

杨佳瑶, 杨越, 薛雨蒙, 等. 谣言判别器的设计与实现[J]. 智能计算机与应用, 2024, 14(4): 76-82. DOI: 10.20169/j.issn.2095-2163.240410

## 谣言判别器的设计与实现

杨佳瑶, 杨越, 薛雨蒙, 王鑫淼, 杨香云, 乔秀明

(北京信息科技大学 计算机学院, 北京 100096)

**摘要:** 在互联网时代, 信息传播的速度和范围都得到了极大的提升, 使得各种信息能够在极短的时间内迅速扩散至广泛的受众, 然而这也带来了一些挑战, 谣言的传播是其中之一, 自动谣言判别可以大大降低谣言传播率。本文构建了中文谣言数据集, 设计实现了谣言自动检测的小程序: 用户输入一段言论, 小程序的前端页面将数据传送到后端, 基于卷积神经网络的模型进行自然语言处理, 模型采用深度学习算法对该言论进行语义建模和分类, 最终得出这段话是谣言的概率并返回给用户。

**关键词:** 中文谣言数据集; 卷积神经网络; 谣言判别

中图分类号: TP311.1

文献标志码: A

文章编号: 2095-2163(2024)04-0076-07

### Design and implementation of rumor discriminator

YANG Jiayao, YANG Yue, XUE Yumeng, WANG Xinmiao, YANG Xiangyun, QIAO Xiuming

(School of Computer, Beijing Information Science and Technology University, Beijing 100096, China)

**Abstract:** In the Internet era, the speed and scope of information dissemination have been greatly improved, which enables all kinds of information to rapidly spread to a wide range of audiences in a very short period of time, however, this also brings some challenges, the spread of rumors is one of them, and automatic rumor discernment can greatly reduce the rate of rumor dissemination. In this paper, we constructed a Chinese rumor dataset, and designed and implemented an applet for automatic rumor detection; the user inputs a piece of speech, the front-end page of the applet transmits the data to the back-end, the model based on convolutional neural network carries out the natural language processing, and the model adopts a deep learning algorithm to semantically model and classify the speech, and ultimately arrives at the probability of the piece of speech being a rumor and returns it to the user.

**Key words:** Chinese rumour dataset; convolutional neural network; rumor discrimination

## 0 引言

随着互联网的发展, 网络上充斥着各种各样的言论, 其中有许多是虚假的、未经过官方所证实的信息。面对这些谣言, 个人的判断往往因知识面有限等原因并不可靠, 通过借助计算机去自动收集知识和特征并判断言论的可信度, 即时的捕捉谣言, 从而可以大大减少谣言传播所带来的危害。

近年来国内外学者对谣言判别进行了许多的研究, 目前谣言检测方面的研究工作有4类: 基于传统特征的建模、基于聚类的建模、基于情感分析的建

模、基于传播结构的建模。Castillo 等<sup>[1]</sup>针对谣言的文本特性、用户传播行为和话题内容等3个关键方面开展了深入研究, 并在此基础上运用决策树算法成功实现了对谣言的识别和检测; Blei 等<sup>[2]</sup>提出了一种谣言检测的方法, 主要关注与时间相关的特征, 建立随机森林分类器对谣言进行检测; 吕胜<sup>[3]</sup>搭建基于情感分析的 CNN-LSTM 模型, 该模型结合了卷积神经网络(CNN)和长短期记忆网络(LSTM), 旨在更准确地识别和分析文本数据中的情感倾向; Gao W 等<sup>[4]</sup>提出了一个双向的树状递归神经网络模型, 一个方向是自上而下, 另一个方向是自下而上, 该模型用于学习谣言传播结构的表示并进行分

**基金项目:** 北京市自然科学基金青年项目(4224090)。

**作者简介:** 杨佳瑶(2002-), 女, 本科生, 主要研究方向: 计算机应用技术; 杨越(2003-), 男, 本科生, 主要研究方向: 计算机应用技术; 薛雨蒙(2003-), 女, 本科生, 主要研究方向: 计算机应用技术; 王鑫淼(2003-), 女, 本科生, 主要研究方向: 计算机应用技术; 杨香云(2000-), 女, 本科生, 主要研究方向: 计算机应用技术。

**通讯作者:** 乔秀明(1989-), 女, 博士, 讲师, 主要研究方向: 自然语言处理。Email: qiaoxiuming@bistu.edu.cn

收稿日期: 2023-12-03

类,在两个公开的 Twitter 数据集上的结果表明模型不仅具有更好的性能,而且可以在比较早的阶段发现谣言。目前谣言判别技术已经在实际场景中得到应用,例如,一些社交媒体平台采用定向标注的方法来识别和过滤虚假信息,但是这些判别和过滤措施通常是被动的,现有的与谣言有关的小程序也仅能搜索特定的谣言或者向用户展示与搜索相关的谣言信息。而本文的谣言识别是用户主动输入信息,小程序及时向用户,反馈用户所输入的信息是谣言的概率。为了提高谣言判别的准确性,通过标注30 000条中文谣言数据集,训练基于卷积神经网络的谣言判别模型,小程序判别谣言的准确率达到95%。

## 1 相关工作

研究人员构建了多个数据集用于谣言判别,如:FakeNewsNet等,这些数据集在训练和测试模型时起到了重要作用<sup>[5]</sup>。然而,这些数据集大部分是英文谣言数据集,中文谣言数据集相对匮乏,不仅数据量小,还存在数据不完整的问题。本文通过数据清洗,完成中文谣言数据集的处理工作,构建了自己的中文谣言数据集,一共有31 669条谣言,5 219条非谣言,并将这些数据用于模型的训练。

最初的词袋模型(Bag of Word,简称BOW)应用于文本处理领域,并用于对文档进行分类和识别。在此模型中,文本(例如句子或文档)表示为无序的单词集合,而忽略了语法甚至词序。1927年,Sparck等<sup>[6]</sup>研究出一种新的算法(Term Frequency-Inverse Document Frequency,TF-IDF)用于探索文本特征。此外深度学习也被广泛用于文本特征的探索;1989年,Müller<sup>[7]</sup>提出了循环神经网络(RNN);次年,Elman<sup>[8]</sup>探讨了RNN在语言建模和序列数据处理中的应用;1998年,LeCun Y等<sup>[9]</sup>首次将卷积神经网络(CNN)用于自然语言处理,卷积神经网络(CNN)具有特征提取、平移不变性、参数共享、局部感知、适应不同尺度、逐层抽象的特点。本文采用基于上下文无关的词向量表示的模型(Text CNN)逐层提取抽象特征,更有效地表示数据,并利用参数共享的特征,减少模型的参数量,降低过拟合的风险,提高训练效率;利用Text CNN的局部感知特征,更深入的理解数据的内在结构和内容;通过灵活调整卷积核的大小和池化层的设置,使模型能够更好地适应不同比例数据集的特征,增强泛化能力。

随着人们对谣言传播机制认识的加深,以及技术手段的不断提升,谣言判别技术也在不断完善和

进步。机器学习方面,2018年,罗嗣卿<sup>[10]</sup>对组合优化决策树算法(Combined Optimization Decision Tree, CODT)做了相应改进,并提出一种E-CODT算法,改进后的算法有效提高了分类判别的精度,减小了数据结构的不同对决策树判别结果产生的影响,提升了谣言判别的准确率,体现了普适性以及现实应用价值。传播模式方面,2019年,陈铜<sup>[11]</sup>提出三元要素特征的时序聚合表示方法,用于表示事件的传播模式,综合考虑了多个方面的特征,能够更好地理解和防控社交网络中谣言信息的传播,为相关领域的研究提供了新的思路和方法。知识表达方面,传统的谣言检测忽略谣言的知识表示;2020年,陈香熠<sup>[12]</sup>使用知识图谱嵌入的方法检测谣言,从知识的角度模拟谣言文本形成的过程,对谣言数据进行增强,并对知识图谱进行编码,进一步实现对谣言的分析和挖掘,提高谣言检测性能。文本特征方面,2021年,李亦轩<sup>[13]</sup>提出将基于情感分析的虚假信息识别模型(Label BERT)和TRRM模型结合在一起做特征融合,从情感分析和时序特征分析两个角度对谣言进行检测,获取更加全面的特征,让模型能够更好地预测文本情感特征信息。

## 2 模型设计

### 2.1 基于上下文无关的词向量表示的模型(Text CNN)

本文采用的第一个模型是Text CNN模型,利用该模型进行文本分类与谣言判别任务,其模型结构如图1所示。

(1)嵌入层(Embedding):用于将整数序列转换为固定长度的随机词向量表示,输入维度为词汇表的大小,输出维度为词向量的维度,输入长度为最大序列长度;

(2)卷积层(Conv1D):采用128个大小为5的卷积核进行卷积操作,激活函数为ReLU;

(3)池化层(MaxPooling1D):对每个卷积核的输出进行最大池化操作,池化窗口大小为2;

(4)再次卷积层(Conv1D):采用128个大小为5的卷积核进行卷积操作,激活函数为ReLU;

(5)全局最大池化层(GlobalMaxPooling1D):对所有卷积核的输出进行全局最大池化操作,提取出整个文本序列中的最显著特征;

(6)全连接层(Dense):包含128个神经元,激活函数为ReLU;

(7)输出层(Dense):包含1个神经元,激活函数为Sigmoid,用于进行二分类任务的预测。

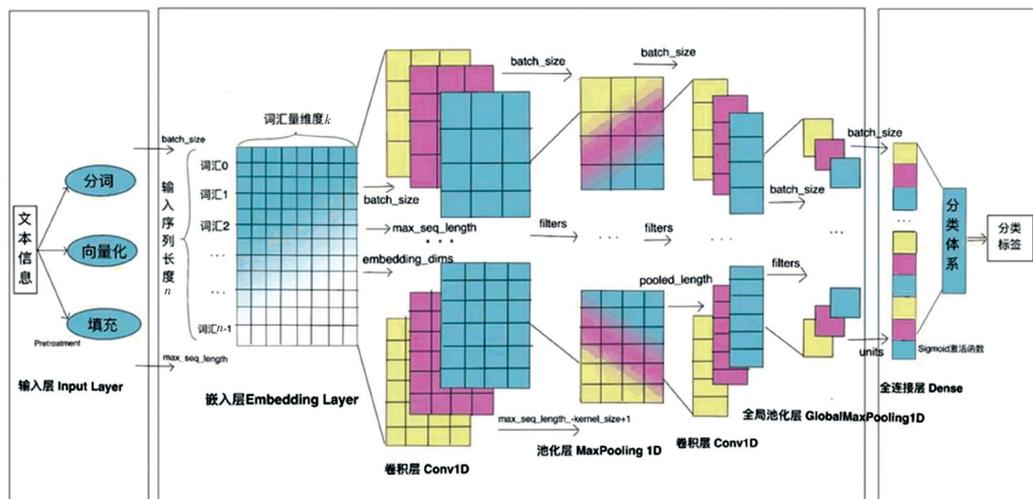


图1 Text CNN 模型结构

Fig. 1 Text CNN model structure

## 2.2 基于上下文相关词向量的谣言判别模型

本文的第二个模型是基于上下文相关词向量的谣言判别模型 (BERT + TextCNN + KG)。BERT (Bidirectional Encoder Representations from Transformers) 是一种基于 Transformer 架构的预训练语言表示模型,用于捕捉文本中的上下文信息;KG (Knowledge Graph) 是知识图谱,可以让模型更好地

理解和处理复杂的谣言。本文的基于上下文相关词向量的谣言判别模型 (BERT+TextCNN+KG) 是一个高度集成和创新的模型,结合了 BERT、TextCNN 和知识图谱 (KG) 的优势,能够综合利用上下文信息、文本特征和背景知识,更准确地识别和判断谣言内容,提高谣言判别的准确性和可靠性,模型结构如图 2 所示。

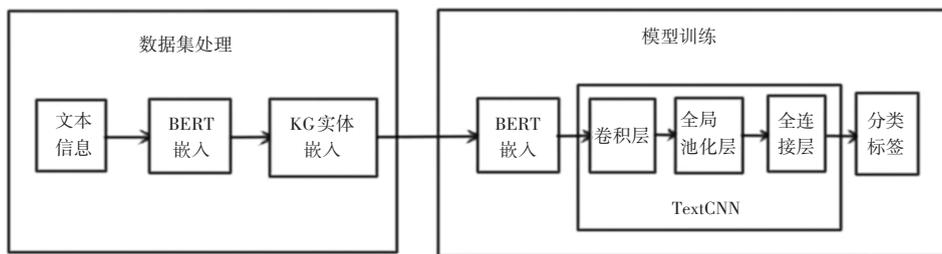


图2 BERT+TextCNN+KG 模型架构

Fig. 2 BERT+TextCNN+KG model architecture

模型的主要架构:

(1) BERT 嵌入: 使用预训练的中文 BERT 模型对输入的文本进行处理, 浓缩成单词或字词; 调整内容保持长度统一, 将填充项标记出来, 确保模型仅关注有效内容; 若文本包含多句, 标识出句子边界, 区分不同单词所属的句子; 最终生成上下文嵌入表示, 其中包含单词自身信息, 又融合了文本上下文。

(2) 知识图谱 (KG) 实体嵌入: 筛选出实体字典中的实体, 获取与输入句子相关的实体和邻接矩阵, 将线性层应用于输入邻接子矩阵进行转换, 用以匹

配 BERT 嵌入维度的大小。

(3) BERT 嵌入: BERT 模型进一步将输入文本编码为包含丰富语义信息的向量表示。

(4) 卷积层: 通过一维卷积操作使模型能够捕捉实体嵌入和 BERT 池化输出之间的复杂关系, 提取文本中的局部特征。

(5) 全局池化层: 将卷积层输出的宽度维度压缩, 同时保留其它维度信息, 整合全局信息并减少参数数量, 提高模型的泛化能力。

(6) 全连接层: 通过线性变换将全局池化层的

输出映射到最终的输出类别,全连接层的输出大小设置为 2,适用于二分类问题。

(7)分类标签(输出层):生成的结果包含了模型对每个类别的预测得分,这些得分是模型在未经过 softmax 归一化操作前的原始预测值,反映了模型对于每个类别的置信度。

### 3 数据集

#### 3.1 数据集和预处理

##### 3.1.1 数据收集

本文收集的实验数据来自中文社交媒体谣言统计网站,小组成员从该网站上爬取新闻信息。由于此部分数据字数过多、信息复杂,将爬取的新闻时间、新闻发布者、评论等过滤,最终收集到了从 2009 年 9 月 4 日至 2017 年 6 月 12 日的 31 669 条谣言和 5 000 条非谣言。数据清洗之后再两种数据进行分类,非谣言标注为 0,谣言标注为 1。

##### 3.1.2 文本预处理

在基于上下文无关键词向量的模型(Text CNN)中,已对文本数据进行序列化处理,利用训练集中的文本数据进行训练,建立词汇表,将训练集和测试集中的文本数据转换为整数序列。由于每次输入的数据不同,将所有的文本序列进行标准化处理,选取最长的文本序列作为标准长度,为了所有的输入序列长度一致,对序列进行填充或截断处理。将数据集划分为训练集和测试集,其中测试集占总数据集的 20%。

基于上下文相关词向量的谣言判别模型利用 BERT 分词器对文本内容进行细致的分割,将每个词或词组分割成一个个标记,并将这些标记转换为对应的唯一标识,构建起词汇表。在获取数据集后将其随机打乱,按照 80%、10%、10%的比例划分为训练集、测试集和开发集。

## 4 实验

### 4.1 基于上下文无关键词向量的模型(Text CNN)的实验

#### 4.1.1 实验设置

为了验证改进模型分类效果,在获取数据集后将其随机打乱,其中 80%作为训练集,20%作为测试集,在模型训练中指定迭代次数,实验的 CNN 卷积核个数为 128,卷积核大小为 5,在不同的数据比例和迭代次数下,观察验证集合中的准确率和损失函数。当选取的模型在训练集中的数据损失率最低时对测试集再进行预测,最后将测试集的准确率与

损失率作为本实验模型的验证指标。

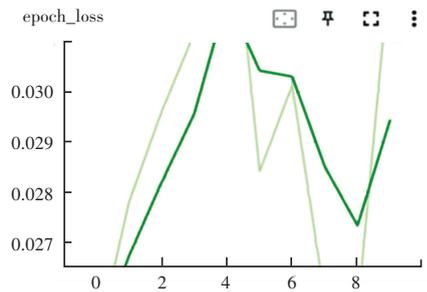
#### 4.1.2 实验结果

在本实验中,主要采用控制变量的方式进行数据之间的预测对比,有利于对比不同的数据比例间能达到最好效果的比例,推测得到这种效果的原因以及对于后续需要进行的完善,5 组实验结果见表 1。

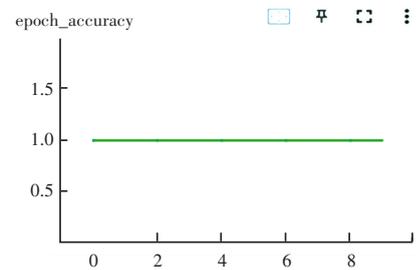
表 1 不同参数实验结果

Table 1 Experimental results with different parameters					
实验组	第一组	第二组	第三组	第四组	第五组
数据集比例	300 : 1	30 : 1	6 : 1	5 : 3	1 : 1
谣言数据集数量	30 000	30 000	30 000	5 000	5 000
非谣言数据集数量	100	1 000	5 000	3 000	5 000
迭代次数	10	30	10	20	20
测试集准确率	0.995 8	0.995 8	0.625 2	0.775 3	0.756 7

第一组实验数据集比例(谣言数量:非谣言数量)为 300 : 1,迭代次数为 10 的测试集损失率与准确率如图 3 所示;第二组实验数据集比例(谣言数量:非谣言数量)为 30 : 1,迭代次数为 30 的测试集损失率与准确率如图 4 所示。对比第一、二组实验结果,很明显比例过大无论是 300 : 1,还是 30 : 1 谣言数据集对于模型都是过拟合的,非谣言数据集在此模型训练的迭代当中,影响力微乎其微,使得训练集准确率高达 99.58%。因此过拟合的数据对于目标效果来说是无效的。



(a) 测试集损失率



(b) 测试集准确率

图 3 迭代次数为 10 的测试集损失率与准确率

Fig. 3 Loss rate and accuracy for a test set with 10 iterations

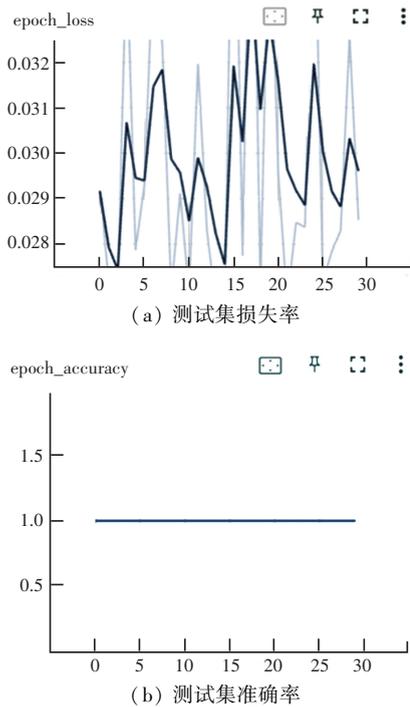


图4 迭代次数为30的测试集损失率与准确率

Fig. 4 Loss rate and accuracy for a test set with 30 iterations

第三组实验数据集比例(谣言数量:非谣言数量)为6:1,迭代次数为10的测试集损失率与准确率如图5所示。可见在提高比例后准确率和损失率明显到了一个合适的范畴,训练集的准确率达到了0.625 2,损失率为0.327 2,比例接近会使准确率提高,但是比例接近并不是唯一影响因素。

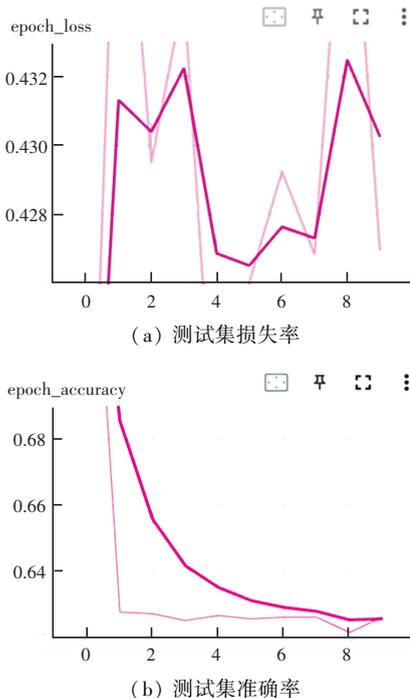


图5 迭代次数为10的测试集损失率与准确率

Fig. 5 Loss rate and accuracy for a test set with 10 iterations

第四组实验数据集比例(谣言数量:非谣言数量)为5:3,迭代次数为20的测试集损失率与准确率如图6所示,第五组实验数据集比例(谣言数量:非谣言数量)为1:1,迭代次数为20的测试集损失率与准确率如图7所示。对比第四、五组实验,可见第四组中比例更加接近但测试值的损失率却高达2.16,这是因为数据只有8 000条,虽然接近但是数据太少也造成了过拟合现象。当谣言数据与非谣言数据之间的比例达到1:1时,测试集的准确率与损失率都在目标结果内,最终也将采用这个结果作为本实验的结论,以及后续小程序的系统内核数据。

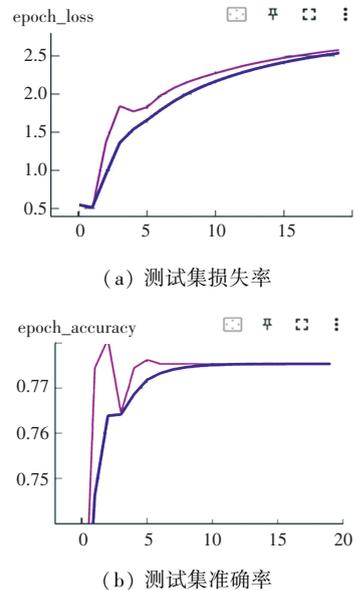


图6 迭代次数为20的测试集损失率与准确率

Fig. 6 Loss rate and accuracy for a test set with 20 iterations

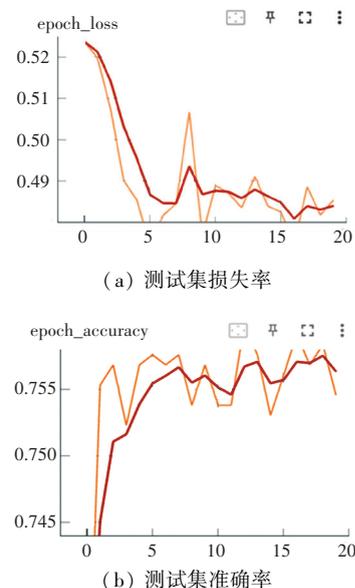


图7 迭代次数为20的测试集损失率与准确率

Fig. 7 Loss rate and accuracy for a test set with 20 iterations

### 4.1.3 数据集的数量级以及迭代次数对测试集和训练集准确率的影响

测试集准确率与数据集数量级关系如图 8 所示,可以看出当数据集中非谣言数据量远远大于谣言数据量时(例如 300 : 1, 100 : 1, 30 : 1),测试集准确率较高,超过了 0.995。但是随着谣言数据量逐渐增加,数据集数量级逐渐减小(15 : 1, 6 : 1, 5 : 3, 1 : 1),测试集准确率也逐渐下降,最低的甚至只有 0.625 2,表明对于谣言检测模型来说,数据集中非谣言数据量相对较大时,模型的准确率会比较高,但是当谣言数据量增加时,模型的准确率会下降。因此,在构建谣言检测模型时,需要注意数据集的平衡,尽量保持谣言和非谣言数据量的平衡,以提高模型的准确率。



图 8 测试集准确率与数据集数量级关系折线图

Fig. 8 Line chart of the relationship between the test set accuracy and the order of magnitude of the dataset

迭代次数与测试集准确率关系如图 9 所示。根据折线图可以看出,随着迭代次数的增加,测试集准确率也逐渐提高。从迭代次数 10 开始,准确率为 0.625 2,随着迭代次数增加到 30,准确率提高到了 0.995 8,表明 Text CNN 模型在更多的迭代训练后,可以提高测试集的准确率。然而,随着迭代次数的增加,模型可能会出现过拟合的情况,因此在选择迭代次数时需要进行适当的调参和验证集验证来避免过拟合。

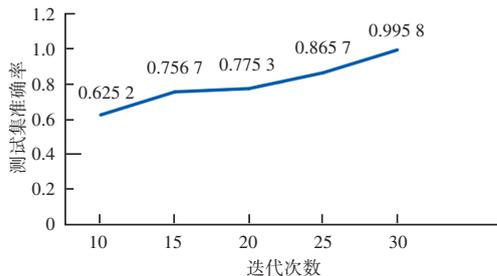


图 9 迭代次数与测试集准确率关系折线图

Fig. 9 Line chart of the relationship between the number of iterations and the test set accuracy

## 4.2 基于上下文相关词向量的谣言判别模型的实验

### 4.2.1 实验设置

由于 Text CNN 和 BERT 的融合能够提高模型判

别谣言的性能,实验记录 Text CNN 卷积核个数和 BERT 嵌入大小改变时模型判别谣言的准确率,分别记录开发集和测试集的准确率,实验中使用的数据集总量为 9 190 条,谣言和非谣言的数据集比例为 1 : 1。

### 4.2.2 实验结果

实验采用了控制变量的方法,以深入探究 Text CNN 卷积核个数和 BERT 嵌入大小这两个关键变量对谣言判别模型性能的影响。对比了不同参数设置下的性能差异,并以开发集准确率作为核心评估指标来衡量各实验组的表现。不同 Text CNN 卷积核个数和 BERT 嵌入大小设置下实验结果见表 2。

表 2 不同 Text CNN 卷积核个数和 BERT 嵌入大小设置下实验结果  
Table 2 Experimental results with different Text CNN convolutional kernel numbers and BERT embedding size settings

实验组	TextCNN 卷积核个数	BERT 嵌入大小	开发集准确率
1	5	512	0.960
2	11	512	0.962
3	5	128	0.932

开发集准确率与 Text CNN 卷积核大小的关系如图 10 所示,可见在卷积核个数为 9 个时,准确率相比之前有一点下降,11 到 25 个之间准确率并没有明显提高,35 个时得到较高的准确率,但是总体上可以得出 Text CNN 卷积核个数的适当增加能够提高模型开发集准确率的结论。

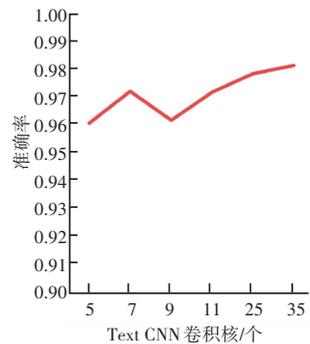


图 10 开发集准确率与 Text CNN 卷积核大小的关系

Fig. 10 Relationship between the development set accuracy and the CNN kernel size

表 2 的实验组 1、2 和图 10 中的实验将模型输出维度设置为 512,学习率为 5e-06,最大训练轮数为 100,分词器最大长度为 512,批处理大小为 32。Text CNN 卷积核较多时,开发集的谣言判别准确率较高。

多次实验记录准确率,得到谣言判别模型的准确率与 BERT 嵌入维度大小之间的关系折线统计图如图 11 所示。由图 11 可知,BERT 嵌入维度为 32 时,准确率值最低;BERT 嵌入维度为 512 时,准确

率值最大;BERT 嵌入维度为 374 时,其准确率几乎与取值为 256 时的一样。总体上,随着 BERT 嵌入维度的增大,模型开发集的准确率也在提高。

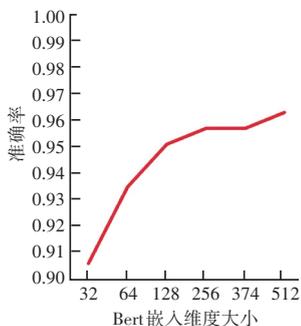


图 11 开发集准确率与 BERT 嵌入维度的大小关系

Fig. 11 Relationship between the development set accuracy and the BERT embedding dimension size

#### 4.2.3 实验结论

实验结果表明,增加 Text CNN 卷积核数和 BERT 嵌入维度可以提高模型性能。当 Text CNN 卷积核数适当增加和 BERT 嵌入维度较大时,模型的融合策略在文本分类任务中表现出更高的性能,在开发集上获得较高的准确率。

## 5 小程序设计及效果展示

### 5.1 前端设计

前端页面设计使用基于 JavaScript 的微信小程序框架。用 wxml 描述微信小程序的页面结构;用 wxss 定义微信小程序的页面样式;使用 API 方法以及网络请求来实现页面的数据处理和与后端服务器的数据交互;使用 flask 框架作为后端框架,用于接收并处理前端发送的请求。

### 5.2 小程序效果展示

用户输入一段文字,确认后程序调用谣言判别程序进行判断,判别完成后显示预测结果和预测是谣言的概率,如图 12 和图 13 所示。



图 12 小程序搜索界面  
Fig. 12 Mini-program search interface



图 13 小程序搜索结果界面  
Fig. 13 Mini-program search results interface

## 6 结束语

本文设计和实现一个谣言判别小程序。经过对网络新闻信息的爬取、数据清洗与预处理,成功收集了一个 3 万条数据的数据集;在此基础上,构建了两个模型:Text CNN 模型和创新的 BERT+Text CNN+KG 模型,两个模型经过不同参数的训练实验,达到了 95% 的准确率。谣言判别小程序与本文模型结合,能够实时检测谣言,并输出其是谣言的概率。展望未来,本文可在模型方面进一步改进,例如尝试在谣言判别模型中引入注意力机制和图神经网络等,以进一步提升谣言判别准确度。

## 参考文献

- [1] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20<sup>th</sup> International Conference on World Wide Web. 2011: 675-684.
- [2] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [3] 吕胜. 基于评论情感和 CNN-LSTM 结合模型的微博谣言检测 [D]. 乌鲁木齐:新疆师范大学,2019.
- [4] GAO W, MA J, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]//Proceedings of IJCAI International Joint Conference on Artificial Intelligence. 2016: 3818-3824.
- [5] SHU K, MAHUESWARAN D, WANG S, et al. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media [J]. Big Data, 2020, 8(3): 171-188.
- [6] SPÄRCK J K. A statistical interpretation of term specificity and its application in retrieval [J]. Journal of Documentation, 2004, 60(5): 493-502.
- [7] MÜLLER H. Learning to predict sequence: A recurrent neural network approach [C]//Proceedings of the IEEE International Conference on Neural Networks. Washington DC, USA: IEEE, 1989: 896-901.
- [8] ELMAN J L. Finding structure in time [J]. Cognitive Science, 1990, 14(2): 179-211.
- [9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [10] 罗嗣卿,王佳玉,李冰珂. 改进的组合优化决策树谣言判别方法研究 [J]. 计算机仿真, 2018, 35(2): 219-223.
- [11] 陈铜. 基于三元特征的深度注意力谣言鉴别模型 [D]. 武汉:武汉大学, 2020.
- [12] 陈香焯. 基于图知识嵌入的谣言检测技术研究 [D]. 厦门:厦门大学, 2020.
- [13] 李亦轩. 基于文本多特征融合的谣言检测模型研究 [D]. 成都:西华大学, 2021.