

文章编号: 2095-2163(2022)08-0033-10

中图分类号: TP391

文献标志码: A

基于神经网络的舆情情感分析研究热点与趋势 ——基于 CiteSpace 的可视化分析

谭坤彦, 杨孔雨

(北京信息科技大学 信息管理学院, 北京 100192)

摘要: 本文以中国知网数据库收录的 877 篇文献和国外 Web of Science 核心合集数据库收录的 423 篇文献作为研究对象, 运用 CiteSpace 软件构建了研究热点和研究前沿主题的可视化知识图谱, 对国内外基于神经网络的网络舆情情感分析研究现状进行分析, 并且对其差异进行对比。研究发现, 国内外对该领域日趋关注但各有侧重, 国内作者及机构存在相互合作较少, 学术交流不足, 研究主题较集中等问题。国内研究热点为情感分类、词向量、深度学习等; 研究前沿趋势主要为特征融合表示方法、预训练模型的探索和词嵌入层的设计, 为研究发展提供理论导向。

关键词: 神经网络; 网络舆情; 情感分析; CiteSpace

Research hotspot and trend of public opinion emotion analysis based on neural network ——visual analysis based on CiteSpace

TAN Kunyan, YANG Kongyu

(School of Information Management, Beijing Information Science and Technology University, Beijing 100192, China)

[Abstract] Taking 877 documents included in the domestic HowNet database and 423 documents included in the foreign web of science core collection database as the research object, this paper constructs a visual knowledge map of research hotspots and cutting-edge topics by using CiteSpace software, analyzes the research status of network public opinion and emotion analysis based on neural network at home and abroad, and compares their differences. It is found that more and more attention has been paid to this field at home and abroad, but each has its own emphasis. Domestic authors and institutions have problems such as less mutual cooperation, insufficient academic exchanges, and more concentrated research topics. Domestic research focuses on emotion classification, word vector, deep learning and so on; The research frontier trends mainly include the representation method of feature fusion, the exploration of pre training model and the design of word embedding layer, which provides a theoretical guidance for research and development.

[Key words] neural network; network public opinion; sentiment analysis; CiteSpace

0 引言

随着人工智能技术的迅速发展和各类社交平台的出现, 互联网用户可以随时随地在网络平台上传播信息。大量的信息往往蕴含着用户的情感倾向, 获取这些信息中的情感倾向或观点就是情感分析。对于决策制定者来说, 可以理解民意, 掌握舆论, 判断其走向^[1]。早期的情感分析主要基于情感词典和传统机器学习, 随着机器学习技术的飞速进步, 深度学习已经成为情感分析的主流。在 Mikolov 等人(2013)^[2]首先提出了 Word2Vec 工具中的模型后, 文本可以表示为词向量矩阵, 为引入各种神经网络模型打下了基础。基于神经网络的舆情情感分析研究具有重要的意义与价值, 本文利用知识图谱软

件 CiteSpace, 运用文献计量法, 通过将结果可视化的方式, 从发文量趋势、作者、机构、关键词的多元角度进行研究, 分析国内外结合神经网络技术进行舆情情感分析的文献发展动态、研究热点、演变过程、前沿研究方向等, 期寄能够给相关领域的研究者带来一定帮助。

1 分析方法与数据来源

CiteSpace 是基于 Java 开发的一款信息可视化软件, 由美国 Drexel University 计算机与情报学教授陈超美研发, 基于共引(co-citation)分析理论和寻径(PathFinder)网络算法等, 对特定领域文献计量, 以探寻出学科领域演化的关键路径及其知识转折点, 并在相关领域中, 查究到具有开创性、标志性、关

基金项目: 北京市社会科学基金重点项目(15ZHA004); 北京信息科技大学“促进高校分类发展-学科建设与研究生教育”项目资助。

作者简介: 谭坤彦(1997-), 女, 硕士研究生, 主要研究方向: 舆情分析; 杨孔雨(1967-), 男, 博士, 教授, 主要研究方向: 舆情智能监测与管理。

收稿日期: 2022-02-08

键性和相互关联的文献,同时研究该领域的演变。

本文以 CiteSpace 软件为主,Excel 等工具为辅,国内中文文献来源于中国知网(CNKI),经过人工进行剔除和筛选后,得到 2011~2021 年的 877 篇样本相关文献。国外英文文献来源于 Web of Science (WOS)的核心合集数据库,人工剔除图书阅读笔记、广告等不相关内容后,导入 CiteSpace 进行文献去重,最终得到 423 篇样本相关英文文献。

2 基于 CiteSpace 的文献可视化分析

2.1 文献计量分析

年度发文量变化趋势图能够反映研究领域的总体情况,在时间维度上反映每年的不同发文量和活跃度。根据检索获得的中英文相关文献绘制发文量



图1 发文量随年份变化趋势图

Fig. 1 Variation trend of documents volume with years

2.2 作者共现分析

通过作者合作共现图谱可以分析该领域各个作者的重要性、中心性和合作关系。对 CNKI 中文文献作者进行分析处理后,得到国内作者共现图谱如图 2 所示。图 2 中,作者节点数为 214,连线数为 30,密度为 0.001 3,连线代表有合作关系,线条粗细代表合作关系的密切程度。由此可见,各个作者之间的联系和合作关系极为分散,多处于个人独立研究状态,未能形成明显的合作网络图谱,发文量最多的作者为张英、黄道英、范涛、王昊,发表文献数均为 3 篇。



图2 国内作者共现图谱(节选)

Fig. 2 Co-occurrence graph of domestic authors (excerpt)

趋势图,各年份不同变化如图 1 所示。国内舆情情感分析结合神经网络技术开始于 2011 年,此后发文量逐渐增加,于 2020 年达到最高峰,2011~2015 年诞生阶段结束、进入发展阶段,2017~2020 年发文量迅速增长,虽然 2021 年发表中文文献 161 篇,数值有所回落,但仍持续受到关注。结合研究背景来看,情感分析的概念由 Nasukawa 等人^[3]在 2003 年最先提出,同年 Bengio 等人^[4]提出了神经网络语言模型,国外相关文献最早发表于 2006 年,初期发展趋势和国内基本一致,2018 年结束快速上升期、进入平缓发展阶段,同样也于 2021 年发表 Web of Science 核心合集 81 篇,较上年发文量最高峰有所回落,相关理论研究也将更加完善。

对 WOS 核心合集进行可视化共现图谱分析,时间为 2006~2021 年,作者节点数为 247,连线数为 162,密度为 0.005 3,部分作者形成了小规模的合作网络,其中发文量最多的作者是 Erik Cambria (10 篇),汇聚成以其为核心的网状合作图谱。国外作者共现图谱如图 3 所示。由图 3 可见,国外作者的节点和连线数都比国内多,可知国外发文作者间的联系与团队合作比国内更加密切,说明国内应该加强作者间的共同合作关系。



图3 国外作者共现图谱(节选)

Fig. 3 Co-occurrence graph of foreign authors (excerpt)

2.3 机构共现分析

为分析该研究领域中的研究力量分布,对 CNKI 中文文献机构进行处理分析,得到国内合作机构共现图谱如图 4 所示。网络裁剪方法同作者分析,发

文量前三的机构分别为国家统计局 8 篇、南京理工大学 7 篇、南京大学信息管理学院 6 篇,机构与机构间联系较为分散,尚未建立明显合作体系。



图 4 国内机构分布共现图谱

Fig. 4 Co-occurrence graph of domestic institutions

对国外发文机构可视化分析发现,发文章量前五的机构分别是新加坡南洋理工大学 (Nanyang Technological University) 13 篇、中国科学院 (Chinese Acad Sci) 10 篇、北京航空航天大学 (Beihang

University) 5 篇、哈尔滨工业大学 (Harbin Institute of Technology) 5 篇、清华大学 (Tsinghua University) 5 篇,国外合作机构共现图谱如图 5 所示。

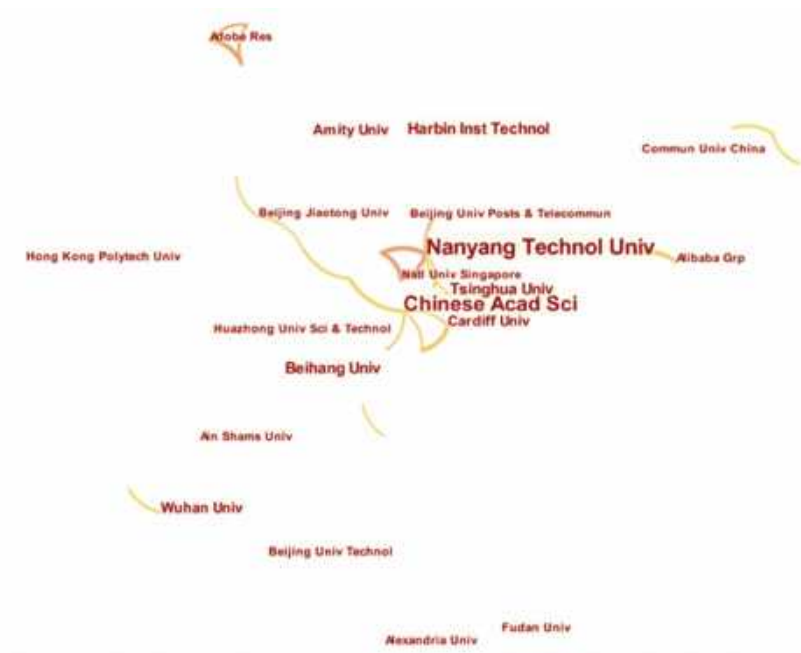


图 5 国外机构分布共现图谱

Fig. 5 Co-occurrence graph of foreign institutions

从国内外研究机构分析发现,发文章最多的都是各地高校及学院、科研机构,对神经网络和网络舆情情感分析做出了主要贡献,但是相互之间的合作仍然亟待加强。

3 关键词可视化分析

3.1 关键词聚类分析

关键词是一篇文章的核心和主题,CiteSpace 在使用关键词共现分析的基础上,使用对数似然比(Log-Likelihood Ratio, LLR)算法进行聚类,中文文

献共生成 14 个聚类,分别对每个类的主题进行归纳总结,前 8 类见表 1;英文文献共生成 9 个聚类,总结主题,选取 7 类见表 2。国内外共同研究热点有数据挖掘、深度学习、文本分类、情绪分析等,国外主要关注情绪倾向分析及预测,国内更关注于舆情文本分析。同时,聚类不仅可以集中于验证研究热点,还能够指出其在各种领域上的具体应用,如表 2 中聚类 0 是电影预测领域的应用,聚类 6 是电动汽车评论领域的应用,在研究计算机相关应用热点的时候,也要重视聚类关键词在非计算机类领域的应用。

表 1 中文文献聚类表

Tab. 1 Chinese literatures clustering table

聚类编号	节点数量	年份	聚类关键词(LLR)	主题归纳
0	39	2017	情感分析;人工智能;情感特征;情感分类;强化学习	人工智能情感分析
1	34	2018	情感分类;情感分析;跨领域;微博评论;域对抗	跨领域文本情感分析
2	33	2017	微博;词向量;胶囊网络;句法分析;在线评论	在线评论中文文本分类
3	32	2017	深度学习;RNN;细粒度情感分析;情感强度;情感词向量	深度学习情感分析
4	27	2018	LSTM;CNN;大数据;可视化;GRU	可视化神经网络、大数据
5	26	2017	机器学习;数据挖掘;电商评论;分类算法;社交媒体	社交媒体评论的数据挖掘和分类
6	25	2016	文本分类;情感词典;短文本;情感倾向;商品评论	短文本评论的情感分类
7	24	2018	社交网络;主题模型;观点挖掘;意见领袖;数据分析	主题模型观点挖掘

表 2 英文文献聚类表

Tab. 2 English literatures clustering table

聚类编号	节点数量	年份	聚类关键词(LLR)	主题归纳
0	23	2017	movie sales; forecast; SA; strength; WOM	电影销售领域情感分析与口碑预测
1	23	2018	web; event detection; natural language processing; text analytics; sentiment analysis	网络文本情感分析
2	20	2018	convolutional neural networks; multimedia data; segmentation; cloud machine learning; visualization	可视化神经网络与数据挖掘
3	17	2017	swivel; multicriteria decision making; multi-party context; annual reports; political leaning	多标准决策、政治倾向分析
4	17	2018	CNN; LSTM; NLP; RNN; ensemble algorithm	神经网络与自然语言处理
5	14	2017	online customer reviews; time series data prediction; opinion leader; information retrieval systems; wordnet	在线评论文本检索与数据预测
6	12	2019	electric vehicles; review mining; social media; deep learning	电动汽车领域社交媒体评论挖掘
8	10	2017	multimodal sentiment analysis; SVM; elm; feature	多模态情绪分析与极限学习机

表 1 中聚类#3 深度学习、#4 长短期记忆网络 LSTM(Long Short-Term Memory)、#5 机器学习主要围绕情感分析的技术展开。

早期的情感分析主要基于情感词典(#6 号聚类),其关键是情感权重及情感计算规则。情感词

典最早由 Whissell 于 1998 年提出^[5],王科等人(2016)将情感词典自动构建方法归纳为三大类^[6]:基于知识库的方法、基于语料库的方法和基于知识库和语料库相结合的方法。目前最常用的 3 个中文开源情感词典有 HowNet、台湾大学的 NTUSD 和大连理工大学的词汇本体库,在此基础上,阳爱民等人

(2013)选用若干个情感种子词,通过改进的 PMI (Pointwise Mutual Information) 算法计算情感词的情感权值,从而构建出一种分类性能更加稳定的情感词典^[7]。陈国兰(2016)通过与具体社交网络(#7号聚类)结合,针对微博(#2号聚类)文本的特点,构建了一个包括微博网络新词和表情符号的情感词典,与传统方法相比,判断正、负面情感的准确率更高^[8]。

随着人工智能技术的发展,情感分析逐渐与机器学习(#5号聚类)相结合。传统机器学习大多是有监督学习,分类器主要有朴素贝叶斯(Naive Bayes, NB)和支持向量机(Support Vector Machine, SVM)等^[9]。

基于情感词典和有监督学习的情感分析人力资源消耗过多,效率低下。为了提高文本分类的准确率,深度学习(#3号聚类)通过构建神经网络模型对文本进行特征抽取和自动学习优化模型输出^[10]。#4号聚类中包括“卷积神经网络(Convolutional Neural Networks, CNN)”关键词,利用卷积层, CNN 可以学习局部特征。梁斌等人(2017)提出一种情感分类方法,通过结合3种注意力机制来构造多注意力卷积神经网络模型,准确表示每一个词在句子中的重要性^[11]。#3号聚类中包括关键词“循环神经网络(Recurrent Neural Network, RNN)”, #4号聚类还包含“LSTM”关键词, LSTM 是一种特殊的 RNN。梁军等人(2015)提出了一种利用文本上下文信息、结构化信息和情感语义信息的模型,来解决情感极性转移问题^[12]。

国外的聚类还围绕具体应用领域展开研究,主要是#0号聚类电影销售和#6号聚类电动汽车, #0号聚类主要包括关键词预测、情感分析、口碑。口碑是由有经验的消费者创造的有用且可信的信息,也就是用户生成内容(User-Generated Content, UGC)。互联网技术的发展使人们能够在网络平台上与他人分享、交换意见,通过在线渠道实时传播评论信息等。电子口碑的早期研究倾向于只使用数量和评分,而不分析内容的语法,不足以解释在线交流的动态。在电子口碑中,情感应该作为一个重要参数, Liu(2006)^[5]分析了从雅虎电影的留言板上收集的40部电影的评论,结果表明电影上映前的电子口碑对票房收入有显著影响。Duan 等人^[6]研究电影销售与在线评论和收视率之间关系,结果显示2个领域之间存在影响,但高的评价不能确保高的销售额。Rui 等人(2013)^[7]分析电影的电子口碑,表明消费

者观看电影的意愿受电子口碑影响很大。

#6号聚类电动汽车包含关键词 social media(社交媒体)、review mining(评论挖掘)和 deep learning(深度学习)。公众对电动汽车的看法和使用电动汽车的普遍意愿对推广电动汽车起到了重要作用, XU 等人(2018)^[8]研究并确定了影响电动汽车接受和拒绝的因素。消费者对电动汽车技术特征的态度和对电动汽车效用的看法是影响是否选用电动汽车重要因素^[9]。消费者情绪和感受也会影响电动汽车的选择^[10],了解消费者的情绪有助于改进客户关系管理(CRM), Jena(2020)^[11]通过情感分析,对电动汽车的情感进行分类。帮助用户获得对各种产品功能的高层次的看法概述,并大幅缩短用户阅读和提取有关产品、服务意见的文本长度。

#8号聚类多模态情绪分析包含的关键词有:支持向量机、极限学习机、特征。情感分析中文本、声音和视觉模式的融合最近引起了越来越多的关注,多模态情感分析通过分析在线视频中的口语情态、听觉情态和视觉情态三种模式,扩展了传统的文本情感分析。Chaturvedi 等人(2017)^[12]提出了一种基于贝叶斯网络的 ELM (Extreme Learning Machine), 克服了传统 ELM 不能推广到非线性数据集的缺点。

通过对该领域研究热点的分析,得到关键词聚类图谱,节点与节点之间的连线颜色越浅代表出现得越晚。国内关键词聚类图谱如图6所示。图6中共有347个节点,622条连线,密度为0.0104, Q值为0.6315, S值为0.8744,国内关键词聚类图谱如图7所示。图7中共有173个节点,454条连线,密度为0.0305, Q值为0.8866, S值为0.7572。模块值 Q (Modularity Q) 用于评价聚类的有效性,平均轮廓值 S (Mean Silhouette) 用于衡量聚类的同质性。当 $Q > 0.3$ 且 $S > 0.5$ 时认为该聚类模块性显著、合理。



图6 国内关键词聚类图谱

Fig. 6 Domestic keywords clustering map

表4 英文关键词词频表

Tab. 4 English keywords frequency table

词频	中心性	年份	关键词
30	0.13	2015	classification
26	0.15	2016	sentiment analysis
22	0.07	2018	LSTM
22	0.48	2011	neural network
20	0.10	2016	CNN
19	0.36	2014	model
9	0.08	2018	system
8	0.05	2015	review
8	0.07	2018	extraction
7	0.03	2015	algorithm
7	0.08	2015	impact
6	0.03	2018	machine
6	0.05	2018	twitter



图7 国外关键词聚类图谱

Fig. 7 Foreign keywords clustering map

中文文献关键词聚类后按词频排序前13位见表3,高频关键词代表了该领域研究的相关热点,关键词出现的频率越高,说明领域越热;关键词与其他关键词之间的联系越多,说明中心度越大^[13]。中文关键词聚类词频表见表3,通过分析可以看出,国内词频出现较高的关键词为“情感分析”、“深度学习”、“情感分类”、“词向量”、“神经网络”和“机器学习”。

表3 中文关键词词频表

Tab. 3 Chinese keywords frequency table

词频	中心性	年份	关键词
417	0.84	2013	情感分析
227	0.42	2014	深度学习
85	0.22	2012	情感分类
69	0.13	2014	词向量
61	0.17	2014	神经网络
44	0.12	2014	机器学习
44	0.10	2016	LSTM
39	0.09	2014	情感词典
34	0.17	2014	微博
33	0.14	2014	文本分类
24	0.06	2015	网络舆情
21	0.13	2015	社交网络
21	0.04	2018	特征融合

英文高频关键词,见表4,国外词频出现较高的关键词为“classification(分类)”、“sentiment analysis(情感分析)”、“LSTM(长短期记忆网络)”、“neural network(神经网络)”、“CNN(卷积神经网络)”和“model(模型)”。国内的研究热点和国外也存在差异,国内更重视舆情中文文本的自然语言处理和中文文本分类方面,国外更重视神经网络模型的研究和具体领域的应用。

3.2 关键词时间线图谱分析

在 CiteSpace 中对数据样本进行分析处理后,得到国内外时间线图谱如图8和图9所示,展现聚类关键词之间的关系,以及热点关键词随着时间变化的研究历程和前沿趋势。每个点对应的是关键词首次出现的年份,圆圈越大表示词频越大,连线表示2个关键词在同一篇文章中出现。连线数量多且密集,说明研究集中程度较好。

通过对图8分析可得,国内基于神经网络的情感分析研究热点从2011年出现,2011~2012年处于起始状态,研究热点为“语义理解文本倾向”、“观点挖掘”和“情感分类”。2013年对情感分类进行进一步探讨,在情感分析的基础上,逐渐传承。2014年开始结合机器学习和神经网络,研究热点为“词向量”、“情感词典”、“文本分类”、“深度学习”等。随着深度学习的研究不断发展完善,2015~2018年,LSTM、门控循环单元(Gated Recurrent Unit,GRU)、CNN等神经网络模型逐渐成为了热点研究对象,社交媒体逐渐成为研究舆情分析的平台。随着神经网络技术的发展,2019年又出现了以“BERT”(Bidirectional Encoder Representation from Transformers)、“词嵌入”、“胶囊网络”为首的研究热点。

国外虽然节点数量少于国内,但连线比国内更加密集,联系更加密切,如图9所示。研究热点从2010年“RFM”(Recency, Frequency, Monetary)模型起步,2011年的“神经网络”开始发展,2014~2015年的研究热点为“模型”、“算法”、“词典”、“分

类”。2016 年后在“情感分析”的基础上, 出现了更加细化的多元分类, 如“特征”、“LSTM”、“CNN”、

“框架”。“BERT”热点稍晚于国内一年。

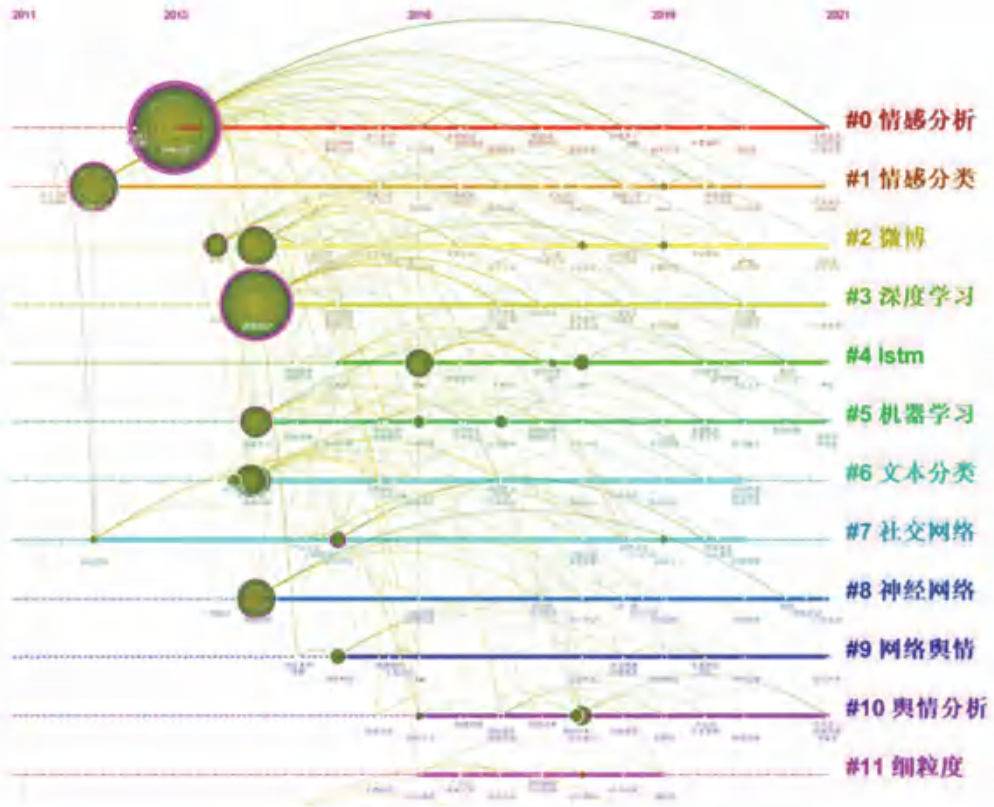


图 8 国内关键词时间线图
Fig. 8 Domestic keywords timeline map

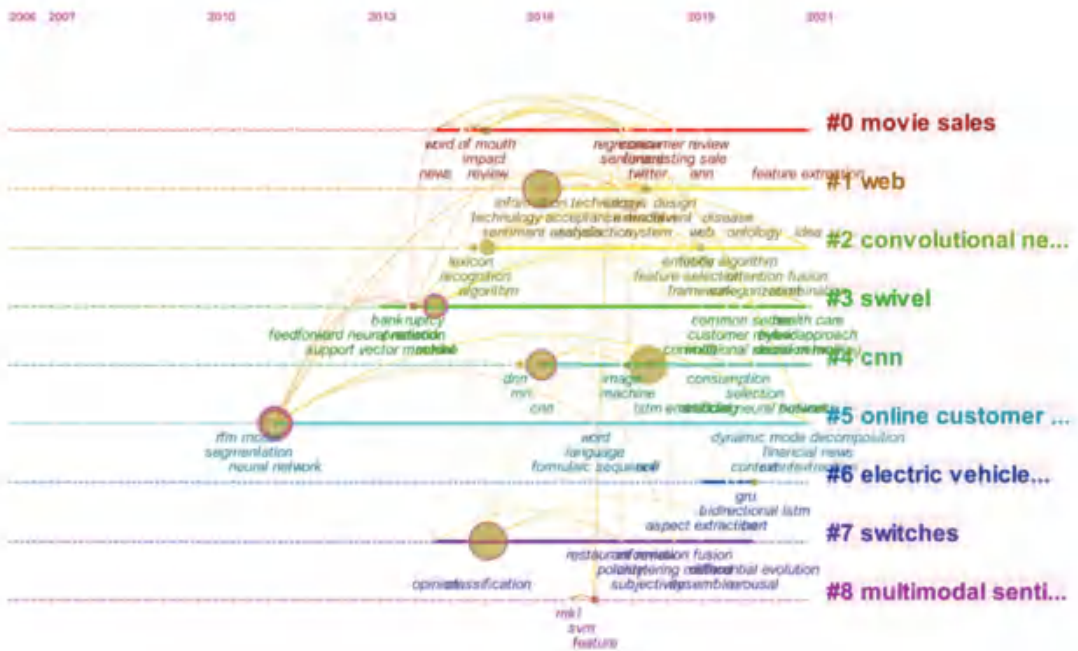


图 9 国外关键词时间线图
Fig. 9 Foreign keywords timeline map

3.3 关键词突现分析

通过 CiteSpace 的 Burstness(突现)功能进行突现词分析,突现词代表了不同时期开始的研究热点,得到领域关键词的起始爆发时间、结束时间和强度等,选择前 13 个关键词展示如图 10、图 11 所示。图 10、图 11 中,Year 代表原文发表年份,Begin 代表 Burst 的开始年份,End 代表结束年份,Strength 代表突现强度,用来表示该领域在某些年份的文献中突现词的剧烈程度。由图 10 可见,在国内,“机器学习”、“文本挖掘”和“中文微博”突现词出现在 2016 年,“深度学习”、“LSTM”、“记忆网络”、“细粒度”、“词嵌入”出现在 2017~2019 年,“深度学习”和“LSTM”在这 3 年中的强度值最大,机器学习的持续发展为舆情情感分析奠定了基础,情感分析进一步与神经网络技术结合。随着深度学习研究的日趋深入与发展,2020~2021 年出现的关键词“特征融合”、“BERT”、“舆情分析”、“销量预测”,可见研究人员已经将深度学习运用于舆情情感分析中,并且使用神经网络模型对销量进行预测。国内外既有联系,又有区别。由图 11 可知,在国外,早期突现词主

要由 2013 年开始突现的“支持向量机”、2016 年开始突现的“卷积神经网络”和“情感分析”组成,随着人工智能的发展,在 2018 年又出现了“主观性”、“社交媒体”和 2019 年出现的“网络”、“模型集成”等新突现词,其中突现强度最高的是卷积神经网络。

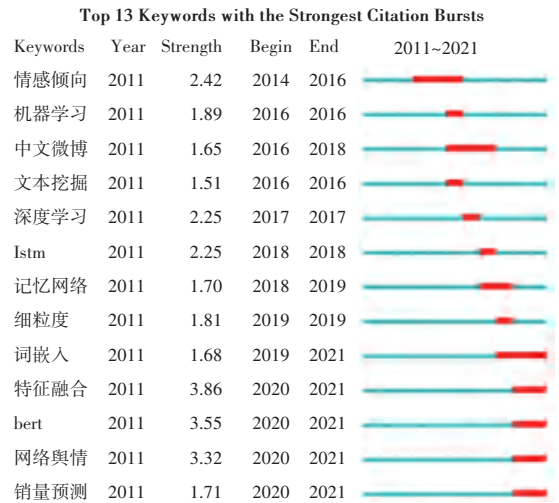


图 10 国内突现强度前 13 的突现词

Fig. 10 Domestic top 13 keywords with the Strongest Citation Bursts

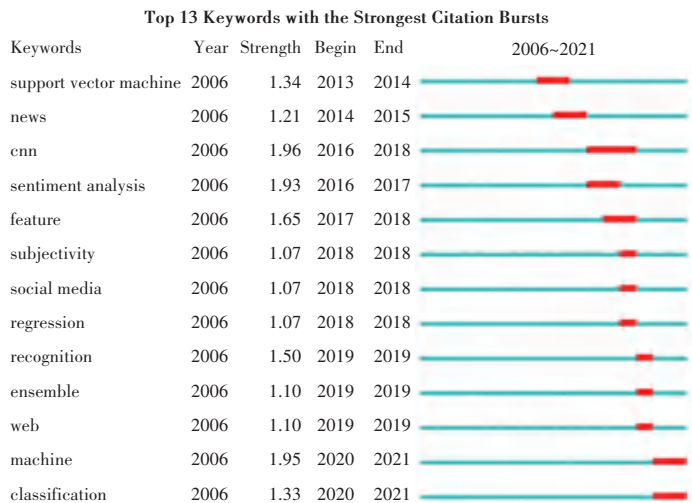


图 11 国外突现强度前 13 的突现词

Fig. 11 Foreign top 13 keywords with the Strongest Citation Bursts

对突现词进行研究分析,总结现阶段国内结合神经网络的舆情情感分析技术研究的前沿方向主要有:

(1) 实现预训练模型的突破,如 BERT、ALBERT (A Lite BERT)、XLNET、ELMo (Embeddings from Language Models) 等模型。目前在舆情情感分析领域有许多经典的深度学习模型,Devlin 等人^[14]提出的 BERT 模型,能够根据指定目标,捕捉到文本中相对应的情感信息,更加具有灵活性。BERT 模型可以动态获取各种语境下的词语表达向量,通过上下

文信息充分解决词语多义性问题。由图 6 看出 BERT 的强度值位于所有年份中第二名,处于发展状态中。史振杰等人(2020)^[15]提出了一种结合卷积神经网络的 BERT-CNN 模型,该网络模型在用 BERT 表示词向量语义信息的基础上,使用 CNN 提取文本特征,实现了对文本信息语义的深层理解。由于 BERT 模型参数量较大、复杂度高问题,Lan 等人(2019)^[16]对 BERT 模型进行改进,提出了 ALBERT 模型。房京珂(2021)^[17]将 ALBERT 模型

与2种双向递归神经网络结构结合,全面提取模型中的语义信息,使其更加适合舆情情感分析。梁淑蓉(2021)^[18]提出一种融合情感词典的XLNet预训练模型和一种基于LSTM+Attention网络层的XLNet预训练模型,前一种模型可以根据不同的上下文语言环境计算词向量,精准提高了情感倾向,解决情感词典构建数量和收录新词工作量的问题,后一种模型在一定程度上解决了中文文本的语义问题。李铮等人(2021)^[19]提出一种基于ELMo和双向自注意力网络(Bidirectional Self-Attention Network, Bi-SAN)的中文文本情感分析模型,利用ELMo模型抽取词向量,更好地表示了中文文本的一词多义性,提高了情感分类的准确性。未来,还能够通过增加计算机基础硬件算力,更广泛地使用其他模型,后续对最新模型的研究将会逐渐成为热点。

(2)特征融合表示方法的探索。多特征融合向量可以从短文本中充分学习词性特征信息和位置特征信息。韩普(2021)等人^[20]将词性特征向量和位置特征向量进行融合,一方面可以充分地学习特征间的语义关系,另一方面也可以有效地利用词性信息和位置信息,提升情感分析模型的效果。郭可心等人(2021)^[21]从多个不同角度挖掘图文之间的情感共现,增强了神经网络模型捕捉情感语义的能力。祁瑞华等人(2020)^[22]融合BERT词向量和跨领域词向量生成跨领域统一特征空间,提高了情感分类的准确率。目前,在图文融合方面还存在许多问题,如数据集规模较小、情感数据集的建立困难、不同模型特征获取不一致、媒体数据类型单调、模型实验效果一般、准确率低、未考虑更多信息^[23]。在今后的研究中,进行情感分析的同时融入图文关系分析,是一个重要的研究方向。

(3)词嵌入层的设计。词嵌入已经成为各任务模型在预处理阶段的事实标准。目前,词嵌入方法是利用较长的上下文,通过神经网络模型来解决自然语言问题^[24]。韩旭丽等人(2019)^[25]提出了一种词嵌入辅助机制的注意力神经网络模型,该模型在原本词嵌入层的基础上加入辅助层,用来提取词向量的特征表示,降低了模型训练的复杂性和训练时间。赵亚欧等人(2020)^[26]提出一种基于语言模型的字嵌入的情感分析模型,将获取的词语向量作为网络输入,融合了词语本身的语义特征及其所在的上下文语义,可以很好地表示一词多义性。目前,基于大规模语料的预训练模型是最热门的研究方向,新型词嵌入越来越通用,语义功能越来越强大。

4 结束语

本文通过CiteSpace、Excel等软件,以中国知网和WoS核心合集文献为研究对象,采用文献计量和可视化知识图谱的方法,对神经网络的网络舆情情感分析进行了研究。通过发文量趋势分析、作者共现分析、研究机构共现分析、关键词聚类分析、关键词时间线图分析 and 关键词突现分析,研究了该领域的热点主题和前沿趋势。得出如下结论:

(1)从基于神经网络的舆情情感分析领域的发文量年度趋势来看,国内发表第一篇相关文献的起始时间比国外晚,虽然增长速度快,但却存在着一些问题,比如作者及机构间相互合作较少、学术交流不足,以及研究主题较集中,缺少领军人物等问题,国内研究成果主要由世界各地的大学和科研机构推广,应该加强作者或机构间的合作与沟通,共同发展。

(2)通过研究热点和研究前沿对比分析,国内外共同研究热点主题为“情感分类”、“LSTM”、“深度学习”、“神经网络”等,但国内外侧重点存在着差异,国内文献更重视文本的语义表达和文本情感分类方面,国外则更重视情感分析模型的应用及预测,更加具有实用性。

(3)国内研究前沿趋势主要为网络舆情情感分析的特征融合表示方法、预训练模型的探索和词嵌入层的设计。

本文分析对比了国内外神经网络和舆情情感分析研究的演变脉络,对进一步研究国内外概况,预测未来发展趋势具有现实意义。由于CiteSpace软件的限制,并未收集所有英文期刊等文献,也未获取重要会议记录等数据,本文内容不能完全描述该领域的研究现状。CiteSpace软件中具有更多功能可以分析WoS核心合集集中的文献,本文也并未全部运用,在后续的工作中拟做更进一步的研究完善。

参考文献

- [1] 陈兴蜀,常天祐,王海舟,等. 基于微博数据的“新冠肺炎疫情”舆情演化时空分析[J]. 四川大学学报(自然科学版),2020,57(2): 409-416.
- [2] MIKOLOV T, CHEN K, CORRADO G S, et al. Efficient Estimation of Word Representations in Vector Space [J]. arXiv preprint arXiv:1301.3781, 2013.
- [3] NASUKAWA T, YI J. Sentiment analysis: Capturing favorability using natural language processing [C]// Proc of the 2nd International Conference on Knowledge Capture. New York: ACM, 2003: 70-77.

- [4] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model [J]. *Journal of Machine Learning Research*, 2003, 3: 1137-1155.
- [5] LIU Yong. Word of mouth for movies: Its dynamics and impact on box office revenue [J]. *Journal of Marketing*, 2006, 70(3): 74-89.
- [6] DUAN Weijing, GU Bin, ANDREW B, et al. Do Online Reviews Matter? -an Empirical Investigation Of Panel Data [J]. *Decision Support Systems*, 2008, 45(4): 1007-1016.
- [7] RUI H, LIU Y, WHINSTON A. Whose and what chatter matters? The effect of tweets on movie sales [J]. *Decision Support Systems*, 2013, 55(4): 863-870.
- [8] XU Min, MENG Qiang, LIU Yisi. Public's perception of adopting electric vehicles: a case study of Singapore [J]. *Journal of the Eastern Asia Society for Transportation Studies*, 2017, 12: 285-298.
- [9] SINGER M R. The barriers to acceptance of plug-in electric vehicles; 2017 update [R]. *National Renewable Energy Lab. (NREL)*, Golden, CO (United States), 2017.
- [10] CHRISTIDIS P, FOCAS C. Factors affecting the uptake of hybrid and electric vehicles in the European Union [J]. *Energies*, 2019, 12(18): 3414.
- [11] JENA R. An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach [J]. *Industrial Marketing Management*, 2020, 90: 605-616.
- [12] CHATURVEDI I, RAGUSA E, GASTALDO P, et al. Bayesian network based extreme learning machine for subjectivity detection [J]. *Journal of the Franklin Institute*, 2017, 355(4): 1780-1797.
- [13] 李诗琦, 刘俊雁, 严贤春, 等. 基于文献计量法的国内外 SWAT 模型研究趋势对比分析 [J/OL]. *西华师范大学学报(自然科学版)*; 1-8 [2022-04-20]. <http://kns.cnki.net/kcms/detail/51.1699.N.20211112.1821.018.html>.
- [14] DEVLIN J, CHANG M-W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [15] 史振杰, 董兆伟, 庞超逸, 等. 基于 BERT-CNN 的电商评论情感分析 [J]. *智能计算机与应用*, 2020, 10(02): 7-11.
- [16] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations [J]. *arXiv preprint arXiv:1909.11942*, 2019.
- [17] 房京珂. 基于 NLP 的微博情感分析研究 [D]. 北京: 中央民族大学, 2021.
- [18] 梁淑蓉. 基于 XLNet 预训练语言模型的短文本情感分析研究 [D]. 桂林: 桂林理工大学, 2021.
- [19] 李铮, 陈莉, 张爽. 基于 ELMo 和 Bi-SAN 的中文文本情感分析 [J]. *计算机应用研究*, 2021, 38(08): 2303-2307.
- [20] 韩普, 张伟, 张展鹏, 等. 基于特征融合和多通道的突发公共卫生事件微博情感分析 [J]. *数据分析与知识发现*, 2021, 5(11): 68-79.
- [21] 郭可心, 张宇翔. 基于多层次空间注意力的图文评论情感分析方法 [J]. *计算机应用*, 2021, 41(10): 2835-2841.
- [22] 祁瑞华, 简悦, 郭旭, 等. 融合特征与注意力的跨领域产品评论情感分析 [J]. *数据分析与知识发现*, 2020, 4(12): 85-94.
- [23] 孟祥瑞, 杨文忠, 王婷. 基于图文融合的情感分析研究综述 [J]. *计算机应用*, 2021, 41(02): 307-317.
- [24] 刘胜杰, 许亮. 基于词嵌入技术的文本表示研究现状综述 [J]. *现代计算机*, 2020(01): 40-43.
- [25] 韩旭丽, 曾碧卿, 曾锋, 等. 基于词嵌入辅助机制的情感分析 [J]. *计算机科学*, 2019, 46(10): 258-264.
- [26] 赵亚欧, 张家重, 李始斌, 等. 融合基于语言模型的词嵌入和多尺度卷积神经网络的情感分析 [J]. *计算机应用*, 2020, 40(03): 651-657.

(上接第 32 页)

参考文献

- [1] BASER E, BALASUBRAMANIAN V, BHATTACHARYYA P, et al. FANTrack: 3D multi-object tracking with feature association network [C] // 2019 IEEE Intelligent Vehicles Symposium (IV). Paris, France: IEEE, 2019: 1426-1433.
- [2] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs [J]. *arXiv preprint arXiv:1704.00028*, 2017.
- [3] SHARMA V, MIR R N. A comprehensive and systematic look up into deep learning based object detection techniques: A review [J]. *Computer Science Review*, 2020, 38: 100301.
- [4] LAW H, DENG Jia. Cornernet: Detecting objects as paired keypoints [J]. *International Journal of Computer Vision*, 2020, 128: 642-656.
- [5] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. Centernet: Keypoint triplets for object detection [C] // Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019: 6568-6577.
- [6] HA D B, ZHU Guangxi, ZHAO Guangzhou. Vehicle tracking method based on corner feature and mean-shift [J]. *Computer Engineering*, 2010, 36(06): 196-197.
- [7] CAO Xianbin, GAO Changcheng, Lan Jinhe, et al. Ego motion guided particle filter for vehicle tracking in airborne videos [J]. *Neurocomputing*, 2014, 124: 168-177.
- [8] LI Zhiyong, HE Shuang, HASHEM M. Robust object tracking via multi-feature adaptive fusion based on stability; contrast analysis [J]. *The Visual Computer*, 2015, 31(10): 1319-1337.
- [9] WU Sunyong, LIAO Guisheng, YANG Zhiwei, et al. Improved track-before-detect algorithm based on particle filter [J]. *Control and Decision*, 2010, 25(12): 1843-1847.
- [10] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.
- [11] GORDON D, FARHADI A, FOX D. Re3: Real-time recurrent regression networks for visual tracking of generic objects [J]. *IEEE Robotics and Automation Letters*, 2018, 3(2): 788-795.