



(12) 发明专利

(10) 授权公告号 CN 115982473 B

(45) 授权公告日 2023.06.23

(21) 申请号 202310277860.9

G06F 16/9538 (2019.01)

(22) 申请日 2023.03.21

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 115982473 A

AU 2020100710 A4, 2020.06.11

CN 104504150 A, 2015.04.08

CN 107239439 A, 2017.10.10

(43) 申请公布日 2023.04.18

CN 109101933 A, 2018.12.28

CN 111694959 A, 2020.09.22

(73) 专利权人 环球数科集团有限公司

地址 518063 广东省深圳市南山区粤海街

道高新南九道10号深圳湾科技生态园

10栋B座17层01-03号

CN 111985243 A, 2020.11.24

CN 112861541 A, 2021.05.28

CN 113065577 A, 2021.07.02

CN 113627550 A, 2021.11.09

CN 115392232 A, 2022.11.25

JP W02022024956 A1, 2022.02.03

US 11194972 B1, 2021.12.07

US 2017140240 A1, 2017.05.18

US 2022405480 A1, 2022.12.22

(72) 发明人 张卫平 丁焯 吴茜 丁园 向荣

(74) 专利代理机构 北京清控智云知识产权代理

事务所(特殊普通合伙)

11919

专利代理师 马肃

令狐阳. 基于图文融合的情感分析研究与应用. 中国优秀硕士学位论文全文数据库电子期刊. 2022, 1-78.

审查员 岳翠莹

权利要求书2页 说明书6页 附图1页

(51) Int. Cl.

G06F 16/9535 (2019.01)

G06F 40/30 (2020.01)

G06N 3/0464 (2023.01)

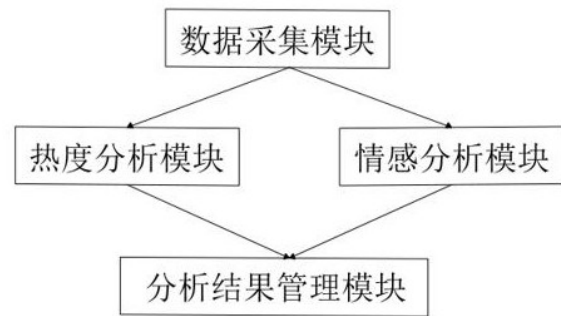
G06N 3/08 (2023.01)

(54) 发明名称

一种基于AIGC的舆情分析编排系统

(57) 摘要

本发明提供了一种基于AIGC的舆情分析编排系统,其特征在在于,包括数据采集模块、热度分析模块、情感分析模块和分析结果管理模块;所述数据采集模块用于采集网络信息数据并在网络信息数据中筛选出各关键词组和关键信息;所述热度分析模块用于根据关键信息分析所述各关键词组在网络中的影响热度;所述情感分析模块用于根据关键信息对各关键词组的舆情正反面进行分析,所述分析结果管理模块用于向用户可视化展示各关键词组热度和舆情影响;本发明采用多维度的情感分析方式,可最大程度上了解到用户的真实情感。



1. 一种基于AIGC的舆情分析编排系统,其特征在于,包括数据采集模块、热度分析模块、情感分析模块和分析结果管理模块;

所述数据采集模块用于采集网络信息数据并在网络信息数据中筛选出各关键词组和关键信息;所述热度分析模块用于根据关键信息分析所述各关键词组在网络中的影响热度;所述情感分析模块用于根据关键信息分析所述各关键词组的舆情正负面,所述分析结果管理模块用于根据所述影响热度和舆情正负面分析向用户可视化展示各关键词组热度和舆情影响;

所述数据采集模块包括采集子模块、预处理模块和预存储模块,所述采集子模块用于在设定的采集周期内对互联网上各数据平台上公开的网络信息数据进行采集,采集的所述网络信息数据包括通过各数据平台的搜索引擎获取的关键词组,所述预处理模块用于对网络信息数据进行清洗筛选获取关键信息,所述预存储模块用于存储关键词组和关键信息;

所述热度分析模块对各关键词组在网络中影响热度的分析包括对单数据平台上关键词组的热度分析和全网络平台上关键词组的热度分析,所述单数据平台上关键词组的热度分析如下:

$$S_i = \sum_{j=1}^n \frac{p_j d_j}{p_{\text{总}}} \times \log_2 \frac{T-t_i}{T_{\text{周}}};$$

其中, S_i 为某关键词组在第*i*个数据平台上的热度值, p_j 为此关键词组相关的第*j*篇文章的所有评论数量,设此关键词组 S_i 下共有*n*篇相关文章,满足 $1 \leq j \leq n$, d_j 为此关键词组相关的第*j*篇文章的所有评论点赞数, $p_{\text{总}}$ 为该平台上所有关键词组相关评论数量, T 为当前时间, t_i 为第*j*篇文章的发布时间, $T-t_i$ 为第*j*篇文章的发布时间与当前时间相差的时间, $T_{\text{周}}$ 为采集周期;

所述全网络平台上关键词组的热度分析满足:

$$C = \sum_{i=1}^m S_i \times \gamma_i;$$

其中, C 为某关键词组在全网络平台上的热度值,设本系统共采集了*m*个数据平台的网络信息数据,满足 $1 \leq i \leq m$; γ_i 为第*i*个数据平台在全网络平台上的权重值;

所述情感分析模块包括第一分析模块、第二分析模块和分析结果融合模块,所述第一分析模块对文章文本、文章评论文本和文章评论图片进行图文融合生成融合特征,并对融合特征进行情感分析;所述第二分析模块根据文章文本和文章评论文本生成文本特征,并对文本特征进行情感分析;所述分析结果融合模块用于融合第一分析模块和第二分析模块的分析结果生成最终分析结果;

所述第一分析模块包括融合词获取单元,所述融合词获取单元用于获取文章文本和文章评论文本中的融合词,所述融合词包括程度副词及该程度副词后一词汇、情感词及该情感词后一词汇、评价词及该评价词后一词汇、否定词及该否定词后一词汇;

所述第一分析模块通过word2vec训练语料库将融合词转化为融合词词向量的形式以便机器学习,通过建立一个图片CNN情感分析模型对图片特征进行提取,所述图片CNN情感分析模型包括多个卷积层、多个池化层和两个全连接层,所述图片CNN情感分析模型的输入为多个 $32 \times 32 \times 3$ 的图片,其中 $32 \times 32 \times 3$ 代表图片的宽度 \times 高度 \times 图片维度,所述图片CNN情感分析模型的输出为图片的情绪分析结论,即所述两个全连接层中的后一全连接层包括代表图片的情绪为积极或消极的两个维度;使用图片CNN情感分析模型仅用于提取表达图

片的特征向量,即提取出两个全连接层中的前一全连接层的特征进行向量化作为图片的特征向量,所述图片的特征向量如下式所示:

$$U_j = f[w(\text{image}_j) + b];$$

其中, U_j 为第 j 张图片的特征向量, f 为激活函数, image_j 为两全连接层中的前一全连接层的特征, w 为图片CNN情感分析模型的权重矩阵, b 为偏置项;

所述第一分析模块建立一个融合CNN情感分析模型来提取融合特征,所述融合CNN情感分析模型的输入为图片特征向量和融合词向量的融合,其表达方式如下:

$$Z_i = V_i \oplus V_{i+1} \oplus U_i;$$

其中, Z_i 为第 i 个融合CNN情感分析模型的输入, V_i 和 V_{i+1} 表示第 i 组融合词特征的向量, V_i 表示第 i 组融合词的前半部分的特征向量, V_{i+1} 表示第 i 组融合词的后半部分的特征向量,每组融合词为某个程度副词及该程度副词后一词汇、或情感词及该情感词后一词汇、或评价词及该评价词后一词汇、或否定词及该否定词后一词汇, U_i 为第 i 组融合词对应图片的特征向量, \oplus 为拼接操作;

所述融合CNN情感分析模型包括多个卷积层、多个池化层和一个全连接层,将融合CNN情感分析模型的全连接层特征向量化即可得到融合特征,将融合特征放入第一分类器中训练继而得到第一分析模块的分析结果,所述第一分析模块的分析结果包括情绪为积极或消极的投票数;

所述第二分析通过word2vec训练语料库将文章文本和文章评论文本转化为文本词向量的形式以便机器学习,通过建立一个文本CNN情感分析模型来对文本特征提取,所述文本CNN情感分析模型的输入为文本词向量,其表达方式如下:

$$V_w = V_1 \oplus V_2 \oplus \dots \oplus V_{n-1} \oplus V_n;$$

其中, V_w 为某个文本CNN情感分析模型的输入, n 为文本中句子的最大长度, V_k 表示文本句子中的某个文本词向量,满足 $1 \leq k \leq n$;

所述文本CNN情感分析模型包括多个卷积层、多个池化层和一个全连接层,将所述文本CNN情感分析模型的全连接层特征向量化即可得到文本特征,将文本特征放入第二分类器中训练继而得到第二分析模块的分析结果,所述第二分析模块的分析结果包括情绪为积极或消极的投票数;

所述分析结果融合模块通过将第一分类器和第二分类器中投票为同类型情绪的票数进行相加,将获得票数多的类型作为最终分析结果。

2. 根据权利要求1所述的一种基于AIGC的舆情分析编排系统,其特征在于,所述预处理模块包括数据清洗模块和分类模块,所述数据清洗模块用于去除网络信息数据中文本的无效信息,所述分类模块用于将去除无效信息后的网络信息数据根据关键词组进行分类生成关键信息,分类后生成的关键信息包括与各关键词组相关的文章文本、文章发布时间、文章评论文本、文章评论图片和文章评论点赞数量。

3. 根据权利要求2所述的一种基于AIGC的舆情分析编排系统,其特征在于,所述分析结果管理模块向用户展示的各关键词组热度和舆情影响包括:展示各数据平台上按热度大小顺序排列的关键词组及其最终分析结果和全网络平台上按热度大小顺序排列的关键词组及其最终分析结果。

一种基于AIGC的舆情分析编排系统

技术领域

[0001] 本发明涉及舆情分析技术领域,尤其涉及一种基于AIGC的舆情分析编排系统。

背景技术

[0002] 随着互联网技术的不断发展,越来越多的用户在网上表达自己的观点和情感,如何高效快速的分析这些带有情感的观点和言论,已得到越来越多研究人员的关注,AIGC即AI Generated Content,是指利用人工智能技术来生成内容,AIGC技术可通过人工智能算法和训练的数据,对多模态的数据(如文本+图像+语音)进行分析处理生成所需新的内容,因而AIGC技术在舆情分析处理中具备广阔应用前景。

[0003] 查阅相关已公开技术方案,如CN107016107A现有技术公开了一种舆情分析方法及系统,方法包括:每间隔预设时长从互联网获取包括第一预设特征词的文本信息;针对获取到的每一文本信息,提取该文本信息中的第二预设特征词,并统计得到一词频文档矩阵;根据所述词频文档矩阵在所述各第二预设特征词中选取合适的分类特征,并根据预设特征权重算法计算得到各分类特征的特征权重;将各分类特征的特征权重输入随机森林分类模型进行分类,得到该文本信息的情感倾向类别;在该文本信息的情感倾向类别为预设情感倾向类别时,将该文本信息推送至舆情展示终端;接收舆情展示终端发送的反馈信息,并根据反馈信息对第一预设特征词进行调整;另一种典型的公开号为CN114385890B的现有技术公开了一种互联网舆情监控系统,包括:数据采集模块,用于采集获取待监控的网络舆情信息;通信部或者传输接口,用于连接数据采集模块并将所述数据采集模块采集到的网络舆情信息发送至处理装置;处理装置具有:比较模块、分析模块、机器学习系统、主分类器、辅分类器、动态分配器、配置模块以及监测模块,通过将监控过程中所积累得到的关键词进行输出,以监控这些关键词是否能够成为热点话题的可能;上述方案均为基于互联网上单一文本信息对舆情进行分析监测,现如今随网络发展,网民在网络社交平台发表内容多样化,包括文本、图片、语音和视频等等,仅根据单一文本并不能保证分析舆情的准确性。

发明内容

[0004] 本发明的目的在于,针对目前所存在的不足,提出了一种基于AIGC的舆情分析编排系统。

[0005] 本发明采用如下技术方案:

[0006] 一种基于AIGC的舆情分析编排系统,其特征在于,包括数据采集模块、热度分析模块、情感分析模块和分析结果管理模块;

[0007] 所述数据采集模块用于采集网络信息数据并在网络信息数据中筛选出各关键词组和关键信息;所述热度分析模块用于根据关键信息分析所述各关键词组在网络中的影响热度;所述情感分析模块用于根据关键信息分析所述各关键词组的舆情正负面进行分析,所述分析结果管理模块用于根据所述影响热度和舆情正负面分析向用户可视化展示各关键词组热度和舆情影响;

[0008] 所述数据采集模块包括采集子模块、预处理模块和预存储模块,所述采集子模块用于在设定的采集周期内对互联网上各数据平台上公开的网络信息数据进行采集,采集的所述网络信息数据包括通过各数据平台的搜索引擎获取的关键词组,所述预处理模块用于对网络信息数据进行清洗筛选获取关键信息,所述预存储模块用于存储关键词组和关键信息;

[0009] 进一步的,所述预处理模块包括数据清洗模块和分类模块,所述数据清洗模块用于去除网络信息数据中文本的无效信息,所述分类模块用于将去除无效信息后的网络信息数据根据关键词组进行分类生成关键信息,分类后生成的关键信息包括与各关键词组相关的文章文本、文章发布时间、文章评论文本、文章评论图片和文章评论点赞数量;

[0010] 进一步的,所述热度分析模块对各关键词组在网络中影响热度的分析包括对单数据平台上关键词组的热度分析和全网络平台上关键词组的热度分析,所述单数据平台上关键词组的热度分析如下:

$$[0011] \quad S_i = \sum_{j=1}^n \frac{p_j d_j}{p_{\text{总}}} \times \log_2 \frac{T-t_i}{T_{\text{周}}};$$

[0012] 其中, S_i 为某关键词组在第*i*个数据平台上的热度值, p_j 为此关键词组相关的第*j*篇文章的所有评论数量,设此关键词组 S_i 下共有 n 篇相关文章,满足 $1 \leq j \leq n$, d_j 为此关键词组相关的第*j*篇文章的所有评论点赞数, $p_{\text{总}}$ 为该平台上所有关键词组相关评论数量, T 为当前时间, t_i 为第*j*篇文章的发布时间, $T-t_i$ 为第*j*篇文章的发布时间与当前时间相差的时间, $T_{\text{周}}$ 为采集周期,满足 $T_{\text{周}} = 72h$;

[0013] 所述全网络平台上关键词组的热度分析满足:

$$[0014] \quad C = \sum_{i=1}^m S_i \times \gamma_i;$$

[0015] 其中, C 为某关键词组在全网络平台上的热度值,设本系统共采集了 m 个数据平台的网络信息数据,满足 $1 \leq i \leq m$; γ_i 为第*i*个数据平台在全网络平台上的权重值;

[0016] 进一步的,所述情感分析模块包括第一分析模块、第二分析模块和分析结果融合模块,所述第一分析模块对文章文本、文章评论文本和文章评论图片进行图文融合生成融合特征,并对融合特征进行情感分析;所述第二分析模块根据文章文本和文章评论文本生成文本特征,并对文本特征进行情感分析;所述分析结果融合模块用于融合第一分析模块和第二分析模块的分析结果生成最终分析结果;

[0017] 进一步的,所述第一分析模块包括融合词获取单元,所述融合词获取单元用于获取文章文本和文章评论文本中的融合词,所述融合词包括程度副词及程度副词后一词汇、情感词及程度副词后一词汇、评价词及评价词后一词汇和否定词及否定词后一词汇;

[0018] 进一步的,所述分析结果管理模块向用户展示的各关键词组热度热度和舆情影响包括:展示各数据平台上按热度大小顺序排列的关键词组及其最终分析结果和全网络平台上按热度大小顺序排列的关键词组及其最终分析结果。

[0019] 本发明所取得的有益效果是:

[0020] 本发明通过数据采集模块采集并筛选寻找出各数据平台上的关键词组,通过热度分析模块根据各个数据平台上关键词组相关的文本、评论和点赞数量对各个关键词组的热度进行分析,生成各关键词组在单数据平台和全网络平台上的热度排行;通过情感分析模块一方面对融合词与图片结合进行情感分析,另一方面对文本进行情感分析,并将两方面

的分析结果结合生成最终分析结果,保证了对于舆情分析的准确性;通过分析结果管理模块向用户可视化展示各关键词组热度和舆情影响,方便用户查看并做出相应举措。

附图说明

[0021] 从以下结合附图的描述可以进一步理解本发明。图中的部件不一定按比例绘制,而是将重点放在示出实施例的原理上。在不同的视图中,相同的附图标记指定对应的部分。

[0022] 图1为本发明整体模块示意图。

[0023] 图2为本发明情感分析流程示意图。

具体实施方式

[0024] 为了使得本发明的目的、技术方案及优点更加清楚明白,以下结合其实施例,对本发明进行进一步详细说明;应当理解,此处所描述的具体实施例仅用于解释本发明,并不用于限定本发明;对于本领域技术人员而言,在查阅以下详细描述之后,本实施例的其它系统、方法和/或特征将变得显而易见;旨在所有此类附加的系统、方法、特征和优点都包括在本说明书内;包括在本发明的范围内,并且受所附权利要求书的保护;在以下详细描述描述了所公开的实施例的另外的特征,并且这些特征根据以下将详细描述将是显而易见的。

[0025] 本发明实施例的附图中相同或相似的标号对应相同或相似的部件;在本发明的描述中,需要理解的是,若有术语“上”、“下”、“左”、“右”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或组件必须具有特定的方位,以特定的方位构造和操作,因此附图中描述位置关系的用语仅用于示例性说明,不能理解为对本专利的限制,对于本领域的普通技术人员而言,可以根据具体情况理解上述术语的具体含义。

[0026] 实施例一:

[0027] 如图1和图2所示,本实施例提供一种基于AIGC的舆情分析编排系统,其特征在于,包括数据采集模块、热度分析模块、情感分析模块和分析结果管理模块;

[0028] 所述数据采集模块用于采集网络信息数据并在网络信息数据中筛选出各关键词组和关键信息;所述热度分析模块用于根据关键信息分析所述各关键词组在网络中的影响热度;所述情感分析模块用于根据关键信息分析所述对各关键词组的舆情正负面进行分析,所述分析结果管理模块用于根据所述影响热度和舆情正负面分析向用户可视化展示各关键词组热度和舆情影响;

[0029] 所述数据采集模块包括采集子模块、预处理模块和预存储模块,所述采集子模块用于在设定的采集周期内对互联网上各数据平台上公开的网络信息数据进行采集,采集的所述网络信息数据包括通过各数据平台的搜索引擎获取的关键词组,所述预处理模块用于对网络信息数据进行清洗筛选获取关键信息,所述预存储模块用于存储关键词组和关键信息;

[0030] 所述各数据平台包括微博、贴吧、各论坛和各新闻网络平台;

[0031] 所述采集子模块对各数据平台上的网络信息数据的采集方式为通过接入各数据平台上的开放API接口获取数据或通过基于python的scrapy框架获取数据;

[0032] 所述预处理模块包括数据清洗模块和分类模块,所述数据清洗模块用于去除网络

信息数据中文本的无效信息,所述分类模块用于将去除无效信息后的网络信息数据根据关键词组进行分类生成关键信息,分类后生成的关键信息包括与各关键词组相关的文章文本、文章发布时间、文章评论文本、文章评论图片和文章评论点赞数量;

[0033] 所述无效信息包括标点符号、停用词和无具体意义的词汇;

[0034] 所述分类模块通过对去除无效信息后的网络信息数据的文章中词频进行检测,并根据检测后生成的特征词与各关键词组的相关性进行判断以将文章归类到各关键词组,检测的算法为已知常用的信息检索的方法,如TF-IDF算法;

[0035] 所述热度分析模块对各关键词组在网络中影响热度的分析包括对单数据平台上关键词组的热度分析和全网络平台上关键词组的热度分析,所述单数据平台上关键词组的热度分析如下:

$$[0036] \quad S_i = \sum_{j=1}^n \frac{p_j d_j}{p_{\text{总}}} \times \log_2 \frac{T-t_i}{T_{\text{周}}};$$

[0037] 其中, S_i 为某关键词组在第*i*个数据平台上的热度值, p_j 为此关键词组相关的第*j*篇文章的所有评论数量,设此关键词组 S_i 下共有*n*篇相关文章,满足 $1 \leq j \leq n$, d_j 为此关键词组相关的第*j*篇文章的所有评论点赞数, $p_{\text{总}}$ 为该平台上所有关键词组相关评论数量, T 为当前时间, t_i 为第*j*篇文章的发布时间, $T-t_i$ 为第*j*篇文章的发布时间与当前时间的的时间差值, $T_{\text{周}}$ 为采集周期,优选的,满足 $T_{\text{周}} = 72h$;

[0038] 所述全网络平台上关键词组的热度分析满足:

$$[0039] \quad C = \sum_{i=1}^m S_i \times \gamma_i;$$

[0040] 其中, C 为某关键词组在全网络平台上的热度值,设本系统共采集了*m*个数据平台的网络信息数据,满足 $1 \leq i \leq m$; γ_i 为第*i*个数据平台在全网络平台上的权重值;

[0041] 所述情感分析模块包括第一分析模块、第二分析模块和分析结果融合模块,所述第一分析模块对文章文本、文章评论文本和文章评论图片进行图文融合生成融合特征,并对融合特征进行情感分析;所述第二分析模块根据文章文本和文章评论文本生成文本特征,并对文本特征进行情感分析;所述分析结果融合模块用于融合第一分析模块和第二分析模块的分析结果生成最终分析结果;本实施例的所述融合具体可采用现有的特征融合算法实现,所述的分析具体可根据提取的特征基于预设的神经网络进行分析得出结果,也可根据提取的特征与预设的经验数据库进行比对得出分析结果,在此不再赘述;

[0042] 所述分析结果管理模块向用户展示的各关键词组热度和舆情影响包括:展示各数据平台上按热度大小顺序排列的关键词组及其最终分析结果和全网络平台上按热度大小顺序排列的关键词组及其最终分析结果。

[0043] 本实施例通过数据采集模块获取并筛选出互联网上的出现热点相关的关键词组及其相关信息;通过热度分析模块对各热点相关的关键词组进行分析运算,从而完成单数据平台和全网络平台的关键词组的热度分析。

[0044] 实施例二:

[0045] 本实施例应当理解为至少包含前述任一个实施例的全部特征,并在其基础上进一步改进;

[0046] 本实施例提供一种基于AIGC的舆情分析编排系统,其特征在于,所述情感分析模块包括第一分析模块、第二分析模块和分析结果融合模块,所述第一分析模块对文章文本、

文章评论文本和文章评论图片进行图文融合生成融合特征,并对融合特征进行情感分析;所述第二分析模块根据文章文本和文章评论文本生成文本特征,并对文本特征进行情感分析;所述分析结果融合模块用于融合第一分析模块和第二分析模块的分析结果生成最终分析结果;

[0047] 所述第一分析模块包括融合词获取单元,所述融合词获取单元用于获取文章文本和文章评论文本中的融合词,所述融合词包括程度副词及该程度副词后一词汇、情感词及该情感词后一词汇、评价词及该评价词后一词汇、否定词及该否定词后一词汇。

[0048] 所述融合词获取单元包括一个预先建立的融合词词库,融合词词库中包含有常见的程度副词(如很、非常、相当)、情感词(如高兴、愉快、伤心)、评价词(如年轻、优雅)和否定词(如不、没);所述融合词获取单元通过将获取到的所述文章文本和文章评论文本分别与融合词词库相互匹配,匹配方式为将词语一一配对,将文章文本和文章评论文本中与融合词词库匹配相似度较高的词语及其后一词汇提取处理作为融合词,其中,匹配相似度较高的词语作为融合词的前半部分,对应的将上述融合词前半部分中匹配相似度较高的词语的后一词汇提取出来作为融合词的后半部分,将融合词的前半部分与后半部分结合从而获取到文章文本和文章评论文本中完整的融合词;

[0049] 现如今网络语言更新迭代速度快,单一图片的分析并不能准确的获取用户所表达的情感,用户可能会发表反讽的文本或图片;所述融合词通常会表达出用户的真实情绪,故通过将融合词与图片结合到一起进行分析可以更准确的获取用户的情感;

[0050] 所述第一分析模块通过word2vec训练语料库将融合词转化为融合词词向量的形式以便机器学习,通过建立一个图片CNN情感分析模型对图片特征进行提取,所述图片CNN情感分析模型包括多个卷积层、多个池化层和两个全连接层,所述图片CNN情感分析模型的输入为多个 $32 \times 32 \times 3$ 的图片,其中 $32 \times 32 \times 3$ 代表图片的宽度 \times 高度 \times 图片维度,所述图片CNN情感分析模型的输出为图片的情绪分析结论,即所述两个全连接层中的后一全连接层包括代表图片的情绪为积极或消极的两个维度;本实施例使用图片CNN情感分析模型仅用于提取表达图片的特征向量,即提取出两个全连接层中的前一全连接层的特征进行向量化作为图片的特征向量,所述图片的特征向量如下式所示:

$$[0051] \quad U_j = f[w(\text{image}_j) + b];$$

[0052] 其中, U_j 为第 j 张图片的特征向量, f 为激活函数, image_j 为两全连接层中的前一全连接层的特征, w 为图片CNN情感分析模型的权重矩阵, b 为偏置项;

[0053] 所述第一分析模块建立一个融合CNN情感分析模型来提取融合特征,所述融合CNN情感分析模型的输入为图片特征向量和融合词向量的融合,其表达方式如下:

$$[0054] \quad Z_i = V_i \oplus V_{i+1} \oplus U_i;$$

[0055] 其中, Z_i 为第 i 个融合CNN情感分析模型的输入, V_i 和 V_{i+1} 表示第 i 组融合词特征的向量, V_i 表示第 i 组融合词的前半部分的特征向量, V_{i+1} 表示第 i 组融合词的后半部分的特征向量,所述每组融合词为某个程度副词及该程度副词后一词汇、或情感词及该情感词后一词汇、或评价词及该评价词后一词汇、或否定词及该否定词后一词汇, U_i 为第 i 组融合词对应图片的特征向量, \oplus 为拼接操作;

[0056] 所述融合CNN模型包括多个卷积层、多个池化层和一个全连接层,将融合CNN模型的全连接层特征向量化即可得到融合特征,将融合特征放入第一分类器中训练继而得到第

一分析模块的分析结果,所述第一分析模块的分析结果包括情绪为积极或消极的投票数;

[0057] 所述第二分析通过word2vec训练语料库将文章文本和文章评论文本转化为文本词向量的形式以便机器学习,通过建立一个文本CNN情感分析模型来对文本特征提取,所述文本CNN情感分析模型的输入为文本词向量,其表达方式如下:

[0058] $V_w = V_1 \oplus V_2 \oplus \dots \oplus V_{n-1} \oplus V_n$;

[0059] 其中, V_w 为某个文本CNN情感分析模型的输入, n 为文本中句子的最大长度, V_k 表示文本句子中的某个文本词向量,满足 $1 \leq k \leq n$;

[0060] 所述文本CNN情感分析模型包括多个卷积层、多个池化层和一个全连接层,将所述文本CNN情感分析模型的全连接层特征向量化即可得到文本特征,将文本特征放入第二分类器中训练继而得到第二分析模块的分析结果,所述第二分析模块的分析结果包括情绪为积极或消极的投票数;

[0061] 所述分析结果融合模块通过将第一分类器和第二分类器中投票为同类型情绪的票数进行相加,将获得票数多的类型作为最终分析结果;

[0062] 所述情感分析模块对于各关键词组的舆情正负面的分析包括生成单数据平台上最终分析结果和全网络平台上的最终分析结果,所述情感分析模块通过对单数据平台上的文章文本、文章评论文本和文章评论图片进行分析获取单数据平台上的最终分析结果,通过对全网络平台上的文章文本、文章评论文本和文章评论图片进行分析获取全网络平台上的最终分析结果;

[0063] 所述分析结果管理模块向用户展示的各关键词组热度和舆情影响包括:展示各数据平台上按热度大小顺序排列的关键词组及其最终分析结果和全网络平台上按热度大小顺序排列的关键词组及其最终分析结果。

[0064] 本实施例通过设置融合词,并将融合词与图片结合后对其进行情感分析,可减少对于用户情感的误判;将融合词与图片结合的分析结果与文本的分析结果相融合得到最终的分析结果,可最大程度上判断出用户的真实情感。

[0065] 以上所公开的内容仅为本发明的优选可行实施例,并非因此局限本发明的保护范围,所以凡是运用本发明说明书及附图内容所做的等效技术变化,均包含于本发明的保护范围内,此外,随着技术发展其中的元素可以更新的。

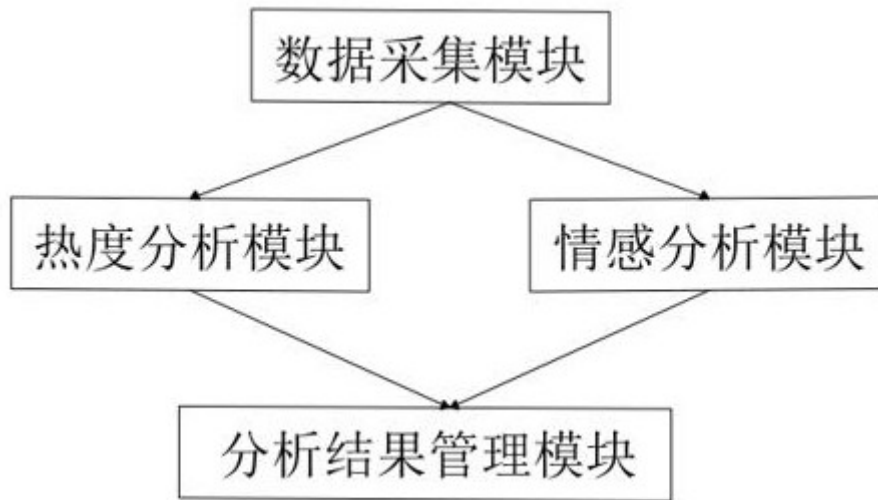


图 1

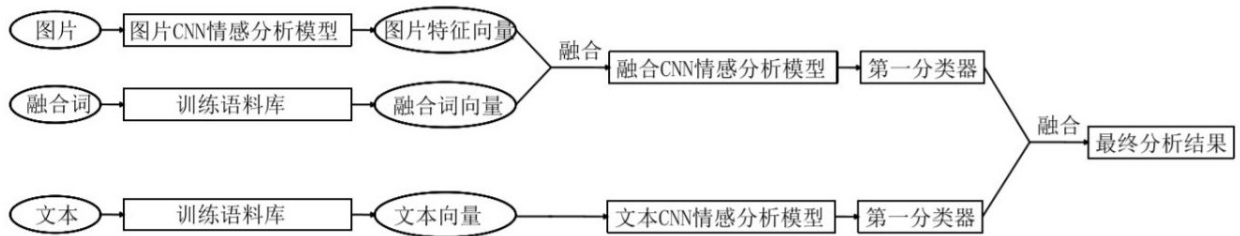


图 2