验结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Percents | Accuracy | Precision | Recall |  |
| 5% | 0.7738 | 0.7914 | 0.8541 | 0.8216 |
| 15% | 0.8435 | 0.8517 | 0.8675 | 0.8595 |
| 25% | 0.8644 | 0.8936 | 0.8907 | 0.8921 |
| 35% | 0.8775 | 0.9002 | 0.9125 | 0.9063 |
| 45% | 0.8789 | 0.8973 | 0.9230 | 0.9100 |
| 55% | 0.8850 | 0.9143 | 0.9257 | 0.9200 |
| 65% | 0.9025 | 0.9216 | 0.8759 | 0.8982 |
| 75% | 0.9083 | 0.9329 | 0.9148 | 0.9238 |
| 85% | 0.9145 | 0.9310 | 0.9089 | 0.9198 |
| 95% | 0.9153 | 0.9230 | 0.9356 | 0.9293 |



**图4.4 CNN-GRU模型早期检测实验Loss与Accuracy曲线**

通过以上4个实验，分别对本文构建的谣言识别与早期检测性能进行了测试，从模型设计上，CNN-GRU模型以GRU网络模型为基础进行了改进，实验结果也表明CNN-GRU模型在谣言识别准确率及早期检测性能上都要优于GRU网络模型。但是两种模型在输入数据处理、序列划分方式以及GRU网络层的参数设计上有一定的差异，此外早期检测实验中两种模型采用了不同角度来模拟谣言传播的“早期”，因此将两种模型直接对比分析并不合理。因此需要未来对实验进行改进，控制两种模型GRU网络层变量，统一序列划分方式，并在两种早期检测角度对两种模型进行测试，并将结果进行对比分析。此外，对于性能更优的CNN-GRU模型，需要设计相关实验以对其模型的影响参数进行分析验证。

基于本文所进行的研究，对于未来工作可进行的改进，总结如下：

1. 在模型设计上，本文所提出的两种模型采用的神经网络结构较为简单，未来可以此为基础设计更加复杂的、能够提取更高层次的特征的模型结构来改进性能。
2. 在序列划分、输入数据处理方式上本文只是进行简单的时间划分或区间划分，可从此方面入手，尝试设计更加科学合理的序列划分算法。
3. 在优化目标上，本文采用经典的分类任务的交叉熵作为损失函数来进行优化，下一步可结合模型结构设计更加合理、多层次的优化目标。
4. 在早期谣言检测方面，寻找更严谨、更合理的评估标准来评估模型的早期谣言检测性能。

结 论

微博谣言检测己成为社交网络内容安全中不可忽视的课题，谣言传播具有突发性且传播速度极快，对正常的社会秩序造成不良影响，研究微博谣言检测技术能够快速、准确地识别微博平台上的谣言信息并在早期制止其传播，降低其社会危害。

本文首先介绍了并分析了国内外对于谣言检测技术的研究与发展，并对本文构建微博谣言检测模型所涉及的相关技术原理及方法进行了详细的介绍。之后利用门控循环单元来设计基本的谣言检测模型，并以此为基础构建更加有效的卷积-门控循环单元网络模型来实现微博谣言的检测。实验结果表明，两种模型在中文微博数据集上的谣言识别准确率分别达到89.36%和92.43%，并在早期谣言检测实验中取得了不错的实验结果。

本文所构建的微博谣言检测模型虽然在微博谣言识别以及早期谣言检测方面取得了较好的效果，但在模型的整体设计方面仍然存在一些可以改进的地方：首先是模型的结构上，本文设计的模型仅仅利用了微博数据的文本序列信息，可以考虑结合谣言传播过程中的其他特征进行特征聚合来改进模型性能；然后在序列数据处理上，本文仅仅是采用了简单的时间划分和区间划分方式，没有科学的理论依据，因此可以探索更加合理有效的序列划分与聚合算法；此外，在优化目标方面，本文选择了经典的交叉熵作为优化目标，在对模型的整体优化上不够完善和具体，因此未来需要结合模型结构来设计相应的优化函数，对模型的每一层进行加权优化；最后，由于微博谣言的早期检测的相关研究还较少，对“早期”的评估标准各不相同，因此需要探索科学合理的早期谣言检测评估标准，对模型的早期检测性能进行科学评估。

致 谢

大学四年的学习生活随着这最后一门课程——毕业设计的完结而告一段落，回首在南京理工大学度过的这四年时光，丰富而充实，短暂而令人难忘。一路走来，我遇到了许许多多的良师益友，是他们的一路相伴，使我的大学生活如此多彩。

首先，要感谢我的指导老师夏睿教授。在毕业设计初期，夏睿老师便对我研究工作的开展提出了明确的要求与悉心的指导；安排我参加每周的组会与记录工作周记使我养成了良好的研究习惯。是夏睿老师的高要求与严格，使我几个月来完全没有松懈，不断去学习去探索，从而能够快速进入研究主题，更好的开展实验。这对我来说是一个很好地开始，不仅在知识储备上，更在心理上做好准备以面对即将到来的研究生生活。在此，对夏睿老师的关怀与指导再次表示由衷的感谢！

其次，感谢对我提供帮助的学长学姐们，特别是负责指导我开展课题研究的宣凯洲学长。从开始时的一无所知到慢慢地打好基础，再到后来的开展实验，宣凯洲学长对我提供了莫大的帮助，总是能够耐心地解答我的各种问题，悉心地指导我进行研究。

同时，感谢大学四年来为我授业解惑的所有老师，感谢四年来为我提供帮助的同学与朋友；感谢在一起努力的校大学生科协的伙伴们……

最后，要特别感谢我的父母、我的家人，感谢他们教会我善良与勇敢，感谢他们教会我做人、做事、做学问，感谢他们一直以来对我的支持与鼓励！

感谢！

参 考 文 献

1. 中国互联网信息中心.第43次中国互联网络发展状况统计报告[EB/OL].[2019-02-28]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201902/P020190318523029756345.pdf
2. 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(10):2799-2805.
3. 李巍胤. 基于情感分析的微博谣言识别模式研究[D]. 重庆大学, 2016.
4. 王彦本,蔡皖东.社交网络中考虑遗忘机制的谣言传播[J].西北工业大学学报, 2016(34):349-355.
5. 王丹磊.基于特征聚合的端到端谣言鉴别技术研究[D].武汉大学, 2018.
6. 微博辟谣.2018年度微博辟谣数据报告[EB/OL].新浪微博, https://ww3.sinaimg.cn/bmiddle/6f3f16a9ly1fziuegsinxj20u0c5z7wr.jpg
7. Castillo C, Mendoza M, Poblete B. Information credibility on Twitter[C]. Proceedings of the 20th International Conference on World Wide Web.New York:ACL,2011:675-684.
8. Qazvinian V, Rosengren E, Radev D R, et al. Rumor has it: Identifying misinformation in microblogs[C]. Proceedings of the 2011 Coference on Empirical Methods in Natural Language Processing. Edinburgh: ACL, 2011:1589-1599.
9. Mendoza M, Pdblete B, Castillo C. Twitter under crisis: Can we trust what we RT ?[C]. Proceedings of the First Workshop on Social Media Analytics. New York: ACL, 2010: 71-79.
10. Takahashi T, Igata N. Rumor detection on Twitter[C].2012 Joint 6th International Conference on Soft Computing and Intelligent Systems(SCIS) and 13 th International Symposium on Advanced Intelligent Systems(ISIS). Kobe: IEEE, 2012: 452-457.
11. 程亮,邱云飞,孙鲁.微博谣言检测方法研究[J].计算机应用与软件, 2013,30(2): 226-228.
12. 许晓东,肖银涛,朱士瑞.微博社区的谣言传播仿真研究[J].计算机工程, 2011,37(10): 272-274.
13. Yang Fan, Liu Y, Yu X, et al. Automatic detection of rumor on Sina Weibo [C].Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. Beijing: ACM, 2012: 1-7.
14. 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.
15. Adrien Friggeri, Lada A Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor cascades. In Proceedings of ICWSM.
16. Aniko Hannak, Drew Margolin, Brian Keegan, and Ingmar Weber. 2014. Get back! you don’t know me like that: The social mediation of fact checking interventions in twitter conversations. In Proceedings of ICWSM.
17. Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang.2013. Prominent features of rumor propagation in online social media. In Proceedings of ICDM. pages 1103–1108.
18. Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. CIKM ’15, pages 1751–1754.
19. Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web. WWW ’15, pages 1395–1405.
20. Ma J, Gao W, Mitra P, et al. Detecting Rumors from Microblogs with Recurrent Neural Networks[C].IJCAI. 2016: 3818-3824.
21. Ma J, Gao W, Wong K F. Rumor detection on twitter with tree-structured recursive neural networks[C].Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018, 1: 1980-1989.
22. Ma J, Gao W, Wong K F. Detect rumor and stance jointly by neural multi-task learning[C].Companion of the The Web Conference 2018 on The Web Conference 2018. International World Wide Web Conferences Steering Committee, 2018: 585-593.
23. Ke Wu, Song Yang, and Kenny Q Zhu. 2015. False rumors detection on sina weibo by propagation structures. In Data Engineering (ICDE),2015 IEEE 31st International Conference on. IEEE, pages 651–662.
24. Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). volume 1, pages 708–717.
25. He Y, Li J, Song Y, et al. Time-evolving Text Classification with Deep Neural Networks[C].IJCAI. 2018: 2241-2247.
26. Yavary A, Sajedi H. Rumor detection on Twitter using extracted patterns from conversational tree[C].2018 4th International Conference on Web Research (ICWR). IEEE, 2018: 78-85.
27. 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. ACM, 2018: 943-951.
28. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. Proceedings of Workshop at 1st International Conference on Learning Representations (ICLR), 2013.
29. N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences,” arXiv preprint arXiv:1404.2188, 2014.
30. Y. Kim, “Convolutional neural networks for sentence classiﬁcation” arXiv preprint arXiv:1408.5882, 2014.
31. A. Tamar, Y. Wu, G. Thomas, S. Levine, and P. Abbeel, “Value iteration networks” in Advances in Neural Information Processing Systems, 2016, pp. 2154–2162.
32. Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.