



(12) 发明专利

(10) 授权公告号 CN 109739989 B

(45) 授权公告日 2021.05.18

(21) 申请号 201811653926.5

(22) 申请日 2018.12.29

(65) 同一申请的已公布的文献号  
申请公布号 CN 109739989 A

(43) 申请公布日 2019.05.10

(73) 专利权人 奇安信科技集团股份有限公司  
地址 100088 北京市西城区新街口外大街  
28号102号楼3层332号

(72) 发明人 李斌 禹庆华

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

代理人 杨静

(51) Int. Cl.

G06F 16/35 (2019.01)

(56) 对比文件

CN 105630827 A, 2016.06.01

US 8856123 B1, 2014.10.07

CN 106874253 A, 2017.06.20

CN 108446388 A, 2018.08.24

CN 108399164 A, 2018.08.14

审查员 谢宜瑾

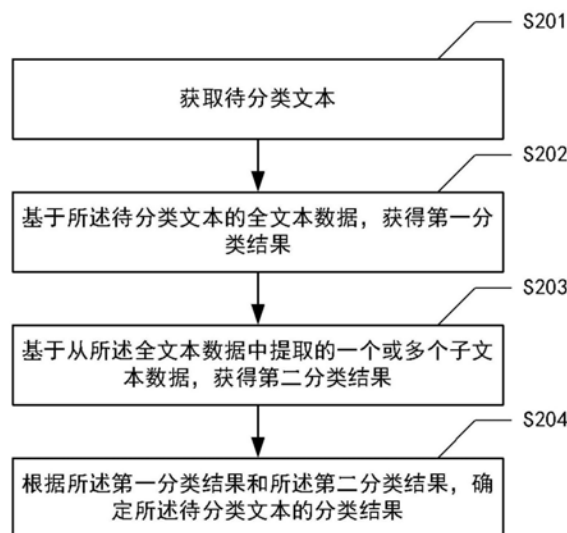
权利要求书2页 说明书14页 附图6页

(54) 发明名称

文本分类方法和计算机设备

(57) 摘要

本公开提供了一种文本分类方法,包括:获取待分类文本;基于所述待分类文本的全文本数据,获得第一分类结果;基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果;以及,根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果。本公开还提供了一种计算机设备。



1. 一种文本分类方法,包括:

获取待分类文本;

基于所述待分类文本的全文本数据,获得第一分类结果,所述第一分类结果包括所述全文本数据关于各预置类别的第一得分;

基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果,所述第二分类结果包括所述一个或多个子文本数据关于各预置类别的第三得分;

从所述全文本数据中提取一个或多个子文本数据包括:

利用预设关键词集合中的关键词与所述全文本数据进行匹配;

对于匹配成功的第一关键词,从所述全文本数据中提取所述第一关键词之前的第一预设长度的字符串和/或所述第一关键词之后的第二预设长度的字符串;以及

将提取出的字符串和所述第一关键词按照在所述全文本数据中的位置顺序组合为一个子文本数据;

对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于所述多个预置类别中的各预置类别的第二得分;以及

基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算所述一个或多个子文本数据关于各预置类别的第三得分,将第三得分最高的预置类别作为与所述一个或多个子文本数据对应的类别;以及

根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果,包括:

根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分,将综合得分最高的预置类别作为与所述待分类文本对应的类别。

2. 根据权利要求1所述的方法,其中,所述基于所述待分类文本的全文本数据,获得第一分类结果包括:

将所述全文本数据输入对应于多个预置类别的全文本分类模型,基于所述全文本分类模型确定所述全文本数据关于所述多个预置类别中的各预置类别的第一得分,将第一得分最高的预置类别作为与所述全文本数据对应的类别。

3. 根据权利要求1所述的方法,其中,所述基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算一个或多个子文本数据关于各预置类别的第三得分包括:

对于所述多个预置类别中的任一预置类别,将各子文本数据关于该预置类别的第二得分进行加权求和,得到所述一个或多个子文本数据关于该预置类别的第三得分。

4. 根据权利要求1所述的方法,其中,所述根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分包括:

设置与所述全文本数据对应的第一权重和与所述一个或多个子文本数据对应的第二权重;以及

对于所述多个预置类别中的任一预置类别,根据所述第一权重和所述第二权重,对所述全文本数据关于该预置类别的第一得分和所述一个或多个子文本数据关于该预置类别

的第三得分进行加权求和,得到所述待分类文本关于该预置类别的综合得分。

5. 根据权利要求2所述的方法,其中,所述基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果包括:

对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于各预置类别的得分,将得分最高的预置类别作为与所述第一子文本数据对应的类别;

当与所述一个或多个子文本中的各子文本数据对应的类别中存在第一类别时,确定与所述一个或多个子文本数据对应的类别为第一类别;以及

当与所述一个或多个子文本中的各子文本数据对应的类别均为第二类别时,确定与所述一个或多个子文本数据对应的类别为第二类别。

6. 根据权利要求1或5所述的方法,其中:

所述预置类别包括第一类别和第二类别;

所述根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果包括:

当与所述全文本数据对应的类别和与所述一个或多个子文本数据对应的类别均为第二类别时,确定与所述待分类文本对应的类别为第二类别;以及

当与所述全文本数据对应的类别和/或与所述一个或多个子文本数据对应的类别为第一类别时,确定与所述待分类文本对应的类别为第一类别。

7. 一种计算机设备,所述计算机设备包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序时用于实现如权利要求1~6中任一项所述的文本分类方法。

## 文本分类方法和计算机设备

### 技术领域

[0001] 本公开涉及一种文本分类方法和计算机设备。

### 背景技术

[0002] 文本分类是在给定分类体系下,根据文本内容确定文本类别的过程。文本分类是自然语言处理的一个重要部分,具有广泛的应用,包括新闻分类、邮件分类、垃圾邮件识别、违规网页识别等。

[0003] 现有的文本分类方案基于文本的整体内容进行分类,由于文本的整体内容中存在大量与文本分类无关的干扰信息,可能导致文本分类的判别性特征湮没在干扰信息中,进而无法得到准确的分类结果。

### 发明内容

[0004] 本公开的一个方面提供了一种文本分类方法,包括:获取待分类文本;基于所述待分类文本的全文本数据,获得第一分类结果;基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果;以及,根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果。

[0005] 可选地,上述基于所述待分类文本的全文本数据,获得第一分类结果包括:将所述全文本数据输入对应于多个预置类别的全文本分类模型,基于所述全文本分类模型确定所述全文本数据关于所述多个预置类别中的各预置类别的第一得分,将第一得分最高的预置类别作为与所述全文本数据对应的类别。

[0006] 可选地,在上述基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果之前,上述方法还包括:从所述全文本数据中提取一个或多个子文本数据。上述从所述全文本数据中提取一个或多个子文本数据包括:利用预设关键词集合中的关键词与所述全文本数据进行匹配;对于匹配成功的第一关键词,从所述全文本数据中提取所述第一关键词之前的第一预设长度的字符串和/或所述第一关键词之后的第二预设长度的字符串;以及,将提取出的字符串和所述第一关键词按照在所述全文本数据中的位置顺序组合为一个子文本数据。

[0007] 可选地,上述基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果包括:对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于所述多个预置类别中的各预置类别的第二得分;以及,基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算所述一个或多个子文本数据关于各预置类别的第三得分,将第三得分最高的预置类别作为与所述一个或多个子文本数据对应的类别。

[0008] 可选地,上述基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算一个或多个子文本数据关于各预置类别的第三得分包括:对于多个预置类别中的任一预置类别,将各子文本数据关于该预置类别的第二得分进行加权求和,得到所

述一个或多个子文本数据关于该预置类别的第三得分。

[0009] 可选地,上述根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果包括:根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分,将综合得分最高的预置类别作为与所述待分类文本对应的类别。

[0010] 可选地,上述根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分包括:设置与所述全文本数据对应的第一权重和与所述一个或多个子文本数据对应的第二权重;以及,对于所述多个预置类别中的任一预置类别,根据所述第一权重和所述第二权重,对所述全文本数据关于该预置类别的第一得分和所述一个或多个子文本数据关于该预置类别的第三得分进行加权求和,得到所述待分类文本关于该预置类别的综合得分。

[0011] 可选地,上述基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果包括:对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于各预置类别的得分,将得分最高的预置类别作为与所述第一子文本数据对应的类别。当与所述一个或多个子文本中的各子文本数据对应的类别中存在第一类别时,确定与所述一个或多个子文本数据对应的类别为第一类别;以及,当与所述一个或多个子文本中的各子文本数据对应的类别均为第二类别时,确定与所述一个或多个子文本数据对应的类别为第二类别。

[0012] 可选地,所述预置类别包括第一类别和第二类别。上述根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果包括:当与所述全文本数据对应的类别和与所述一个或多个子文本数据对应的类别均为第二类别时,确定与所述待分类文本对应的类别为第二类别;以及,当与所述全文本数据对应的类别和/或与所述一个或多个子文本数据对应的类别为第一类别时,确定与所述待分类文本对应的类别为第一类别。

[0013] 本公开的另一方面提供了一种文本分类装置,包括:获取模块、第一分类模块、第二分类模块和综合分类模块。获取模块用于获取待分类文本。第一分类模块用于基于所述待分类文本的全文本数据,获得第一分类结果。第二分类模块用于基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果。以及,综合分类模块用于根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果。

[0014] 可选地,第一分类模块用于将所述全文本数据输入对应于多个预置类别的全文本分类模型,基于所述全文本分类模型确定所述全文本数据关于所述多个预置类别中的各预置类别的第一得分,将第一得分最高的预置类别作为与所述全文本数据对应的类别。

[0015] 可选地,所述装置还包括子文本提取模块,用于在所述第二分类模块基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果之前,从所述全文本数据中提取一个或多个子文本数据。其中,子文本提取模块包括匹配子模块、提取子模块和组合子模块。匹配子模块用于利用预设关键词集合中的关键词与所述全文本数据进行匹配。提取子模块用于对于匹配成功的第一关键词,从所述全文本数据中提取所述第一关键词之前的第一预设长度的字符串和/或所述第一关键词之后的第二预设长度的字符串。以及,组合子模块用于将提取出的字符串和所述第一关键词按照在所述全文本数据中的位置顺序组合为一个子文本数据。

[0016] 可选地,所述第二分类模块包括第一预测子模块和计算子模块。第一预测子模块用于对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于所述多个预置类别中的各预置类别的第二得分。以及,计算子模块用于基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算所述一个或多个子文本数据关于各预置类别的第三得分,将第三得分最高的预置类别作为与所述一个或多个子文本数据对应的类别。

[0017] 可选地,计算子模块具体用于对于所述多个预置类别中的任一预置类别,将各子文本数据关于该预置类别的第二得分进行加权求和,得到所述一个或多个子文本数据关于该预置类别的第三得分。

[0018] 可选地,综合分类模块包括综合计算子模块,用于根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分,将综合得分最高的预置类别作为与所述待分类文本对应的类别。

[0019] 可选地,综合计算子模块用于设置与所述全文本数据对应的第一权重和与所述一个或多个子文本数据对应的第二权重;以及,对于所述多个预置类别中的任一预置类别,根据所述第一权重和所述第二权重,对所述全文本数据关于该预置类别的第一得分和所述一个或多个子文本数据关于该预置类别的第三得分进行加权求和,得到所述待分类文本关于该预置类别的综合得分。

[0020] 可选地,第二分类模块包括第二预测子模块和第一确定子模块。第二预测子模块用于对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于各预置类别的得分,将得分最高的预置类别作为与所述第一子文本数据对应的类别。第一确定子模块用于当与所述一个或多个子文本中的各子文本数据对应的类别中存在第一类别时,确定与所述一个或多个子文本数据对应的类别为第一类别;以及,当与所述一个或多个子文本中的各子文本数据对应的类别均为第二类别时,确定与所述一个或多个子文本数据对应的类别为第二类别。

[0021] 可选地,预置类别包括第一类别和第二类别。综合分类模块包括第二确定子模块,用于当与所述全文本数据对应的类别和与所述一个或多个子文本数据对应的类别均为第二类别时,确定与所述待分类文本对应的类别为第二类别;以及,当与所述全文本数据对应的类别和/或与所述一个或多个子文本数据对应的类别为第一类别时,确定与所述待分类文本对应的类别为第一类别。

[0022] 本公开的另一方面提供了一种计算机设备,所述计算机设备包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序时用于实现如上所述的方法。

[0023] 本公开的另一方面提供了一种计算机可读存储介质,存储有计算机可执行指令,所述指令在被执行时用于实现如上所述的方法。

[0024] 本公开的另一方面提供了一种计算机程序,所述计算机程序包括计算机可执行指令,所述指令在被执行时用于实现如上所述的方法。

## 附图说明

- [0025] 为了更完整地理解本公开及其优势,现在将参考结合附图的以下描述,其中:
- [0026] 图1示意性示出了根据本公开的实施例的文本分类方法和计算机设备的应用场景;
- [0027] 图2示意性示出了根据本公开的实施例的文本分类方法的流程图;
- [0028] 图3A示意性示出了根据本公开的实施例的文本分类过程的示意图;
- [0029] 图3B示意性示出了根据本公开的另一实施例的文本分类过程的示意图;
- [0030] 图4示意性示出了根据本公开的实施例的文本分类装置的框图;
- [0031] 图5示意性示出了根据本公开的另一实施例的文本分类装置的框图;以及
- [0032] 图6示意性示出了根据本公开的实施例的计算机设备的框图。

## 具体实施方式

[0033] 以下,将参照附图来描述本公开的实施例。但是应该理解,这些描述只是示例性的,而并非要限制本公开的范围。在下面的详细描述中,为便于解释,阐述了许多具体的细节以提供对本公开实施例的全面理解。然而,明显地,一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外,在以下说明中,省略了对公知结构和技术的描述,以避免不必要地混淆本公开的概念。

[0034] 在此使用的术语仅仅是为了描述具体实施例,而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、步骤、操作和/或部件的存在,但是并不排除存在或添加一个或多个其他特征、步骤、操作或部件。

[0035] 在此使用的所有术语(包括技术和科学术语)具有本领域技术人员通常所理解的含义,除非另外定义。应注意,这里使用的术语应解释为具有与本说明书的上下文相一致的含义,而不应以理想化或过于刻板的方式来解释。

[0036] 在使用类似于“A、B和C等中至少一个”这样的表述的情况下,一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如,“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。在使用类似于“A、B或C等中至少一个”这样的表述的情况下,一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释(例如,“具有A、B或C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等)。

[0037] 附图中示出了一些方框图和/或流程图。应理解,方框图和/或流程图的一些方框或其组合可以由计算机程序指令来实现。这些计算机程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器,从而这些指令在由该处理器执行时可以创建用于实现这些方框图和/或流程图中所说明的功能/操作的装置。本公开的技术可以硬件和/或软件(包括固件、微代码等)的形式来实现。另外,本公开的技术可以采取存储有指令的计算机可读存储介质上的计算机程序产品的形式,该计算机程序产品可供指令执行系统使用或者结合指令执行系统使用。

[0038] 本公开的实施例提供了一种文本分类方法以及能够应用该方法的计算机设备。该方法包括待分类文本获取阶段、分类阶段和综合处理阶段。在待分类文本获取阶段获取待

分类文本。在分类阶段,分别获得与待分类文本的全文本数据对应的第一分类结果以及与从所述全文本数据中提取的一个或多个子文本数据对应的第二分类结果。然后进入综合处理阶段,根据第一分类结果和第二分类结果,确定待分类文本的分类结果。

[0039] 图1示意性示出了根据本公开的实施例的文本分类方法和计算机设备的应用场景。需要注意的是,图1所示仅为可以应用本公开实施例的场景的示例,以帮助本领域技术人员理解本公开的技术内容,但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

[0040] 如图1所示,该应用场景可以包括终端设备101、102、103,网络104和服务器/服务器集群105。网络104用以在终端设备101、102、103和服务器/服务器集群105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0041] 用户可以使用终端设备101、102、103通过网络104与服务器/服务器集群105交互,以接收或发送消息等。终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

[0042] 服务器/服务器集群105可以是提供各种服务的服务器或服务器集群,后台管理服务器或服务器集群可以对接收到的用户请求等数据进行分析等处理,并将处理结果反馈给终端设备。

[0043] 需要说明的是,本公开实施例所提供的文本分类方法一般可以由服务器/服务器集群105执行。相应地,本公开实施例所提供的文本分类装置一般可以设置于服务器/服务器集群105中。本公开实施例所提供的文本分类方法也可以由不同于服务器/服务器集群105且能够与终端设备101、102、103、和/或、服务器/服务器集群105通信的服务器或服务器集群执行。相应地,本公开实施例所提供的文本分类装置也可以设置于不同于服务器/服务器集群105且能够与终端设备101、102、103、和/或、服务器/服务器集群105通信的服务器或服务器集群中。

[0044] 应该理解,图1中的终端设备、网络和服务器/服务器集群的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器/服务器集群。

[0045] 图2示意性示出了根据本公开的实施例的文本分类方法的流程图。

[0046] 如图2所示,该方法包括操作S201~S204。

[0047] 在操作S201,获取待分类文本。

[0048] 在操作S202,基于所述待分类文本的全文本数据,获得第一分类结果。

[0049] 在操作S203,基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果。

[0050] 在操作S204,根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果。

[0051] 可见,图2所示的方法对于待分类文本,一方面基于该待分类文本的全文本数据获得第一分类结果,另一方面基于该待分类文本的全文本数据中的子文本数据获得第二分类结果,结合第一分类结果和第二分类结果得到待处理文本的分类结果。不仅从待分类文本整体角度出发考虑其分类,还从待分类文本中的关键信息角度出发考虑其分类,二者结合能够有效提高文本分类结果的召回率和准确率。

[0052] 在本公开的一个实施例中,操作S202基于所述待分类文本的全文本数据,获得第



一分类结果包括：将所述全文本数据输入对应于多个预置类别的全文本分类模型，基于所述全文本分类模型确定所述全文本数据关于所述多个预置类别中的各预置类别的第一得分，将第一得分最高的预置类别作为与所述全文本数据对应的类别。

[0053] 其中，获取到待分类文本后，对待分类文本进行预处理得到全文本数据，全文本分类模型对应于多个预置类别。将该全文本数据输入全文本分类模型，分别得到该全文本数据关于各个预置类别的第一得分。全文本数据关于一个预置类别X的第一得分表示预测全文本数据属于此预置类别X的评分。当全文本数据关于一个预置类别X的第一得分大于等于该全文本数据关于其他任一预置类别的第一得分时，将预置类别X作为与全文本数据对应的类别。利用全文本分类模型对全文本数据进行分类的优点在于分类过程中能够从整体考虑全文本数据中的特征并考虑特征之间的关联，但由于全文本数据中还存在大量与文本分类无关的干扰信息，可能导致文本分类的判别性特征湮没在干扰信息中，进而无法得到准确的第一分类结果。因此后续要对第一分类结果与第二分类结果进行综合处理。

[0054] 在本公开的一个实施例中，在操作S203基于从所述全文本数据中提取的一个或多个子文本数据，获得第二分类结果之前，图2所示的方法还包括：从所述全文本数据中提取一个或多个子文本数据。

[0055] 具体地，上述从所述全文本数据中提取一个或多个子文本数据包括：利用预设关键词集合中的关键词与所述全文本数据进行匹配；对于匹配成功的第一关键词，从所述全文本数据中提取所述第一关键词之前的第一预设长度的字符串和/或所述第一关键词之后的第二预设长度的字符串；以及，将提取出的字符串和所述第一关键词按照在所述全文本数据中的位置顺序组合为一个子文本数据。

[0056] 其中，预设关键词集合包括对文本分类起到关键作用的一个或多个关键词。基于预设关键词集合对全文本数据进行匹配，对于每个匹配位置，该匹配位置上匹配成功的关键词与该关键词前后的字符串组合得到该关键词在该匹配位置的上下文文本数据，将该上下文文本数据作为一个子文本数据。由于子文本数据筛选出具有关键作用的关键词信息，去除了全文本数据中冗余信息的干扰，使得基于子文本数据进行分类的过程相比于基于全文本数据进行分类的过程会更关注与文本分类相关的判别性特征。进一步地，子文本数据不仅包括关键词信息，还包括该关键词的上下文信息，将关键词放在具体的上下文情境中才能体现关键词在文本分类中起到的关键性作用，使得第二分类结果更加准确。

[0057] 在本公开的一个实施例中，操作S203基于从所述全文本数据中提取的一个或多个子文本数据，获得第二分类结果包括：对于第一子文本数据，将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型，基于所述子文本分类模型确定所述第一子文本数据关于所述多个预置类别中的各预置类别的第二得分；以及，基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算所述一个或多个子文本数据关于各预置类别的第三得分，将第三得分最高的预置类别作为与所述一个或多个子文本数据对应的类别。

[0058] 其中，子文本分类模型与全文本分类模型所对应的预置类别一致。对于任一子文本数据，将该子文本数据输入子文本分类模型，分别得到该子文本数据关于各个预置类别的第二得分。子文本数据关于一个预置类别X的第二得分表示预测子文本数据属于此预置类别X的评分。

[0059] 当子文本数据仅有一个时,该子文本数据关于各个预置类别的第二得分等于子文本数据整体关于各个预置类别的第三得分。当子文本数据关于一个预置类别X的第二得分大于等于该子文本数据关于其他任一预置类别的第二得分时,将预置类别X作为与子文本数据对应的类别。

[0060] 当子文本数据有多个时,根据各子文本数据关于各个预置类别的第二得分计算得到子文本数据整体关于各个预置类别的第三得分。当子文本数据整体关于一个预置类别X的第三得分大于等于子文本数据整体关于其他任一预置类别的第三得分时,将预置类别X作为与子文本数据整体对应的类别。

[0061] 上述过程逐子文本数据地利用子文本分类模型进行分类,得到每个子文本数据的分类结果,再结合各子文本数据的分类结果得到第二分类结果。由于不同的子文本数据包含处于不同上下文情景中的相同或不同的关键词信息,不同的子文本数据能够表达全文本数据中与文本分类相关的不同特征,子文本分类模型对于每个子文本数据的分类过程相当于对与每个子文本数据对应的特征的提取和分类过程,综合考虑到各种与文本分类相关的特征,得到第二分类结果。

[0062] 具体地,上述基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算一个或多个子文本数据关于各预置类别的第三得分包括:对于所述多个预置类别中的任一预置类别,将各子文本数据关于该预置类别的第二得分进行加权求和,得到所述一个或多个子文本数据关于该预置类别的第三得分。

[0063] 其中,预设关键词集合中的关键词可以具有不同的等级,该等级表示相应关键词在文本分类中所起作用的程度。例如,如果文本中包含某个关键词时该文本一定为为违规文本,则该关键词具有最高等级。可以根据各子文本数据所包含的关键词的等级分别为各子文本数据设置相应的权重,基于与各子文本数据对应的权重,对各子文本数据关于同一预置类别X的第二得分进行加权求和,得到子文本数据整体关于该预置类别X的第三得分。

[0064] 在此基础上,在本公开的一个实施例中,操作S204根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果包括:根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分,将综合得分最高的预置类别作为与所述待分类文本对应的类别。

[0065] 具体地,上述根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分包括:设置与所述全文本数据对应的第一权重和与所述一个或多个子文本数据对应的第二权重;以及,对于所述多个预置类别中的任一预置类别,根据所述第一权重和所述第二权重,对所述全文本数据关于该预置类别的第一得分和所述一个或多个子文本数据关于该预置类别的第三得分进行加权求和,得到所述待分类文本关于该预置类别的综合得分。

[0066] 本实施例的方案基于第一分类结果和第二分类结果得到待分类文本的分类结果,由于第一分类结果以全文本数据中的特征整体及特征间关联为分类基础,第二分类结果以各与文本分类相关的判别性特征为分类基础,二者结合,既能把握整体分类趋势的正确,也能关注到判别性特征以使分类结果正确。具体地,可以根据全文本分类过程和子文本分类过程的重要程度,分别为第一分类结果和第二分类结果设置第一权重和第二权重,基于第

一权重和第二权重,对关于同一预置类别X的第一得分和第三得分进行加权求和,得到关于该预置类别X的综合得分,根据关于各预置类别的综合得分确定最终分类结果。

[0067] 在本公开的另一个实施例中,操作S203基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果包括:对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于各预置类别的得分,将得分最高的预置类别作为与所述第一子文本数据对应的类别。当与所述一个或多个子文本中的各子文本数据对应的类别中存在第一类别时,确定与所述一个或多个子文本数据对应的类别为第一类别;以及,当与所述一个或多个子文本中的各子文本数据对应的类别均为第二类别时,确定与所述一个或多个子文本数据对应的类别为第二类别。

[0068] 其中,子文本分类模型与全文本分类模型所对应的预置类别一致,预置类别包括第一类别和第二类别,第一类别的风险等级最高,第二类别的风险等级最低。对于任一子文本数据,将该子文本数据输入子文本分类模型,分别得到该子文本数据关于各个预置类别的第二得分,子文本数据关于一个预置类别X的第二得分表示预测子文本数据属于此预置类别X的评分。当子文本数据关于一个预置类别X的第二得分大于等于该子文本数据关于其他任一预置类别的第二得分时,将预置类别X作为与子文本数据对应的类别。

[0069] 当子文本数据仅有一个时,与该子文本数据对应的类别即为与子文本数据整体对应的类别。当子文本数据有多个时,当与子文本数据对应的类别中存在第一类别时,将第一类别作为与子文本数据整体对应的类别。当与子文本数据对应的类别均为第二类别时,将第二类别作为与子文本数据整体对应的类别。

[0070] 上述过程逐子文本数据地利用子文本分类模型进行分类,得到与每个子文本数据对应的类别,关于各子文本数据所对应的类别中是否存在风险等级最高的第一类别,是则直接将第一类别作为与子文本数据整体对应的类别,即第二分类结果。该过程适用于预置类别中包含需要特别关注的类别(如第一类别)的情景,给该需要特别关注的类别以最高的判断权重,以消除该需要特别关注的类别的文本可能导致的安全威胁。

[0071] 在此基础上,在本公开的一个实施例中,操作S204根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果包括:当与所述全文本数据对应的类别和与所述一个或多个子文本数据对应的类别均为第二类别时,确定与所述待分类文本对应的类别为第二类别;以及,当与所述全文本数据对应的类别和/或与所述一个或多个子文本数据对应的类别为第一类别时,确定与所述待分类文本对应的类别为第一类别。

[0072] 本实施例的方案基于第一分类结果和第二分类结果得到待分类文本的分类结果,由于第一分类结果以全文本数据中的特征整体及特征间关联为分类基础,第二分类结果以各与文本分类相关的判别性特征为分类基础,二者结合,既能把握整体分类趋势的正确,也能关注到判别性特征以使分类结果正确。具体地,当与全文本数据对应的类别和与子文本数据对应的类别中存在风险等级最高的第一类别时,待分类文本的分类结果是第一类别,当与全文本数据对应的类别和与子文本数据对应的类别均为风险等级最低的第二类别时,待分类文本的分类结果是第二类别。该过程适用于预置类别中包含需要特别关注的类别(如第一类别)的情景,给该需要特别关注的类别以最高的判断权重,以消除该需要特别关注的类别的文本可能导致的安全威胁。

[0073] 下面参考图3A~图3B,结合具体实施例对图2所示的方法进行说明:

[0074] 图3A示意性示出了根据本公开的实施例的文本分类过程的示意图。

[0075] 如图3A所示,在文本分类过程中,首先获取待分类文本,对待分类文本进行预处理,得到全文本数据。其中,预处理包括一些常规的处理方法,如去除待分类文本中的HTML标签、对待分类文本进行分词、去除冗余词等操作。进行预处理后的全文本基于预设关键词集合进行匹配,在全文本找到匹配成功的关键词,共有 $N(N>0)$ 个匹配位置,提取各匹配位置前后一定长度范围作为相应关键词的上下文文本,称为子文本数据,本例中共提取出 $N$ 个子文本数据。其中,预设关键词集合包括根据具体分类任务收集的重要词语。采用关键词匹配就是为了能够提取到全文本中与文本分类任务相关的关键内容,剔除无关文本的干扰。

[0076] 一方面,将全文本数据输入全文本分类模型,使用全文本分类模型预测与全文本数据对应的分类结果。全文本分类模型在训练阶段采用各种语料库中的文本全文作为训练集,以训练全文本分类模型针对文本整体内容进行分类的能力。全文本分类模型的选择有多种,如基于卷积神经网络的文本分类(Text-CNN)、递归神经网络(RNN)等分类模型。全文本分类模型输出与全文本数据对应的分类结果,对应于图2中的第一分类结果。具体地,全文本分类模型对应于3个预置类别:类别1、类别2和类别3。与全文本数据对应的分类结果包括:全文本数据关于类别1的第一得分 $a(1)$ ,全文本数据关于类别2的第一得分 $a(2)$ ,以及全文本数据关于类别3的第一得分 $a(3)$ 。

[0077] 另一方面,将 $N$ 个子文本数据中的每个子文本数据输入子文本分类模型,使用子文本分类模型预测与每个子文本数据对应的分类结果。子文本分类模型在训练阶段采用预设关键词集合的关键词在各种语料库中的上下文文本作为训练集,以训练子文本分类模型针对子文本数据进行分类的能力。子文本分类模型的选择有多种,如基于卷积神经网络的文本分类(Text-CNN)、递归神经网络(RNN)等分类模型。子文本分类模型输出与每个子文本数据对应的分类结果,共得到与 $N$ 个子文本数据对应的 $N$ 个分类结果,对应于图2中的第二分类结果。具体地,子文本分类模型对应于3个预置类别:类别1、类别2和类别3。对于第 $i$ 个子文本数据,与该子文本数据对应的分类结果包括:第 $i(1 \leq i \leq N, i$ 为整数)个子文本数据关于类别1的第二得分 $b_i(1)$ ,第 $i$ 个子文本数据关于类别2的第二得分 $b_i(2)$ ,以及第 $i$ 个子文本数据关于类别3的第二得分 $b_i(3)$ 。

[0078] 根据具体分类任务为中模型的重要程度为所用到的模型分配权重,用分配权重方案结合全文本分类模型和子文本分类模型分类结果得到最终分类结果。本例中,为输出与全文本数据对应的分类结果的全文本分类模型分配权重 $\alpha$ ,为输出与第 $i$ 个子文本数据对应的分类结果的子文本分类模型分配权重 $\beta_i$ ,则得到待分类文本关于类别1的综合得分:

$$f(1) = \alpha a(1) + \sum_{i=1}^N \beta_i b_i(1), \text{待分类文本关于类别2的综合得分:}$$

$$f(2) = \alpha a(2) + \sum_{i=1}^N \beta_i b_i(2), \text{待分类文本关于类别3的综合得分:}$$

$f(3) = \alpha a(3) + \sum_{i=1}^N \beta_i b_i(3)$ 。选取综合得分最高的类别作为待分类文本的类别,即得到最终分类结果,对应于图2中的待分类文本的分类结果。

[0079] 本实施例的方案与现有技术相比,精确定位关键文本内容,避免了无关文本内容的干扰,有效的提高了文本分类的召回率和准确率。

[0080] 图3B示意性示出了根据本公开的另一实施例的文本分类过程的示意图。

[0081] 如图3B所示,该文本分类过程具体为判断网页是否是违规内容(色情、赌博、恐怖主义等相关内容)的分类过程。首先由爬虫系统爬取网页源码,该网页源码带有HTML标签,作为待分类文本。对网页源码进行一些预处理操作,比如去除HTML标签、切词等操作。经过预处理后得到网页原文,作为全文本数据。预设关键词集合包括对文本分类起到关键作用的一些关键词,比如“香港赛马”、“澳门赌博”、“一码中特”、“一夜情”等。基于预设关键词集合对网页原文进行匹配,具体利用多模匹配算法进行匹配,找到匹配成功的关键词,共有N( $N > 0$ )个匹配位置,提取各关键词匹配位置前后各60个字符长度的字符串,将提取到的字符串和匹配关键词本身按照在网页原文中的顺序组合得到大约120个字符长度的关键词上下文文本,作为子文本数据,本例中共提取出N个子文本数据。

[0082] 一方面,将全文本数据输入全文本分类模型,使用全文本分类模型预测与全文本数据对应的分类结果。本例中全文本分类模型采用基于卷积神经网络的文本分类(Text-CNN)模型,对应于两个预置类别:“违规”和“正常”。全文本分类模型输出与全文本数据对应的分类结果,对应于图2中的第一分类结果,该第一分类结果为“违规”或“正常”。

[0083] 另一方面,将N个子文本数据中的每个子文本数据输入子文本分类模型,使用子文本分类模型预测与每个子文本数据对应的分类结果。本例中,子文本分类模型也采用基于卷积神经网络的文本分类(Text-CNN)模型。子文本分类模型输出与每个子文本数据对应的分类结果,共得到与N个子文本数据对应的N个分类结果,对应于图2中的第二分类结果,与每个子文本数据对应的分类结果为“违规”或“正常”。

[0084] 为了结合两个模型输出的分类结果,采用的权重分配方案是:当所有分类结果中存在“违规”时,为输出分类结果“违规”的模型分配权重1,为输出分类结果“正常”的模型分配权重0。当所有分类结果中不存在“违规”时,为输出分类结果“正常”的模型分配权重1。也就是说,如果分类结果中存在“违规”的类别,最终网页的分类结果是“违规”,如果分类结果中均为“正常”的类别,最终网页的分类结果是“正常”,以给予“违规”分类结果以最高的优先级和敏感度,以避免违规网页所带来的安全威胁。

[0085] 经过测试,在这种违规内容分类任务中,本实施例的方案相比于现有技术,分类结果的准确率至少提高10%。

[0086] 图4示意性示出了根据本公开的实施例的文本分类装置的框图。

[0087] 如图4所示,文本分类装置400包括:获取模块410、第一分类模块420、第二分类模块430和综合分类模块440。

[0088] 获取模块410用于获取待分类文本。

[0089] 第一分类模块420用于基于所述待分类文本的全文本数据,获得第一分类结果。

[0090] 第二分类模块430用于基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果。

[0091] 综合分类模块440用于根据所述第一分类结果和所述第二分类结果,确定所述待分类文本的分类结果。

[0092] 图5示意性示出了根据本公开的另一实施例的文本分类装置的框图。

[0093] 如图5所示,文本分类装置500包括:获取模块410、第一分类模块420、第二分类模块430和综合分类模块440。

[0094] 其中,第一分类模块420用于将所述全文本数据输入对应于多个预置类别的全文

本分类模型,基于所述全文本分类模型确定所述全文本数据关于所述多个预置类别中的各预置类别的第一得分,将第一得分最高的预置类别作为与所述全文本数据对应的类别。

[0095] 在本公开的一个实施例中,文本分类装置500还包括子文本提取模块450,用于在所述第二分类模块基于从所述全文本数据中提取的一个或多个子文本数据,获得第二分类结果之前,从所述全文本数据中提取一个或多个子文本数据。

[0096] 其中,子文本提取模块450可以包括:匹配子模块451、提取子模块452和组合子模块453。

[0097] 匹配子模块451用于利用预设关键词集合中的关键词与所述全文本数据进行匹配。提取子模块452用于对于匹配成功的第一关键词,从所述全文本数据中提取所述第一关键词之前的第一预设长度的字符串和/或所述第一关键词之后的第二预设长度的字符串。以及,组合子模块453用于将提取出的字符串和所述第一关键词按照在所述全文本数据中的位置顺序组合为一个子文本数据。

[0098] 在本公开的一个实施例中,第二分类模块430包括第一预测子模块431和计算子模块432。

[0099] 第一预测子模块431用于对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于所述多个预置类别中的各预置类别的第二得分。以及,计算子模块432用于基于所述一个或多个子文本数据中的各子文本数据关于各预置类别的第二得分计算所述一个或多个子文本数据关于各预置类别的第三得分,将第三得分最高的预置类别作为与所述一个或多个子文本数据对应的类别。

[0100] 其中,计算子模块432具体用于对于所述多个预置类别中的任一预置类别,将各子文本数据关于该预置类别的第二得分进行加权求和,得到所述一个或多个子文本数据关于该预置类别的第三得分。

[0101] 在本公开的一个实施例中,综合分类模块440包括综合计算子模块441,用于根据所述全文本数据关于各预置类别的第一得分和所述一个或多个子文本数据关于各预置类别的第三得分,计算得到所述待分类文本关于各预置类别的综合得分,将综合得分最高的预置类别作为与所述待分类文本对应的类别。

[0102] 具体地,作为一个可选的实施例,综合计算子模块441用于设置与所述全文本数据对应的第一权重和与所述一个或多个子文本数据对应的第二权重;以及,对于所述多个预置类别中的任一预置类别,根据所述第一权重和所述第二权重,对所述全文本数据关于该预置类别的第一得分和所述一个或多个子文本数据关于该预置类别的第三得分进行加权求和,得到所述待分类文本关于该预置类别的综合得分。

[0103] 在本公开的一个实施例中,第二分类模块430包括第二预测子模块433和第一确定子模块434。第二预测子模块433用于对于第一子文本数据,将所述第一子文本数据输入对应于所述多个预置类别的子文本分类模型,基于所述子文本分类模型确定所述第一子文本数据关于各预置类别的得分,将得分最高的预置类别作为与所述第一子文本数据对应的类别。第一确定子模块434用于当与所述一个或多个子文本中的各子文本数据对应的类别中存在第一类别时,确定与所述一个或多个子文本数据对应的类别为第一类别;以及,当与所述一个或多个子文本中的各子文本数据对应的类别均为第二类别时,确定与所述一个或多个

个子文本数据对应的类别为第二类别。

[0104] 其中,预置类别包括第一类别和第二类别。在本公开的一个实施例中,综合分类模块440包括第二确定子模块442,用于当与所述全文本数据对应的类别和与所述一个或多个子文本数据对应的类别均为第二类别时,确定与所述待分类文本对应的类别为第二类别;以及,当与所述全文本数据对应的类别和/或与所述一个或多个子文本数据对应的类别为第一类别时,确定与所述待分类文本对应的类别为第一类别。

[0105] 需要说明的是,装置部分实施例中各模块/单元/子单元等的实施方式、解决的技术问题、实现的功能、以及达到的技术效果分别与方法部分实施例中各对应的步骤的实施方式、解决的技术问题、实现的功能、以及达到的技术效果相同或类似,在此不再赘述。

[0106] 根据本公开的实施例的模块、子模块、单元、子单元中的任意多个、或其中任意多个的至少部分功能可以在一个模块中实现。根据本公开实施例的模块、子模块、单元、子单元中的任意一个或多个可以被拆分成多个模块来实现。根据本公开实施例的模块、子模块、单元、子单元中的任意一个或多个可以至少被部分地实现为硬件电路,例如现场可编程门阵列(FPGA)、可编程逻辑阵列(PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路(ASIC),或可以通过对电路进行集成或封装的任何其他的合理方式的硬件或固件来实现,或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者,根据本公开实施例的模块、子模块、单元、子单元中的一个或多个可以至少被部分地实现为计算机程序模块,当该计算机程序模块被运行时,可以执行相应的功能。

[0107] 例如,获取模块410、第一分类模块420、第二分类模块430、综合分类模块440和子文本提取模块450中的任意多个可以合并在一个模块中实现,或者其中的任意一个模块可以被拆分成多个模块。或者,这些模块中的一个或多个模块的至少部分功能可以与其他模块的至少部分功能相结合,并在一个模块中实现。根据本公开的实施例,获取模块410、第一分类模块420、第二分类模块430、综合分类模块440和子文本提取模块450中的至少一个可以至少被部分地实现为硬件电路,例如现场可编程门阵列(FPGA)、可编程逻辑阵列(PLA)、片上系统、基板上的系统、封装上的系统、专用集成电路(ASIC),或可以通过对电路进行集成或封装的任何其他的合理方式等硬件或固件来实现,或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者,获取模块410、第一分类模块420、第二分类模块430、综合分类模块440和子文本提取模块450中的至少一个可以至少被部分地实现为计算机程序模块,当该计算机程序模块被运行时,可以执行相应的功能。

[0108] 图6示意性示出了根据本公开的实施例的适于实现上文描述的方法的计算机设备的框图。图6示出的计算机设备仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0109] 如图6所示,计算机设备600包括处理器610和计算机可读存储介质620。该计算机设备600可以执行根据本公开实施例的方法。

[0110] 具体地,处理器610例如可以包括通用微处理器、指令集处理器和/或相关芯片组和/或专用微处理器(例如,专用集成电路(ASIC)),等等。处理器610还可以包括用于缓存用途的板载存储器。处理器610可以是用于执行根据本公开实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

[0111] 计算机可读存储介质620,例如可以是非易失性的计算机可读存储介质,具体示例

包括但不限于:磁存储装置,如磁带或硬盘(HDD);光存储装置,如光盘(CD-ROM);存储器,如随机存取存储器(RAM)或闪存;等等。

[0112] 计算机可读存储介质620可以包括计算机程序621,该计算机程序621可以包括代码/计算机可执行指令,其在由处理器610执行时使得处理器610执行根据本公开实施例的方法或其任何变形。

[0113] 计算机程序621可被配置为具有例如包括计算机程序模块的计算机程序代码。例如,在示例实施例中,计算机程序621中的代码可以包括一个或多个程序模块,例如包括621A、模块621B、……。应当注意,模块的划分方式和个数并不是固定的,本领域技术人员可以根据实际情况使用合适的程序模块或程序模块组合,当这些程序模块组合被处理器610执行时,使得处理器610可以执行根据本公开实施例的方法或其任何变形。

[0114] 根据本发明的实施例,获取模块410、第一分类模块420、第二分类模块430、综合分类模块440和子文本提取模块450中的至少一个可以实现为参考图6描述的计算机程序模块,其在被处理器610执行时,可以实现上文所述的文本分类方法。

[0115] 本公开还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的;也可以是单独存在,而未装配入该设备/装置/系统中。上述计算机可读存储介质承载有一个或者多个程序,当上述一个或者多个程序被执行时,实现根据本公开实施例的方法。

[0116] 根据本公开的实施例,计算机可读存储介质可以是非易失性的计算机可读存储介质,例如可以包括但不限于:便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0117] 附图中的流程图和框图,图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0118] 本领域技术人员可以理解,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合,即使这样的组合或结合没有明确记载于本公开中。特别地,在不脱离本公开精神和教导的情况下,本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本公开的范围。

[0119] 尽管已经参照本公开的特定示例性实施例示出并描述了本公开,但是本领域技术人员应该理解,在不背离所附权利要求及其等同物限定的本公开的精神和范围的情况下,可以对本公开进行形式和细节上的多种改变。因此,本公开的范围不应该限于上述实施例,



而是应该不仅由所附权利要求来进行确定,还由所附权利要求的等同物来进行限定。

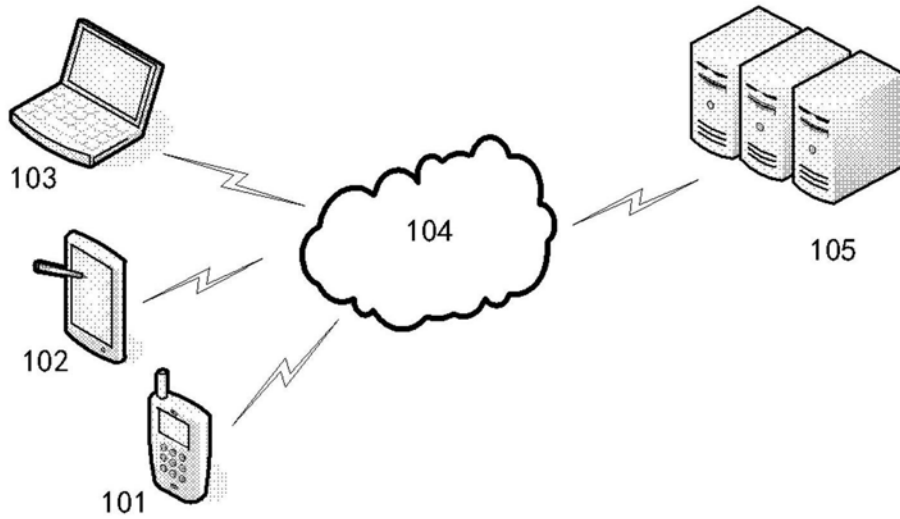


图1

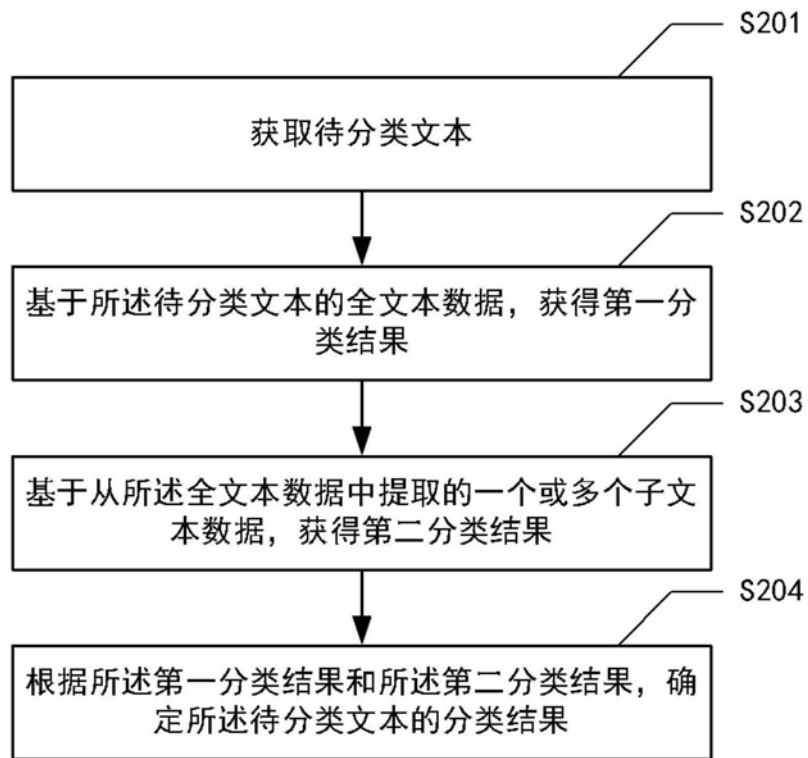


图2

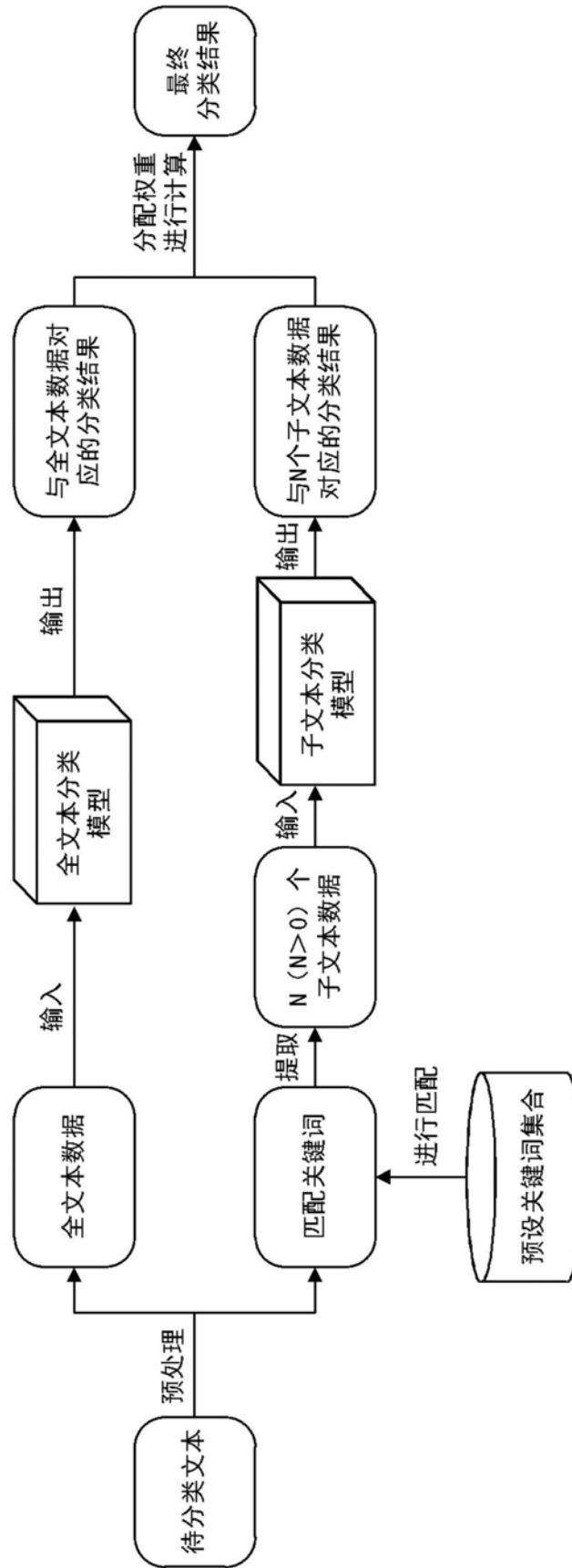


图3A

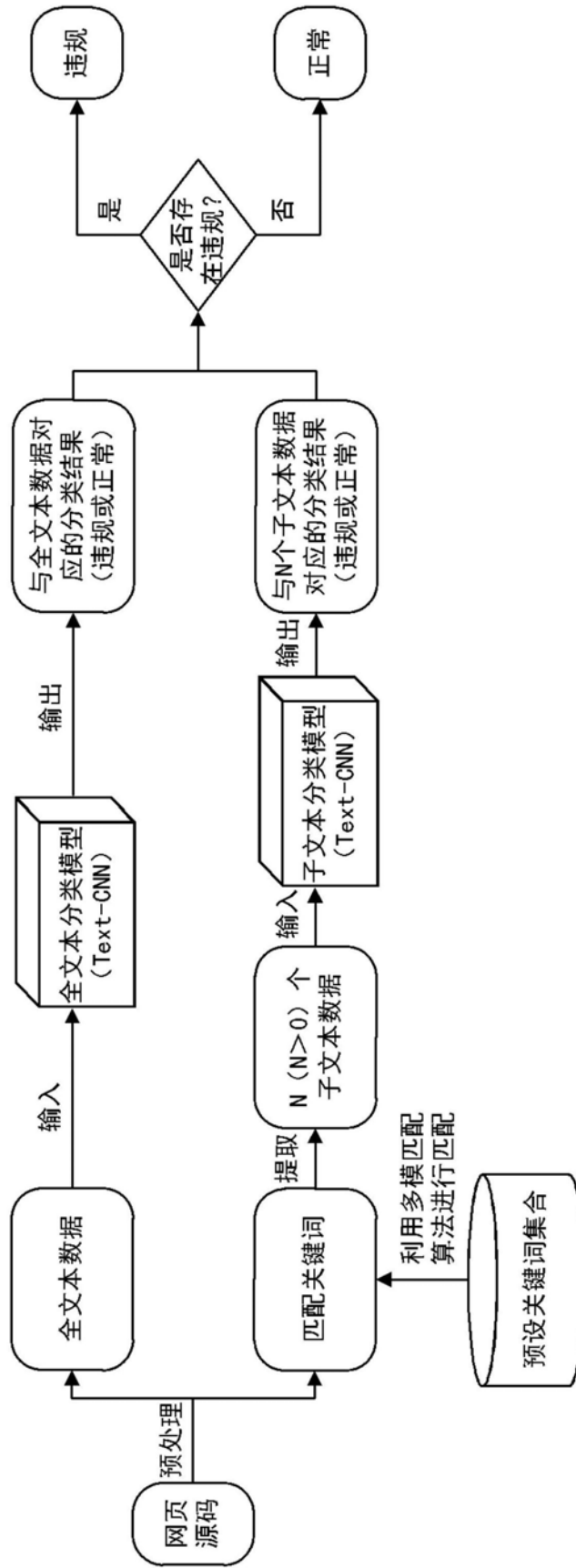


图3B

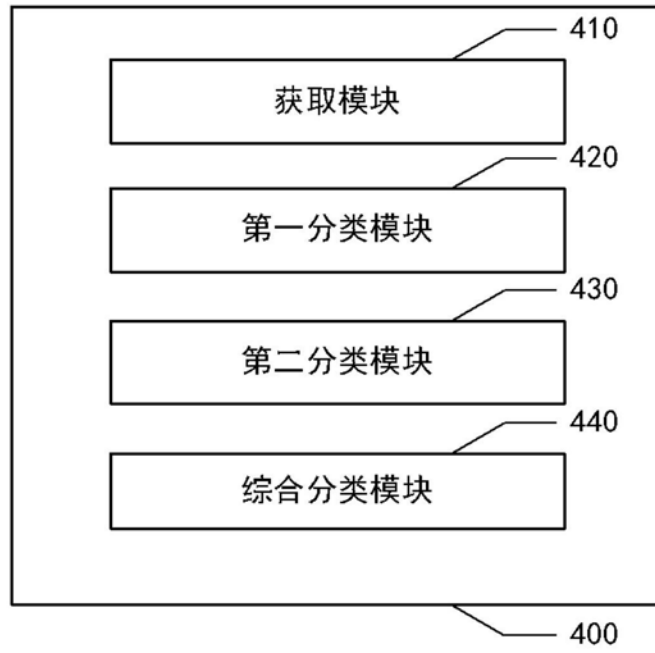


图4

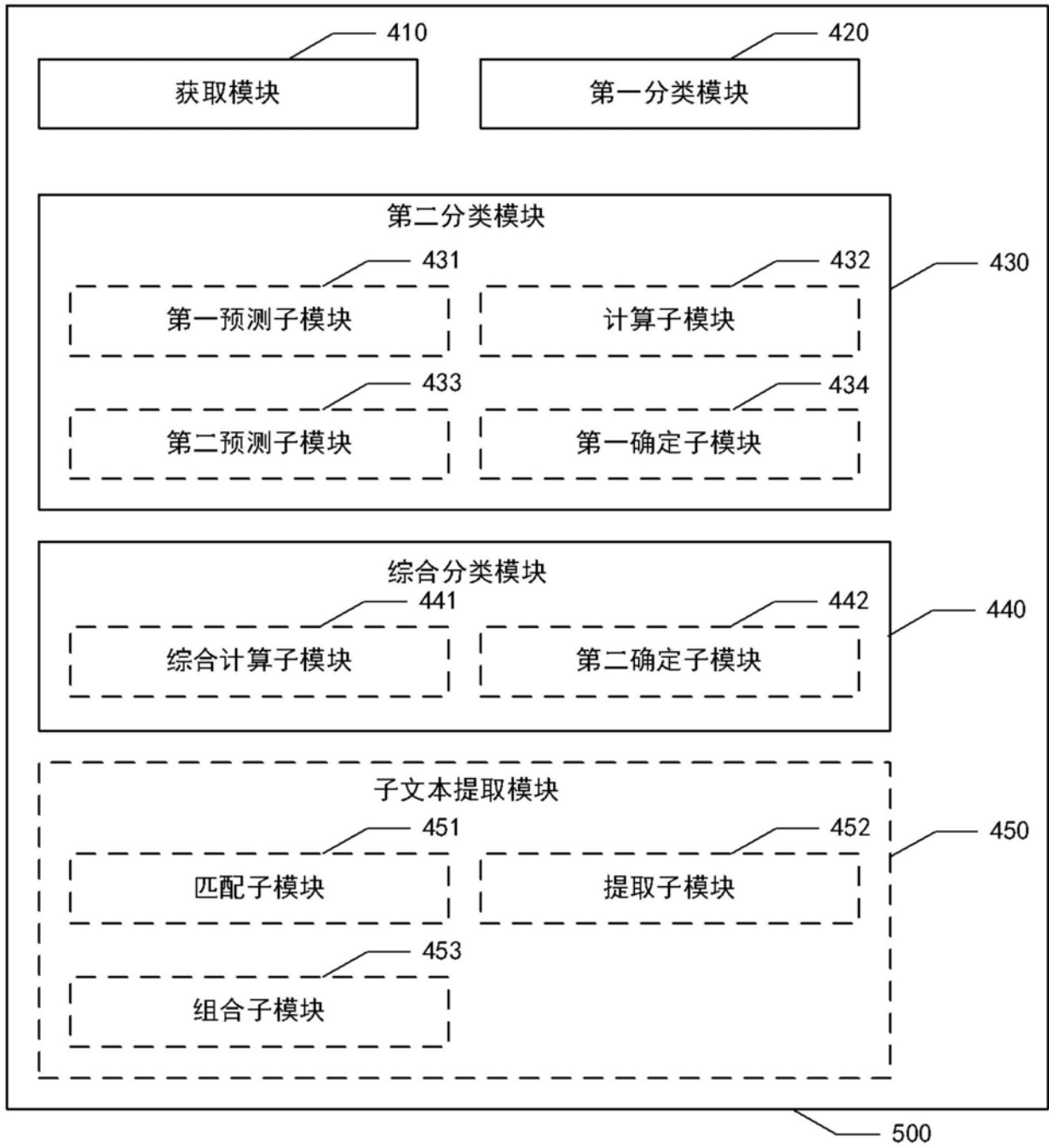


图5

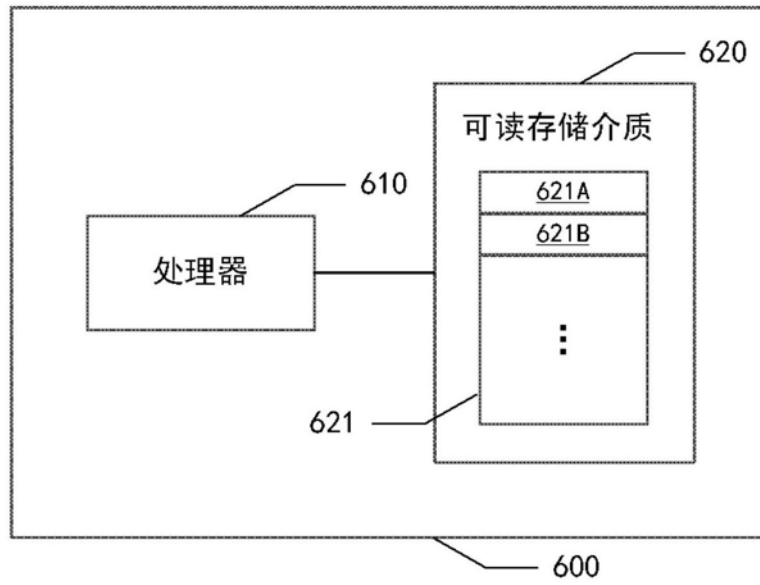


图6