

罗嘉¹ 王乐豪² 涂姗姗¹ 宋鸽¹ 韩莹²

基于 LSTM-BLS 的突发气象灾害事件中 公众情感倾向分析

摘要

近年来长短期记忆网络(LSTM)在文本情感倾向分析方面显示出一定优势,但 LSTM 提取特征时存在语义不完整、精度不高等问题.研究者往往通过引入卷积神经网络(CNN)来弥补这一缺陷,但仍然未考虑到单词之间的句法依存问题.本文将增量学习算法为核心的宽度学习(BLS)与 LSTM 相融合,提出了 LSTM-BLS 文本情感分析模型,并以 2020 断崖式降温事件为例,对突发气象灾害发生时公众情感倾向进行分析.结果表明:与基线模型 K-means 和支持向量机(SVM)相比,LSTM-BLS 模型精度分别提高 17.23 和 13.46 个百分点;与已有深度模型 LSTM、CNN-LSTM 相比,本文模型精度分别提高 7.13 和 4.17 个百分点.

关键词

气象灾害;宽度学习;长短期记忆网络;情感分析

中图分类号 TP391.1;G206.3

文献标志码 A

收稿日期 2021-05-19

资助项目 南方海洋科学与工程广东省实验室(珠海)项目(SML2020SP007);国家自然科学基金(62076136)

作者简介

罗嘉,女,硕士,研究方向为灾害应急建模与分析,气象数据分析.jeeaaan@qq.com

韩莹(通信作者),女,博士,副教授,研究方向为大数据处理方法及其应用.hanyingcs@163.com

1 湖北省公众气象服务中心,武汉,430074

2 南京信息工程大学 自动化学院,南京,210044

0 引言

我国作为气象灾害^[1-2]发生频率较高的国家之一,各种极端天气的频繁发生,给人们带来了严重的经济损失,甚至会诱发灾难造成人员伤亡.气象灾害的发生不仅是对国家应急治理体制的重大考验,影响人们的日常生活,还会在网络上引起社会舆论的爆发^[3],使人产生负面焦虑的情绪^[4].对于灾害引起的舆论^[5],若不加以干预引导,容易在网络上演变成新的突发事件,加剧危机事件本身的负面影响.

网络用户数量在近几年呈几何趋势增长,大大提升了舆论的产生和传播速度,网络舆论本身带有的情感指向性和其内在包含的应用价值也逐渐成为专家们关注的热点.传统的机器学习算法诸如支持向量机(Support Vector Machines, SVM)^[6]、K-means 算法^[7]等方法已被提出用于英语和汉语的情感极性分析,且取得了很大的进展.但是面对大量数据时,传统算法的训练力不从心.深度学习的发展减轻了机器学习模型手动提取特征的负担.长短期记忆网络(Long Short-Term Memory, LSTM)^[8]具有长时记忆功能并且实现起来简单的优势,解决了训练过程中存在的梯度消失和梯度爆炸的问题.文献[9]将 LSTM 模型与 SVM 模型在相同数据集上进行对比,在情感分类方面 LSTM 模型的准确率远高于基线模型.舆情文本分析^[10-11]是通过度量向量空间中单词向量之间的关系进行的,因此单词嵌入的质量直接影响到分类结果.虽然 LSTM 能够从训练数据中挖掘出更抽象的特征,使其具有很好的泛化能力,但它的拟合能力并不理想,而且 LSTM 提取特征时存在语义不完整、精度不高等问题.

研究者通过引入卷积神经网络(Convolutional Neural Networks, CNN)来弥补上述缺陷,CNN-LSTM 模型^[12]的提出使得处理高维数据更加轻松、精度更高.自然语言包含结合词和短语的句法特性使得底层模型不满足应用需求.短文包含的上下文信息往往有限,对其进行情感分类具有一定的挑战性.文献[13]将 10 层 CNN 和 10 层 LSTM 结合起来,使用不同的超参数和不同的预训练策略训练,其产生的模型比单独使用模型的历史最高精度更胜一筹.许杰等^[14]考虑到 CNN 并行计算能力强的优点而将其作为特征提取器,提取到的高层次特征输入到 LSTM 中得到最终结果.较之以往模型,该模型能够在提取局

部特征的同时获取句子的时态语义,提高了情感文本分类的精度.但 CNN-LSTM 模型存在性能过度依赖于标记数据的数量和质量,且未考虑到单词之间的句法依存的问题,情感文本分类的精度仍有提升的空间.

注意到宽度学习(Broad Learning System, BLS)作为一种简单的新型快速增量学习神经网络^[15],是基于随机向量函数连接网络 RVFL(Random Vector Functional Link Network),将原始的输入先通过特征节点学习稀疏的映射特征,然后经由增强节点非线性扩展得到增强特征,并联两种特征表达作为最后的总输入送到输出层进行分类识别,由此可以从训练数据中学习到重要的特征,对训练数据达到高度拟合.

本文爬取了 2020 年末断崖式降温舆情文本并对其进行数据预处理之后,发现文本中大多是短句文本,其包含的上下文信息有限,因此底层模型不满足应用需求.将数据集使用 LSTM 模型进行训练,得到的实验结果在拟合度方面没有达到预期的效果,精确度不高.

考虑到深度学习模型能够从训练数据中挖掘出更抽象的特征,这使得它具有很好的泛化能力,而 BLS 具有能够从训练数据中学习到重要的特征,对训练数据达到高度拟合的特点,因此使用 LSTM 作为本文模型的特征提取层,将文本进行初步分类,再将初步分类的文本输入到 BLS 层中进一步提取重要特征,最终获取高精度分类结果.本文提出的 LSTM-BLS 模型,将不同级别的特征层信息进行融合使用,既有低级纹理信息又有高级语义信息,弥补了 LSTM、CNN-LSTM 模型在舆情文本分析方面提取特征时存在语义不完整、精度不高和未考虑到单词之间的句法依存等问题.实验结果表明,本文提出的模型较之于 K-means、SVM 模型精度分别提高了 17.23 和 13.46 个百分点,与 LSTM 与 CNN-LSTM 模型相比精度分别提高了 7.13 和 4.17 个百分点.文中模型在经过数据集测试后取得良好的效果,并且不依赖特定领域的自然语言库,也可以应用在其他中文文本分析验证上,具有广泛的适用性.

1 模型

1.1 LSTM

LSTM 的核心概念在于细胞状态以及“门”结构.图 1 展示了 LSTM 的基本架构,其包含的三个

门:输入门、遗忘门和输出门.通过三个门的函数和细胞状态,LSTM 可以捕获短期和长期时间序列中的复杂相关性.其中, x_t 是输入向量, i_t 是时间步长 t 中的输入状态, f_t 是时间步长 t 中的遗忘状态, o_t 是时间步长 t 中的输出状态, $h_t(h_{t-1})$ 是时间步长 $t(t-1)$ 中的隐藏状态, $c_t(c_{t-1})$ 是时间步长 $t(t-1)$ 中的单元状态.以 \tanh 和 sigmoid 激活函数 σ 的形式在三个门的顶部添加非线性.

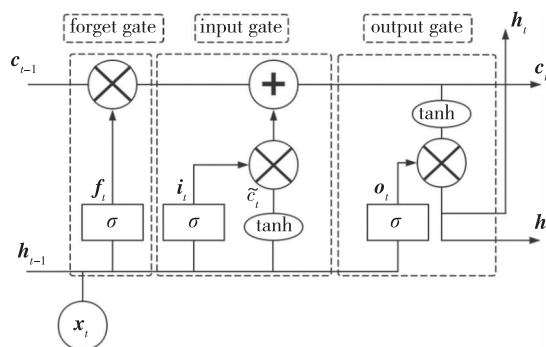


图 1 LSTM 的基本架构

Fig. 1 Basic architecture of LSTM

LSTM 的数学原理在式(1)—(6)中给出:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \quad (1)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \quad (2)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \quad (3)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (4)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t, \quad (5)$$

$$h_t = o_t \otimes \tanh(c_t), \quad (6)$$

其中, $W_{xf}, W_{hf}, W_{xi}, W_{hi}, W_{xo}, W_{ho}, W_{xc}, W_{hc}$ 代表了输入向量与输入门、输出门、记忆单元之间对应的权向量, b_f, b_i, b_o, b_c 是偏置变量, \otimes 是矩阵的 Hadamard 积.

1.2 BLS

BLS 网络结构如图 2 所示.

假设输入数据集 X 包含 N 个样本,每个样本有 M 个维度, Y 是属于 $\mathbf{R}^{N \times C}$ 的输出矩阵.对于 n 个特征通过式(7)映射后生成 k 个节点:

$$Z_i = \phi(XW_{ei} + \beta_{ei}), i = 1, \dots, n. \quad (7)$$

将所有特征节点表示为 $Z_n \equiv [Z_1, \dots, Z_n]$,第 m 组增强节点表示为

$$H_m \equiv \xi(Z^n W_{hj} + \beta_{hj}), j = 1, \dots, m, \quad (8)$$

其中 W_{ei}, W_{hj} 和 β_{ei}, β_{hj} 分别是随机生成的连接权重和偏置值.

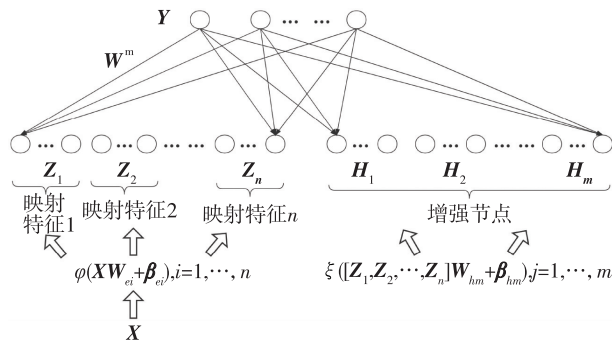


图2 BLS网络结构

Fig. 2 Network structure of BLS

因此,BLS模型可以表示为

$$Y = [Z_1, \dots, Z_n | \xi(Z^n W_{h1} + \beta_{h1}), \dots, \xi(Z^n W_{hm} + \beta_{hm})] W^m = [Z_1, \dots, Z_n | H_1, \dots, H_m] W^m = [Z^n | H^m] W^m \quad (9)$$

1.3 LSTM-BLS 舆情文本情感分类模块

针对舆情文本的歧义性或多义性的问题,考虑到深度学习能够对每个分量进行数据特征识别与特征提取,从而实现分量的单步向前预测,达到获取最终分类的效果的特点,本文在上述基础模型上通过融合 BLS 提出了一种新的的文本情感分析模型.该模型整体结构如图 3 所示.

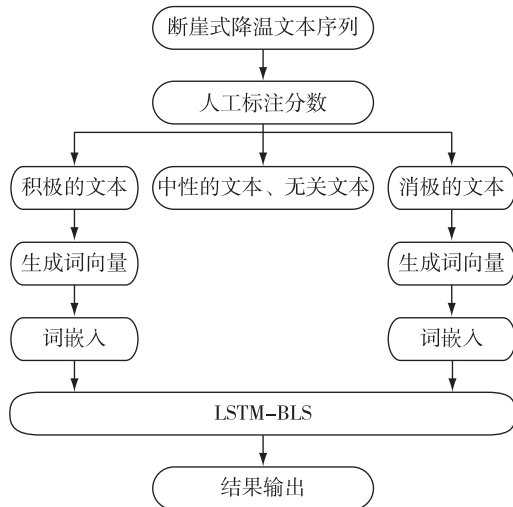


图3 LSTM-BLS的文本情感分析流程

Fig. 3 Flow chart of text sentiment analysis based on LSTM-BLS

LSTM-BLS模型(图4)的主体是3个拼接的LSTM层、2个全连接层(Dense)和1个BLS模型通道,3层LSTM通道主要用来提取句子特征,而加入的BLS层可以和别的机器学习算法灵活地结合,即

利用 LSTM 提取到的特征来训练,并且,其加入的增量学习算法,允许在网络结构中加入新的节点时,以很小的计算开销来更新网络权重.

1)第1层是输入层(input layer),将爬取到的文本进行人工数据集标注后生成的文本导入程序.

2)第2层是嵌入层(embedding layer),作用是把输入每个词语映射成一个向量表示.

3)第3~5层是LSTM层,主要是为了提取句子的特征.本文词向量维度选取300维,选择的过滤器分别为256、128、64个,训练次数(epoch)=50,并在LSTM的每一层都加入了dropout机制,该机制可以在训练过程中通过让部分神经元停止工作,而达到防止过拟合的效果.

4)第6层是Dense层,将LSTM层输出端融合的特征作为全连接层的输入,应用Sigmoid激活函数后的值是介于0到1之间的浮点数,表示概率或者置信度,并加入L2正则化,防止过拟合.

5)第7层为BLS层,利用前面提取到的特征进行计算,调用Softmax进行分类,其中每个映射特征点的个数 $N_1=10$ 、映射特征个数 $N_2=30$ 、增强节点个数 $N_3=280$ 以及正则化参数 $C=0.0001$.

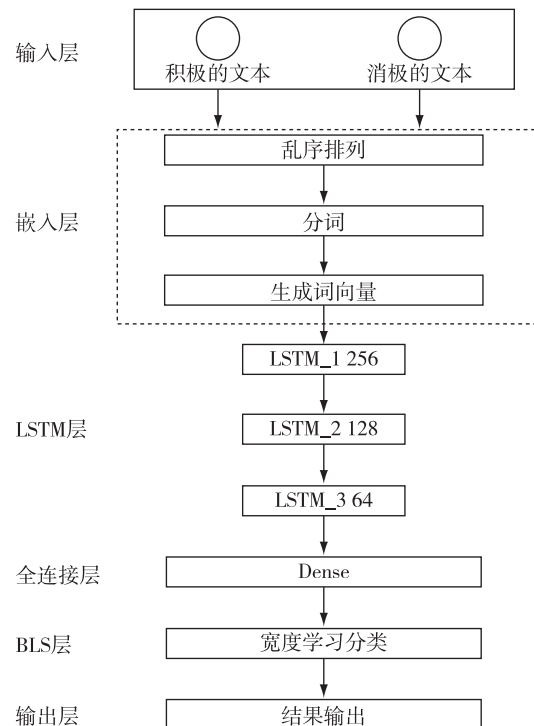


图4 LSTM-BLS模型

Fig. 4 LSTM-BLS public opinion analysis model

2 实证分析

2.1 数据来源

本文主要以 2020 年末两次断崖式降温舆情数据为例.2020 年 12 月 28—30 日我国中东部大部地区迎来大风降温天气,局部地区遭遇了 16 ℃ 以上的断崖式降温.中央气象台迅速发布了最高级别的寒潮橙色预警.此次寒潮影响范围广,以及其伴随而来的剧烈降温、长时间持续性大风,给各行各业,甚至人们日常的穿衣出行都带来了极大的影响,并迅速在网络上掀起了舆论风暴.本文基于 Python 的 Scrapy 模块,爬取微博大 V 及相关媒体共计 37 852 条数据,包括发表评论的用户名称、评论内容、发表时间、转发量、点赞量等多极化数据信息.

在对文本进行训练前,先对爬取到的信息进行了一系列的数据清洗工作,包括繁体字简化、删除垃圾广告、无效评论等,最终得到 32 358 条有效数据.预处理包括以下内容:

- 1) 去除含 HTML 标签的内容;
- 2) 删除纯标点评论;
- 3) 去除表情评论;
- 4) 单词词形还原;
- 5) 删除停用词.

2.2 超参数设置

社交网络的快速发展使得媒体信息在社交平台的传播更加迅速.灾害相关信息会在事件发生短期内迅速传播发酵,在网络上引起舆论风波.舆情结果往往会形成两极分化,积极的舆论引导事件朝好的方向发展,而消极的舆论将会在网上引发二次灾害.因此,舆情文本分析的准确性,大大影响着舆情引导的走向.有鉴于此,本文将数据集分为两个类别:积极的、消极的.

将分类结果生成词云可以直观地反映正面和负面评论中出现频率最高的词语,可以看到图 5、6 中,部分词汇可以清楚地反映用户的积极和消极情绪.然而,网络评论中时常包含褒义贬用以及贬义褒用的词汇,此类词汇经常包含在具有完全相反含义的评论中.因此,仅仅根据是否出现代表积极或者消极情绪的词语而对评论进行简单的分类已经达不到预想的结果时,便需要深度学习技术来分析词语之间的关系,获取语义从而进行分类.

本文使用人工标注的微博语料数据集正向和负向评论各 14 000 条,从中挑取正负向评论各 8 000 条作为训练集,剩下的作为测试集进行测试.实验需



图 5 积极的词汇

Fig. 5 Examples of positive word



图 6 消极的词汇

Fig. 6 Examples of negative word

要通过词袋模型将语料库更改词向量,之后将训练好的词向量输入到搭建好的神经网络中.

在本文搭建的舆情文本分析模型中,输入的词向量的效果很大程度上影响着最终训练结果,多次实验结果表明,将词向量进行预训练可以有效地提高模型的准确率.经过不断调试模型参数,本文最终采用词向量维度为 300.数据训练前,利用 Jieba 分词工具将原文本进行分词、去除停用词后,将得到的词语转换为词向量并进行预训练,得到的结果将作为本文模型的输入.

为了得到更加丰富的情感特征信息,本文在 LSTM-BLS 模型上进行了一系列实验.为了防止过拟合现象,本文使用了 dropout 机制和 L2 正则化.详细超参数设置如表 1 所示.

表 1 模型超参数设置

Table 1 Model super parameter setting

方法	超参数描述	值
LSTM	Batch_size	256
	Max_sentence_length	20
	Dropout	0.4
	Recurrent_dropout	0.4
	Num_classes	2
	Epochs	50
	Filters	216
	Kernel_size	3
	Embedding_dims	ndims
	Dense	Timesteps
Layers		2
BLS	Number of neurons	{10,1}
	Activation	Relu
	N_1	10
	N_2	30
	N_3	280
	C	0.0001

2.3 结果讨论

为了验证本文提出的模型的有效性,本文使用准确率(A)、召回率(R)、F1 值(F_1)作为评估指标.

准确率计算公式:

$$A = \frac{\eta_{TP} + \eta_{TN}}{\eta_{TP} + \eta_{TN} + \eta_{FP} + \eta_{FN}}, \quad (10)$$

召回率计算公式:

$$R = \frac{\eta_{TP}}{\eta_{TP} + \eta_{FN}}, \quad (11)$$

F1 值计算公式:

$$F_1 = \frac{2 \times A \times R}{A + R}, \quad (12)$$

其中: η_{TP} 为正确的匹配数目; η_{FP} 为匹配不正确的数目; η_{FN} 为没有找到正确匹配的数目; η_{TN} 为正确的非匹配数目.

将本文使用的 LSTM-BLS 模型分别与另外的 4 种模型方法在准确率上进行实验对比:

1) K-means. K-means 算法是无监督的聚类算法,因其实现简单、聚类效果好,被广泛应用.本文设置算法模型最大迭代次数为 300, k 值为 2, 容忍度为

0.0001.

2) SVM. SVM 是机器学习中最好的现成的分类器,可以不加修改直接使用,并且能够得到较低的错误率.本次试验将句子中的单词转换为词向量,再将这些向量进行加权平均作为模型的输入进行分类.

3) LSTM. 本文使用的是三层 LSTM 的单通道模型,将词向量作为输入,最后通过全连接层和 Softmax 层输出分类结果.

4) CNN-LSTM. 该模型由 CNN 和 LSTM 组合而成,在上述 LSTM 模型的基础上,使用 CNN 提取特征,作为 LSTM 模型的输入.

5) LSTM-BLS. 本文提出的模型,是在上述三层 LSTM 的基础上拼接了 BLS 层.将爬取到的微博评论经过预处理,分词后经过词嵌入生成词向量,作为本文模型的输入得到最终准确率结果.

实验结果对比如表 2、3 所示.

表 2 基线模型在断崖式降温数据集上的性能对比

Table 2 Performance comparison on cold wave dataset between baseline models

方法	准确率	召回率	F1 值
K-means	80.19	76.38	78.24
SVM	83.96	79.42	81.63

表 3 本文模型和其他深度学习模型在断崖式降温数据集上的性能对比

Table 3 Performance comparison on cold wave dataset between the proposed model and deep learning models

方法	准确率	召回率	F1 值
LSTM	90.29	85.15	87.65
CNN-LSTM	93.25	90.08	91.64
LSTM-BLS	97.42	92.82	95.07

由表 2 和表 3 可知,在断崖式降温数据集上, LSTM-BLS 模型的网络舆情分析准确率相比基础的机器学习模型 K-means 和 SVM,分别提高了 17.23 和 13.46 个百分点,比 LSTM 模型提高 7.13 个百分点,比 CNN-LSTM 复合模型提高 4.17 个百分点.准确率和召回率在实践中会出现矛盾的情况,而 F1 值作为准确率和召回率的调和平均值,往往成为实验最有效的综合评价指标.由表 2、3 可知,本文提出的新模型在综合评价方面表现最优,效果最好.综上, LSTM-BLS 模型在文本情感分析方面具有良好的性能,可以更加准确地对舆情文本进行正负向情感分类,弥补了现有的机器学习模型面对大量数据集时训练吃力的缺陷,解决了深度学习模型拟合能力欠

佳、精度不高的问题。

3 结论

气象舆情分析的主要任务是为了更加精准地掌握网络舆情动态,避免灾害事件在网络舆情方面形成二次灾害。本文提出的 LSTM-BLS 算法模型一方面考虑到单词之间的句法依存问题,另一方面在此次舆情文本分析经过爬取到的数据集验证测试后,准确率、F1 值均超过 K-means、SVM、LSTM 和 CNN-LSTM 模型,在短文本情感分类方面效果优异,经公开数据集验证也同样具有适用性。将宽度学习与深度学习相结合,对纷繁复杂的网络舆情文本进行准确的分类,对短文本数据信息的有效分析和挖掘,提高了对网络舆情的监管能力,有利于开展后续引导工作。

参考文献

References

- [1] 唐黎标.我国重大气象灾害的应急管理研究[J].防灾博览,2021(1):38-41
TANG Libiao. Research on emergency management of major meteorological disasters in China [J]. Overview of Disaster Prevention, 2021 (1): 38-41
- [2] 路永玲,陶风波,周志成,等.气象灾害对江苏电网设备的影响及防御分析[J].南京信息工程大学学报(自然科学版),2015,7(5):469-474
LU Yongling, TAO Fengbo, ZHOU Zhicheng, et al. Analysis of influence and prevention of meteorological disasters on Jiangsu's power grid equipment [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2015, 7 (5): 469-474
- [3] 金志威,曹娟,王博,等.融合多模态特征的社会多媒体谣言检测技术研究[J].南京信息工程大学学报(自然科学版),2017,9(6):583-592
JIN Zhiwei, CAO Juan, WANG Bo, et al. Rumor detection on social media with multimodal feature fusion [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2017, 9 (6): 583-592
- [4] Nakov P, Ritter A, Rosenthal S, et al. SemEval-2016 task 4: sentiment analysis in Twitter [C] // Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016: 1-18
- [5] 张岚,艾文文,罗晓春,等.气象灾害下网络信息传播模式研究:以台风为例[J].气象科学,2020,40(6):868-874
ZHANG Lan, AI Wenwen, LUO Xiaochun, et al. Research on communication of network information under meteorological disaster: case of typhoon [J]. Journal of the Meteorological Sciences, 2020, 40 (6): 868-874
- [6] Li Y, Wang J, Wang S G, et al. Local dense mixed region cutting+global rebalancing: a method for imbalanced text sentiment classification [J]. International Journal of Machine Learning and Cybernetics, 2019, 10(7): 1805-1820
- [7] 贾燕花,徐蔚鸿.K-means 聚类和支持向量机结合的文本分类研究[J].计算机工程与应用,2010,46(22):172-174
JIA Yanhua, XU Weihong. Research on document classification based on k-means and support vector machine [J]. Computer Engineering and Applications, 2010, 46(22): 172-174
- [8] 陈凌,宋衍欣.基于公众情绪上下文的 LSTM 情感分析研究:以台风“利奇马”为例[J].现代情报,2020,40(6):98-105
CHEN Ling, SONG Yanxin. LSTM sentiment analysis based on the context of public emotion: a case study of super typhoon “Lekima” [J]. Modern Information, 2020, 40(6): 98-105
- [9] 郭明强,郭佳明,辛伟彬. Word2Vec+LSTM 多类别情感分类算法优化[J].计算机系统应用,2020,29(1):130-136
WU Mingqiang, WU Jiaming, XIN Weibin. Optimization of Word2Vec and LSTM multi-category sentiment classification algorithm [J]. Computer Systems & Applications, 2020, 29(1): 130-136
- [10] 邢鹏飞,李鑫鑫.重大疫情防控中网络舆情形成机制及引导策略研究:基于新冠肺炎疫情期间网络舆情文本的质性分析[J].情报杂志,2020,39(7):67-74,158
XING Pengfei, LI Xinxin. Research on formation mechanism of online public opinions and guiding strategy in major epidemic prevention and control: qualitative analysis based on online public opinion texts during the outbreak of COVID-19 [J]. Journal of Intelligence, 2020, 39(7): 67-74, 158
- [11] 习海旭,蒋红芬,程志凡,等.特定事件下网络舆情的情感分析与可视化方法[J].情报理论与实践,2020,43(9):132-136,143
XI Haixu, JIANG Hongfen, CHENG Zhifan, et al. Emotional analysis and visualization of Internet public opinion under specific events [J]. Information Studies (Theory & Application), 2020, 43(9): 132-136, 143
- [12] Alotaibi F M, Asghar M Z, Ahmad S. A hybrid CNN-LSTM model for psychopathic class detection from tweeter users [J]. Cognitive Computation, 2021, 13(3): 709-723
- [13] Cliche M. BB_twtr at SemEval-2017 task 4: twitter sentiment analysis with CNNs and LSTMs [J]. arXiv e-print, 2017, arXiv:1704.06125
- [14] 许杰,朱超越,毕成,等.基于 LSTM 和 CNN 文本分类模型实现配网作业智能风险评估[J].现代信息科技,2020,4(18):153-156,161
XU Jie, ZHU Chaoyue, BI Cheng, et al. Intelligent risk assessment of distribution network operation based on LSTM and CNN text classification model [J]. Modern Information Technology, 2020, 4 (18): 153-156, 161
- [15] 任长娥,袁超,孙彦丽,等.宽度学习系统研究进展[J/OL].计算机应用研究:1-12[2021-06-27].https://doi.org/10.19734/j.issn.1001-3695.2020.11.0348
REN Chang'e, YUAN Chao, SUN Yanli, et al. Research

progress of breadth learning system[J].Computer Application Research; 1-12 [2021-06-27]. https://doi.org/

10.19734/j.issn.1001-3695.2020.11.0348

Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS

LUO Jia¹ WANG Lehao² TU Shanshan¹ SONG Ge¹ HAN Ying²

1 Hubei Public Meteorological Service Center, Wuhan 430074

2 School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract Meteorological disasters will not only bring huge economic losses, but also cause the outbreak of public opinion and even social panic. The Internet era brings opportunities as well as challenges to the governance of public opinion. In recent years, Long Short-Term Memory network (LSTM) has shown advantages in text sentiment analysis, yet it has problems such as incomplete semantics and low accuracy in feature extracting. Convolutional Neural Network (CNN) has been introduced to make up for this shortcoming, but it still cannot consider the syntactic dependence between words. Here, we combine LSTM and Broad Learning System (BLS) with incremental learning algorithm as its core, and thus propose an LSTM-BLS text sentiment analysis model. Then the 2020 cold wave event with gale process and great temperature drop in central and eastern China is taken as an example to analyze the public sentiment tendency when sudden meteorological disasters occur. The results show that the proposed LSTM-BLS reaches a high accuracy of 97.42%, which is 17.23 and 13.46 percentage points higher than the baseline models of K-means and Support Vector Machine (SVM), respectively, and 7.13 and 4.17 percentage points higher than the existing deep learning models of LSTM and CNN-LSTM, respectively.

Key words meteorological disaster; broad learning system (BLS); long short-term memory (LSTM) network; sentiment analysis