



(12) 发明专利申请

(10) 申请公布号 CN 113095076 A

(43) 申请公布日 2021.07.09

(21) 申请号 202110426541.0

(22) 申请日 2021.04.20

(71) 申请人 平安银行股份有限公司

地址 518000 广东省深圳市罗湖区深南东路5047号

(72) 发明人 彭晶晶

(74) 专利代理机构 深圳市沃德知识产权代理事

务所(普通合伙) 44347

代理人 高杰 于志光

(51) Int. Cl.

G06F 40/289 (2020.01)

G06F 16/35 (2019.01)

G06F 40/242 (2020.01)

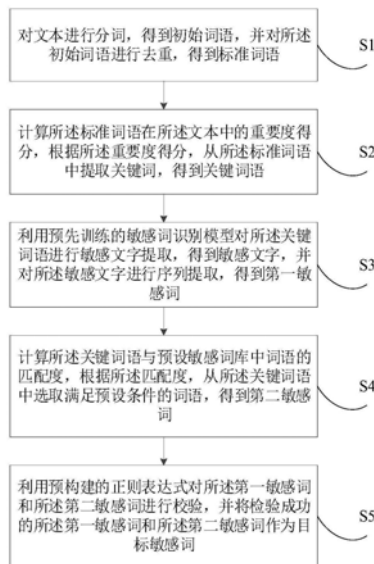
权利要求书2页 说明书11页 附图4页

(54) 发明名称

敏感词识别方法、装置、电子设备及存储介质

(57) 摘要

本发明涉及一种人工智能领域,揭露了一种敏感词识别方法,包括:对文本进行分词和去重,得到标准词语;计算标准词语在文本中的重要度得分,根据重要度得分,从标准词语中提取关键词,得到关键词语;利用预训练的敏感词识别模型对关键词语进行敏感文字提取,得到敏感文字,对敏感文字进行序列提取,得到第一敏感词;计算关键词语与预设敏感词库中词语的匹配度,根据匹配度,从关键词语中选取满足预设条件的词语,得到第二敏感词;利用预构建的正则表达式对第一敏感词和第二敏感词进行校验,将检验成功的第一敏感词和第二敏感词作为目标敏感词。此外,本发明还涉及区块链技术,所述目标敏感词可存储于区块链中。本发明可以提高敏感词识别的准确性。



1. 一种敏感词识别方法,其特征在于,所述方法包括:

对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

2. 如权利要求1所述的敏感词识别方法,其特征在于,所述对所述初始词语进行去重,得到标准词语,包括:

通过遍历操作依次从所述初始词语中选取一个词语作为当前初始词语;

计算所述当前初始词语与所述初始词语中剩余每个初始词语之间的距离值;

在所述初始词语中删除与所述当前初始词语的距离值大于预设阈值的初始词语,并将剩余的初始词语作为标准词语。

3. 如权利要求1所述的敏感词识别方法,其特征在于,所述计算所述标准词语在所述文本中的重要度得分,包括:

计算所述标准词语中任意两个词语在所述文本中的依存关联度;

根据所述依存关联度,计算所述任意两个词语的引力;

根据所述依存关联度和所述引力,计算所述任意两个词语的关联强度;

根据所述关联强度,计算所述任意两个词语的重要度得分。

4. 如权利要求3所述的敏感词识别方法,其特征在于,所述计算所述标准词语中任意两个词语在所述文本中的依存关联度,包括:

计算所述标准词语中任意两个词语在所述文本中的依存路径长度;

根据所述依存路径长度及预设的超参数,得到所述任意两个词语在所述文本中的依存关联度。

5. 如权利要求1中所述的敏感词识别方法,其特征在于,所述利用预先训练的所述敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,包括:

利用所述敏感词识别模型中敏感词分类模块的卷积层对所述关键词语进行卷积操作,得到初始敏感文字;

利用所述敏感词识别模型中敏感词分类模块的池化层对所述初始敏感文字进行降维,得到降维敏感文字;

利用所述敏感词识别模型中敏感词分类模块的全连接层输出所述降维敏感文字,得到敏感文字。

6. 如权利要求1至5中任一项所述的敏感词识别方法,其特征在于,所述对所述敏感文字进行序列提取,生成第一敏感词包括:

利用所述敏感词识别模型中敏感词回归模块的输入门计算所述敏感文字的状态值;

利用所述敏感词识别模型中敏感词回归模块的遗忘门计算所述敏感文字的激活值;

根据所述状态值和激活值计算所述敏感文字的状态更新值；

利用所述敏感词识别模型中敏感词回归模块的输出门计算所述状态更新值的文字位置序列,生成第一敏感词。

7. 如权利要求1所述的敏感词识别方法,其特征在于,所述利用所述敏感词识别模型中敏感词回归模块的输入门计算所述敏感文字的状态值,包括:

利用所述输入门中细胞单元接收所述敏感文字,并确定所述敏感文字在所述输入门中细胞单元的偏置、权重及激活因子;

识别所述敏感文字在所述输入门中细胞单元的文字序列,根据所述输入门中细胞单元的偏置、权重、激活因子以及文字序列,计算所述敏感文字的状态值。

8. 一种敏感词识别装置,其特征在于,所述装置包括:

预处理模块,用于对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

计算模块,用于计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

提取模块,用于利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

选取模块,用于计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

生成模块,用于利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

9. 一种电子设备,其特征在于,所述电子设备包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的计算机程序,所述计算机程序被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1至7中任意一项所述的敏感词识别方法。

10. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至7中任意一项所述的敏感词识别方法。

## 敏感词识别方法、装置、电子设备及存储介质

### 技术领域

[0001] 本发明涉及人工智能领域,尤其涉及一种敏感词识别方法、装置、电子设备及计算机可读存储介质。

### 背景技术

[0002] 敏感词指的是不文明词汇,具有反政、色情、反动及言语不当的词汇,因此,敏感词识别可以理解为查找出文本中不文明词汇的过程,其应用于不同的行业领域中,例如在金融宣传场景当中,金融企业的各个营销单元每天都产出海量宣传文案,这些宣传文案良莠不齐,容易出现金融敏感词,为保障宣传文案的合法合规性,通常企业会在宣传文案发布前,进行宣传文案的审核,找出宣传文案中的敏感词,从而规避企业宣传的非法性。

[0003] 目前文本的敏感词通常是基于字符匹配进行识别,即通过字符匹配算法将文本中每个字符进行扫描,如果当某一个字符是敏感字表中任意一个敏感词的第一个字符,则触发敏感词检索,但由于文本中涉及到大量错综复杂的词语,若是基于字符匹配很容易造成敏感词的模糊匹配,影响敏感词识别的准确率。

### 发明内容

[0004] 本发明提供一种敏感词识别方法、装置、电子设备及计算机可读存储介质,其主要目的在于提高敏感词识别的准确性。

[0005] 为实现上述目的,本发明提供了一种敏感词识别方法,包括:

[0006] 对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

[0007] 计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

[0008] 利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

[0009] 计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

[0010] 利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将校验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0011] 可选地,所述对所述初始词语进行去重,得到标准词语,包括:

[0012] 通过遍历操作依次从所述初始词语中选取一个词语作为当前初始词语;

[0013] 计算所述当前初始词语与所述初始词语中剩余每个初始词语之间的距离值;

[0014] 在所述初始词语中删除与所述当前初始词语的距离值大于预设阈值的初始词语,并将剩余的初始词语作为标准词语。

[0015] 可选地,所述计算所述标准词语在所述文本中的重要度得分,包括:

[0016] 计算所述标准词语中任意两个词语在所述文本中的依存关联度;

[0017] 根据所述依存关联度,计算所述任意两个词语的引力;

- [0018] 根据所述依存关联度和所述引力,计算所述任意两个词语的关联强度;
- [0019] 根据所述关联强度,计算所述任意两个词语的重要度得分。
- [0020] 可选地,所述计算所述标准词语中任意两个词语在所述文本中的依存关联度,包括:
- [0021] 计算所述标准词语中任意两个词语在所述文本中的依存路径长度;
- [0022] 根据所述依存路径长度及预设的超参数,得到所述任意两个词语在所述文本中的依存关联度。
- [0023] 可选地,所述利用预先训练的所述敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,包括:
- [0024] 利用所述敏感词识别模型中敏感词分类模块的卷积层对所述关键词语进行卷积操作,得到初始敏感文字;
- [0025] 利用所述敏感词识别模型中敏感词分类模块的池化层对所述初始敏感文字进行降维,得到降维敏感文字;
- [0026] 利用所述敏感词识别模型中敏感词分类模块的全连接层输出所述降维敏感文字,得到敏感文字。
- [0027] 可选地,所述对所述敏感文字进行序列提取,生成第一敏感词包括:
- [0028] 利用所述敏感词识别模型中敏感词回归模块的输入门计算所述敏感文字的状态值;
- [0029] 利用所述敏感词识别模型中敏感词回归模块的遗忘门计算所述敏感文字的激活值;
- [0030] 根据所述状态值和激活值计算所述感文字的状态更新值;
- [0031] 利用所述敏感词识别模型中敏感词回归模块的输出门计算所述状态更新值的文字位置序列,生成第一敏感词。
- [0032] 可选地,所述利用所述敏感词识别模型中敏感词回归模块的输入门计算所述敏感文字的状态值,包括:
- [0033] 利用所述输入门中细胞单元接收所述敏感文字,并确定所述敏感文字在所述输入门中细胞单元的偏置、权重及激活因子;
- [0034] 识别所述敏感文字在所述输入门中细胞单元的文字序列,根据所述输入门中细胞单元的偏置、权重、激活因子以及文字序列,计算所述敏感文字的状态值。
- [0035] 为了解决上述问题,本发明还提供一种敏感词识别装置,所述装置包括:
- [0036] 预处理模块,用于对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;
- [0037] 计算模块,用于计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;
- [0038] 提取模块,用于利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;
- [0039] 选取模块,用于计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;
- [0040] 生成模块,用于利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进

行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0041] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:

[0042] 至少一个处理器;以及,

[0043] 与所述至少一个处理器通信连接的存储器;其中,

[0044] 所述存储器存储有可被所述至少一个处理器执行的计算机程序,所述计算机程序被所述至少一个处理器执行,以实现上述所述的敏感词识别方法。

[0045] 为了解决上述问题,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一个计算机程序,所述至少一个计算机程序被电子设备中的处理器执行以实现上述所述的敏感词识别方法。

[0046] 本发明实施例首先对获取的文本进行分词、去重以及关键词提取,得到关键词语,可以筛选出所述文本中的一些无用词语,提高后续文本的敏感词识别速度;其次,本发明实施例利用敏感词识别模型和敏感词库分别对所述关键词语进行敏感词识别,得到第一敏感词第二敏感词,可以保障识别的敏感词全面性,避免所述文本中一些敏感词遗漏造成敏感词模糊匹配的情况,提高敏感词识别的准确率;进一步地,本发明实施例利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词,可以进一步保障识别的敏感词准确率。因此,本发明提出的一种敏感词识别方法、装置、电子设备以及存储介质可以提高敏感词识别的准确性。

## 附图说明

[0047] 图1为本发明一实施例提供的敏感词识别方法的流程示意图;

[0048] 图2为本发明第一实施例中图1提供的敏感词识别方法其中一个步骤的详细流程示意图;

[0049] 图3为本发明一实施例提供的敏感词识别装置的模块示意图;

[0050] 图4为本发明一实施例提供的实现敏感词识别方法的电子设备的内部结构示意图;

[0051] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

## 具体实施方式

[0052] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0053] 本申请实施例提供一种敏感词识别方法。所述敏感词识别方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之,所述敏感词识别方法可以由安装在终端设备或服务端设备的软件或硬件来执行,所述软件可以是区块链平台。所述服务端包括但不限于:单台服务器、服务器集群、云端服务器或云端服务器集群等。

[0054] 参照图1所示,为本发明一实施例提供的敏感词识别方法的流程示意图。在本发明实施例中,所述敏感词识别方法包括:

[0055] S1、对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语。

[0056] 本发明实施例中,所述文本包括金融领域中企业的宣传文案。进一步地,本发明通过预设的匹配策略将所述文本与预设词典中的词表进行匹配,得到所述初始词语。其中,所

述预设的匹配策略可以为正向最大匹配法,即按照从左到右的顺序将所述文本中的几个连续字符与词表匹配,如果匹配上,则切分出一个词。所述预设词典包括:停用词词典、统计字典以及前缀词典等。

[0057] 其中,所述停用词词典用于删除所述文本中出现频率高,但无实际意义的词,比如,语气助词、副词、介词以及连词等;所述统计词典用于与文本中句子的字符进行匹配;所述前缀词典用于将匹配成功的词进行切分。

[0058] 例如,文本中的一个句子为“特价理财产品的信息”,通过所述停用词删除句子中的“的”,得到句子“特供理财产品信息”,通过所述统计词典进行字符匹配,得到句子中“特供”“理财”“产品”“信息”为匹配成功的词语,利用所述前缀词典将匹配成功的词进行切分,得到词语为:“特供”、“理财”、“理财产品”、“产品”、“信息”以及“产品信息”。

[0059] 应该了解,在所述初始词语中可能会存在相同词语,为了提高后续初始词语的处理速度,本发明实施对所述初始词语进行去重,以减少后续初始词语的计算量,提高初始词语的处理时效。

[0060] 具体地,所述对所述初始词语进行去重,得到标准词语,包括:通过遍历操作依次从所述初始词语中选取一个词语作为当前初始词语;计算所述当前初始词语与所述初始词语中剩余每个初始词语之间的距离值;在所述初始词语中删除与所述当前初始词语的距离值大于预设阈值的初始词语,并将剩余的初始词语作为标准词语。

[0061] 一个可选实施例中,利用下述公式计算所述当前初始词语与所述初始词语中剩余每个初始词语之间的距离值:

$$[0062] \quad d = \text{sqrt}\left(\sum_{j=1}^n (w_{1j} - w_{2j})^2\right)$$

[0063] 其中,d表示距离值, $w_{1j}$ 表示当前初始词语, $w_{2j}$ 表示初始词语中剩余第2j个词语。

[0064] S2、计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词。

[0065] 应该了解,在金融领域的营销宣传行为中,为了突出相对应的宣传材料主题,会在文本中凸显出一些关键词语,以吸引用户眼球,比如对于理财产品营销宣传行为中,往往会出现“领先行业”、“独创”等初始词语,为了提高后续文本的敏感词识别速度,本发明实施例通过计算标准词语在所述文本中的重要度得分,以从所述标准词语提取关键词。

[0066] 详细地,参阅图2所示,所述计算所述标准词语在所述文本中的重要度得分,包括:

[0067] S20、计算所述标准词语中任意两个词语在所述文本中的依存关联度;

[0068] S21、根据所述依存关联度计算所述任意两个词语的引力;

[0069] S22、根据所述依存关联度和所述引力计算所述任意两个词语的关联强度;

[0070] S23、根据所述关联强度计算所述任意两个词语的重要度得分。

[0071] 一个可选实施例中,所述计算所述标准词语中任意两个词语在所述文本中的依存关联度,包括:计算所述标准词语中任意两个词语在所述文本中的依存路径长度,根据所述依存路径长度及预设的超参数,得到所述任意两个词语在所述文本中的依存关联度,可选的,利用下述公式计算所述标准词语中任意两个词语在所述文本中的依存关联度:

$$[0072] \quad \text{Dep}(W_i, W_j) = \frac{1}{b^{\text{len}(W_i, W_j)}}$$

[0073] 其中,  $\text{Dep}(W_i, W_j)$  表示词语  $W_i$  和  $W_j$  的依存关联度,  $\text{len}(W_i, W_j)$  表示词语  $W_i$  和  $W_j$  之间的依存路径长度,  $b$  是超参数。

[0074] 一个可选实施例中, 利用下述公式计算所述任意两个词语的引力:

$$[0075] \quad f_{\text{grav}}(W_i, W_j) = \frac{\text{tfidf}(W_i) * \text{tfidf}(W_j)}{d^2}$$

[0076] 其中,  $f_{\text{grav}}(W_i, W_j)$  表示词语  $W_i$  和  $W_j$  的引力,  $\text{tfidf}(W_i)$  表示词语  $W_i$  的 TF-IDF 值,  $\text{tfidf}(W_j)$  表示词语  $W_j$  的 TF-IDF 值, TF 表示词频, IDF 表示逆文档频率指数,  $d$  是词语  $W_i$  和  $W_j$  的欧式距离。

[0077] 一个可选实施例中, 利用下述公式计算所述任意两个词语的关联强度:

$$[0078] \quad \text{weight}(W_i, W_j) = \text{Dep}(W_i, W_j) * f_{\text{grav}}(W_i, W_j)$$

[0079] 一个可选实施例中, 利用下述公式计算所述任意两个词语的重要度得分:

$$[0080] \quad \text{WS}(W_i) = 1 - \eta + \eta \sum_{W_j \in C_{W_i}} \frac{\text{weight}(W_i, W_j)}{\sum_{W_k \in C_{W_j}} \text{weight}(W_j, W_k)} \text{WS}(W_j)$$

[0081] 其中,  $\text{WS}(W_i)$  表示词语  $W_i$  的重要度得分,  $C_{W_i}$  是与顶点  $W_i$  有关的集合,  $\eta$  为阻尼系数。

[0082] 进一步地, 所述从所述标准词语中提取关键词, 得到关键词语包括: 从所述标准词语中选取重要度得分大于预设阈值的标准词语, 得到关键词语, 可选的, 所述预设阈值为 0.75, 也可以根据实际场景设置。

[0083] S3、利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取, 得到敏感文字, 并对所述敏感文字进行序列提取, 得到第一敏感词。

[0084] 本发明实施例中, 所述敏感词识别模型包括: 敏感词分类模块和敏感词回归模块, 其中, 所述敏感词分类模块用于识别所述关键词语的敏感文字, 所述敏感词回归模块用于提取敏感文字中文字序列。进一步地, 所述敏感词分类模块包括卷积层、池化层以及全连接层, 所述敏感词回归模块包括: 输入门、遗忘门以及输出门。

[0085] 进一步地, 所述利用预先训练的所述敏感词识别模型对所述关键词语进行敏感文字提取, 得到敏感文字, 包括: 利用所述敏感词识别模型中敏感词分类模块的卷积层对所述关键词语进行卷积操作, 得到初始敏感文字; 利用所述敏感词识别模型中敏感词分类模块的池化层对所述初始敏感文字进行降维, 得到降维敏感文字; 利用所述敏感词识别模型中敏感词分类模块的全连接层输出所述降维敏感文字, 得到敏感文字。

[0086] 一个可选实施例中, 所述卷积操作可以通过所述卷积层中的卷积核实现。

[0087] 一个可选实施例中, 所述初始敏感文字的降维可以通过所述池化层的降维函数实现, 其中, 所述降维函数包括:

$$[0088] \quad x'_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

[0089] 其中,  $x'_i$  为降维敏感文字,  $x_i$  为初始敏感文字,  $\mu$  为初始敏感文字的均值,  $\sigma^2$  为初始



敏感文字的方差,  $\varepsilon$  为无穷小的随机数。

[0090] 进一步地, 所述对所述敏感文字进行序列提取, 生成第一敏感词包括: 利用所述敏感词识别模型中敏感词回归模块的输入门计算所述敏感文字的状态值; 利用所述敏感词识别模型中敏感词回归模块的遗忘门计算所述敏感文字的激活值; 根据所述状态值和激活值计算所述敏感文字的状态更新值; 利用所述敏感词识别模型中敏感词回归模块的输出门计算所述状态更新值的文字位置序列, 生成第一敏感词。

[0091] 其中, 所述状态值用于确定所述敏感文字的文字序列, 所述激活值用于确定所述敏感文字的提取数量, 所述状态更新值用于确定待提取的敏感文字的位置序列, 基于所述敏感词回归模块对敏感文字的序列提取, 可以控制所述敏感文字的信息流动, 以确保最终生成的敏感词的信息完整性。

[0092] 一个可选实施例中, 所述利用所述敏感词识别模块中敏感词回归模块的输入门计算所述敏感文字的状态值, 包括: 利用所述输入门中细胞单元接收所述敏感文字, 并确定所述敏感文字在所述输入门中细胞单元的偏置、权重及激活因子, 识别所述敏感文字在所述输入门中细胞单元的文字序列, 根据所述输入门中细胞单元的偏置、权重、激活因子以及文字序列, 计算所述敏感文字的状态值。其中, 所述偏置和权重是指所述输入门的网络参数, 所述激活因子用于将输入门生成的状态值激活, 可以通过激活函数 (如 Sigmoid) 进行设置, 进一步地, 所述状态值可以通过下述公式计算:

$$[0093] \quad i_t = \sigma(w_i \cdot [h_{t-1}, x_t]) + b_i$$

[0094] 其中,  $i_t$  表示状态值,  $\sigma$  表示输入门中细胞单元的偏置,  $w_i$  表示输入门中细胞单元的激活因子,  $h_{t-1}$  表示敏感文字在输入门中细胞单元  $t-1$  时刻的文字序列,  $x_t$  表示敏感文字在输入门中细胞单元  $t$  时刻的敏感文字,  $b_i$  表示输入门中细胞单元的权重。

[0095] 一个可选实施例中, 所述利用所述敏感词识别模型中敏感词回归模块的遗忘门计算所述敏感文字的激活值包括: 利用所述遗忘门中细胞单元接收所述敏感文字, 并确定所述敏感文字在所述遗忘门中细胞单元的偏置、权重及激活因子, 识别所述敏感文字在所述遗忘门中细胞单元的文字序列, 根据所述遗忘门中细胞单元的偏置、权重、激活因子以及文字序列, 计算所述敏感文字的激活值。其中, 所述偏置和权重是指所述遗忘门的网络参数, 所述激活因子用于将遗忘门生成的激活值激活, 可以通过激活函数 (如 Sigmoid) 进行设置, 进一步地, 所述激活值可以通过下述公式计算:

$$[0096] \quad f_t = \sigma(w_f \cdot [h_{t-1}^{\wedge}, x_t]) + b_f$$

[0097] 其中,  $f_t$  表示激活值,  $\sigma$  表示遗忘门中细胞单元的偏置,  $w_f$  表示遗忘门中细胞单元的激活因子,  $h_{t-1}^{\wedge}$  表示敏感文字在所述遗忘门中细胞单元  $t-1$  时刻的峰值,  $x_t$  表示敏感文字在所述遗忘门中细胞单元  $t$  时刻输入的敏感文字,  $b_f$  表示遗忘门中细胞单元的权重。

[0098] 一个可选实施例中, 所述状态更新值可以通过下述公式计算:

$$[0099] \quad c_t = i_t * h_{t-1} + f_t * h_{t-1}^{\wedge}$$

[0100] 其中,  $c_t$  表示状态更新值,  $h_{t-1}$  表示敏感文字在输入门  $t-1$  时刻的峰值,  $h_{t-1}^{\wedge}$  表示敏感文字在遗忘门  $t-1$  时刻的峰值。

[0101] 一个可选实施例中,所述文字位置序列可以通过下述公式计算:

$$[0102] \quad o_t = \tan h(c_t)$$

[0103] 其中, $o_t$ 表示文字位置序列, $\tan h$ 表示输出门的激活函数, $c_t$ 表示状态更新值。

[0104] S4、计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词。

[0105] 本发明实施例中,所述预设敏感词库中词语从历史业务场景中产生的词语获取,比如在历史金融的理财产品营销场景中,经审核发现,该理财产品的宣传文案中“尽享牛市”和“限时特供”等词语为敏感词语,于是将这些词语存入所述预设敏感词库中,可选的,所述预设敏感词库可以通过数据库构建,用于实现数据的快速的存取。

[0106] 进一步地,本发明实施例中,利用下述公式计算所述关键词语与预设敏感词库中词语的匹配度:

$$[0107] \quad T(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} * \sqrt{\sum y_i^2}}$$

[0108] 其中, $T(x, y)$ 表示匹配度, $x_i$ 表示筛选的所述关键词语中第*i*个初始词语, $y_i$ 表示敏感词库中第*i*个中词语。

[0109] 进一步地,本发明实施例根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词,其中所述预设条件可以为选取所述匹配度大于预设匹配度,即若所述关键词语的匹配度大于预设匹配度,则将对应的关键词语筛选出作为所述第二敏感词。

[0110] S5、利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0111] 应该了解,在获取的所述第一敏感词和所述第二敏感词中可能会存在一些格式错误的词语,比如汉字有误,因此本发明实施例利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0112] 一个可选实施例中,所述正则表达式包括:数字校验表达式(如 $\hat{[0-9]*\$}$ )、汉字校验表达式( $\hat{[\u4e00-\u9fa5]\{0,\}\$}$ )以及特殊需求检验表达式(如日期格式: $\hat{\d\{4\}-\d\{1,2\}-\d\{1,2\}}$ )。

[0113] 进一步地,为保障所述目标敏感词的复用性和可用性,所述目标敏感词还可存储于一区块链节点中。

[0114] 进一步地,本发明另一实施例还包括:对所述目标敏感词进行颜色标记,方便用户可以快速找到文本中的敏感词语。

[0115] 本发明实施例首先对获取的文本进行分词、去重以及关键词提取,得到关键词语,可以筛选出所述文本中的一些无用词语,提高后续文本的敏感词识别速度;其次,本发明实施例利用敏感词识别模型和敏感词库分别对所述关键词语进行敏感词识别,得到第一敏感词第二敏感词,可以保障识别的敏感词全面性,避免所述文本中一些敏感词遗漏造成敏感词模糊匹配的情况,提高敏感词识别的准确率;进一步地,本发明实施例利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词,可以进一步保障识别的敏感词准确率。因此,本发明提出

的一种敏感词识别方法可以提高敏感词识别的准确性。

[0116] 如图3所示,是本发明敏感词识别装置的功能模块图。

[0117] 本发明所述敏感词识别装置100可以安装于电子设备中。根据实现的功能,所述敏感词识别装置可以包括预处理模块101、计算模块102、提取模块103、选取模块104以及生成模块105。本发明所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0118] 在本实施例中,关于各模块/单元的功能如下:

[0119] 所述预处理模块101,用于对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

[0120] 所述计算模块102,用于计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

[0121] 所述提取模块103,用于利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

[0122] 所述选取模块104,用于计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

[0123] 所述生成模块105,用于利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0124] 详细地,本发明实施例中所述敏感词识别装置100中的所述各模块在使用时采用与上述的图1和图2中所述的敏感词识别方法一样的技术手段,并能够产生相同的技术效果,这里不再赘述。

[0125] 如图4所示,是本发明实现敏感词识别方法的电子设备的结构示意图。

[0126] 所述电子设备1可以包括处理器10、存储器11和总线,还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如敏感词识别程序12。

[0127] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备1的内部存储单元,例如该电子设备1的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备1的外部存储设备,例如电子设备1上配备的插接式移动硬盘、智能存储卡(Smart Media Card, SMC)、安全数字(Secure Digital, SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备1的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备1的应用软件及各类数据,例如敏感词识别程序12的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0128] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit, CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如执行敏感词识别程序12等),以及调用存储在所述存储器11内的数据,以执行电子设备1的各种功能和处理数据。

[0129] 所述总线可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。

[0130] 图4仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图4示出的结构并不构成对所述电子设备1的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0131] 例如,尽管未示出,所述电子设备1还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备1还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0132] 进一步地,所述电子设备1还可以包括网络接口,可选地,所述网络接口可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备1与其他电子设备之间建立通信连接。

[0133] 可选地,该电子设备1还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备1中处理的信息以及用于显示可视化的用户界面。

[0134] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0135] 所述电子设备1中的所述存储器11存储的敏感词识别程序12是多个程序的组合,在所述处理器10中运行时,可以实现:

[0136] 对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

[0137] 计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

[0138] 利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

[0139] 计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

[0140] 利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将校验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0141] 具体地,所述处理器10对上述程序的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0142] 进一步地,所述电子设备1集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个非易失性计算机可读取存储介质中。所述计算机可读存储介质可以是易失性的,也可以是非易失性的。例如,所述计算机可读介质可以

包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0143] 本发明还提供一种计算机可读存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0144] 对文本进行分词,得到初始词语,并对所述初始词语进行去重,得到标准词语;

[0145] 计算所述标准词语在所述文本中的重要度得分,根据所述重要度得分,从所述标准词语中提取关键词,得到关键词语;

[0146] 利用预先训练的敏感词识别模型对所述关键词语进行敏感文字提取,得到敏感文字,并对所述敏感文字进行序列提取,得到第一敏感词;

[0147] 计算所述关键词语与预设敏感词库中词语的匹配度,根据所述匹配度,从所述关键词语中选取满足预设条件的词语,得到第二敏感词;

[0148] 利用预构建的正则表达式对所述第一敏感词和所述第二敏感词进行校验,并将检验成功的所述第一敏感词和所述第二敏感词作为目标敏感词。

[0149] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0150] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0151] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0152] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0153] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0154] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0155] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0156] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的

技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

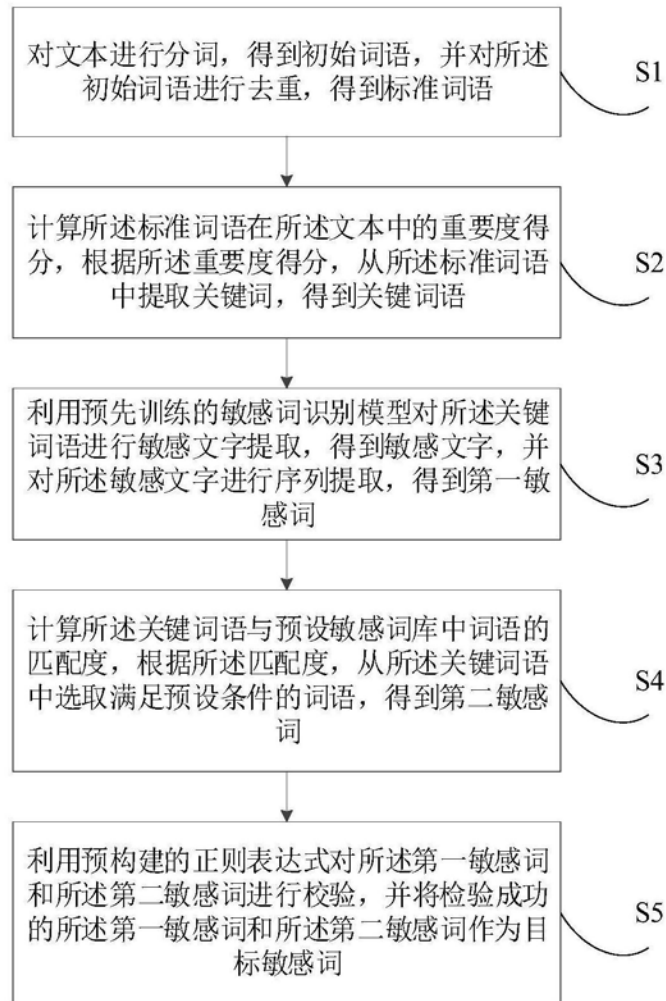


图1

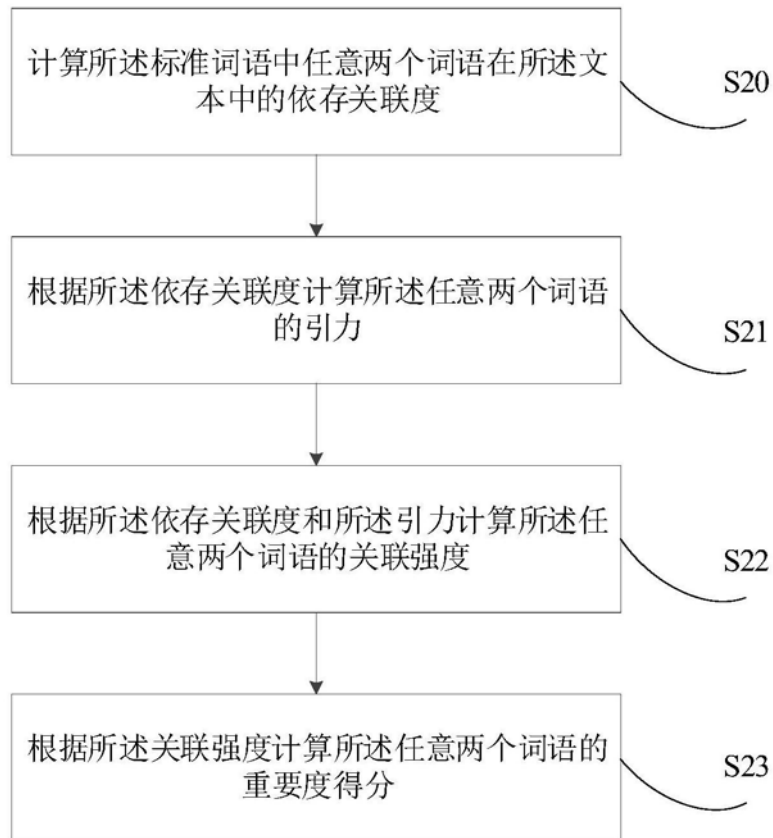


图2





图3



图4