(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 108763202 A
(43)申请公布日 2018.11.06

(21)申请号 201810479412.6

(22)申请日 2018.05.18

(71)申请人 广州腾讯科技有限公司
地址 510000 广东省广州市海珠区新港中路397号品牌区1号

(72)发明人 黄东平

(74)专利代理机构 北京派特恩知识产权代理有限公司 11270
代理人 蒋雅洁 张颖玲

(51)Int.Cl.
G06F 17/27(2006.01)
G06F 17/30(2006.01)

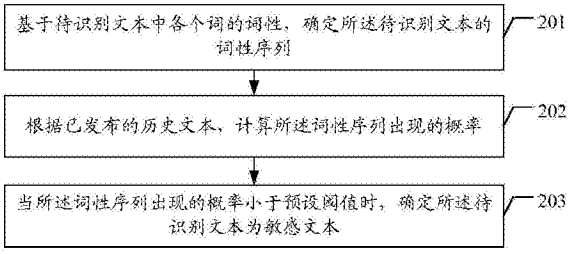
权利要求书2页 说明书10页 附图3页

(54)发明名称

识别敏感文本的方法、装置、设备及可读存储介质

(57)摘要

本发明公开了一种识别敏感文本的方法、装置、设备及可读存储介质，方法包括：基于待识别文本中各个词的词性，确定待识别文本的词性序列，然后根据已发布的历史文本，计算上述词性序列出现的概率，当该词性序列出现的概率小于预设阈值时，确定待识别文本为敏感文本。相较于现有技术而言，本发明实施例通过计算待识别文本的词性序列出现的概率，来确定待识别文本是否为敏感文本，由于不符合语法规定且不常见的文本，其对应的词性序列出现的概率也会较低，因此，当待识别文本的词性序列出现的概率小于预设阈值时，则可以确定待识别文本不符合语法规则且不常见，属于敏感文本。



CN 108763202 A

1.一种识别敏感文本的方法，其特征在于，所述方法包括：

基于待识别文本中各个词的词性，确定所述待识别文本的词性序列；

根据已发布的历史文本，计算所述词性序列出现的概率；

当所述词性序列出现的概率小于预设阈值时，确定所述待识别文本为敏感文本。

2.如权利要求1所述的方法，其特征在于，所述根据已发布的历史文本，计算所述词性序列出现的概率的步骤包括：

采集已发布的非敏感文本，并确定所述非敏感文本中各个词的词性；

基于所述非敏感文本中任意N种词性连续出现的概率，计算所述词性序列出现的概率，N为正整数，且N≥2。

3.如权利要求2所述的方法，其特征在于，所述基于所述非敏感文本中任意N种词性连续出现的概率，计算所述词性序列出现的概率的步骤包括：

利用以下公式计算所述词性序列出现的概率P(T)，

$P(T)=\prod P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$

其中，T表示待识别文本的词性序列，$T=t_1、t_2、\cdots、t_n$，$P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$表示所述非敏感文本中词性$t_{i-N+1}、\cdots、t_{i-2}、t_{i-1}、t_i$连续出现的概率，$t_i$表示所述词性序列中的第i个词性，N表示在非敏感文本中任意N个词性连续出现，$N\leqslant i\leqslant n$。

4.如权利要求1至3任意一项所述的方法，其特征在于，所述基于待识别文本中各个词的词性，确定所述待识别文本的词性序列的步骤包括：

划分所述待识别文本，得到所述待识别文本中的各个词；

查找预置的词性数据库，确定所述各个词的词性；

将所述各个词的词性按照所述各个词在所述待识别文本中的顺序进行排列，得到所述待识别文本的词性序列。

5.如权利要求1所述的方法，其特征在于，所述根据已发布的历史文本，计算所述词性序列出现的概率的步骤包括：

采集历史文本，所述历史文本包括敏感文本与非敏感文本；

将所述历史文本的词性序列作为样本词性序列，训练概率计算模型；

将所述词性序列输入训练后的概率计算模型，输出所述词性序列出现的概率。

6.一种识别敏感文本的装置，其特征在于，所述装置包括：

确定模块，用于基于待识别文本中各个词的词性，确定所述待识别文本的词性序列；

计算模块，用于根据已发布的历史文本，计算所述词性序列出现的概率；

识别模块，用于当所述词性序列出现的概率小于预设阈值时，确定所述待识别文本为敏感文本。

7.如权利要求6所述的装置，其特征在于，所述计算模块包括：

采集模块，用于采集已发布的非敏感文本，并确定所述非敏感文本中各个词的词性；

计算执行模块，用于基于所述非敏感文本中任意N种词性连续出现的概率，计算所述词性序列出现的概率，N为正整数，且N≥2。

8.如权利要求7所述的装置，其特征在于，所述计算执行模块用于：

利用以下公式计算所述词性序列出现的概率P(T)，

$P(T)=\prod P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$

其中，T表示待识别文本的词性序列，$T=t_1、t_2、\cdots、t_n$，$P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$表示所述非敏感文本中词性$t_{i-N+1}、\cdots、t_{i-2}、t_{i-1}、t_i$连续出现的概率，$t_i$表示所述词性序列中的第i个词性，N表示在非敏感文本中任意N个词性连续出现，$N\leqslant i\leqslant n$。

9.如权利要求6至8任意一项所述的装置，其特征在于，所述确定模块包括：

划分模块，用于划分所述待识别文本，得到所述待识别文本中的各个词；

查找模块，用于查找预置的词性数据库，确定所述各个词的词性；

排列模块，用于将所述各个词的词性按照所述各个词在所述待识别文本中的顺序进行排列，得到所述待识别文本的词性序列。

10.如权利要求6所述的装置，其特征在于，所述计算模块还包括：

样本采集模块，用于采集历史文本，所述历史文本包括敏感文本与非敏感文本；

训练模块，用于将所述历史文本的词性序列作为样本词性序列，训练概率计算模型；

输出模块，用于将所述词性序列输入训练后的概率计算模型，输出所述词性序列出现的概率。

11.一种设备，包括存储器、处理器及存储在所述存储器上且在所述处理器上运行的计算机程序，其特征在于，所述计算机程序为识别敏感文本的程序，所述处理器执行所述计算机程序时，实现如权利要求1至5任意一项所述的识别敏感文本的方法中的各个步骤。

12.一种存储介质，其上存储有计算机程序，其特征在于，所述计算机程序为识别敏感文本的程序，所述计算机程序被处理器执行时，实现如权利要求1至5任意一项所述的识别敏感文本的方法中的各个步骤。

识别敏感文本的方法、装置、设备及可读存储介质

技术领域

[0001] 本发明涉及信息处理技术领域，尤其涉及一种识别敏感文本的方法、装置、设备及可读存储介质。

背景技术

[0002] 随着互联网行业的高速发展，网络上的信息得到了极大的丰富，伴随而来 也会产生许多不符合互联网使用环境甚至违反国家法律法规的内容，如政治敏 感话题、不文明言论等等，导致相关的网络平台因违反国家法律法规的强制性 规定，存在安全运营风险；另外，这些负面的信息内容，极大地损害了网络平 台的品牌形象，对用户的使用体验也会造成不良影响。

[0003] 目前，部分网络平台逐渐开发出了面向互联网的敏感信息过滤系统，通过 关键词匹配、语义识别等方式，可以识别互联网中发布的带有暴力、低俗、商 业广告等信息的标题。但是，由于中文的表达方式层出不穷，不法人员为了躲 避打击，经常会在标题中使用生僻字、拼接字体等，导致通用的语义、语法模 型失效，造成审核漏洞，即现有技术中难以识别不符合语法规则的敏感文本。

发明内容

[0004] 本发明的主要目的在于提供一种识别敏感文本的方法、装置、设备及可读 存储介质，可以解决现有技术中难以识别不符合语法规则的敏感文本的技术问 题。

[0005] 为实现上述目的，本发明第一方面提供一种识别敏感文本的方法，该方法 包括：

[0006] 基于待识别文本中各个词的词性，确定所述待识别文本的词性序列；

[0007] 根据已发布的历史文本，计算所述词性序列出现的概率；

[0008] 当所述词性序列出现的概率小于预设阈值时，确定所述待识别文本为敏感 文本。

[0009] 为实现上述目的，本发明第二方面提供一种识别敏感文本的装置，该装置 包括：

[0010] 确定模块，用于基于待识别文本中各个词的词性，确定所述待识别文本的 词性序列；

[0011] 计算模块，用于根据已发布的历史文本，计算所述词性序列出现的概率；

[0012] 识别模块，用于当所述词性序列出现的概率小于预设阈值时，确定所述待 识别文本为敏感文本。

[0013] 为实现上述目的，本发明第三方面提供一种设备，包括存储器、处理器及 存储在所述存储器上且在所述处理器上运行的计算机程序，该计算机程序为识 别敏感文本的程序，所述处理器执行所述计算机程序时，实现本发明第一方面 提供的识别敏感文本的方法中的各个步骤。

[0014] 为实现上述目的，本发明第四方面提供一种存储介质，其上存储有计算机 程序，该计算机程序为识别敏感文本的程序，所述计算机程序被处理器执行时， 实现本发明第一方面提供的识别敏感文本的方法中的各个步骤。

[0015] 本发明实施例所提供的识别敏感文本的方法，包括：基于待识别文本中各 个词的词性，确定待识别文本的词性序列，然后根据已发布的历史文本，计算 上述词性序列出现的概率，当该词性序列出现的概率小于预设阈值时，确定待 识别文本为敏感文本。相较于现有技术而言，本发明实施例通过计算待识别文 本的词性序列出现的概率，来确定待识别文本是否为敏感文本，由于不符合语 法规定且不常见的文本，其对应的词性序列出现的概率也会较低，因此当待识 别文本的词性序列出现的概率小于预设阈值时，则可以确定待识别文本不符合 语法规则且不常见，属于敏感文本。

附图说明

[0016] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施 例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述 中的附图仅仅是本发明的一些实施例，对于本领域技术人员来讲，在不付出创 造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0017] 图1a为本发明实施例中识别敏感文本的方法的运行系统示意图；

[0018] 图1b为本发明实施例中识别敏感文本的方法的实现场景图；

[0019] 图2为本发明实施例中识别敏感文本的方法的步骤流程示意图；

[0020] 图3为本发明实施例中确定待识别文本的词性序列的示意图；

[0021] 图4为本发明实施例中识别敏感文本的方法的另一步骤流程示意图；

[0022] 图5为本发明实施例中识别敏感文本的装置的程序模块示意图；

[0023] 图6为本发明实施例中识别敏感文本的装置的另一程序模块示意图；

[0024] 图7为本发明实施例中设备70的结构示意图。

具体实施方式

[0025] 为使得本发明的发明目的、特征、优点能够更加的明显和易懂，下面将结 合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描 述，显然，所描述的实施例仅仅是本发明一部分实施例，而非全部实施例。基 于本发明中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的 所有其他实施例，都属于本发明保护的范围。

[0026] 由于在现有技术中，当标题中使用生僻字、拼接字体等时，会导致通用的 语义、语法模型失效，通过关键词匹配、语义识别等方式便难以识别，在审核 敏感信息的过程中存在审核漏洞，即现有技术中存在难以识别不符合语法规则 的敏感文本的技术问题。

[0027] 为了解决上述技术问题，本发明提出一种识别敏感文本的方法，通过计算 待识别文本的词性序列出现的概率，来确定待识别文本是否为敏感文本，由于 不符合语法规定且不常见的文本，其对应的词性序列出现的概率也会较低，因 此当待识别文本的词性序列出现的概率小于预设阈值时，则可以确定待识别文 本不符合语法规则且不常见，属于敏感文本。

[0028] 请参阅图1a，图1a为本发明实施例中识别敏感文本的方法的运行系统示 意图，上述识别敏感文本的方法可以应用于如图1a所示的由终端101和服务器 102所构成的硬件环境中。在图1a中，服务器102通过网络与终端101进行连 接，上述网络包括但不限于：广域

网、城域网或局域网，终端101并不限定于 PC((Personal Computer，个人计算机)、手机、平板电脑等，服务器102包括 但不限于单路服务器、多路服务器和分布式服务器。本发明实施例中识别敏感 文本的方法的各个步骤可以由服务器102来执行，也可以由终端101来执行，还可以是由服务器102和终端101共同执行。例如，在一些实施例中，用户通 过终端101中的客户端上传文本至服务器102；服务器102在接收到终端101 上传的文本后，对接收到的文本进行识别，确定该文本是否属于敏感文本。其 中，当服务器102确定终端101上传的文本为敏感文本时，可以向终端101发 送修改提醒消息，或者直接删除/隐藏上述文本。

[0029] 为了更好的理解本发明实施例，参照图1b，图1b为本发明实施例中识别 敏感文本的方法的实现场景图。本发明实施例可以应用于网络平台，帮助网络 平台审核用户上传的文本。在图1b中，网络平台接收用户通过终端上传的文本， 然后对该文本进行识别，确定该文本是否为敏感文本，若该文本为敏感文本， 则向终端反馈修改提醒消息，或者直接删除/隐藏该文本；若该文本为非敏感文 本，则在网络平台中进行发布。

[0030] 请参阅图2，图2为本发明实施例中识别敏感文本的方法的步骤流程示意 图，上述识别敏感文本的方法包括：

[0031] 步骤201、基于待识别文本中各个词的词性，确定所述待识别文本的词性 序列。

[0032] 本发明实施例中，先获取待识别文本，例如文档标题、标题链接、广告语 等，然后确定待识别文本中各个词的词性，并基于待识别文本中各个词的词性， 确定待识别文本的词性序列。例如：假设待识别文本为“我住在广州”，则由 于待识别文本中“我”的词性为代词(r)、“住”的词性为动词(v)、“在” 的词性为介词(p)、“广州”的词性为地名(ns)，因此待识别文本“我住在 广州”的词性序列为“r-v-p-ns”。

[0033] 步骤202、根据已发布的历史文本，计算所述词性序列出现的概率。

[0034] 本发明实施例中，可以随机采集若干数量已发布的历史文本，例如随机抽 取某合法网络平台中已发布的10万篇文章的标题。然后分析历史文本中各个词 的词性，确定历史文本中各个词的词性变化情况，通过历史文本中各个词的词 性变化情况，来计算待识别文本的词性序列出现的概率。

[0035] 步骤203、当所述词性序列出现的概率小于预设阈值时，确定所述待识别 文本为敏感文本。

[0036] 可以理解的是，对于不符合语法规则的敏感文本而言，其对应的词性序列 也不符合语法规则，例如“777％痔cc疮长期不治会怎样这个要一定看”的 词性序列显然不符合语法规则，因此在正常情况下出现的概率会非常低。故本 发明实施例中，当待识别文本的词性序列出现的概率小于预设的阈值(例如1％) 时，则可以确定待识别文本不符合语法规则且不常见，属于敏感文本。

[0037] 本发明实施例所提供的识别敏感文本的方法，包括：基于待识别文本中各 个词的词性，确定待识别文本的词性序列，然后根据已发布的历史文本，计算 上述词性序列出现的概率，当该词性序列出现的概率小于预设阈值时，确定待 识别文本为敏感文本。相较于现有技术而言，本发明实施例通过计算待识别文 本的词性序列出现的概率，来确定待识别文本是否为敏感文本，由于不符合语 法规定且不常见的文本，其对应的词性序列出现的概率也会较低，因此当待识 别文本的词性序列出现的概率小于预设阈值时，则可以确定待识别文本不符合 语法规则且不常见，属于敏感文本。

[0038] 进一步地，基于上述实施例，本发明实施例中，步骤201可以细化为以下 步骤：

[0039] 步骤a、划分待识别文本，得到待识别文本中的各个词；

[0040] 其中，可以采用分词算法来划分待识别文本，例如采用基于字符串匹配的 分词算法、基于理解的分词算法和基于统计的分词算法等。

[0041] 具体的，字符串匹配的分词算法又叫做机械分词方法，它是按照一定的策 略将待分析的汉字串与一个“充分大的”机器词典中的词条进行匹配，若在词典 中找到某个字符串，则匹配成功(识别出一个词)。按照扫描方向的不同，字 符串匹配分词算法可以分为正向匹配和逆向匹配；按照不同长度优先匹配的情 况，可以分为最大(最长)匹配和最小(最短)匹配。

[0042] 基于理解的分词算法是通过让计算机模拟人对句子的理解，达到识别词的 效果。其基本思想就是在分词的同时进行句法、语义分析，利用句法信息和语 义信息来处理歧义现象。它通常包括三个部分：分词子系统、句法语义子系统、 总控部分。在总控部分的协调下，分词子系统可以获得有关词、句子等的句法 和语义信息来对分词歧义进行判断，即它模拟了人对句子的理解过程。

[0043] 基于统计的分词算法的原理是：从形式上看，词是稳定的字的组合，因此 在上下文中，相邻的字同时出现的次数越多，就越有可能构成一个词。因此字 与字相邻共现的频率或概率能够较好的反映成词的可信度。可以对语料中相邻 共现的各个字的组合的频度进行统计，计算它们的互现信息。定义两个字的互 现信息，计算两个汉字W、V的相邻共现概率。互现信息体现了汉字之间结合 关系的紧密程度，当紧密程度高于某一个阈值时，便可认为此字组可能构成了 一个词。

[0044] 步骤b、查找预置的词性数据库，确定各个词的词性；

[0045] 其中，本发明实施例中，可以通过CLAWS(Contituent-Likelihood Automatic Word-tagging System，成分似然性自动词性标注系统)来标注各个词的词性。

[0046] 步骤c、将各个词的词性按照各个词在待识别文本中的顺序进行排列，得 到待识别文本的词性序列。

[0047] 本发明实施例所提供的识别敏感文本的方法，包括：划分待识别文本，得 到待识别文本中的各个词；查找预置的词性数据库，确定各个词的词性；将各 个词的词性按照各个词在待识别文本中的顺序进行排列，得到待识别文本的词 性序列。即本发明实施例可以通过对待识别文本进行分词，来确定待识别文本 的词性序列，从而可以根据待识别文本的词性序列出现的概率，来确定待识别 文本是否为敏感文本。

[0048] 进一步地，基于上述实施例，本发明实施例中，上述步骤202中根据已发 布的历史文本，计算所述词性序列出现的概率，具体包括：

[0049] 步骤a、采集已发布的非敏感文本，并确定所述非敏感文本中各个词的词 性；

[0050] 步骤b、基于所述非敏感文本中任意N种词性连续出现的概率，计算所述 词性序列出现的概率，N为正整数，且$N \geq 2$。

[0051] 可以理解的是，在写作常规的文本时，需要根据各个词的词性来进行文本 的组织，例如，在语法规则中，冠词用来修饰名词，介词用来连接地点、时间 等或用来构成一个短语，动词不能用来修饰其他词，它只能被副词修饰。

[0052] 本发明实施例中，可以基于历史文本中各个词的词性，统计历史文本中任 意N种

词性连续出现的次数，进而得到历史文本中任意N种词性连续出现的概 率。例如，当N＝2时，统计历史文本中任意2种词性t_{i-1}、t_i连续出现的次数X，以及t_i出现的次数Y，进而得到历史文本中词性t_{i-1}、t_i连续出现的概率＝X/Y。

[0053] 其中，在确定历史文本中任意N种词性连续出现的概率之后，可以采用 HMM(Hidden Markov Model，隐马尔可夫模型)，来计算上述词性序列出现 的概率。

[0054] 具体的，可以利用以下公式计算上述词性序列出现的概率：

[0055] $P(T)=\prod P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$

[0056] 其中，T表示待识别文本的词性序列，$T=t_1、t_2、\cdots、t_n$，$P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$ 表示上述非敏感文本中词性t_{i-N+1}、…、t_{i-2}、t_{i-1}、t_i连续出现的概率，t_i表示上述 词性序列中的第i个词性，N≤i≤n。

[0057] 本发明实施例中，假设待识别文本的词性序列为T，$T=t_1、t_2、\cdots、t_n$，t_k表示 待识别文本中的第k个词的词性，1≤k≤n，然后建立语言模型，其中：

[0058] 对于一元语言模型，t_k独立出现，即待识别文本中的各个词的词性相互独立， 则词性序列T出现的概率为：

[0059] $P(t_1、t_2、\cdots、t_n)=P(t_1)*P(t_2)*、\cdots、*P(t_n)$；

[0060] 对于二元语言模型，t_k的出现决定于t_{k-1}，即待识别文本中的两个相邻的词 的词性之间存在联系，则词性序列T出现的概率为：

[0061] $P(t_1、t_2、\cdots、t_n)=P(t_1)*P(t_2|t_1)*P(t_3|t_2)、*\cdots、*P(t_n|t_{n-1})$

[0062] 对于三元语言模型，t_k的出现决定于t_{k-1}与t_{k-2}，即待识别文本中的三个相邻 的词的词性之间存在联系，则词性序列T出现的概率为：

[0063] $P(t_1、t_2、\cdots、t_n)=P(t_1)*P(t_2|t_1)*P(t_3|t_1、t_2)、*\cdots、*P(t_n|t_{n-2}、t_{n-1})$

[0064] 通过上述方式，即可推导出N元语言模型中，词性序列T出现的概率为：

[0065] $P(t_1、t_2、\cdots、t_n)=P(t_1)*P(t_2|t_1)*P(t_3|t_1、t_2)、*\cdots、*P(t_n|t_{n-N+1}、\cdots、t_{n-2}、t_{n-1})$

[0066] 其中，为了便于计算，本实施例中可以对N元语言模型中，词性序列T出 现的概率的计算公式进行简化，得到如下计算公式：

[0067] $P(t_1、t_2、\cdots、t_n)=\prod P(t_i|t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$，N≤i≤n

[0068] 例如：当N＝2时，$P(t_1、t_2、\cdots、t_n)=\prod P(t_i|t_{i-1})$；

[0069] 当N＝3时，$P(t_1、t_2、\cdots、t_n)=\prod P(t_i|t_{i-2}、t_{i-1})$；

[0070] 当N＝5时，$P(t_1、t_2、\cdots、t_n)=\prod P(t_i|t_{i-4}、t_{i-3}、t_{i-2}、t_{i-1})$。

[0071] 具体的，通过上述步骤b中确定的非敏感文本中任意N种词性连续出现的 概率，即可计算出上述词性序列T出现的概率。例如，假设待识别文本“我住 在广州”的词性序列为“r-v-p-ns”，在2元模型中，通过上述步骤b中确定的 历史文本中词性r与词性v连续出现的概率为P(r-v)、词性v与词性p连续 出现的概率为P(v-p)、词性p与词性ns连续出现的概率为P(p-ns)，则词 性序列“r-v-p-ns”出现的概率P(r-v-p-ns)＝P(r-v)*P(v-p)*P(p-ns)。

[0072] 进一步地，考虑到待识别文本中，可能有个别词具有两种或两种以上词性， 这种情况下，待识别文本则会具有两条或两条以上不同的词性序列。本发明实 施例中，可以根据待识别文本中各个词的具有的词性，确定待识别文本所有可 能的词性序列，然后分别计算各条词性序列出现的概率，当各条词性序列出现 的概率均小于预设阈值时，即可确定待识别文本为敏感文本。

[0073] 为了更好的理解本发明实施例，参照图3，图3为本发明实施例中确定待 识别文本的词性序列的示意图，在图3中，假设待识别文本为ABCD，其中A 的词性为a，B的词性为b_1与b_2，C的词性为c_1与c_2、D的词性为d，则待识 别文本ABCD的词性序列包括“a-b_1-c_1-d”、“a-b_2-c_1-d”、“a-b_1-c_2-d”、 “a-b_2-c_2-d”。

[0074] 本发明实施例所提供的识别敏感文本的方法，包括采集已发布的非敏感文 本，确定非敏感文本中各个词的词性，然后基于非敏感文本中任意N种词性连 续出现的概率，计算待识别文本的词性序列出现的概率，当待识别文本的词性 序列出现的概率小于预设阈值时，则可以认为待识别文本的语法与非敏感文本 中所采用的语法不同，从而可以确定待识别文本为敏感文本。

[0075] 进一步地，基于上述实施例，参照图4，图4为本发明实施例中识别敏感 文本的方法的另一步骤流程示意图，本发明实施例中，上述方法包括：

[0076] 步骤401、基于待识别文本中各个词的词性，确定所述待识别文本的词性 序列。

[0077] 其中，步骤401与本发明上述实施例中的步骤201描述的内容一致，可参 照上述实施例中的步骤201，在此不再赘述。

[0078] 步骤402、采集历史文本，该历史文本包括敏感文本与非敏感文本。

[0079] 本发明实施例中，可以采集已发现的敏感文本作为负样本，采集非敏感文 本作为正样本。

[0080] 步骤403、将历史文本的词性序列作为样本词性序列，训练概率计算模型。

[0081] 本发明实施例中，确定影响文本敏感度的影响特征，例如文本中符号的占 比、中英文比例、数字比例等，然后利用历史文本的词性序列作为样本词性序 列，训练概率计算模型，确定概率计算模型中上述影响特征对应的参数。其中， 可以采用采用梯度提升树，xgboost(eXtreme Gradient Boosting，极值梯度提升) 训练算法、LR(Logistic Regression，逻辑回归模型)、神经网络算法等进行训练。

[0082] 步骤404、将所述词性序列输入训练后的概率计算模型，输出所述词性序 列出现的概率。

[0083] 步骤405、当所述词性序列出现的概率小于预设阈值时，确定所述待识别 文本为敏感文本。

[0084] 本发明实施例中，在概率计算模型训练完成之后，即可将确定的待识别文 本的词性序列输入该模型，然后输出待识别文本的词性序列出现的概率，根据 待识别文本的词性序列出现的概率，即可确定待识别文本是否为敏感文本。

[0085] 本发明实施例所提供的识别敏感文本的方法，包括：基于待识别文本中各 个词的词性，确定待识别文本的词性序列；采集历史文本，将历史文本的词性 序列作为样本词性序列，训练概率计算模型；将上述词性序列输入训练后的概 率计算模型，输出上述词性序列出现的概率。相较于现有技术而言，本发明实 施例通过训练概率计算模型，来计算待识别文本的词性序列出现的概率，由于 不符合语法规定且不常见的文本，其对应的词性序列出现的概率也会较低，因 此当待识别文本的词性序列出现的概率小于预设阈值时，则可以确定待识别文 本不符合语法规则且不常见，属于敏感文本。

[0086] 为了解决上述技术问题，本发明实施例还提供一种识别敏感文本的装置， 参照图5，图5为本发明实施例中识别敏感文本的装置的程序模块示意图。本 发明实施例中，上述

装置包括：

[0087] 确定模块501，用于基于待识别文本中各个词的词性，确定待识别文本的 词性序列。

[0088] 计算模块502，用于根据已发布的历史文本，计算词性序列出现的概率。

[0089] 识别模块503，用于当词性序列出现的概率小于预设阈值时，确定待识别 文本为敏感文本。

[0090] 本发明实施例所提供的识别敏感文本的装置，包括：确定模块501，用于 基于待识别文本中各个词的词性，确定待识别文本的词性序列；计算模块502， 用于根据已发布的历史文本，计算上述词性序列出现的概率；识别模块503用 于当该词性序列出现的概率小于预设阈值时，确定待识别文本为敏感文本。相 较于现有技术而言，本发明实施例通过计算待识别文本的词性序列出现的概率， 来确定待识别文本是否为敏感文本，由于不符合语法规定且不常见的文本，其 对应的词性序列出现的概率也会较低，因此当待识别文本的词性序列出现的概 率小于预设阈值时，则可以确定待识别文本不符合语法规则且不常见，属于敏 感文本。

[0091] 进一步地，基于上述实施例，本发明实施例中，确定模块501包括：

[0092] 划分模块，用于划分待识别文本，得到待识别文本中的各个词。

[0093] 查找模块，用于查找预置的词性数据库，确定各个词的词性。

[0094] 排列模块，用于将各个词的词性按照各个词在待识别文本中的顺序进行排 列，得到待识别文本的词性序列。

[0095] 本发明实施例所提供的识别敏感文本的装置，确定模块501包括：划分模 块，用于划分待识别文本，得到待识别文本中的各个词；查找模块，用于查找 预置的词性数据库，确定各个词的词性；排列模块，用于将各个词的词性按照 各个词在待识别文本中的顺序进行排列，得到待识别文本的词性序列。即本发 明实施例可以通过对待识别文本进行分词，来确定待识别文本的词性序列，从 而可以根据待识别文本的词性序列出现的概率，来确定待识别文本是否为敏感 文本。

[0096] 进一步地，基于上述实施例，本发明实施例中，计算模块502包括：

[0097] 采集模块，用于采集已发布的非敏感文本，并确定非敏感文本中各个词的 词性。

[0098] 计算执行模块，用于基于非敏感文本中任意N种词性连续出现的概率，计 算上述词性序列出现的概率，N为正整数，且$N \geq 2$。

[0099] 其中，计算执行模块具体用于：

[0100] 利用以下公式计算上述词性序列出现的概率，

[0101] $P(T) = \prod P(t_i | t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$

[0102] 其中，T表示待识别文本的词性序列，$T = t_1、t_2、\cdots、t_n$，$P(t_i | t_{i-N+1}、\cdots、t_{i-2}、t_{i-1})$ 表示非敏感文本中词性$t_{i-N+1}、\cdots、t_{i-2}、t_{i-1}、t_i$连续出现的概率，$t_i$表示所述词性 序列中的第i个词性，N表示在非敏感文本中任意N个词性连续出现，$N \leq i \leq n$。

[0103] 本发明实施例所提供的识别敏感文本的装置，计算模块502包括：采集模 块，用于采集已发布的非敏感文本，确定非敏感文本中各个词的词性；计算执 行模块，用于基于非敏感文本中任意N种词性连续出现的概率，计算待识别文 本的词性序列出现的概率，当待识别文本的词性序列出现的概率小于预设阈值 时，则可以认为待识别文本的语法与非敏

感文本所采用的语法不同，即可以确 定待识别文本为敏感文本。

[0104] 进一步地，基于上述实施例，参照图6，图6为本发明实施例中识别敏感 文本的装置的另一程序模块示意图。本发明实施例中，上述装置包括：

[0105] 确定模块501，用于基于待识别文本中各个词的词性，确定所述待识别文 本的词性序列。

[0106] 计算模块502，用于根据已发布的历史文本，计算所述词性序列出现的概 率。

[0107] 识别模块503，用于当所述词性序列出现的概率小于预设阈值时，确定所 述待识别文本为敏感文本。

[0108] 具体的，计算模块502包括：

[0109] 样本采集模块601，用于采集历史文本，所述历史文本包括敏感文本与非 敏感文本。

[0110] 训练模块602，用于将所述历史文本的词性序列作为样本词性序列，训练 概率计算模型。

[0111] 输出模块603，用于将所述词性序列输入训练后的概率计算模型，输出所 述词性序列出现的概率。

[0112] 本发明实施例所提供的识别敏感文本的装置中，计算模块502包括：样本 采集模块601，用于采集历史文本；训练模块602，用于将历史文本的词性序列 作为样本词性序列，训练概率计算模型；输出模块603，用于将待识别文本的 词性序列输入训练后的概率计算模型，输出待识别文本的词性序列出现的概率。相较于现有技术而言，本发明实施例通过训练概率计算模型，来计算待识别文 本的词性序列出现的概率，由于不符合语法规定且不常见的文本，其对应的词 性序列出现的概率也会较低，因此当待识别文本的词性序列出现的概率小于预 设阈值时，则可以确定待识别文本不符合语法规则且不常见，属于敏感文本。

[0113] 可以理解的是，在本发明实施例中，上述识别敏感文本的装置可以是一种 设备，请参阅图7，图7为本发明实施例中设备70的结构示意图。该设备70 包括处理器701、存储器702和收发器703，存储器702可以包括只读存储器和 随机存取存储器，并向处理器701提供操作指令和数据。存储器702的一部分 还可以包括非易失性随机存取存储器(NVRAM)。

[0114] 在一些实施方式中，存储器702存储了如下的元素：可执行模块或者数据 结构，或者他们的子集，或者他们的扩展集。

[0115] 在本发明实施例中，通过调用存储器702存储的操作指令(该操作指令可存 储在操作系统中)，执行以下过程：

[0116] 基于待识别文本中各个词的词性，确定待识别文本的词性序列；根据已发 布的历史文本，计算上述词性序列出现的概率；当上述词性序列出现的概率小 于预设阈值时，确定待识别文本为敏感文本。

[0117] 与现有技术相比，本发明实施例通过计算待识别文本的词性序列出现的概 率，来确定待识别文本是否为敏感文本，由于不符合语法规定且不常见的文本， 其对应的词性序列出现的概率也会较低，因此当待识别文本的词性序列出现的 概率小于预设阈值时，则可以确定待识别文本不符合语法规则且不常见，属于 敏感文本。

[0118] 其中，处理器701控制设备70的操作，处理器701还可以称为CPU(Central

Processing Unit,中央处理单元)。存储器702可以包括只读存储器和随机存取 存储器,并向处理器701提供指令和数据。存储器702的一部分还可以包括非 易失性随机存取存储器(NVRAM)。具体的应用中设备70的各个组件通过总线 系统704耦合在一起,其中总线系统704除包括数据总线之外,还可以包括电 源总线、控制总线和状态信号总线等。但是为了清楚说明起见,在图中将各种 总线都标为总线系统704。

[0119] 上述本发明实施例揭示的方法可以应用于处理器701中,或者由处理器701 实现。处理器701可能是一种集成电路芯片,具有信号的处理能力。在实现过 程中,上述方法的各步骤可以通过处理器701中的硬件的集成逻辑电路或者软 件形式的指令完成。上述的处理器701可以是通用处理器、数字信号处理器 (DSP)、专用集成电路(ASIC)、现成可编程门阵列(FPGA)或者其他可编程逻 辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发 明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或 者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的 步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及 软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器, 可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介 质中。该存储介质位于存储器702,处理器701读取存储器702中的信息,结 合其硬件完成上述方法的步骤。

[0120] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,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,例如多个模块或组件可以结合或者可以集成到另一个系统,或一些特 征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合 或通信连接可以是通过一些接口,装置或模块的间接耦合或通信连接,可以是 电性,机械或其它的形式。

[0121] 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模块中, 也可以是各个模块单独物理存在,也可以两个或两个以上模块集成在一个模块 中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的 形式实现。

[0122] 所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或 使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明 的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或 部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质 中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或 者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的 存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以 存储程序代码的介质。

[0123] 需要说明的是,对于前述的各方法实施例,为了简便描述,故将其都表述 为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的 动作顺序的限制,因为依据本发明,某些步骤可以采用其它顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施 例,所涉及的动作和模块并不一定都是本发明所必须的。

[0124] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中没有详 述的部

分，可以参见其它实施例的相关描述。

[0125] 以上为对本发明所提供的一种识别敏感文本的方法、装置、设备及可读存 储介质的描述，对于本领域的技术人员，依据本发明实施例的思想，在具体实 施方式及应用范围上均会有改变之处，综上，本说明书内容不应理解为对本发 明的限制。

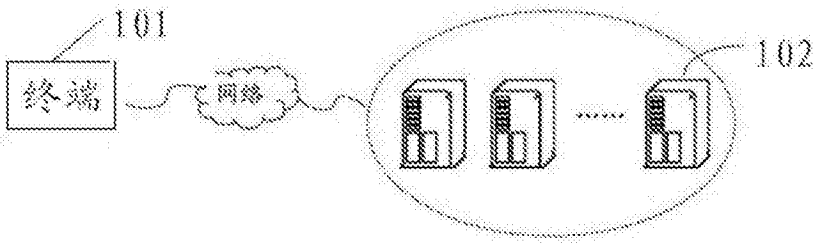

101
终端
网络
102

图1a

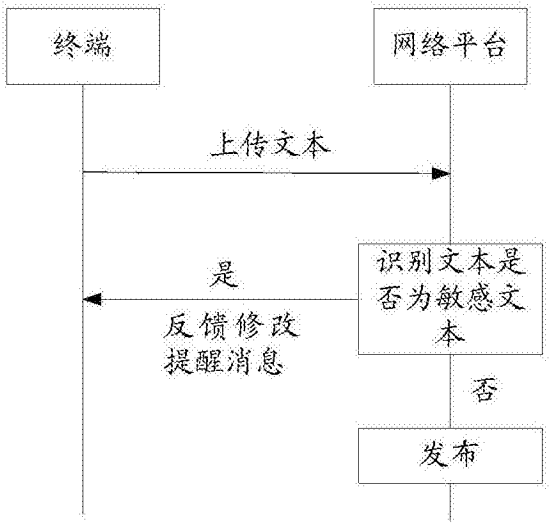

终端
网络平台
上传文本
识别文本是否为敏感文本
是
反馈修改提醒消息
否
发布

图1b

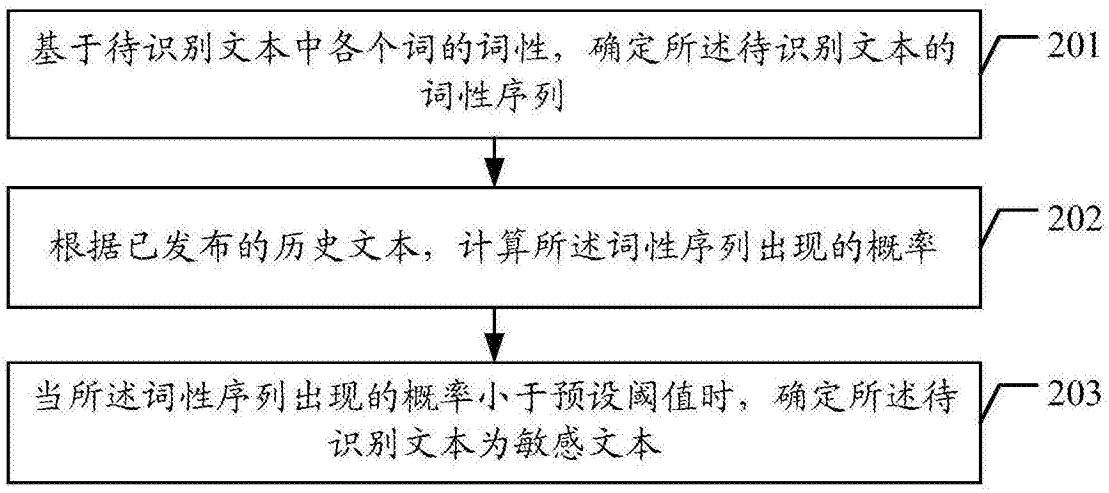

基于待识别文本中各个词的词性，确定所述待识别文本的词性序列
201
根据已发布的历史文本，计算所述词性序列出现的概率
202
当所述词性序列出现的概率小于预设阈值时，确定所述待识别文本为敏感文本
203

图2

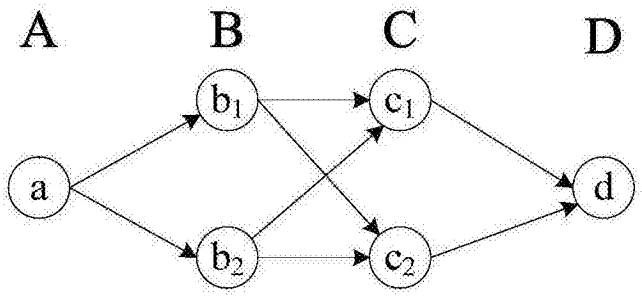

A
B
C
D
a
b1
b2
c1
c2
d

图3

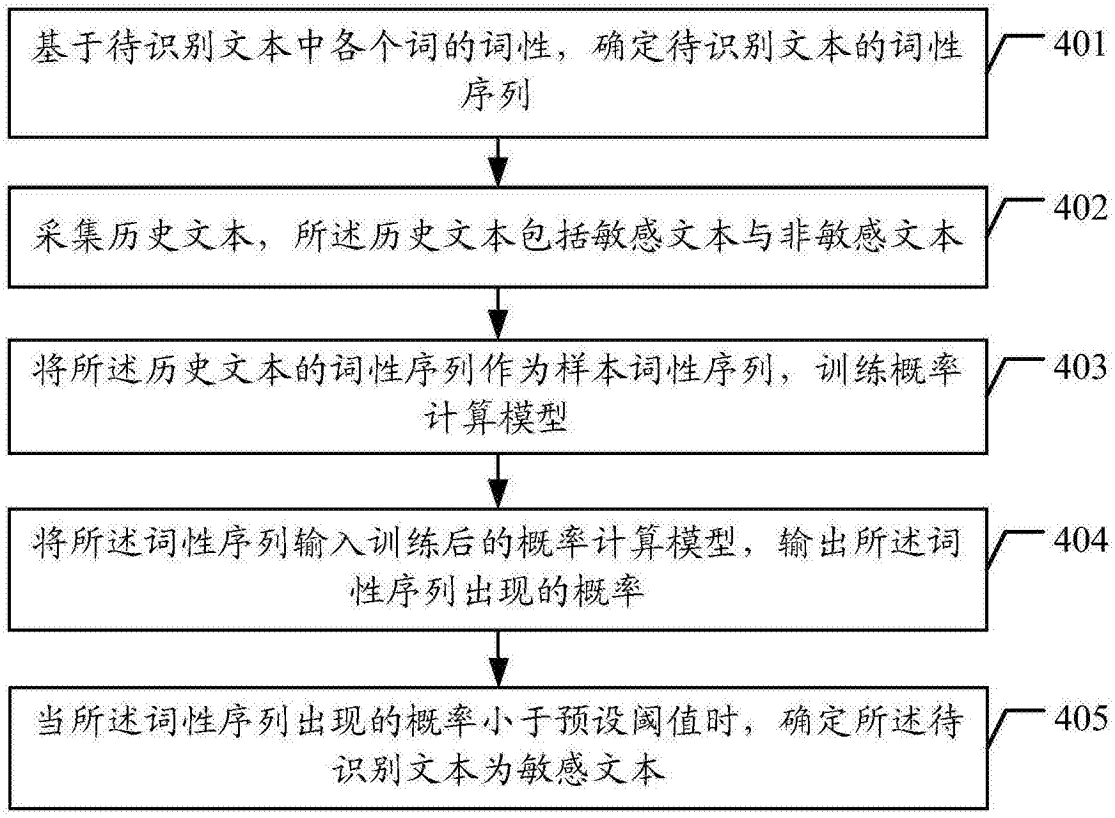

基于待识别文本中各个词的词性，确定待识别文本的词性序列
401
采集历史文本，所述历史文本包括敏感文本与非敏感文本
402
将所述历史文本的词性序列作为样本词性序列，训练概率计算模型
403
将所述词性序列输入训练后的概率计算模型，输出所述词性序列出现的概率
404
当所述词性序列出现的概率小于预设阈值时，确定所述待识别文本为敏感文本
405

图4

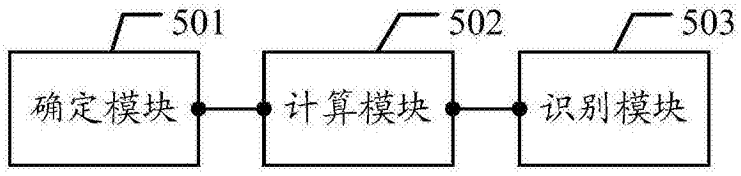

501
502
503
确定模块
计算模块
识别模块

图5

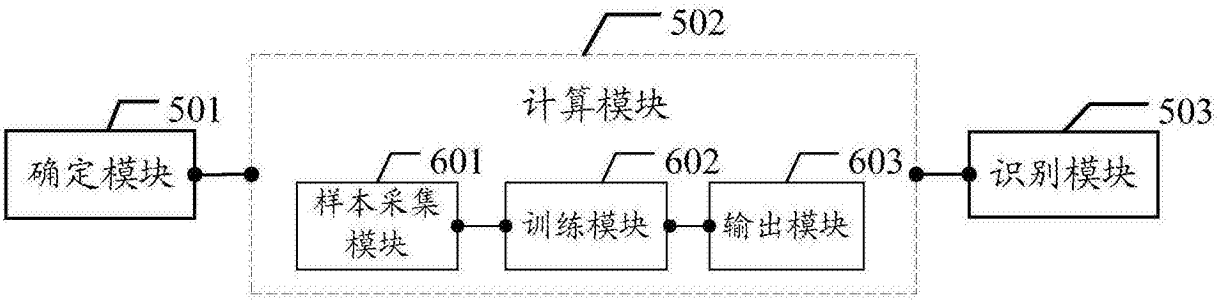

502
计算模块
501
确定模块
601
样本采集模块
602
训练模块
603
输出模块
503
识别模块

图6

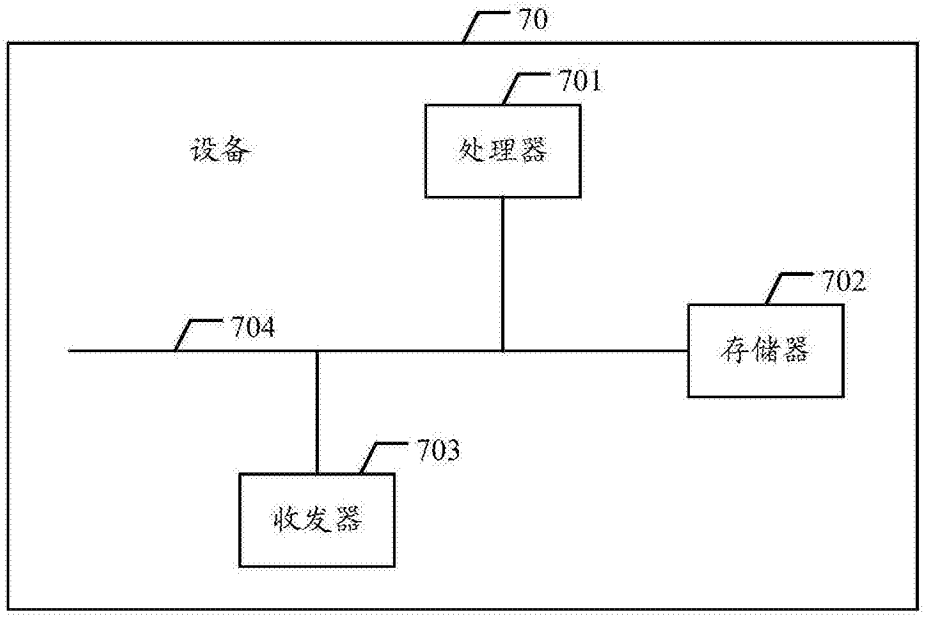

70
设备
701
处理器
702
存储器
704
703
收发器

图7