(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112395414 B
(45)授权公告日 2024.06.04

(21)申请号 201910759761.8
(22)申请日 2019.08.16
(65)同一申请的已公布的文献号
申请公布号 CN 112395414 A
(43)申请公布日 2021.02.23
(73)专利权人 北京地平线机器人技术研发有限公司
地址 100080 北京市海淀区中关村大街1号3层318
(72)发明人 马腾岳 周蕾蕾
(74)专利代理机构 北京思源智汇知识产权代理有限公司 11657
专利代理师 毛丽琴
(51)Int.Cl.
G06F 16/35(2019.01)
G06F 16/33(2019.01)

(56)对比文件
CN 109858035 A,2019.06.07
CN 109918680 A,2019.06.21
CN 109918682 A,2019.06.21
CN 108170733 A,2018.06.15
CN 110119786 A,2019.08.13
CN 108415897 A,2018.08.17
CN 109918673 A,2019.06.21
CN 110019782 A,2019.07.16
US 2017286399 A1,2017.10.05

审查员 刘林林

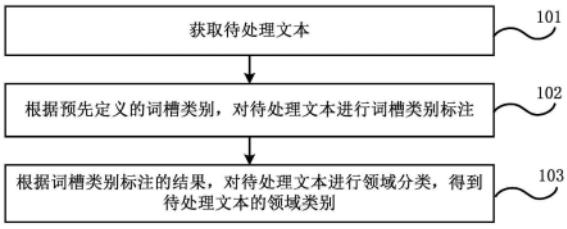
权利要求书3页 说明书13页 附图5页

(54)发明名称

文本分类方法和分类模型的训练方法、装置、介质和设备

(57)摘要

本公开实施例公开了一种文本分类方法和分类模型的训练方法、装置、介质和设备。其中，文本分类方法包括：获取待处理文本；根据预先定义的词槽类别，对所述待处理文本进行词槽类别标注；根据所述词槽类别标注的结果，对所述待处理文本进行领域分类，得到所述待处理文本的领域类别。本公开实施例可以实现对句子准确的领域分类，从而提高了领域分类的准确率。



CN 112395414 B

1.一种文本分类方法,包括:

获取待处理文本;

根据预先定义的词槽类别,对所述待处理文本进行词槽类别标注;

根据所述词槽类别标注的结果,对所述待处理文本进行领域分类,得到所述待处理文本的领域类别;

其中,所述根据所述词槽类别标注的结果,对所述待处理文本进行领域分类,包括:

根据所述词槽类别标注的结果,确定所述待处理文本对应的句式;

基于所述待处理文本对应的句式,确定所述待处理文本的领域类别。

2.根据权利要求1所述的方法,其中,所述基于所述待处理文本对应的句式,确定所述待处理文本的领域类别,包括:

提取所述待处理文本对应的句式中的特征,得到所述待处理文本的文本特征;

基于所述待处理文本的文本特征,对所述待处理文本进行领域分类,得到所述待处理文本的领域类别。

3.根据权利要求2所述的方法,其中,所述基于所述待处理文本的文本特征,对所述待处理文本进行领域分类,包括:

将所述待处理文本的文本特征输入领域分类模型,通过所述领域分类模型对所述待处理文本进行领域分类,得到所述待处理文本的领域类别。

4.根据权利要求1至3中任意一项所述的方法,其中,所述根据预先定义的词槽类别,对所述待处理文本进行词槽类别标注,包括:

将所述待处理文本输入序列标注模型,通过所述序列标注模型标注所述待处理文本的词槽类别。

5.一种分类模型的训练方法,包括:

获取第一数据集,所述第一数据集中的样本标注有领域类别信息;

根据预先定义的词槽类别,对所述第一数据集中的样本进行词槽类别标注;

根据所述词槽类别标注的结果,利用所述第一数据集训练领域分类模型;

其中,所述根据所述词槽类别标注的结果,利用所述第一数据集训练领域分类模型,包括:

根据所述词槽类别标注的结果,确定所述第一数据集中的样本对应的句式;

基于所述第一数据集中的样本对应的句式训练所述领域分类模型。

6.根据权利要求5所述的方法,其中,所述基于所述第一数据集中的样本对应的句式训练所述领域分类模型,包括:

提取所述第一数据集中的样本对应的句式中的特征,得到所述第一数据集中的样本的文本特征;

基于所述第一数据集中的样本的文本特征训练所述领域分类模型。

7.根据权利要求6所述的方法,其中,所述基于所述第一数据集中的样本的文本特征训练所述领域分类模型,包括:

将所述第一数据集中的样本的文本特征输入所述领域分类模型,通过所述领域分类模型对所述第一数据集中的样本进行领域预测,得到所述第一数据集中的样本的领域类别预测信息;

根据所述第一数据集中的样本的领域类别预测信息与所述第一数据集中的样本标注的领域类别信息之间的差异,对所述领域分类模型进行训练。

8.根据权利要求5至7中任意一项所述的方法,其中,所述根据预先定义的词槽类别,对所述第一数据集中的样本进行词槽类别标注,包括:

将所述第一数据集中的样本输入序列标注模型,通过所述序列标注模型标注所述第一数据集中的样本的词槽类别。

9.根据权利要求8所述的方法,其中,所述将所述第一数据集中的样本输入序列标注模型,通过所述序列标注模型标注所述第一数据集中的样本的词槽类别之前,还包括:

获取第二数据集,所述第二数据集中的样本根据所述预先定义的词槽类别标注有词槽类别信息;

利用所述第二数据集训练所述序列标注模型。

10.根据权利要求9所述的方法,其中,所述利用所述第二数据集训练所述序列标注模型,包括:

将所述第二数据集中的样本输入所述序列标注模型,通过所述序列标注模型对所述第二数据集中的样本进行词槽类别预测,得到所述第二数据集中的样本的词槽类别预测信息;

根据所述第二数据集中的样本的词槽类别预测信息与所述第二数据集中的样本标注的词槽类别信息之间的差异,对所述序列标注模型进行训练。

11.一种领域分类装置,包括:

第一获取模块,用于获取待处理文本;

标注模块,用于根据预先定义的词槽类别,对所述第一获取模块获取的所述待处理文本进行词槽类别标注;

分类模块,用于根据所述标注模块得到的所述词槽类别标注的结果,对所述待处理文本进行领域分类,得到所述待处理文本的领域类别;

所述分类模块对所述待处理文本进行领域分类时,具体用于根据所述词槽类别标注的结果,确定所述待处理文本对应的句式;基于所述待处理文本对应的句式,确定所述待处理文本的领域类别。

12.一种分类模型的训练装置,包括:

第二获取模块,用于获取第一数据集,所述第一数据集中的样本标注有领域类别信息;

标注模块,用于根据预先定义的词槽类别,对所述第二获取模块获取的所述第一数据集中的样本进行词槽类别标注;

第一训练模块,用于根据所述标注模块得到的所述词槽类别标注的结果,利用所述第一数据集训练领域分类模型;

所述第一训练模块利用所述第一数据集训练领域分类模型时,具体用于根据所述词槽类别标注的结果,确定所述第一数据集中的样本对应的句式;基于所述第一数据集中的样本对应的句式训练所述领域分类模型。

13.一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序用于执行上述权利要求1至10中任意一项所述的方法。

14.一种电子设备,所述电子设备包括:

处理器；

用于存储所述处理器可执行指令的存储器；

所述处理器，用于从所述存储器中读取所述可执行指令，并执行所述指令以实现上述权利要求1至10中任意一项所述的方法。

文本分类方法和分类模型的训练方法、装置、介质和设备

技术领域

[0001] 本公开涉及语音技术,尤其是一种文本分类方法和分类模型的训练方法、装置、介质和设备。

背景技术

[0002] 语音识别技术,也被称为自动语音识别(Automatic Speech Recognition,ASR),是将人类的语音转换为计算机可读的输入形式的技术。在语音识别的过程中,在将人类的语音转换为文本后,需要对文本进行语义理解,才能够将文本转换为计算机可读的输入形式。

[0003] 其中,短文本分类是语义理解的关键步骤。短文本分类是指确定文本中句子属于的领域类别信息,例如:“播放儿歌”,属于“音乐”领域;“今天的天气”,属于“天气”领域。

发明内容

[0004] 为了解决现有技术中的至少一个技术问题,本公开实施例提供了一种文本分类的技术方案和分类模型训练的技术方案。

[0005] 根据本公开实施例的一个方面,提供了一种文本分类方法,包括:

[0006] 获取待处理文本;

[0007] 根据预先定义的词槽类别,对所述待处理文本进行词槽类别标注;

[0008] 根据所述词槽类别标注的结果,对所述待处理文本进行领域分类,得到所述待处理文本的领域类别。

[0009] 根据本公开实施例的另一个方面,提供了一种分类模型的训练方法,包括:

[0010] 获取第一数据集,所述第一数据集中的样本标注有领域类别信息;

[0011] 根据预先定义的词槽类别,对所述第一数据集中的样本进行词槽类别标注;

[0012] 根据所述词槽类别标注的结果,利用所述第一数据集训练领域分类模型。

[0013] 根据本公开实施例的又一个方面,提供了一种文本分类装置,包括:

[0014] 第一获取模块,用于获取待处理文本;

[0015] 标注模块,用于根据预先定义的词槽类别,对所述第一获取模块获取的所述待处理文本进行词槽类别标注;

[0016] 分类模块,用于根据所述标注模块得到的所述词槽类别标注的结果,对所述待处理文本进行领域分类,得到所述待处理文本的领域类别。

[0017] 根据本公开实施例的再一个方面,提供了一种分类模型的训练装置,包括:

[0018] 第二获取模块,用于获取第一数据集,所述第一数据集中的样本标注有领域类别信息;

[0019] 标注模块,用于根据预先定义的词槽类别,对所述第二获取模块获取的所述第一数据集中的样本进行词槽类别标注;

[0020] 第一训练模块,用于根据所述标注模块得到的所述词槽类别标注的结果,利用所

述第一数据集训练领域分类模型。

[0021] 根据本公开实施例的再一个方面,提供了一种计算机可读存储介质,所述存储介质存储有计算机程序,所述计算机程序用于执行上述任一实施例所述的方法。

[0022] 根据本公开实施例的再一个方面,提供了一种电子设备,所述电子设备包括:

[0023] 处理器;

[0024] 用于存储所述处理器可执行指令的存储器;

[0025] 所述处理器,用于从所述存储器中读取所述可执行指令,并执行所述指令以实现上述任一实施例所述的方法。

[0026] 基于本公开上述实施例提供的文本分类方法和装置、计算机可读存储介质和电子设备,根据预先定义的词槽类别,对待处理文本进行词槽类别标注,根据词槽类别标注的结果,对待处理文本进行领域分类。由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,可以实现对句子准确的领域分类,从而提高了领域分类的准确率。

[0027] 基于本公开上述实施例提供的分类模型的训练方法和装置、计算机可读存储介质和电子设备,根据预先定义的词槽类别,对第一数据集中的样本进行词槽类别标注,第一数据集中的样本标注有领域类别信息,然后根据词槽类别标注的结果,利用第一数据集训练领域分类模型。利用本实施例的方法训练好的领域分类模型对待处理文本进行领域分类时,由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,对于待处理文本中存在的未出现在训练样本中的词语的句子,仍然可以根据句子的词槽类别,对句子进行准确的领域分类,从而提高领域分类的准确率。

附图说明

[0028] 通过结合附图对本公开实施例进行更详细的描述,本公开的上述以及其他目的、特征和优势将变得更加明显。附图用来提供对本公开实施例的进一步理解,并且构成说明书的一部分,与本公开实施例一起用于解释本公开,并不构成对本公开的限制。在附图中,相同的参考标号通常代表相同部件或步骤。

[0029] 图1是本公开所适用的场景图。

[0030] 图2是本公开一示例性实施例提供的文本分类方法的流程示意图。

[0031] 图3是本公开另一示例性实施例提供的文本分类方法的流程示意图。

[0032] 图4是本公开又一示例性实施例提供的文本分类方法的流程示意图。

[0033] 图5是本公开一示例性实施例提供的分类模型的训练方法的流程示意图。

[0034] 图6是本公开另一示例性实施例提供的分类模型的训练方法的流程示意图。

[0035] 图7是本公开又一示例性实施例提供的分类模型的训练方法的流程示意图。

[0036] 图8是本公开一示例性实施例提供的文本分类装置的结构示意图。

[0037] 图9是本公开另一示例性实施例提供的文本分类装置的结构示意图。

[0038] 图10是本公开一示例性实施例提供的分类模型的训练装置的结构示意图。

[0039] 图11是本公开另一示例性实施例提供的分类模型的训练装置的结构示意图。

[0040] 图12是本公开一示例性实施例提供的电子设备的结构图。

具体实施方式

[0041] 下面,将参考附图详细地描述根据本公开的示例实施例。显然,所描述的实施例仅仅是本公开的一部分实施例,而不是本公开的全部实施例,应理解,本公开不受这里描述的示例实施例的限制。

[0042] 应注意到:除非另外具体说明,否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

[0043] 本领域技术人员可以理解,本公开实施例中的“第一”、“第二”等术语仅用于区别不同步骤、设备或模块等,既不代表任何特定技术含义,也不表示它们之间的必然逻辑顺序。

[0044] 还应理解,在本公开实施例中,“多个”可以指两个或两个以上,“至少一个”可以指一个、两个或两个以上。

[0045] 还应理解,对于本公开实施例中提及的任一部件、数据或结构,在没有明确限定或者在前后文给出相反启示的情况下,一般可以理解为一个或多个。

[0046] 另外,本公开中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本公开中字符“/”,一般表示前后关联对象是一种“或”的关系。

[0047] 还应理解,本公开对各个实施例的描述着重强调各个实施例之间的不同之处,其相同或相似之处可以相互参考,为了简洁,不再一一赘述。

[0048] 同时,应当明白,为了便于描述,附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

[0049] 以下对至少一个示例性实施例的描述实际上仅仅是说明性的,决不作为对本公开及其应用或使用的任何限制。

[0050] 对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论,但在适当情况下,所述技术、方法和设备应当被视为说明书的一部分。

[0051] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步讨论。

[0052] 本公开实施例可以应用于终端设备、计算机系统、服务器等电子设备,其可与众多其它通用或专用计算系统环境或配置一起操作。适于与终端设备、计算机系统、服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子包括但不限于:个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境,等等。

[0053] 终端设备、计算机系统、服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令(诸如程序模块)的一般语境下描述。通常,程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等,它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施,分布式云计算环境中,任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中,程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

[0054] 申请概述

[0055] 在实现本发明的过程中,本发明人通过研究发现,现有的领域分类方法是基于直接对文本中的原始句子进行特征提取、然后通过领域分类模型来进行短文本分类,这样,在训练样本为小样本(即:训练样本的数量较少、覆盖类别不全面)的情况下,对于存在未出现在训练样本中的词语的句子,句子的领域分类很容易出错,从而影响领域分类的准确率。

[0056] 例如,假设训练样本中存在“刘德华的冰雨”,则基于上述现有技术的分类方法可以对文本“刘德华的冰雨”正确分类,但是对于不存在于训练样本中的文本“周杰伦的青花瓷”,则基于上述现有技术的分类方法就容易分类出错。

[0057] 本公开实施例对文本进行领域分类时,根据对文本标注的词槽类别来完成,并不考虑具体的词语,这样,对于文本中存在的未出现在训练样本中的词语的句子,仍然可以根据句子的词槽类别,对句子进行准确的领域分类,从而提高领域分类的准确率。

[0058] 示例性系统

[0059] 本公开实施例可以应用于与机器人、儿童玩具、音响等有语音交互的场景,也可以应用于搜索等场景。图1是本公开所适用的一个场景图。如图1所示,本公开实施例应用于语音交互场景时,由音频采集模块(例如麦克风等)采集原始音频信号,经前端信号处理模块处理后的语音,进行语音识别,得到文本信息;对文本信息进行语义理解和领域分类,并基于领域分类结果在相应领域的信息库进行搜索后输出搜索结果。例如,针对用户的语音“周杰伦的青花瓷”,基于本公开实施例可以分类到音乐领域,从音乐数据库中搜索“周杰伦的青花瓷”并进行返回。

[0060] 另外,本公开实施例应用于搜索场景时,用户可以输入文本信息,例如“李白的静思夜”,服务器对该文本信息进行语义理解和领域分类,并基于分类结果在相应类别的信息库进行搜索后输出搜索结果,例如“李白的静思夜”被分到诗歌领域,服务器在诗歌数据库通过关键字“李白的静思夜”搜索诗歌,并返回给用户。

[0061] 示例性方法

[0062] 图2是本公开一示例性实施例提供的文本分类的流程示意图。本实施例可应用在电子设备上,如图2所示,本实施例的文本分类方法包括如下步骤:

[0063] 步骤101,获取待处理文本。

[0064] 其中的待处理文本,可以是用户输入的文本,例如“我要听周杰伦的歌”;或者,也可以是对用户输入的语音进行语音识别得到的文本信息。其中,用户输入的语音可以是由音频采集模块(例如麦克风等)采集的原始音频信号,也可以是该原始音频信号经前端信号处理模块处理后的语音。

[0065] 其中,前端信号处理模块对音频信号的处理例如可以包括但不限于:语音活动检测(Voice Activity Detection,VAD)、降噪、声学回声消除(Acoustic Echo Cancellaction,AEC)、去混响处理、声源定位、波束形成(Beam Forming,BF)等。

[0066] 语音活动检测(Voice Activity Detection,VAD)又称语音端点检测、语音边界检,是指在噪声环境中检测音频信号中语音的存在与否,准确的检测出音频信号中语音段起始位置,通常用于语音编码、语音增强等语音处理系统中,起到降低语音编码速率、节省通信带宽、减少移动设备能耗、提高识别率等作用。

[0067] 步骤102,根据预先定义的词槽(slot)类别,对待处理文本进行词槽类别标注。

[0068] 步骤103,根据词槽类别标注的结果,对待处理文本进行领域分类,得到待处理文

本的领域类别。

[0069] 基于本公开上述实施例提供的文本分类方法,根据预先定义的词槽类别,对待处理文本进行词槽类别标注,根据词槽类别标注的结果,对待处理文本进行领域分类。由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,可以实现对句子准确的领域分类,从而提高了领域分类的准确率。

[0070] 本公开实施例中,可以预先定义全领域类别的词槽。例如,如下表1所示,为本公开实施例定义的词槽示例:

[0071] 表1

[0072]

Slot(词槽)类别	含义	举例
artist	人名	周杰伦、刘德华……
title	作品名	青花瓷、冰雨……
poi	位置	中关村、西直门……
time	时间	今天、十月三号、星期二……
location	地点	北京、南京……
……	……	……

[0073] 在图2所示实施例的步骤102中,根据预先定义的词槽类别,对待处理文本进行词槽类别标注,例如可以是:

[0074] 对于待处理文本“刘德华的冰雨”,基于步骤102进行词槽类别标注,得到:[刘德华:artist]的[冰雨:title];

[0075] 对于待处理文本“导航到中关村”,基于步骤102进行词槽类别标注,得到:导航到[中关村:poi];

[0076] 对于待处理文本“今天的天气”,基于步骤102进行词槽类别标注,得到:[今天:time]的天气。

[0077] 在其中一些实施方式中,步骤102中,可以将待处理文本输入序列标注模型,通过序列标注模型标注待处理文本的词槽类别。

[0078] 在一些可选示例中,序列标注模型可以通过隐马尔可夫模型(Hidden Markov Model,HMM)、最大熵模型(Maximum Entropy Model,MaxEnt)、条件随机场算法(conditional random field algorithm,CRF)、神经网络等实现,其中的神经网络例如可以是卷积神经网络(CNN)、循环神经网络(RNN)等,本公开实施例对序列标注模型的实现方式不做限制。

[0079] 本实施例中,通过预先训练好的序列标注模型对待处理文本进行词槽类别标注,序列标注模型通过HMM、MaxEnt、CRF等对待处理文本进行词槽类别标注时,由于序列标注模型输出的是一个序列,输出序列本身会带有一些上下文关联,利用这些上下文关联,序列标注模型在对作为输入序列的待处理文本的标注上可以达到比传统分类方法更高的性能,因此提高了词槽类别的准确性和效率,从而提高了整个文本分类的效率。图3是本公开另一示例性实施例提供的文本分类的流程示意图。如图3所示,在上述图2所示实施例的基础上,步骤103可包括如下步骤:

[0080] 步骤1031,根据词槽类别标注的结果,确定待处理文本对应的句式。

[0081] 在其中一些实施方式中,可以直接以通过步骤102得到的词槽类别标注的结果,作

为待处理文本对应的句式。例如,可以直接以词槽类别标注的结果“[刘德华:artist]的[冰雨:title]”、“导航到[中关村:poi]”、“[今天:time]的天气”,作为待处理文本对应的句式。

[0082] 在另一些实施方式中,可以针对通过步骤102得到的词槽类别标注的结果,分别以标注的词槽类别代替待处理文本中相应的字词,得到待处理文本对应的句式。例如,针对词槽类别标注的结果“[刘德华:artist]的[冰雨:title]”、“导航到[中关村:poi]”、“[今天:time]的天气”,分别以标注的词槽类别代替待处理文本中相应的字词,得到“[artist]的[title]”、“导航到[poi]”、“[time]的天气”,作为待处理文本对应的句式。

[0083] 本公开实施例对待处理文本对应的句式的形式不做限制,只要可以体现标注的词槽类别即可。

[0084] 步骤1032,基于待处理文本对应的句式,确定待处理文本的领域类别。

[0085] 本实施例根据词槽类别标注的结果,确定待处理文本对应的句式,由于待处理文本对应的句式包括了标注的结果以及整个待处理文本的结构关系,基于待处理文本对应的句式来确定待处理文本的领域类别,可以得到较精确的领域类别,提高了整个文本分类的准确性和效率。

[0086] 图4是本公开又一示例性实施例提供的文本分类的流程示意图。如图4所示,在上述图3所示实施例的基础上,步骤1032可包括如下步骤:

[0087] 步骤10321,提取待处理文本对应的句式中的特征,得到待处理文本的文本特征。

[0088] 本公开实施例中的文本特意可以以特征向量或者特征图等的方式表示,本公开实施例对文本特征的表示方式不做限制。

[0089] 步骤10322,基于待处理文本的文本特征,对待处理文本进行领域分类,得到待处理文本的领域类别。

[0090] 例如,基于待处理文本的文本特征“[artist]的[title]”、“[artist]的”、“的[title]”,对待处理文本进行领域分类,可以得到待处理文本分别分类到各领域的分数,例如,音乐领域类别:0.95;导航领域类别:0.10;天气领域类别:0.10;…;选取一个分数最高的领域类别作为待处理文本的领域类别。

[0091] 在其中一些实施方式中,在步骤10321中,可以基于按字的N元模型(N-gram),采用固定长度为n的滑动窗口对待处理文本对应的句式进行切分、然后再提取特征,其中n的取值为大于0的整数,例如,2、3、4等。其中,标注的每个词槽类别算作1个字。例如,对于句式“导航到[poi]”,n=4;“导航到”,n=3;“导航”,n=2。

[0092] 例如,采用N元模型、n=2~4对待处理文本对应的句式进行切分、然后再进行提取特征,对于句式“[artist]的[title]”进行特征提取,可以得到以下几种可能的文本特征:[artist]的[title],[artist]的,的[title];对于句式“导航到[poi]”进行特征提取,可以得到以下几种可能的文本特征:导航到[poi],导航到,导航,航到[poi],到[poi];对于句式“[time]的天气”进行特征提取,可以得到以下几种可能的文本特征:[time]的天气,[time]的天,[time]的,的天气,的天,天气。

[0093] 另外,在训练样本为小样本的情况下,例如在音乐领域类别下,经常出现“刘德华”这样的词,如果采用现有的领域分类方法是基于直接对文本中的原始句子进行特征提取、然后通过领域分类模型来进行短文本分类,那么“刘德华有多高”这种不属于音乐领域类别的文本,也会被分到音乐类别,从而对短文本的领域分类产生比较严重的过拟合现象。本公

开实施例将待处理文本抽象为句式,变成“[artist]有多高”,文本特征只与句式有关,与具体的“刘德华”无关,从而可以实现对待处理文本的正确分类,降低了短文本的领域分类的过拟合现象。

[0094] 在其中一些实施方式中,该步骤10322中,可以将待处理文本的文本特征输入领域分类模型,通过领域分类模型对待处理文本进行领域分类,得到待处理文本的领域类别。

[0095] 在一些可选示例中,领域分类模型可以通过支持向量机(Support Vector Machine,SVM)、最大熵模型(Maximum Entropy Model,MaxEnt)、神经网络等实现,其中的神经网络例如可以是卷积神经网络(CNN)、循环神经网络(RNN)等,本公开实施例对领域分类模型的实现方式不做限制。

[0096] 本实施例中,通过预先训练好的领域分类模型对待处理文本进行领域分类,提高了领域分类结果的准确性和效率,从而提高了整个文本分类效率。

[0097] 在本公开上述各实施例的文本分类方法之前,还可以预先对领域分类模型和序列标注模型进行训练,然后基于训练好的领域分类模型和序列标注模型执行上述相应的操作。

[0098] 图5是本公开一示例性实施例提供的分类模型的训练方法的流程示意图。本实施例可应用在电子设备上,如图5所示,本实施例的分类模型的训练方法包括如下步骤:

[0099] 步骤201,获取第一数据集。

[0100] 其中,第一数据集中包括至少一个领域类别的样本,每个样本标注有领域类别信息。其中,每个样本标注的领域类别信息为相对准确的领域类别信息。

[0101] 步骤202,根据预先定义的词槽类别,对第一数据集中的样本进行词槽类别标注。

[0102] 步骤203,根据词槽类别标注的结果,利用第一数据集训练领域分类模型。

[0103] 基于本公开上述实施例提供的分类模型的训练方法,根据预先定义的词槽类别,对第一数据集中的样本进行词槽类别标注,第一数据集中的样本标注有领域类别信息,然后根据词槽类别标注的结果,利用第一数据集训练领域分类模型。利用本实施例的方法训练好的领域分类模型对待处理文本进行领域分类时,由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,在训练样本为小样本的情况下,对于待处理文本中存在的未出现在训练样本中的词语的句子,仍然可以根据句子的词槽类别,对句子进行准确的领域分类,从而提高领域分类的准确率。

[0104] 在其中一些实施方式中,步骤203包括:根据词槽类别标注的结果,确定第一数据集中的样本对应的句式;基于第一数据集中的样本对应的句式训练领域分类模型。

[0105] 在其中一些可选示例中,基于第一数据集中的样本对应的句式训练领域分类模型,可以包括:提取第一数据集中的样本对应的句式中的特征,得到第一数据集中的样本的文本特征;基于第一数据集中的样本的文本特征训练领域分类模型。

[0106] 图6是本公开另一示例性实施例提供的分类模型的训练方法的流程示意图。如图6所示,本实施例的分类模型的训练方法包括如下步骤:

[0107] 步骤301,获取第一数据集,该第一数据集中的样本标注有领域类别信息。

[0108] 其中,第一数据集中的样本,例如可以是:刘德华的冰雨,播放刘德华的歌,…,导航到中关村,我要去西直门导航,…,今天的天气,北京今天有雨吗,…,等等。样本标注的领域类别信息用于标识该样本的领域类别,样本的领域类别例如可以包括但不限于:音乐、诗

歌、导航、天气等等。利用标注有领域类别信息的样本的对领域分类模型训练完成后，领域分类模型便可以对属于上述样本的领域类别的文本进行分类。

[0109] 步骤302，根据预先定义的词槽类别，对第一数据集中的样本进行词槽类别标注。

[0110] 其中，词槽类别标注，例如：[刘德华:artist]的[冰雨:title]，导航到[中关村:poi]，[今天:time]的天气，等等。

[0111] 步骤303，根据词槽类别标注的结果，确定第一数据集中的样本对应的句式。

[0112] 其中的句式，例如可以是词槽类别标注的结果：[刘德华:artist]的[冰雨:title]，导航到[中关村:poi]，[今天:time]的天气；或者，也可以是分别以标注的词槽类别代替待处理文本中相应的字词得到的结果：[artist]的[title]，导航到[poi]，[time]的天气，等等。

[0113] 步骤304，提取第一数据集中的样本对应的句式中的特征，得到第一数据集中的样本的文本特征。

[0114] 其中，样本对应的句式中的特征，，可以是采用预设特征提取方式对样本对应的句式进行特征提取得到的文本特征。例如，采用N元模型、n＝2～4，对样本对应的句式“[artist]的[title]”进行特征提取，可以得到以下几种可能文本特征：[artist]的[title]，[artist]的，的[title]；对样本对应的句式“导航到[poi]”进行特征提取，可以得到以下几种可能的文本特征：导航到[poi]，导航到，导航，航到[poi]，到[poi]；对样本对应的句式“[time]的天气”进行特征提取，可以得到以下几种可能的文本特征：[time]的天气，[time]的天，[time]的，的天气，的天，天气。

[0115] 步骤305，基于第一数据集中的样本的文本特征训练领域分类模型。

[0116] 本实施例中，根据预先定义的词槽类别，对第一数据集中的样本进行词槽类别标注，根据词槽类别标注的结果，确定第一数据集中的样本对应的句式，然后提取第一数据集中的样本对应的句式中的特征，基于第一数据集中的样本的文本特征训练领域分类模型，利用本实施例的方法训练好的领域分类模型对待处理文本进行领域分类时，由于对待处理文本的领域分类并不需要考虑具体的词语，而是根据对待处理文本标注的词槽类别、对应的句式中的特征来完成，对于待处理文本中存在的未出现在训练样本中的词语的句子，仍然可以根据句子的词槽类别，对句子进行准确的领域分类，提高了领域分类结果的准确性和效率，从而提高了整个文本分类效率。

[0117] 在其中一些实施方式中，步骤305可以包括：将第一数据集中的样本的文本特征输入领域分类模型，通过领域分类模型对第一数据集中的样本进行领域预测，得到第一数据集中的样本的领域类别预测信息；根据第一数据集中的样本的领域类别预测信息与第一数据集中的样本标注的领域类别信息之间的差异，对领域分类模型进行训练。

[0118] 上述步骤305可以为一个迭代执行的过程。在其中一些可选示例中，可以根据第一数据集中的样本的领域类别预测信息与第一数据集中的样本标注的领域类别信息之间的差异，对领域分类模型的参数进行调整，直到满足训练完成条件，例如，第一数据集中的样本的领域类别预测信息与第一数据集中的样本标注的领域类别信息之间的差异小于预设阈值，或者对领域分类模型的训练次数达到预设次数。

[0119] 在其中一些实施方式中，上述图5-图6所示实施例的步骤202或302中，可以将第一数据集中的样本输入训练好的序列标注模型，通过训练好的序列标注模型标注第一数据集

中的样本的词槽类别。

[0120] 本实施例中,通过训练好的序列标注模型标注第一数据集中的样本的词槽类别,提高了词槽类别标注的准确性和效率。

[0121] 上述图6所示实施例中的领域分类模型训练完成后,即可用于实现上述图2-图5所示实施例中103中根据词槽类别标注的结果,对待处理文本进行领域分类,得到待处理文本的领域类别的操作,相关之处可以参见上述图2-图4所示实施例中的记载,此处不再赘述。

[0122] 图7是本公开又一示例性实施例提供的分类模型的训练方法的流程示意图。如图7所示,本实施例的分类模型的训练方法包括如下步骤:

[0123] 步骤401,获取第一数据集和第二数据集。

[0124] 其中,第一数据集中的样本标注有领域类别信息;第二数据集中的样本根据预先定义的词槽类别标注有词槽类别信息。

[0125] 其中,第一数据集中的样本及其标注的领域类别信息,可参见图6所示实施例中301的记载,此处不再赘述。

[0126] 第二数据集中的样本,例如可以是:刘德华的冰雨,播放刘德华的歌,…,导航到中关村,我要去西直门导航,…,今天的天气,北京今天有雨吗,…,等等。第二数据集中的样本标注的词槽类别信息可以是预先定义全领域类别的词槽,例如可以是artist、title、poi、time、location等等,具体可以参见上述表1。利用标注有词槽类别信息的样本的对序列标注模型训练完成后,序列标注模型便可以对文本进行相应的词槽类别标注。

[0127] 步骤402,利用第二数据集训练序列标注模型。

[0128] 步骤403,将第一数据集中的样本输入序列标注模型,通过序列标注模型标注第一数据集中的样本的词槽类别。

[0129] 步骤406,根据词槽类别标注的结果,确定第一数据集中的样本对应的句式。

[0130] 步骤405,提取第一数据集中的样本对应的句式中的特征,得到第一数据集中的样本的文本特征。

[0131] 步骤406,基于第一数据集中的样本的文本特征训练领域分类模型。

[0132] 本实施例中,预先利用样本数据集对序列标注模型进行训练,通过训练好的序列标注模型对待处理文本进行槽位信息标注,提高了槽位信息标注的准确性和效率,从而提高了整个文本分类效率。

[0133] 在其中一些实施方式中,步骤402可以包括:

[0134] 将第二数据集中的样本输入序列标注模型,通过序列标注模型对第二数据集中的样本进行词槽类别预测,得到第二数据集中的样本的词槽类别预测信息;

[0135] 根据第二数据集中的样本的词槽类别预测信息与第二数据集中的样本标注的词槽类别信息之间的差异,对序列标注模型进行训练。

[0136] 上述步骤402可以为一个迭代执行的过程。在其中一些可选示例中,可以根据第二数据集中的样本的词槽类别预测信息与第二数据集中的样本标注的词槽类别信息之间的差异,对序列标注模型的参数进行调整,直到满足训练完成条件,例如,第二数据集中的样本的词槽类别预测信息与第二数据集中的样本标注的词槽类别信息之间的差异小于预设阈值,或者对序列标注模型的训练次数达到预设次数。

[0137] 上述图7所示实施例中的序列标注模型和领域分类模型训练完成后,即可用于对

应实现上述图2-图5所示实施例中102和103的操作,相关之处可以参见上述图2-图4所示实施例中的记载,此处不再赘述。

[0138] 本公开实施例提供的任一种方法可以由任意适当的具有数据处理能力的设备执行,包括但不限于:终端设备和服务器等。或者,本公开实施例提供的任一种方法可以由处理器执行,如处理器通过调用存储器存储的相应指令来执行本公开实施例提及的任一种方法。下文不再赘述。

[0139] 示例性装置

[0140] 图8是本公开一示例性实施例提供的文本分类装置的结构示意图。该文本分类装置可以设置于终端设备、服务器等电子设备中,执行本公开上述任一实施例的文本分类方法。如图8所示,该文本分类装置包括:第一获取模块501,标注模块502,分类模块503。其中:

[0141] 第一获取模块501,用于获取待处理文本。

[0142] 标注模块502,用于根据预先定义的词槽类别,对第一获取模块501获取的待处理文本进行词槽类别标注。

[0143] 在其中一些实施方式中,标注模块502可以包括序列标注模型,用于将待处理文本输入序列标注模型,通过序列标注模型标注待处理文本的词槽类别。

[0144] 分类模块503,用于根据标注模块502得到的词槽类别标注的结果,对待处理文本进行领域分类,得到待处理文本的领域类别。

[0145] 基于本公开上述实施例提供的文本分类装置,根据预先定义的词槽类别,对待处理文本进行词槽类别标注,根据词槽类别标注的结果,对待处理文本进行领域分类。由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,可以实现对句子准确的领域分类,从而提高了领域分类的准确率。

[0146] 图9是本公开另一示例性实施例提供的文本分类装置的结构示意图。在上述图8所示实施例的基础上,分类模块503包括:第一确定单元5031,根据词槽类别标注的结果,确定待处理文本对应的句式;第二确定单元5032,基于待处理文本对应的句式,确定待处理文本的领域类别。

[0147] 在其中一些实施方式中,第二确定单元5032可以包括:提取子单元,用于提取待处理文本对应的句式中的特征,得到待处理文本的文本特征;分类子单元,用于基于待处理文本的文本特征,对待处理文本进行领域分类,得到待处理文本的领域类别。

[0148] 在其中一些可选示例中,分类子单元可以包括领域分类模型,用于将待处理文本的文本特征输入领域分类模型,通过领域分类模型对待处理文本进行领域分类,得到待处理文本的领域类别。

[0149] 图10是本公开一示例性实施例提供的分类模型的训练装置的结构示意图。该分类模型的训练装置可以设置于终端设备、服务器等电子设备中,执行本公开上述任一实施例的文本分类方法。如图10所示,该分类模型的训练装置包括:第二获取模块601,标注模块602,第一训练模块603。其中:

[0150] 第二获取模块601,用于获取第一数据集,第一数据集中的样本标注有领域类别信息。

[0151] 标注模块602,用于根据预先定义的词槽类别,对第二获取模块获取的第一数据集中的样本进行词槽类别标注。

[0152] 第一训练模块603,用于根据标注模块得到的词槽类别标注的结果,利用第一数据集训练领域分类模型。

[0153] 基于本公开上述实施例提供的分类模型的训练装置,根据预先定义的词槽类别,对第一数据集中的样本进行词槽类别标注,第一数据集中的样本标注有领域类别信息,然后根据词槽类别标注的结果,利用第一数据集训练领域分类模型。利用本实施例的方法训练好的领域分类模型对待处理文本进行领域分类时,由于对待处理文本的领域分类并不需要考虑具体的词语,而是根据对待处理文本标注的词槽类别来完成,在训练样本为小样本的情况下,对于待处理文本中存在的未出现在训练样本中的词语的句子,仍然可以根据句子的词槽类别,对句子进行准确的领域分类,从而提高领域分类的准确率。

[0154] 图11是本公开另一示例性实施例提供的分类模型的训练装置的结构示意图。在上述图10所示实施例的基础上,第一训练模块603包括:第三确定单元6031,用于根据词槽类别标注的结果,确定第一数据集中的样本对应的句式;第一训练单元6032,用于基于第一数据集中的样本对应的句式训练领域分类模型。

[0155] 再参见图11,在其中一些实施方式中,第一训练单元6032可以包括:提取子单元,用于提取第一数据集中的样本对应的句式中的特征,得到第一数据集中的样本的文本特征;训练子单元,用于基于第一数据集中的样本的文本特征训练领域分类模型。

[0156] 在其中一些可选示例中,训练子单元具体用于:将第一数据集中的样本的文本特征输入领域分类模型,通过领域分类模型对第一数据集中的样本进行领域预测,得到第一数据集中的样本的领域类别预测信息;根据第一数据集中的样本的领域类别预测信息与第一数据集中的样本标注的领域类别信息之间的差异,对领域分类模型进行训练。

[0157] 在其中一些可选示例中,标注模块602具体用于:将第一数据集中的样本输入序列标注模型,通过序列标注模型标注第一数据集中的样本的词槽类别。

[0158] 再参见图11,在又一示例性实施例提供的分类模型的训练装置中,还包括:第二获取模块604,用于获取第二数据集,第二数据集中的样本根据预先定义的词槽类别标注有词槽类别信息;第二训练模块605,用于利用第二数据集训练序列标注模型。

[0159] 在其中一些实施方式中,第二训练模块605可以包括:预测单元6051,用于将第二数据集中的样本输入序列标注模型,通过序列标注模型对第二数据集中的样本进行词槽类别预测,得到第二数据集中的样本的词槽类别预测信息;第二训练单元6052,用于根据第二数据集中的样本的词槽类别预测信息与第二数据集中的样本标注的词槽类别信息之间的差异,对序列标注模型进行训练。

[0160] 示例性电子设备

[0161] 下面,参考图12来描述根据本公开实施例的电子设备。该电子设备可以是第一设备和第二设备中的任一个或两者、或与它们独立的单机设备,该单机设备可以与第一设备和第二设备进行通信,以从它们接收所采集到的输入信号。

[0162] 图12图示了根据本公开实施例的电子设备的框图。如图12所示,电子设备包括一个或多个处理器701和存储器702。

[0163] 处理器701可以是中央处理单元(CPU)或者具有数据处理能力和/或指令执行能力的其他形式的处理单元,并且可以控制电子设备中的其他组件以执行期望的功能。

[0164] 存储器702可以包括一个或多个计算机程序产品,所述计算机程序产品可以包括

各种形式的计算机可读存储介质,例如易失性存储器和/或非易失性存储器。所述易失性存储器例如可以包括随机存取存储器(RAM)和/或高速缓冲存储器(cache)等。所述非易失性存储器例如可以包括只读存储器(ROM)、硬盘、闪存等。在所述计算机可读存储介质上可以存储一个或多个计算机程序指令,处理器701可以运行所述程序指令,以实现上文所述的本公开的各个实施例的方法以及/或者其他期望的功能。在所述计算机可读存储介质中还可以存储诸如输入信号、信号分量、噪声分量等各种内容。

[0165] 在一个示例中,电子设备还可以包括:输入装置703和输出装置704,这些组件通过总线系统和/或其他形式的连接机构(未示出)互连。

[0166] 例如,在该电子设备是第一设备或第二设备时,该输入装置703可以是上述的麦克风或麦克风阵列,用于捕捉声源的输入信号。在该电子设备是单机设备时,该输入装置703可以是通信网络连接器,用于从第一设备和第二设备接收所采集的输入信号。

[0167] 此外,该输入设备703还可以包括例如键盘、鼠标等等。

[0168] 该输出装置704可以向外部输出各种信息,包括确定出的距离信息、方向信息等。该输出设备704可以包括例如显示器、扬声器、打印机、以及通信网络及其所连接的远程输出设备等等。

[0169] 当然,为了简化,图12中仅示出了该电子设备中与本公开有关的组件中的一些,省略了诸如总线、输入/输出接口等等的组件。除此之外,根据具体应用情况,电子设备还可以包括任何其他适当的组件。

[0170] 示例性计算机程序产品和计算机可读存储介质

[0171] 除了上述方法和设备以外,本公开的实施例还可以是计算机程序产品,其包括计算机程序指令,所述计算机程序指令在被处理器运行时使得所述处理器执行本说明书上述“示例性方法”部分中描述的根据本公开各种实施例的方法中的步骤。

[0172] 所述计算机程序产品可以以一种或多种程序设计语言的任意组合来编写用于执行本公开实施例操作的程序代码,所述程序设计语言包括面向对象的程序设计语言,诸如Java、C++等,还包括常规的过程式程序设计语言,诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。

[0173] 此外,本公开的实施例还可以是计算机可读存储介质,其上存储有计算机程序指令,所述计算机程序指令在被处理器运行时使得所述处理器执行本说明书上述“示例性方法”部分中描述的根据本公开各种实施例的方法中的步骤。

[0174] 所述计算机可读存储介质可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以包括但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。

[0175] 以上结合具体实施例描述了本公开的基本原理,但是,需要指出的是,在本公开中提及的优点、优势、效果等仅是示例而非限制,不能认为这些优点、优势、效果等是本公开的

各个实施例必须具备的。另外，上述公开的具体细节仅是为了示例的作用和便于理解的作用，而非限制，上述细节并不限制本公开为必须采用上述具体的细节来实现。

[0176] 本说明书中各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言，由于其与方法实施例基本对应，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0177] 本公开中涉及的器件、装置、设备、系统的方框图仅作为例示性的例子并且不意图要求或暗示必须按照方框图示出的方式进行连接、布置、配置。如本领域技术人员将认识到的，可以按任意方式连接、布置、配置这些器件、装置、设备、系统。诸如“包括”、“包含”、“具有”等等的词语是开放性词汇，指“包括但不限于”，且可与其互换使用。这里所使用的词汇“或”和“和”指词汇“和/或”，且可与其互换使用，除非上下文明确指示不是如此。这里所使用的词汇“诸如”指词组“诸如但不限于”，且可与其互换使用。

[0178] 可能以许多方式来实现本公开的方法和装置。例如，可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法和装置。用于所述方法的步骤的上述顺序仅是为了进行说明，本公开的方法的步骤不限于以上具体描述的顺序，除非以其它方式特别说明。此外，在一些实施例中，还可将本公开实施为记录在记录介质中的程序，这些程序包括用于实现根据本公开的方法的机器可读指令。因而，本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

[0179] 还需要指出的是，在本公开的装置、设备和方法中，各部件或各步骤是可以分解和/或重新组合的。这些分解和/或重新组合应视为本公开的等效方案。

[0180] 提供所公开的方面的以上描述以使本领域的任何技术人员能够做出或者使用本公开。对这些方面的各种修改对于本领域技术人员而言是非常显而易见的，并且在此定义的一般原理可以应用于其他方面而不脱离本公开的范围。因此，本公开不意图被限制到在此示出的方面，而是按照与在此公开的原理和新颖的特征一致的最宽范围。

[0181] 为了例示和描述的目的已经给出了以上描述。此外，此描述不意图将本公开的实施例限制到在此公开的形式。尽管以上已经讨论了多个示例方面和实施例，但是本领域技术人员将认识到其某些变型、修改、改变、添加和子组合。

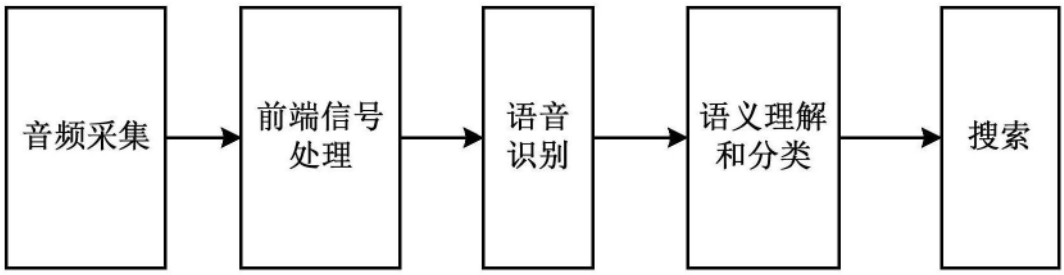

音频采集
前端信号处理
语音识别
语义理解和分类
搜索

图1

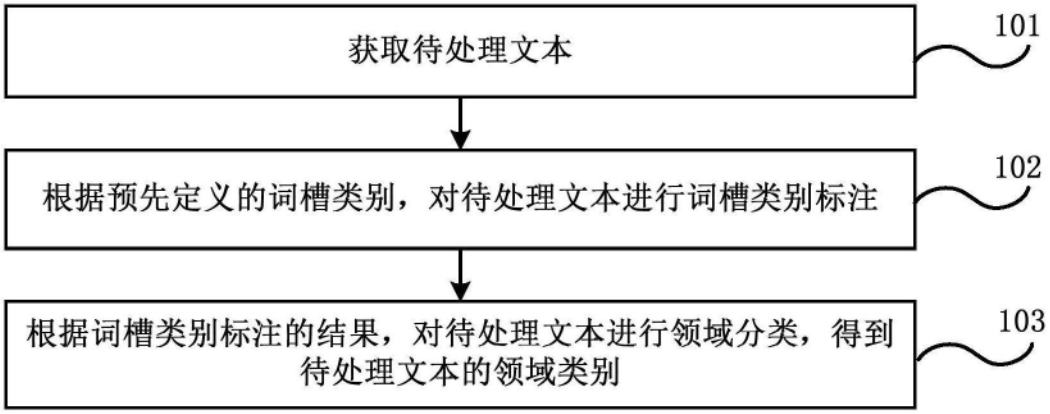

获取待处理文本
101
根据预先定义的词槽类别，对待处理文本进行词槽类别标注
102
根据词槽类别标注的结果，对待处理文本进行领域分类，得到待处理文本的领域类别
103

图2

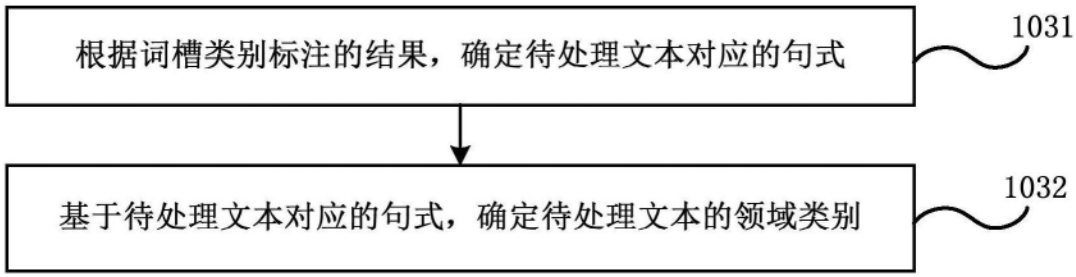

根据词槽类别标注的结果，确定待处理文本对应的句式
1031
基于待处理文本对应的句式，确定待处理文本的领域类别
1032

图3

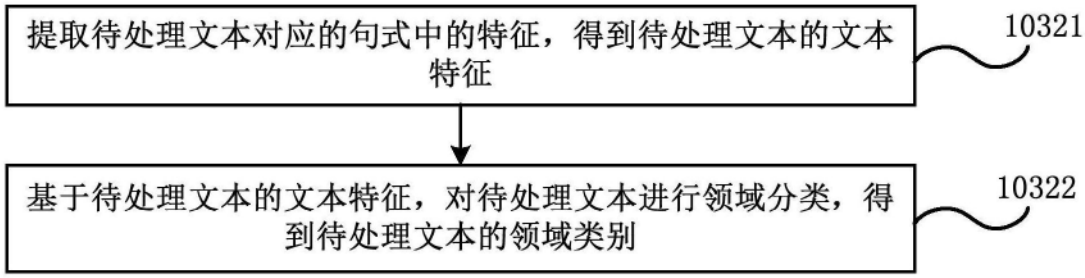

提取待处理文本对应的句式中的特征，得到待处理文本的文本特征
10321
基于待处理文本的文本特征，对待处理文本进行领域分类，得到待处理文本的领域类别
10322

图4

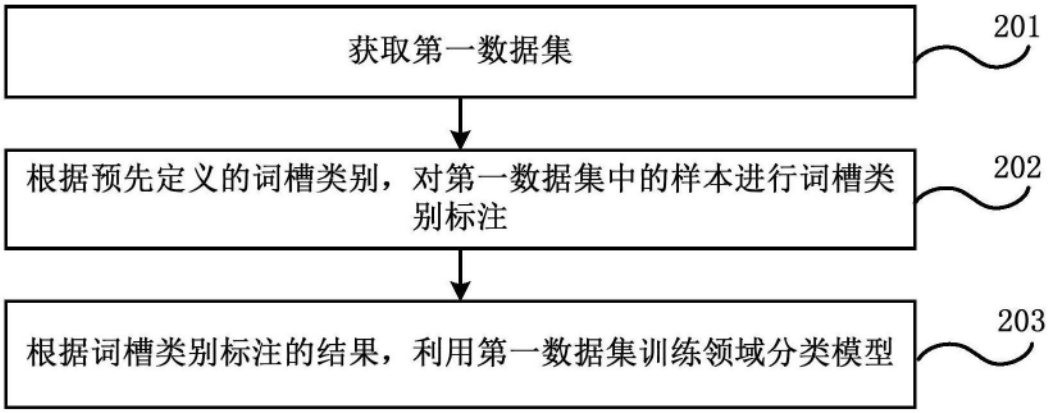

获取第一数据集
201
根据预先定义的词槽类别，对第一数据集中的样本进行词槽类别标注
202
根据词槽类别标注的结果，利用第一数据集训练领域分类模型
203

图5

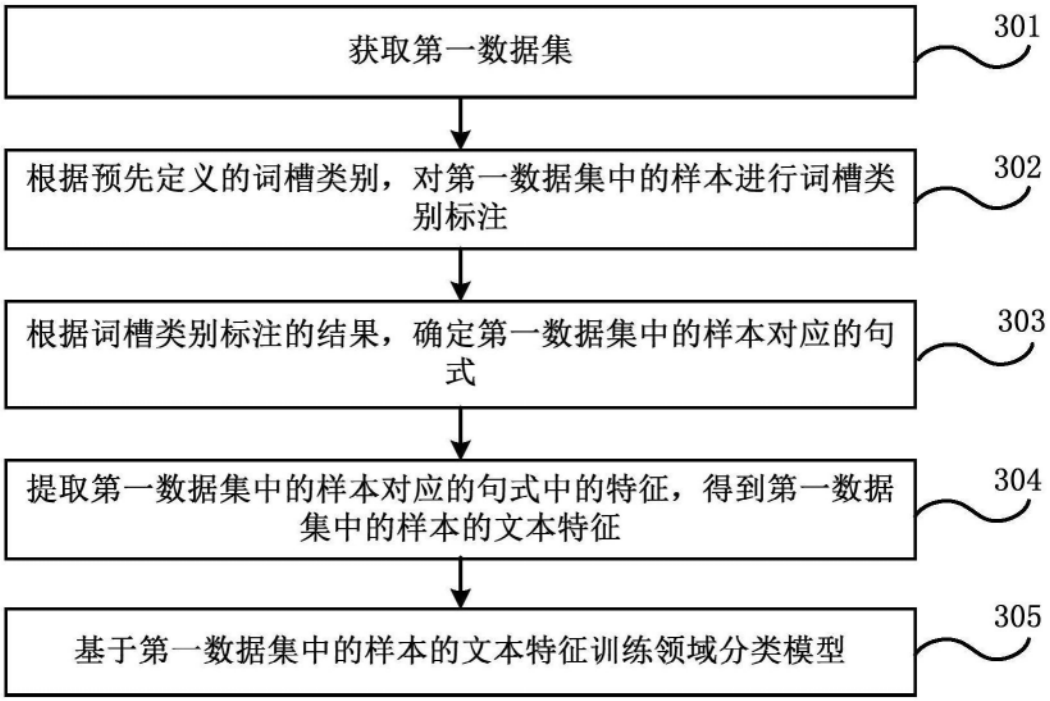

获取第一数据集
301
根据预先定义的词槽类别，对第一数据集中的样本进行词槽类别标注
302
根据词槽类别标注的结果，确定第一数据集中的样本对应的句式
303
提取第一数据集中的样本对应的句式中的特征，得到第一数据集中的样本的文本特征
304
基于第一数据集中的样本的文本特征训练领域分类模型
305

图6

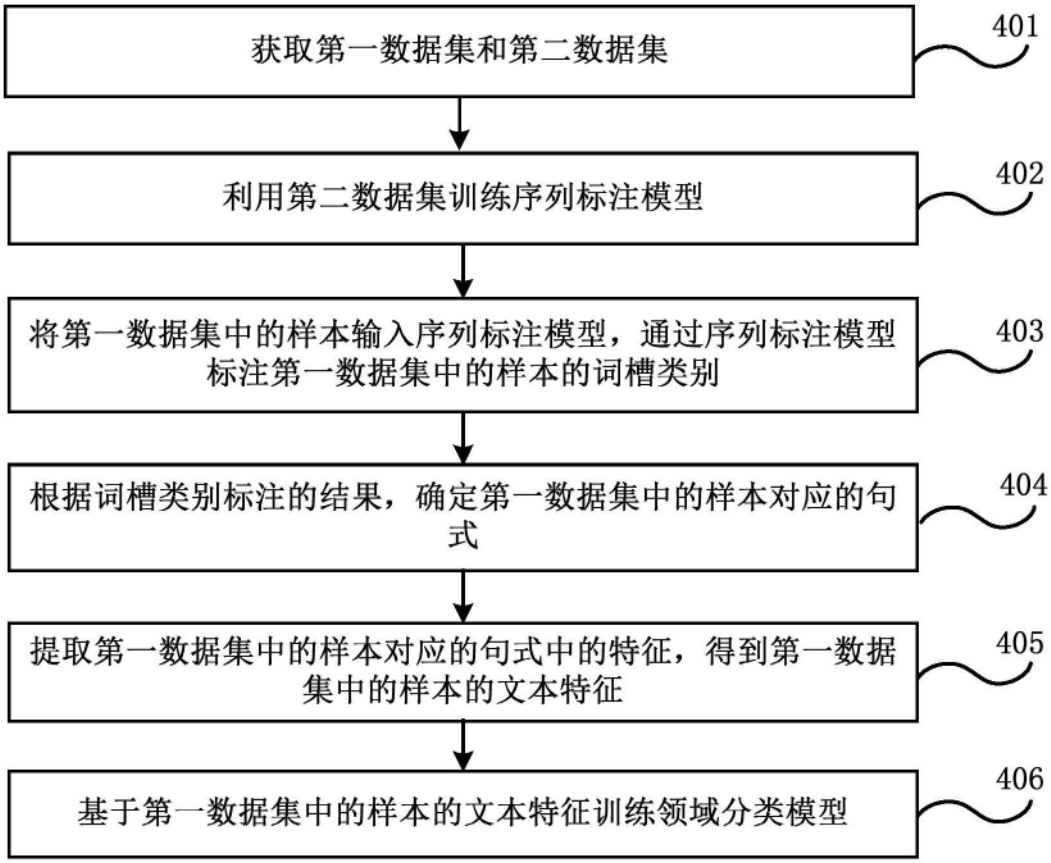

获取第一数据集和第二数据集
401
利用第二数据集训练序列标注模型
402
将第一数据集中的样本输入序列标注模型，通过序列标注模型标注第一数据集中的样本的词槽类别
403
根据词槽类别标注的结果，确定第一数据集中的样本对应的句式
404
提取第一数据集中的样本对应的句式中的特征，得到第一数据集中的样本的文本特征
405
基于第一数据集中的样本的文本特征训练领域分类模型
406

图7

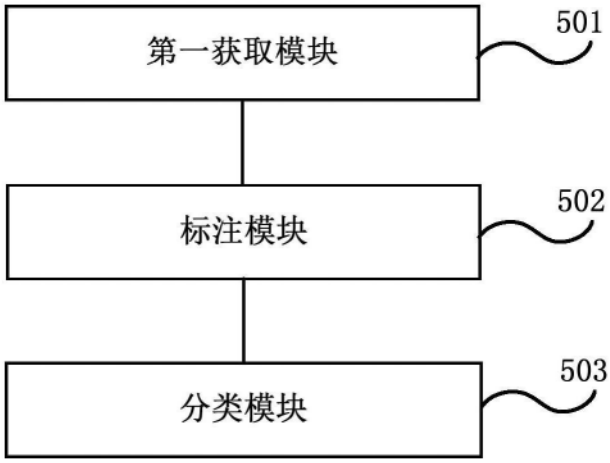

第一获取模块
501
标注模块
502
分类模块
503

图8

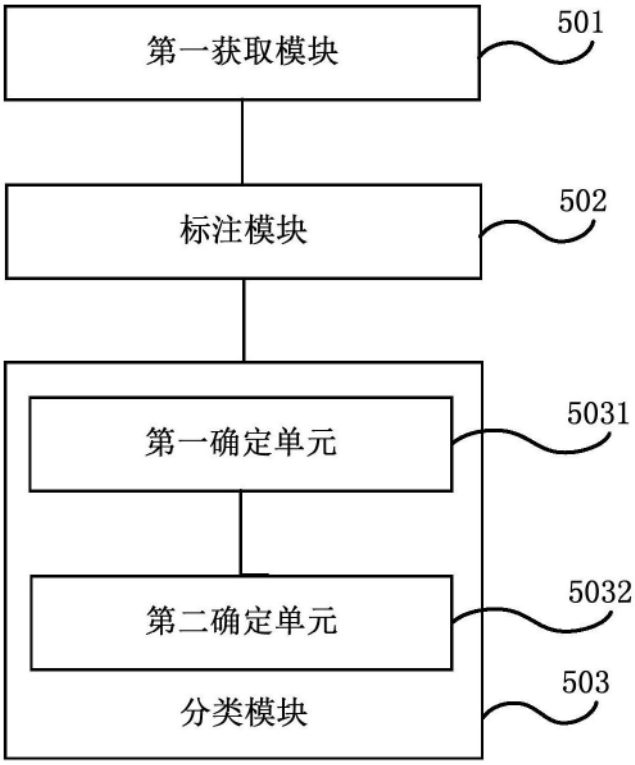

第一获取模块
501
标注模块
502
第一确定单元
5031
第二确定单元
5032
分类模块
503

图9

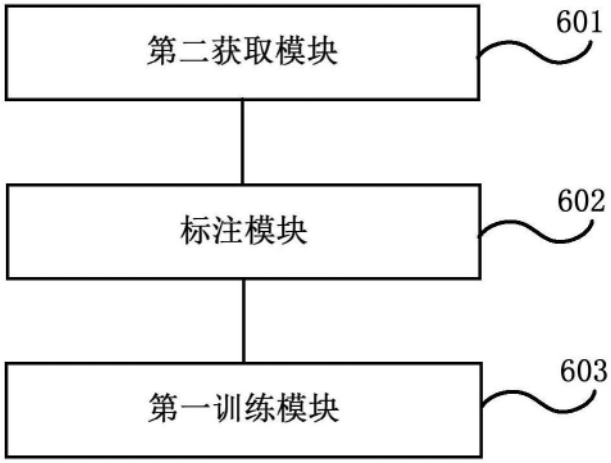

第二获取模块
601
标注模块
602
第一训练模块
603

图10

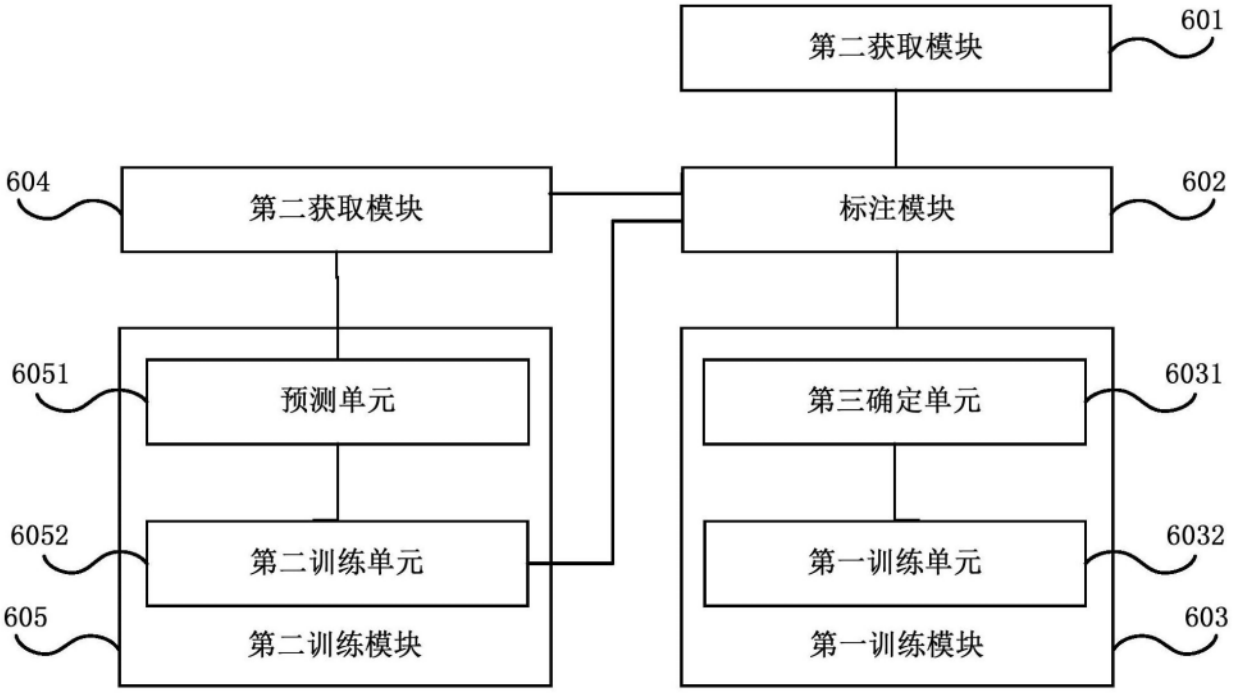

第二获取模块
601
604
第二获取模块
标注模块
602
6051
预测单元
第三确定单元
6031
6052
第二训练单元
第一训练单元
6032
605
第二训练模块
第一训练模块
603

图11

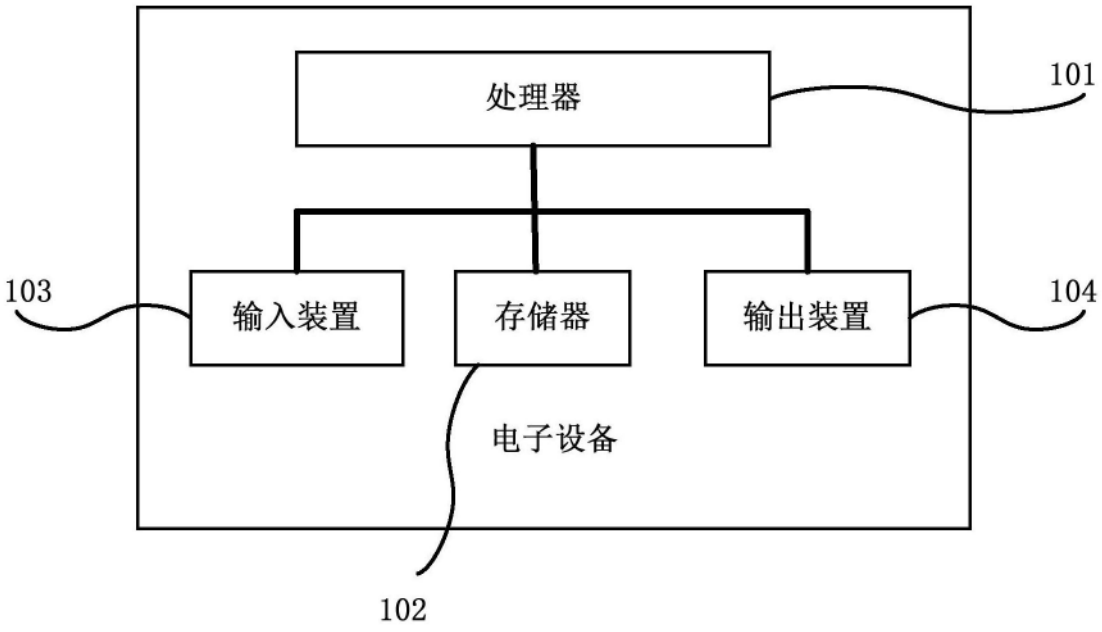

处理器
101
103
输入装置
存储器
输出装置
104
电子设备
102

图12