



(12) 发明专利

(10) 授权公告号 CN 116521827 B

(45) 授权公告日 2024. 07. 23

(21) 申请号 202310575102.5

G06F 16/35 (2019.01)

(22) 申请日 2023.05.19

G06F 16/387 (2019.01)

(65) 同一申请的已公布的文献号

G06F 40/242 (2020.01)

申请公布号 CN 116521827 A

G06F 40/30 (2020.01)

(43) 申请公布日 2023.08.01

(56) 对比文件

(73) 专利权人 北京百度网讯科技有限公司

CN 109492103 A, 2019.03.19

地址 100085 北京市海淀区上地十街10号

CN 111444344 A, 2020.07.24

百度大厦二层

审查员 韩洁

(72) 发明人 陈珺仪

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

专利代理师 石茵汀

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/31 (2019.01)

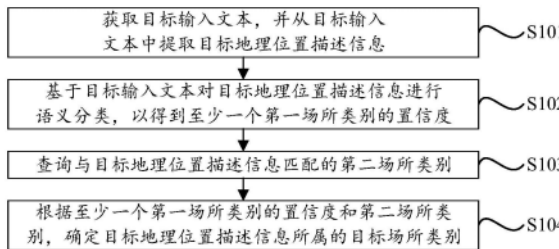
权利要求书3页 说明书16页 附图9页

(54) 发明名称

地理位置的场所类别确定方法、装置、电子设备和介质

(57) 摘要

本公开提供了一种地理位置的场所类别确定方法、装置、电子设备和介质,涉及自然语言处理、大数据、知识图谱、数据挖掘技术,可应用在智慧城市、城市治理、应急管理场景下。具体实现方案为:从目标输入文本中提取目标地理位置描述信息,并基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;查询与目标地理位置描述信息匹配的第二场所类别;根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。由此,综合各第一场所类别的置信度和第二场所类别,确定地理位置描述信息所属的最终场所类别,可以提升场所类别确定的准确性和可靠性。



1. 一种地理位置的场所类别确定方法,包括:

获取目标输入文本,并从所述目标输入文本中提取目标地理位置描述信息;

基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度,其中,采用经过训练的语义分类模型,基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到所述至少一个第一场所类别的置信度,所述语义分类模型已学习到地理位置描述信息到场所类别的映射关系;

查询与所述目标地理位置描述信息匹配的第二场所类别;

根据所述至少一个第一场所类别的置信度和所述第二场所类别,确定所述目标地理位置描述信息所属的目标场所类别;

其中,所述查询与所述目标地理位置描述信息匹配的第二场所类别,包括:

根据所述目标地理位置描述信息查询第一映射字典,以得到与所述目标地理位置描述信息匹配的目标兴趣点POI位置标签;

根据所述目标POI位置标签查询第二映射字典,以得到与所述目标POI位置标签匹配的第二场所类别;

其中,所述第一映射字典中包括至少一个POI位置标签与地理位置描述信息之间的映射关系;

其中,所述第二映射字典中包括至少一个POI位置标签与场所类别之间的映射关系。

2. 根据权利要求1所述的方法,其中,所述基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度,包括:

对所述目标输入文本和所述目标地理位置描述信息进行拼接,以得到拼接文本信息;

对所述拼接文本信息进行语义分类,以得到所述至少一个第一场所类别的置信度。

3. 根据权利要求1所述的方法,其中,所述根据所述目标地理位置描述信息查询第一映射字典,以得到与所述目标地理位置描述信息匹配的目标兴趣点POI位置标签,包括:

根据所述目标地理位置描述信息查询第一映射字典,以确定所述第一映射字典中是否包含所述目标地理位置描述信息;

在所述第一映射字典中包含所述目标地理位置描述信息的情况下,从所述第一映射字典中获取与所述目标地理位置描述信息具有映射关系的目标POI位置标签;

在所述第一映射字典中未包含所述目标地理位置描述信息的情况下,从所述第一映射字典中查询与所述目标地理位置描述信息语义相似的参考地理位置描述信息;

将所述第一映射字典中与所述参考地理位置描述信息具有映射关系的POI位置标签,作为所述目标POI位置标签。

4. 根据权利要求3所述的方法,其中,所述从所述第一映射字典中查询与所述目标地理位置描述信息语义相似的参考地理位置描述信息,包括:

获取所述目标地理位置描述信息对应的第一索引;

获取索引树,其中,所述索引树是根据所述第一映射字典中各地理位置描述信息对应的第二索引生成的;

根据所述第一索引和所述索引树,从各所述地理位置描述信息中确定参考地理位置描述信息。

5. 根据权利要求1所述的方法,其中,所述从所述目标输入文本中提取目标地理位置描

述信息,包括:

对所述目标输入文本进行序列标注,以得到所述目标输入文本中各字符的标注标签;

根据各所述字符的标注标签,从所述目标输入文本中确定至少一个实体词;

从所述至少一个实体词中,确定与地理位置关联的所述目标地理位置描述信息。

6. 根据权利要求1-5中任一项所述的方法,其中,所述根据所述至少一个第一场所类别的置信度和所述第二场所类别,确定所述目标地理位置描述信息所属的目标场所类别,包括:

根据所述至少一个第一场所类别的置信度,生成置信度向量;

对所述第二场所类别进行编码,以得到编码向量;

对所述置信度向量和所述编码向量进行融合,以得到融合向量;

对所述融合向量进行分类,以得到所述目标地理位置描述信息所属的目标场所类别。

7. 根据权利要求6所述的方法,其中,所述对所述融合向量进行分类,以得到所述目标地理位置描述信息所属的目标场所类别,包括:

对所述融合向量进行分类,以得到至少一个候选场所类别的预测概率;

根据所述至少一个候选场所类别的预测概率,从所述至少一个候选场所类别中确定所述目标场所类别。

8. 一种地理位置的场所类别确定装置,包括:

获取模块,用于获取目标输入文本;

提取模块,用于从所述目标输入文本中提取目标地理位置描述信息;

分类模块,用于基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度,其中,采用经过训练的语义分类模型,基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到所述至少一个第一场所类别的置信度,所述语义分类模型已学习到地理位置描述信息到场所类别的映射关系;

查询模块,用于查询与所述目标地理位置描述信息匹配的第二场所类别;

确定模块,用于根据所述至少一个第一场所类别的置信度和所述第二场所类别,确定所述目标地理位置描述信息所属的目标场所类别;

其中,所述查询模块,用于:

根据所述目标地理位置描述信息查询第一映射字典,以得到与所述目标地理位置描述信息匹配的目标兴趣点POI位置标签;

根据所述目标POI位置标签查询第二映射字典,以得到与所述目标POI位置标签匹配的第二场所类别;

其中,所述第一映射字典中包括至少一个POI位置标签与地理位置描述信息之间的映射关系;

其中,所述第二映射字典中包括至少一个POI位置标签与场所类别之间的映射关系。

9. 根据权利要求8所述的装置,其中,所述分类模块,用于:

对所述目标输入文本和所述目标地理位置描述信息进行拼接,以得到拼接文本信息;

对所述拼接文本信息进行语义分类,以得到所述至少一个第一场所类别的置信度。

10. 根据权利要求8所述的装置,其中,所述查询模块,用于:

根据所述目标地理位置描述信息查询第一映射字典,以确定所述第一映射字典中是否

包含所述目标地理位置描述信息；

在所述第一映射字典中包含所述目标地理位置描述信息的情况下,从所述第一映射字典中获取与所述目标地理位置描述信息具有映射关系的目标POI位置标签；

在所述第一映射字典中未包含所述目标地理位置描述信息的情况下,从所述第一映射字典中查询与所述目标地理位置描述信息语义相似的参考地理位置描述信息；

将所述第一映射字典中与所述参考地理位置描述信息具有映射关系的POI位置标签,作为所述目标POI位置标签。

11. 根据权利要求10所述的装置,其中,所述查询模块,用于:

获取所述目标地理位置描述信息对应的第一索引；

获取索引树,其中,所述索引树是根据所述第一映射字典中各地理位置描述信息对应的第二索引生成的；

根据所述第一索引和所述索引树,从各所述地理位置描述信息中确定参考地理位置描述信息。

12. 根据权利要求8所述的装置,其中,所述提取模块,用于:

对所述目标输入文本进行序列标注,以得到所述目标输入文本中各字符的标注标签；

根据各所述字符的标注标签,从所述目标输入文本中确定至少一个实体词；

从所述至少一个实体词中,确定与地理位置关联的所述目标地理位置描述信息。

13. 根据权利要求8-12中任一项所述的装置,其中,所述确定模块,用于:

根据所述至少一个第一场所类别的置信度,生成置信度向量；

对所述第二场所类别进行编码,以得到编码向量；

对所述置信度向量和所述编码向量进行融合,以得到融合向量；

对所述融合向量进行分类,以得到所述目标地理位置描述信息所属的目标场所类别。

14. 根据权利要求13所述的装置,其中,所述确定模块,用于:

对所述融合向量进行分类,以得到至少一个候选场所类别的预测概率；

根据所述至少一个候选场所类别的预测概率,从所述至少一个候选场所类别中确定所述目标场所类别。

15. 一种电子设备,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指令,所述指令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权利要求1-7中任一项所述的地理位置的场所类别确定方法。

16. 一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-7中任一项所述的地理位置的场所类别确定方法。

17. 一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-7中任一项所述地理位置的场所类别确定方法的步骤。

## 地理位置的场所类别确定方法、装置、电子设备和介质

### 技术领域

[0001] 本公开涉及人工智能技术领域,具体涉及自然语言处理、大数据、知识图谱、数据挖掘技术,可应用在智慧城市、城市治理、应急管理场景下,尤其涉及地理位置的场所类别确定方法、装置、电子设备和介质。

### 背景技术

[0002] 在一些场景中,比如智慧城市、城市治理、应急管理等场景,可能需要对地理位置描述信息进行上位类别(或称为上位类型、场所类别、用途类别,后续统称为场所类别)挖掘。比如,当地理位置描述信息为“某某佳苑”、“某某家园”时,场所类别可为“城镇居民小区”,当地理位置描述信息为“某某大厦A座”时,场所类别可为“企事业单位”。

[0003] 举例而言,在公安场景中,对地理位置描述信息进行场所类别挖掘,可以为后续案件侦办过程中涉及的涉案重点场所的挖掘分析,打下数据基础。

[0004] 因此,如何确定地理位置描述信息所属的场所类别是非常重要的。

### 发明内容

[0005] 本公开提供了一种用于地理位置的场所类别确定方法、装置、电子设备和介质。

[0006] 根据本公开的一方面,提供了一种地理位置的场所类别确定方法,包括:

[0007] 获取目标输入文本,并从所述目标输入文本中提取目标地理位置描述信息;

[0008] 基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;

[0009] 查询与所述目标地理位置描述信息匹配的第二场所类别;

[0010] 根据所述至少一个第一场所类别的置信度和所述第二场所类别,确定所述目标地理位置描述信息所属的目标场所类别。

[0011] 根据本公开的另一方面,提供了一种地理位置的场所类别确定装置,包括:

[0012] 获取模块,用于获取目标输入文本;

[0013] 提取模块,用于从所述目标输入文本中提取目标地理位置描述信息;

[0014] 分类模块,用于基于所述目标输入文本对所述目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;

[0015] 查询模块,用于查询与所述目标地理位置描述信息匹配的第二场所类别;

[0016] 确定模块,用于根据所述至少一个第一场所类别的置信度和所述第二场所类别,确定所述目标地理位置描述信息所属的目标场所类别。

[0017] 根据本公开的又一方面,提供了一种电子设备,包括:

[0018] 至少一个处理器;以及

[0019] 与所述至少一个处理器通信连接的存储器;其中,

[0020] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令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本公开上述一方面提出的地理位置的场

所类别确定方法。

[0021] 根据本公开的再一方面,提供了一种计算机指令的非瞬时计算机可读存储介质,所述计算机指令用于使所述计算机执行本公开上述一方面提出的地理位置的场所类别确定方法。

[0022] 根据本公开的还一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现本公开上述一方面提出的地理位置的场所类别确定方法。

[0023] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

## 附图说明

[0024] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0025] 图1为本公开实施例一所提供的地理位置的场所类别确定方法的流程示意图;

[0026] 图2为本公开实施例二所提供的地理位置的场所类别确定方法的流程示意图;

[0027] 图3为本公开实施例三所提供的地理位置的场所类别确定方法的流程示意图;

[0028] 图4为本公开实施例四所提供的地理位置的场所类别确定方法的流程示意图;

[0029] 图5为本公开实施例五所提供的地理位置的场所类别确定方法的流程示意图;

[0030] 图6为本公开实施例所提供的序列标注模型的模型结构示意图;

[0031] 图7为本公开实施例六所提供的地理位置的场所类别确定方法的流程示意图;

[0032] 图8为本公开实施例所提供的总体框架结构示意图;

[0033] 图9为本公开实施例所提供的语义上位挖掘模块的原理示意图;

[0034] 图10为本公开实施例所提供的地图POI tag上位挖掘模块的原理示意图;

[0035] 图11为本公开实施例所提供的多源信息融合模块的原理示意图;

[0036] 图12为本公开实施例七所提供的地理位置的场所类别确定装置的结构示意图;

[0037] 图13示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。

## 具体实施方式

[0038] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0039] 目前,可以基于地理位置描述信息所属的POI (Point Of Interest,兴趣点,在实际应用场景中可理解为目标位置) 位置标签 (POI tag), 来确定地理位置描述信息所属的场所类别。具体地,可以通过构建两种映射关系,根据用户输入的地理位置描述信息查询上述两种映射关系,来确定地理位置描述信息对应的场所类别。

[0040] 其中,第一种映射关系为:POI tag与地理位置描述之间的映射关系。可以借助地理信息系统中自带的POI tag,构建地理位置描述到POI tag的映射字典,当输入某一地理位置描述信息时,该映射字典输出与该地理位置描述信息对应的POI tag。

[0041] 第二种映射关系为:POI tag与场所类别之间的映射关系。由于地理信息系统中自带的POI tag并不能完全满足实际应用场景中的场所类别需求,因此,需要根据实际应用需

求,构建从POI tag到场所类别的映射字典,当输入某一POI tag时,该映射字典输出与该POI tag对应的场所类别。

[0042] 上述确定与地理位置描述信息对应的场所类别的方式,主要优势和劣势如下所示:

[0043] 优势包括:原理简单,且省去大规模计算,系统平均响应时长能很好地满足项目交付部署需求。

[0044] 劣势包括以下几点:

[0045] 第一点,需人工配置相关映射字典:地理位置描述到POI tag的映射字典,以及POI tag到场所类别的映射字典。

[0046] 第二点,强规则式的判断机制,不够灵活。

[0047] 第三点,当前地理信息系统中的POI tag存在较多覆盖面不全以及误分类的问题。

[0048] 第四点,场所类别挖掘结果的准确程度,受限于POI tag到场所类别的映射字典的规模和覆盖度,只有对各个省市县区不同地理位置描述的覆盖程度足够的情况下,该方法才能给出足够准确的场所类别,对于没有覆盖到的地理位置描述,系统无法给出准确的场所类别。

[0049] 针对上述存在的至少一项问题,本公开提出一种地理位置的场所类别确定方法、装置、电子设备和存储介质。

[0050] 下面参考附图描述本公开实施例的地理位置的场所类别确定方法、装置、电子设备和存储介质。

[0051] 图1为本公开实施例一所提供的地理位置的场所类别确定方法的流程示意图。

[0052] 本公开实施例以该地理位置的场所类别确定方法被配置于地理位置的场所类别确定装置中来举例说明,该地理位置的场所类别确定装置可以应用于任一电子设备中,以使该电子设备可以执行地理位置的场所类别确定功能。

[0053] 其中,电子设备可以为任一具有计算能力的设备,例如可以为个人电脑(Personal Computer,简称PC)、移动终端、服务器等,移动终端例如可以为车载设备、手机、平板电脑、个人数字助理、穿戴式设备等具有各种操作系统、触摸屏和/或显示屏的硬件设备。

[0054] 如图1所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0055] 步骤S101,获取目标输入文本,并从目标输入文本中提取目标地理位置描述信息。

[0056] 在本公开实施例中,对目标输入文本的获取方式不作限制,比如,目标输入文本可以为用户输入的文本片段,其中,对目标输入文本的输入方式不作限制,比如输入方式可以包括但不限于触摸输入(如滑动、点击等)、键盘输入、语音输入等;或者,目标输入文本也可以通过其他方式获取,比如,目标输入文本可以为在线采集的文本片段,如通过网络爬虫技术,在线采集目标输入文本;或者,目标输入文本可以为随机生成的文本片段,等等,本公开对此并不做限制。

[0057] 在本公开实施例中,可以从目标输入文本中提取地理位置描述信息。

[0058] 步骤S102,基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0059] 其中,第一场所类别包括但不限于:餐饮场所、城镇居民小区、党政机关、公园广场、交通道路、交通枢纽、企事业单位、商业场所、洗浴休闲、乡村野外、娱乐场所、住宿场所、

宗教场所等。

[0060] 在本公开实施例中,可以基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0061] 作为一种示例,可以采用经过训练的语义分类模型(如BERT (Bidirectional Encoder Representations from Transformers,来自转换器的双向编码器的表示)+FC (Fully Connected layers,全连接层)结构的语义分类模型),基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0062] 其中,语义分类模型已学习到地理位置描述信息到场所类别的映射关系。

[0063] 步骤S103,查询与目标地理位置描述信息匹配的第二场所类别。

[0064] 其中,第二场所类别与第一场所类别可以相同,或者也可以不同,本公开实施例对此并不做限制。

[0065] 在本公开实施例中,可以通过查询映射字典的方式,获取与目标地理位置描述信息匹配的第二场所类别。

[0066] 需要说明的是,对步骤S102和S103的执行时序不作限制,本公开仅以步骤S103在步骤S102之后执行进行示例,实际应用时,步骤S103也可以在步骤S102之前执行,或者,步骤S103可以与步骤S102并列执行。

[0067] 步骤S104,根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0068] 在本公开实施例中,可以综合各第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0069] 本公开实施例的地理位置的场所类别确定方法,通过从目标输入文本中提取目标地理位置描述信息,并基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;查询与目标地理位置描述信息匹配的第二场所类别;根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。由此,可以实现基于地理位置描述信息与其上下文的语义信息,确定至少一个第一场所类别的置信度,并通过查询方式,得到第二场所类别,综合各第一场所类别的置信度和第二场所类别,确定地理位置描述信息所属的最终场所类别,可以提升场所类别确定的准确性和可靠性。

[0070] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均在征得用户同意的前提下进行,并且均符合相关法律法规的规定,且不违背公序良俗。

[0071] 为了清楚说明本公开实施例中是如何基于目标输入文本对目标地理位置描述信息进行语义分类的,本公开还提出一种地理位置的场所类别确定方法。

[0072] 图2为本公开实施例二所提供的地理位置的场所类别确定方法的流程示意图。

[0073] 如图2所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0074] 步骤S201,获取目标输入文本,并从目标输入文本中提取目标地理位置描述信息。

[0075] 步骤S201的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0076] 步骤S202,对目标输入文本和目标地理位置描述信息进行拼接,以得到拼接文本信息。



[0077] 在本公开实施例中,可以对目标输入文本和目标地理位置描述信息进行拼接,以得到拼接文本信息。

[0078] 作为一种示例,可以在目标输入文本之后拼接目标地理位置描述信息,得到拼接文本信息。

[0079] 作为另一种示例,可以在目标地理位置描述信息之后拼接目标输入文本,得到拼接文本信息。

[0080] 步骤S203,对拼接文本信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0081] 在本公开实施例中,可以对拼接文本信息进行语义分类,得到至少一个第一场所类别的置信度。

[0082] 作为一种示例,可以采用经过训练的语义分类模型(如BERT+FC结构的语义分类模型),对拼接文本信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0083] 其中,语义分类模型已学习到拼接文本信息到场所类别的映射关系。

[0084] 例如,语义分类模型的训练过程可以包括:构造训练样本:<文本text:地理位置描述信息+文本片段,标注标签label:场所类别>,其中,文本片段(或称为原文本段)中包含地理位置描述信息,将text输入至语义分类模型进行语义分类,得到语义分类模型输出的各个场所类别对应的置信度,根据语义分类模型的输出和标注标签label,对语义分类模型进行训练,以使语义分类模型学习到text到label的映射关系。

[0085] 步骤S204,查询与目标地理位置描述信息匹配的第二场所类别。

[0086] 步骤S205,根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0087] 步骤S204至S205的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0088] 本公开实施例的地理位置的场所类别确定方法,可以实现基于地理位置描述信息的上下文语境,对地理位置描述信息进行语义分类,提升分类结果的准确性和可靠性。

[0089] 为了清楚说明本公开实施例中是如何查询与目标地理位置描述信息匹配的第二场所类别的,本公开还提出一种地理位置的场所类别确定方法。

[0090] 图3为本公开实施例三所提供的地理位置的场所类别确定方法的流程示意图。

[0091] 如图3所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0092] 步骤S301,获取目标输入文本,并从目标输入文本中提取目标地理位置描述信息。

[0093] 步骤S302,基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0094] 步骤S301至S302的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0095] 步骤S303,根据目标地理位置描述信息查询第一映射字典,以得到与目标地理位置描述信息匹配的目标POI位置标签。

[0096] 其中,第一映射字典中包括至少一个POI位置标签与地理位置描述信息之间的映射关系。

[0097] 在本公开实施例中,可以根据目标地理位置描述信息查询第一映射字典,以得到

与目标地理位置描述信息匹配的POI位置标签(本公开中记为目标POI位置标签)。

[0098] 举例而言,当目标地理位置描述信息为“某某佳苑”时,目标POI位置标签可为“居民区”。

[0099] 步骤S304,根据目标POI位置标签查询第二映射字典,以得到与目标POI位置标签匹配的第二场所类别。

[0100] 其中,第二映射字典中包括至少一个POI位置标签与场所类别之间的映射关系。

[0101] 在本公开实施例中,可以根据目标POI位置标签查询第二映射字典,以得到与目标POI位置标签匹配的第二场所类别。

[0102] 仍以上述例子进行示例,当目标POI位置标签为“居民区”时,第二场所类别可为“城镇居民小区”。

[0103] 步骤S305,根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0104] 步骤S305的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0105] 本公开实施例的地理位置的场所类别确定方法,通过查询映射字典的方式,得到与地理位置描述信息匹配的场所类别,操作简单,易于实现。并且,无需大规模计算,可以提升处理效率。

[0106] 为了清楚说明本公开上述实施例中是如何获取与目标地理位置描述信息匹配的目标POI位置标签的,本公开还提出一种地理位置的场所类别确定方法。

[0107] 图4为本公开实施例四所提供的地理位置的场所类别确定方法的流程示意图。

[0108] 如图4所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0109] 步骤S401,获取目标输入文本,并从目标输入文本中提取目标地理位置描述信息。

[0110] 步骤S402,基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0111] 步骤S401至S402的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0112] 步骤S403,根据目标地理位置描述信息查询第一映射字典,以确定第一映射字典中是否包含目标地理位置描述信息,若是,则执行步骤S404,若否,则执行步骤S405至S406。

[0113] 其中,第一映射字典中包括至少一个POI位置标签与地理位置描述信息之间的映射关系。

[0114] 需要说明的是,受限于第一映射字典中地理位置描述信息的覆盖程度,第一映射字典中可能包含目标地理位置描述信息,也可能未包含目标地理位置描述信息,在第一映射字典中包含目标地理位置描述信息的情况下,可以执行步骤S404,而在第一映射字典中未包含目标地理位置描述信息的情况下,可以执行步骤S405至S406。

[0115] 步骤S404,从第一映射字典中获取与目标地理位置描述信息具有映射关系的POI位置标签。

[0116] 在本公开实施例中,在第一映射字典中包含目标地理位置描述信息的情况下,可以从第一映射字典中获取与目标地理位置描述信息具有映射关系的POI位置标签,并作为目标POI位置标签。

[0117] 步骤S405,从第一映射字典中查询与目标地理位置描述信息语义相似的参考地理

位置描述信息。

[0118] 在本公开实施例中,在第一映射字典中未包含目标地理位置描述信息的情况下,可以从第一映射字典中查询与目标地理位置描述信息语义相似的参考地理位置描述信息。

[0119] 需要说明的是,第一映射字典中可能包括大量的地理位置描述信息,如果将目标地理位置描述信息与第一映射字典中全量的地理位置描述信息进行语义相似度匹配,计算量巨大,查询效率较低。

[0120] 因此,针对该问题,在本公开实施例的一种可能的实现方式中,可以基于索引技术,来提升参考地理位置描述信息的查询效率。

[0121] 作为一种示例,参考地理位置描述信息的查询方式,例如为:构建目标地理位置描述信息对应的第一索引,并获取索引树,其中,索引树是根据第一映射字典中各地理位置描述信息对应的第二索引生成的,从而可以根据第一索引和索引树,从第一映射字典中各地理位置描述信息中确定参考地理位置描述信息。

[0122] 比如,可以从索引树中查询与第一索引相似或距离最近的第二索引,并将该第二索引所对应的地理位置描述信息作为参考地理位置描述信息。

[0123] 步骤S406,将第一映射字典中与参考地理位置描述信息具有映射关系的POI位置标签,作为目标POI位置标签。

[0124] 在本公开实施例中,可以从第一映射字典中获取与参考地理位置描述信息具有映射关系的POI位置标签,并作为目标POI位置标签。

[0125] 举例而言,虽然“某某大厦A座”与“某某大厦B座”在描述上存在差异,但是主体是相似的,不会影响最终对场所类别的判别,两者所属的场所类别都是“企事业单位”。假设第一映射字典中仅包含“某某大厦”或“某某大厦A座”,则当目标地理位置描述信息为“某某大厦B座”、“某某大厦C座”或“某某大厦D座”时,可以将第一映射字典中的“某某大厦”或“某某大厦A座”作为参考地理位置描述信息,并将第一映射字典中与参考地理位置描述信息具有映射关系的POI位置标签,作为与目标地理位置描述信息匹配的目标POI位置标签。

[0126] 步骤S407,根据目标POI位置标签查询第二映射字典,以得到与目标POI位置标签匹配的第二场所类别。

[0127] 其中,第二映射字典中包括至少一个POI位置标签与场所类别之间的映射关系。

[0128] 步骤S408,根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0129] 步骤S407至S408的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0130] 本公开实施例的地理位置的场所类别确定方法,可以实现无论第一映射字典中是否包含目标地理位置描述信息,均可以基于目标地理位置描述信息的语义,确定与目标地理位置描述信息匹配的目标POI位置标签,提升目标POI位置标签确定的有效性和准确性。

[0131] 为了清楚说明本公开任一实施例中是如何从目标输入文本中提取目标地理位置描述信息的,本公开还提出一种地理位置的场所类别确定方法。

[0132] 图5为本公开实施例五所提供的地理位置的场所类别确定方法的流程示意图。

[0133] 如图5所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0134] 步骤S501,获取目标输入文本。

- [0135] 步骤S501的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。
- [0136] 步骤S502,对目标输入文本进行序列标注,以得到目标输入文本中各字符的标注标签。
- [0137] 在本公开实施例中,可以基于序列标注方法,对目标输入文本进行序列标注,以得到目标输入文本中各字符的标注标签。
- [0138] 其中,常见的序列标注方法包括:基于规则、基于统计以及CRF (Conditional Random Fields,条件随机场)、Bi-LSTM(Bi-directional Long Short-Term Memory,双向长短期记忆)+CRF、带上预训练BERT的方法。
- [0139] 作为一种示例,可以采用序列标注模型对目标输入文本进行序列标注,得到目标输入文本中各字符的标注标签。例如,假设目标输入文本包含N个字符,则序列标注模型可以输出N个标注标签(即,每个字符均具有对应的标注标签)。
- [0140] 举例而言,假设目标输入文本为“张三工作地位于某某大厦A座26层”,序列标注模型采用BIO标注法,对目标输入文本进行序列标注,得到的各字符的标注标签可以为{BI00000BIIIIIOO},其中,B表示Begin(即实体的开头)、I表示Inside(即实体的中间或结尾)、O表示Outside(不属于实体)。
- [0141] 再例如,假设目标输入文本为“张三工作地位于某某大厦A座26层”,序列标注模型采用BIOES标注法,对目标输入文本进行序列标注,得到的各字符的标注标签可以为{BE00000BIIIIIEOO},其中,B表示Begin(即实体的开始位置),I表示Inside(即实体的中间位置),E表示End(即实体的结束位置),O表示Outside(不属于实体,用于标记无关字符)、S表示Single(单个字符,本身就是一个实体)。
- [0142] 一种示例,以序列标注模型的模型结构如图6所示的Bi-LSTM+CRF进行示例,可以采用Bi-LSTM做特征计算以及序列标签分类,并在Bi-LSTM后面接一个CRF用于条件约束并将各字符的标注标签输出。
- [0143] 步骤S503,根据各字符的标注标签,从目标输入文本中确定至少一个实体词。
- [0144] 在本公开实施例中,可以根据各字符的标注标签,从目标输入文本中确定至少一个实体词。
- [0145] 仍以上述例子进行示例,目标输入文本中的实体词可以包括:张三、某某大厦A座。
- [0146] 步骤S504,从至少一个实体词中,确定与地理位置关联的目标地理位置描述信息。
- [0147] 在本公开实施例中,可以从目标输入文本中的至少一个实体词中,确定与地理位置关联的目标地理位置描述信息。
- [0148] 仍以上述例子进行示例,目标地理位置描述信息可以为“某某大厦A座”。
- [0149] 步骤S505,基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。
- [0150] 步骤S506,查询与目标地理位置描述信息匹配的第二场所类别。
- [0151] 步骤S507,根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。
- [0152] 步骤S505至S507的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。
- [0153] 本公开实施例的地理位置的场所类别确定方法,可以实现基于序列标注方法,从

目标输入文本中提取实体词,并从各实体词中确定与地理位置关联的目标地理位置描述信息,可以提升目标地理位置描述信息提取的有效性和准确性。

[0154] 为了清楚说明本公开任一实施例是如何根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别的,本公开还提出一种地理位置的场所类别确定方法。

[0155] 图7为本公开实施例六所提供的地理位置的场所类别确定方法的流程示意图。

[0156] 如图7所示,该地理位置的场所类别确定方法可以包括以下步骤:

[0157] 步骤S701,获取目标输入文本,并从目标输入文本中提取目标地理位置描述信息。

[0158] 步骤S702,基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0159] 步骤S703,查询与目标地理位置描述信息匹配的第二场所类别。

[0160] 步骤S701至S703的解释说明可以参见本公开任一实施例中的相关描述,在此不做赘述。

[0161] 步骤S704,根据至少一个第一场所类别的置信度,生成置信度向量。

[0162] 在本公开实施例中,可以根据上述至少一个第一场所类别的置信度,生成置信度向量。

[0163] 步骤S705,对第二场所类别进行编码,以得到编码向量。

[0164] 在本公开实施例中,可以对第二场所类别进行编码,得到编码向量。

[0165] 作为一种示例,可以基于独热编码(onehot编码)方式,对第二场所类别进行编码,得到编码向量。

[0166] 步骤S706,对置信度向量和编码向量进行融合,以得到融合向量。

[0167] 在本公开实施例中,可以对置信度向量和编码向量进行拼接或融合,得到融合向量。

[0168] 步骤S707,对融合向量进行分类,以得到目标地理位置描述信息所属的目标场所类别。

[0169] 在本公开实施例中,可以对融合向量进行分类,得到目标地理位置描述信息所属的目标场所类别。

[0170] 在本公开的任意一个实施例之中,目标场所类别的获取方式,例如为:对融合向量进行分类,得到至少一个候选场所类别的预测概率,根据至少一个候选场所类别的预测概率,从至少一个候选场所类别中确定目标场所类别。

[0171] 作为一种示例,可以将预测概率最大的候选场所类别,作为目标场所类别。

[0172] 作为另一种示例,可以将预测概率高于设定阈值的候选场所类别,作为目标场所类别。

[0173] 作为另一种示例,可以将各候选场所类别按照预测概率的取值由大至小排序,并将排序在前设定个数的候选场所类别,作为目标场所类别。

[0174] 由此,根据各候选场所类别的预测概率,从各候选场所类别中确定目标场所类别,可以提升目标场所类别确定的准确性。

[0175] 本公开实施例的地理位置的场所类别确定方法,可以实现将语义分类结果或语义挖掘结果(即至少一个第一场所类别的置信度)与查询结果(第二场所类别)进行融合后分

类,得到最终的目标场所类别,可以目标场所类别确定的准确性和可靠性。

[0176] 在本公开的任意一个实施例之中,可以结合地理信息系统的POI信息和地理位置描述信息的上下文语义信息,采用多源信息融合的方法来提升场所类别确定的准确率与覆盖度。

[0177] 例如,在当前的应用场景中,常见的地理位置描述信息均会带有明显的语义特征,例如“某某佳苑”、“某某家园”,结合已有的知识,可以从文本语义上判断出对应的场所类别(即“城镇居民小区”)。

[0178] 借鉴相同的思路,本公开中,无需完全依赖于POI tag对应的映射字典,来判断地理位置描述信息所属的场所类别。结合自然语言处理技术,通过大量的训练样本来训练模型,使其具备这种能力,能够构建地理位置描述信息到场所类别的关联关系。

[0179] 通过引入语义信息这个新信息源,优化了已有的实现方案,提升了场所类别挖掘的泛化能力和准确度,使整个系统的输出效果不再受限于POI tag对应的映射字典的规模和覆盖度。相关技术中使用的POI tag对应的映射字典,放在本公开所提供的方案中,作为一个辅助手段对模型无法判断的疑难文本进行辅助挖掘,映射字典中只需更新语义上难以辨别出场所类别的疑难文本,例如“某某大厦”。

[0180] 本公开的总体结构可以如图8所示,主要包括以下四个部分:

[0181] 第一部分:信息抽取。

[0182] 对于用户输入的检索词(或称为原文本段,本公开中记为目标输入文本),需要对检索词(原文本段)做实体识别,以明确每一个词语代表的真实实体含义,从而提升排序的准确度。

[0183] 在当前应用场景中,待抽取要素对象是与地理位置描述相关的实体。例如,当检索词(原文本段)为“张三工作地位于某某大厦A座26层”时,“某某大厦A座”即为待抽取要素对象(即需要抽取识别出来的地理位置描述信息)。

[0184] 其中,可以采用序列标注方法来完成实体识别,比如,可以采用Bi-LSTM做特征计算以及序列标签分类,并在Bi-LSTM的后面接一个CRF用于条件约束并将标注结果输出。

[0185] 第二部分:语义上位挖掘,即基于地理位置描述信息的语义,挖掘地理位置描述信息所属的场所类别。

[0186] 相关技术中的场所类别挖掘方案,完全依赖于POI tag对应的映射字典,属于强规则型的判断方法,不够灵活,缺乏对上下文语境上的考量,且受限于映射字典的规模或准确度,会导致误判的情况发生。

[0187] 例如,当检索词(原文本段)为“张三工作地位于某某大厦A座26层”时,对于地理位置描述信息“某某大厦A座”,如果没有考量上下文中的“工作地”,很难区分出“某某大厦A座”这个地理位置描述信息对应的场所类别为“企事业单位”。而如果完全依赖于映射字典,当映射字典中未覆盖到“某某大厦”时,则很难准确地判断出“某某大厦A座”对应的场所类别。

[0188] 因此,本公开中,可以采用语义分类模型(比如模型结构可为图9所示的BERT+FC结构),将地理位置描述信息所属的场所类别的挖掘场景,转换成一个文本多标签分类问题进行解决。

[0189] 首先,可以构造训练样本:<文本text:地理位置描述信息+文本片段,标注标签

label:场所类别>,其中,文本片段中包含地理位置描述信息。之后,可以经过多轮微调 finetune,让语义分类模型直接学习到地理位置描述信息到场所类别的映射关系。在语义分类模型训练完成后,产出的预测模块的输入输出结构为:

[0190] 输入为:地理位置描述信息+文本片段;

[0191] 输出为:各个场所类别对应的置信度。

[0192] 第三部分:地图POI tag上位挖掘,即挖掘POI tag对应的场所类别。

[0193] 地图POI tag上位挖掘,作为辅助手段对语义上位挖掘模块难以辨别的地理位置描述信息进行补充挖掘。

[0194] 该模块的输入为地理位置描述信息,输出为对应的场所类别。模块内部的核心逻辑主要涉及:两大映射字典的构建、键值匹配阶段的近似词检索:

[0195] 1、两大映射字典的构建。

[0196] 第一映射字典(POI tag映射字典):借助地理信息系统中自带的POI位置标签,构建地理位置描述信息到POI tag的映射字典,输入地理位置描述信息,输出POI tag。

[0197] 第二映射字典(场所类别映射字典):因为地理信息系统中自带的POI tag并不能完全满足实际应用场景中的场所类别需求,因此需要根据实际的应用需求,构建从POI tag到场所类别的映射字典。

[0198] 2、近似词检索。

[0199] 第1步已经构建好了映射字典,接下来需要根据输入的地理位置描述信息,在两个映射字典中依次进行查询,最终得到对应的场所类别。

[0200] 在获取POI tag阶段,考虑到各种地理位置描述信息较为繁杂,在某些情况下,地理位置描述信息表述上存在差异,但实际上指代的主体内容属于同一地理位置场所类别,例如,“某某大厦A座”与“某某大厦B座”虽然在描述上存在差异,但是主体是相似的,不会影响最终对场所类别的判别,两者所属的场所类别都是“企事业单位”。

[0201] 因此,相对于仅简单的使用键值完全匹配来找到对应的场所类别,本公开中,可以引入近似词检索来提升键值匹配阶段中相似词的召回量,避免出现同一指代类型由于表述上存在差异而导致找不到场所类别的情况出现。

[0202] 例如,可以采用FAISS(Facebook AI Similarity Search)第三方库来实现近似词检索:FAISS的工作,就是将候选词向量封装成一个索引index数据库,以加速相似词向量TopK(前K个)的过程,其中有些索引还支持GPU(Graphic Processing Units,图形处理单元)构建,更好的提升了搜索时效。实现过程分为如下几步:训练词向量索引树、生成索引,构建索引数据库、查询索引。

[0203] 在获取POI tag之后,由于地图库中的POI tag是相对固定的,所以只需要在构建好的第二映射字典中进行完全匹配来找到对应的场所类别,并转换成与上位类别体系(或称为场所类别体系)对应的onehot编码(01编码)。

[0204] 作为一种示例,地图POI tag上位挖掘模块的原理可以如图10所示。

[0205] 其中,地图POI tag上位挖掘模块的输入为地理位置描述信息,输出为onehot编码。

[0206] 第四部分:多源信息融合。

[0207] 该模块将语义上位挖掘模块与地图POI上位挖掘模块的输出结果进行融合,得到

最终输出。其中,语义上位挖掘作为主挖掘方法,地图POI上位挖掘作为辅助补充信息对语义上位挖掘中遇到的疑难文本进行补充,因此需要设计一个合理的机制将两者的输出结果进行融合。

[0208] 作为一种示例,融合机制可以考虑采用概率分布值进行设计,因为语义上位挖掘是个多标签分类问题,在各个类别的置信度分布上不存在总和为1的约束,因此,可以将语义上位挖掘模块输出的各个场所类别的置信度,与地图POI上位挖掘模块输出的onehot编码进行融合,给到xgboost(eXtreme Gradient Boosting)决策树模型转换成多分类问题,输出为单标签(即目标场所类别),也就是融合两个输入源后进行处理的最最终的场所类别。

[0209] 作为一种示例,多源信息融合模块的原理可以如图11所示。

[0210] 综上,本公开所提供的地理位置的场所类别确定方法,可以有效支持重点场所的上位分类体系进行地理位置上位挖掘与分析。

[0211] 与上述图1至图7实施例提供的地理位置的场所类别确定方法相对应,本公开还提供一种地理位置的场所类别确定装置,由于本公开实施例提供的地理位置的场所类别确定装置与上述图1至图7实施例提供的地理位置的场所类别确定方法相对应,因此在地理位置的场所类别确定方法的实施方式也适用于本公开实施例提供的地理位置的场所类别确定装置,在本公开实施例中不再详细描述。

[0212] 图12为本公开实施例七所提供的地理位置的场所类别确定装置的结构示意图。

[0213] 如图12所示,该地理位置的场所类别确定装置1200可以包括:获取模块1201、提取模块1202、分类模块1203、查询模块1204以及确定模块1205。

[0214] 其中,获取模块1201,用于获取目标输入文本。

[0215] 提取模块1202,用于从目标输入文本中提取目标地理位置描述信息。

[0216] 分类模块1203,用于基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0217] 查询模块1204,用于查询与目标地理位置描述信息匹配的第二场所类别。

[0218] 确定模块1205,用于根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。

[0219] 在本公开实施例的一种可能的实现方式中,分类模块1203,用于:对目标输入文本和目标地理位置描述信息进行拼接,以得到拼接文本信息;对拼接文本信息进行语义分类,以得到至少一个第一场所类别的置信度。

[0220] 在本公开实施例的一种可能的实现方式中,查询模块1204,用于:根据目标地理位置描述信息查询第一映射字典,以得到与目标地理位置描述信息匹配的目标兴趣点POI位置标签;根据目标POI位置标签查询第二映射字典,以得到与目标POI位置标签匹配的第二场所类别;其中,第一映射字典中包括至少一个POI位置标签与地理位置描述信息之间的映射关系;其中,第二映射字典中包括至少一个POI位置标签与场所类别之间的映射关系。

[0221] 在本公开实施例的一种可能的实现方式中,查询模块1204,用于:根据目标地理位置描述信息查询第一映射字典,以确定第一映射字典中是否包含目标地理位置描述信息;在第一映射字典中包含目标地理位置描述信息的情况下,从第一映射字典中获取与目标地理位置描述信息具有映射关系的目标POI位置标签;在第一映射字典中未包含目标地理位置描述信息的情况下,从第一映射字典中查询与目标地理位置描述信息语义相似的参考地



理位置描述信息;将第一映射字典中与参考地理位置描述信息具有映射关系的POI位置标签,作为目标POI位置标签。

[0222] 在本公开实施例的一种可能的实现方式中,查询模块1204,用于:获取目标地理位置描述信息对应的第一索引;获取索引树,其中,索引树是根据第一映射字典中各地理位置描述信息对应的第二索引生成的;根据第一索引和索引树,从各地理位置描述信息中确定参考地理位置描述信息。

[0223] 在本公开实施例的一种可能的实现方式中,提取模块1202,用于:对目标输入文本进行序列标注,以得到目标输入文本中各字符的标注标签;根据各字符的标注标签,从目标输入文本中确定至少一个实体词;从至少一个实体词中,确定与地理位置关联的目标地理位置描述信息。

[0224] 在本公开实施例的一种可能的实现方式中,确定模块1205,用于:根据至少一个第一场所类别的置信度,生成置信度向量;对第二场所类别进行编码,以得到编码向量;对置信度向量和编码向量进行融合,以得到融合向量;对融合向量进行分类,以得到目标地理位置描述信息所属的目标场所类别。

[0225] 在本公开实施例的一种可能的实现方式中,确定模块1205,用于:对融合向量进行分类,以得到至少一个候选场所类别的预测概率;根据至少一个候选场所类别的预测概率,从至少一个候选场所类别中确定目标场所类别。

[0226] 本公开实施例的地理位置的场所类别确定装置,通过从目标输入文本中提取目标地理位置描述信息,并基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;查询与目标地理位置描述信息匹配的第二场所类别;根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。由此,可以实现基于地理位置描述信息与其上下文的语义信息,确定至少一个第一场所类别的置信度,并通过查询方式,得到第二场所类别,综合各第一场所类别的置信度和第二场所类别,确定地理位置描述信息所属的最终场所类别,可以提升场所类别确定的准确性和可靠性。

[0227] 为了实现上述实施例,本公开还提供一种电子设备,该电子设备可以包括至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行本公开上述任一实施例提出的地理位置的场所类别确定方法。

[0228] 为了实现上述实施例,本公开还提供一种存储有计算机指令的非瞬时计算机可读存储介质,其中,计算机指令用于使计算机执行本公开上述任一实施例提出的地理位置的场所类别确定方法。

[0229] 为了实现上述实施例,本公开还提供一种计算机程序产品,该计算机程序产品包括计算机程序,计算机程序在被处理器执行时实现本公开上述任一实施例提出的地理位置的场所类别确定方法。

[0230] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0231] 图13示出了可以用来实施本公开的实施例的示例电子设备的示意性框图。其中,电子设备可以包括上述实施例中的服务端、客户端。电子设备旨在表示各种形式的数字计

算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0232] 如图13所示,电子设备1300包括计算单元1301,其可以根据存储在ROM(Read-Only Memory,只读存储器)1302中的计算机程序或者从存储单元1308加载到RAM(Random Access Memory,随机访问/存取存储器)1303中的计算机程序,来执行各种适当的动作和处理。在RAM 1303中,还可存储电子设备1300操作所需的各种程序和数据。计算单元1301、ROM 1302以及RAM 1303通过总线1304彼此相连。I/O(Input/Output,输入/输出)接口1305也连接至总线1304。

[0233] 电子设备1300中的多个部件连接至I/O接口1305,包括:输入单元1306,例如键盘、鼠标等;输出单元1307,例如各种类型的显示器、扬声器等;存储单元1308,例如磁盘、光盘等;以及通信单元1309,例如网卡、调制解调器、无线通信收发机等。通信单元1309允许电子设备1300通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0234] 计算单元1301可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元1301的一些示例包括但不限于CPU(Central Processing Unit,中央处理单元)、GPU(Graphic Processing Units,图形处理单元)、各种专用的AI(Artificial Intelligence,人工智能)计算芯片、各种运行机器学习模型算法的计算单元、DSP(Digital Signal Processor,数字信号处理器)、以及任何适当的处理器、控制器、微控制器等。计算单元1301执行上文所描述的各个方法和处理,例如上述地理位置的场所类别确定方法。例如,在一些实施例中,上述地理位置的场所类别确定方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元1308。在一些实施例中,计算机程序的部分或者全部可以经由ROM 1302和/或通信单元1309而被载入和/或安装到电子设备1300上。当计算机程序加载到RAM 1303并由计算单元1301执行时,可以执行上文描述的地理位置的场所类别确定方法的一个或多个步骤。备选地,在其他实施例中,计算单元1301可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行上述地理位置的场所类别确定方法。

[0235] 本文中以上描述的系统和技术和各种实施方式可以在数字电子电路系统、集成电路系统、FPGA(Field Programmable Gate Array,现场可编程门阵列)、ASIC(Application-Specific Integrated Circuit,专用集成电路)、ASSP(Application Specific Standard Product,专用标准产品)、SOC(System On Chip,芯片上系统的系统)、CPLD(Complex Programmable Logic Device,复杂可编程逻辑设备)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0236] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码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

理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0237] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、RAM、ROM、EPROM(Electrically Programmable Read-Only-Memory,可擦除可编程只读存储器)或快闪存储器、光纤、CD-ROM(Compact Disc Read-Only Memory,便捷式紧凑盘只读存储器)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0238] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(Cathode-Ray Tube,阴极射线管)或者LCD(Liquid Crystal Display,液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0239] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:LAN(Local Area Network,局域网)、WAN(Wide Area Network,广域网)、互联网和区块链网络。

[0240] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务(Virtual Private Server,虚拟专用服务器)中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0241] 其中,需要说明的是,人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术;人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

[0242] 深度学习是机器学习领域中一个新的研究方向。它是学习样本数据的内在规律和表示层次,这些学习过程中获得的信息对诸如文字,图像和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等

数据。

[0243] 根据本公开实施例的技术方案,通过从目标输入文本中提取目标地理位置描述信息,并基于目标输入文本对目标地理位置描述信息进行语义分类,以得到至少一个第一场所类别的置信度;查询与目标地理位置描述信息匹配的第二场所类别;根据至少一个第一场所类别的置信度和第二场所类别,确定目标地理位置描述信息所属的目标场所类别。由此,可以实现基于地理位置描述信息与其上下文的语义信息,确定至少一个第一场所类别的置信度,并通过查询方式,得到第二场所类别,综合各第一场所类别的置信度和第二场所类别,确定地理位置描述信息所属的最终场所类别,可以提升场所类别确定的准确性和可靠性。

[0244] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开提出的技术方案所期望的结果,本文在此不进行限制。

[0245] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

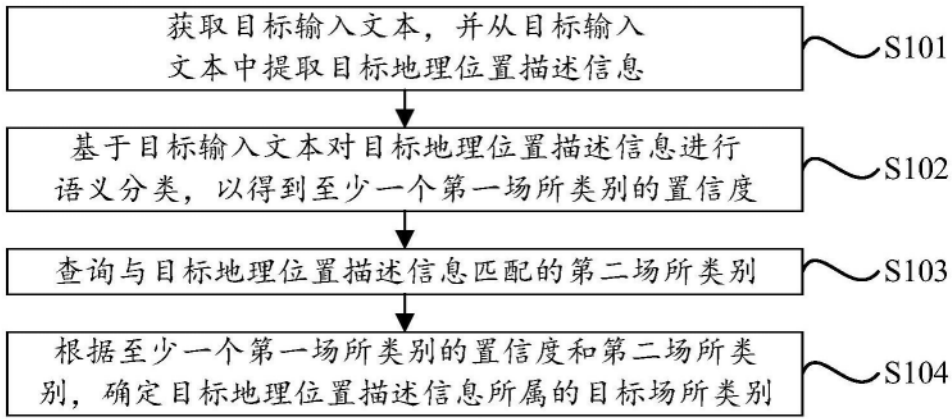


图1

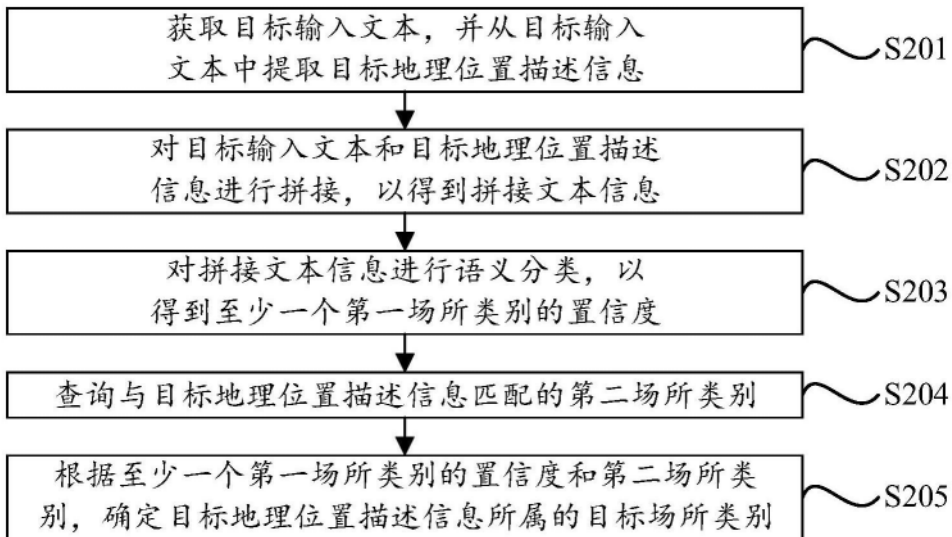


图2

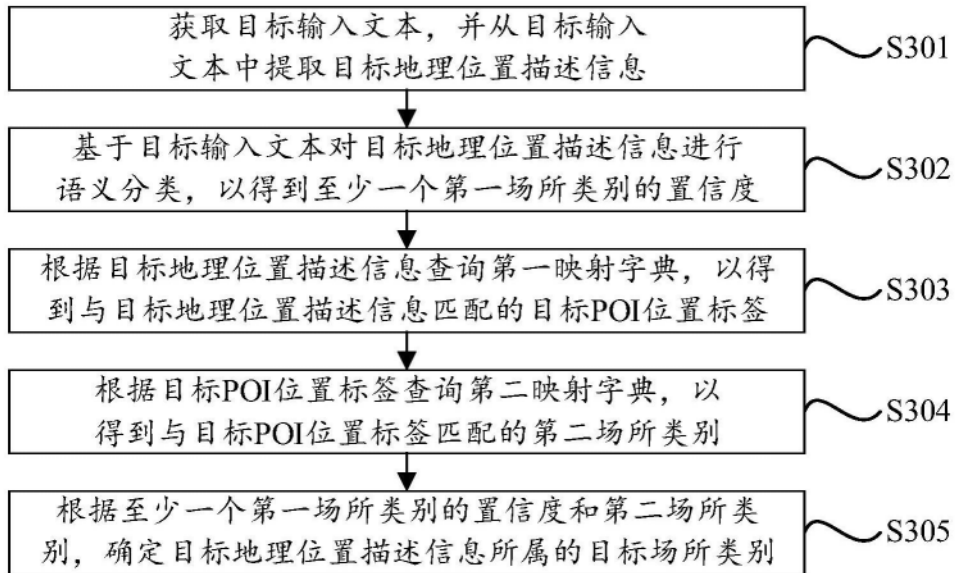


图3

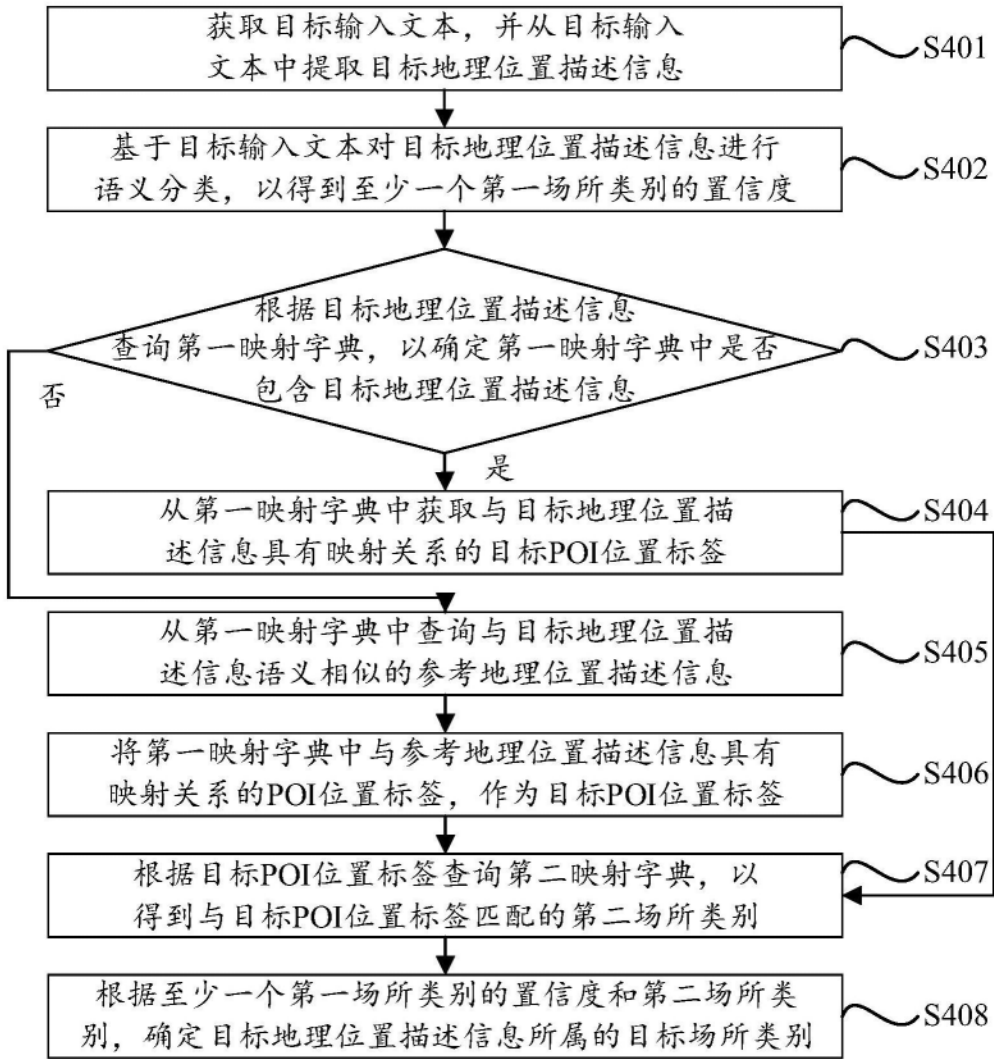


图4



图5



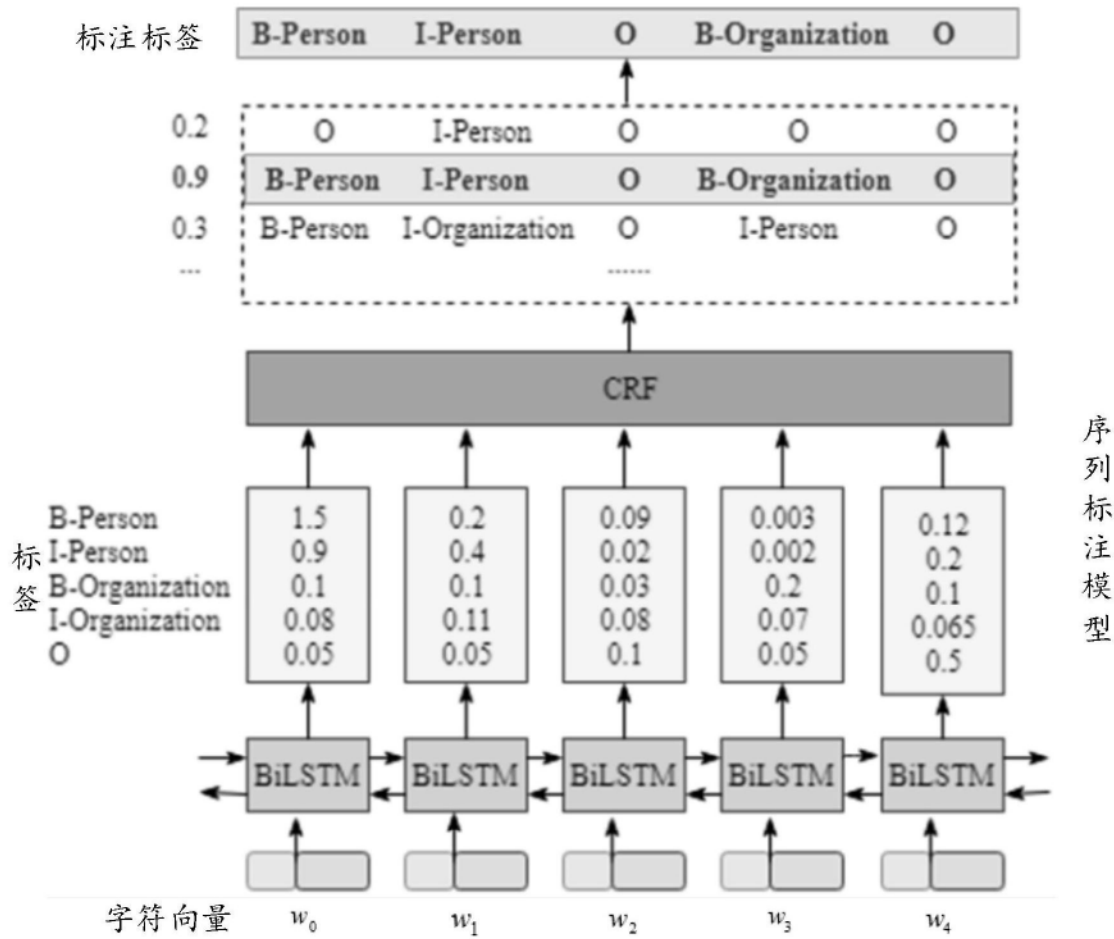


图6

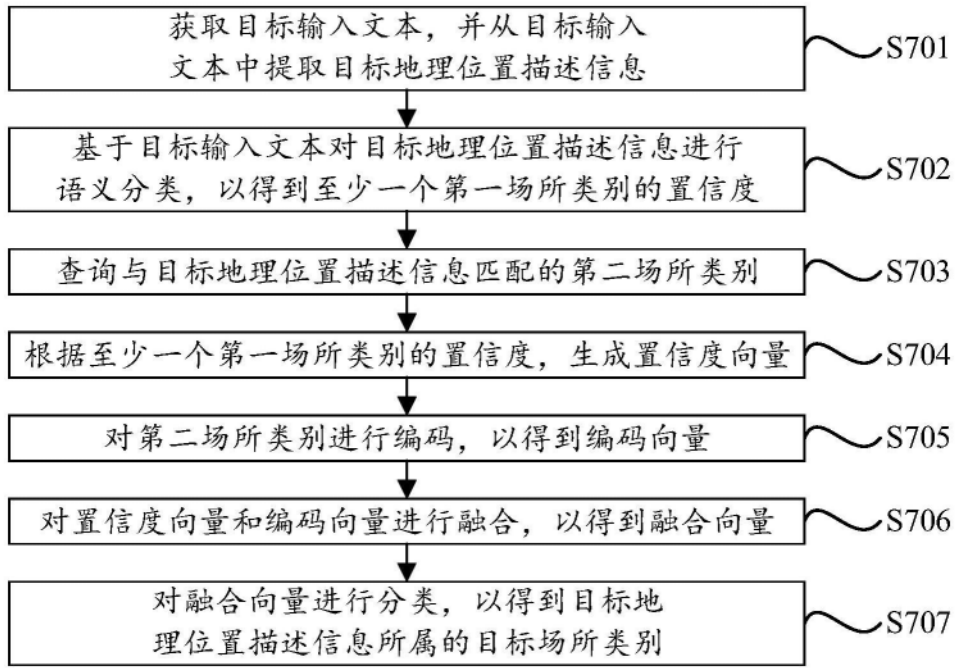


图7

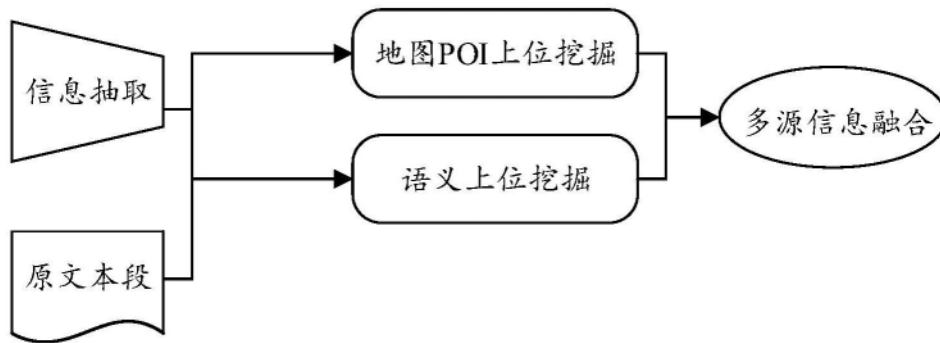


图8

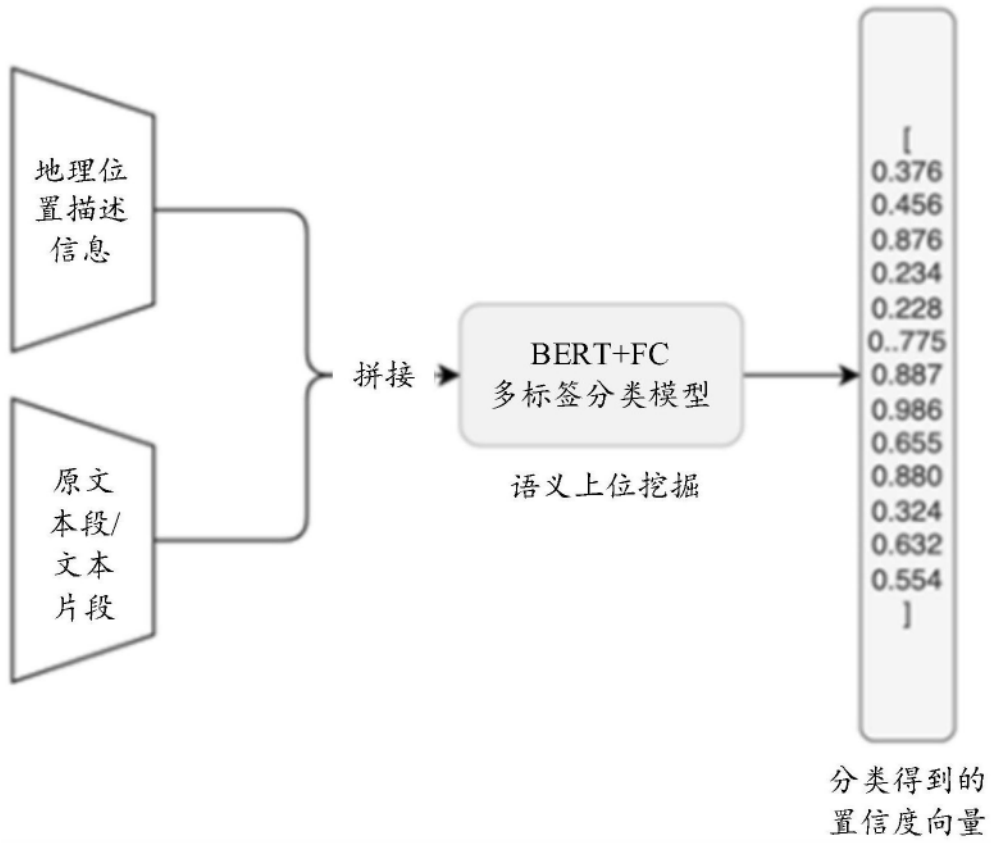


图9

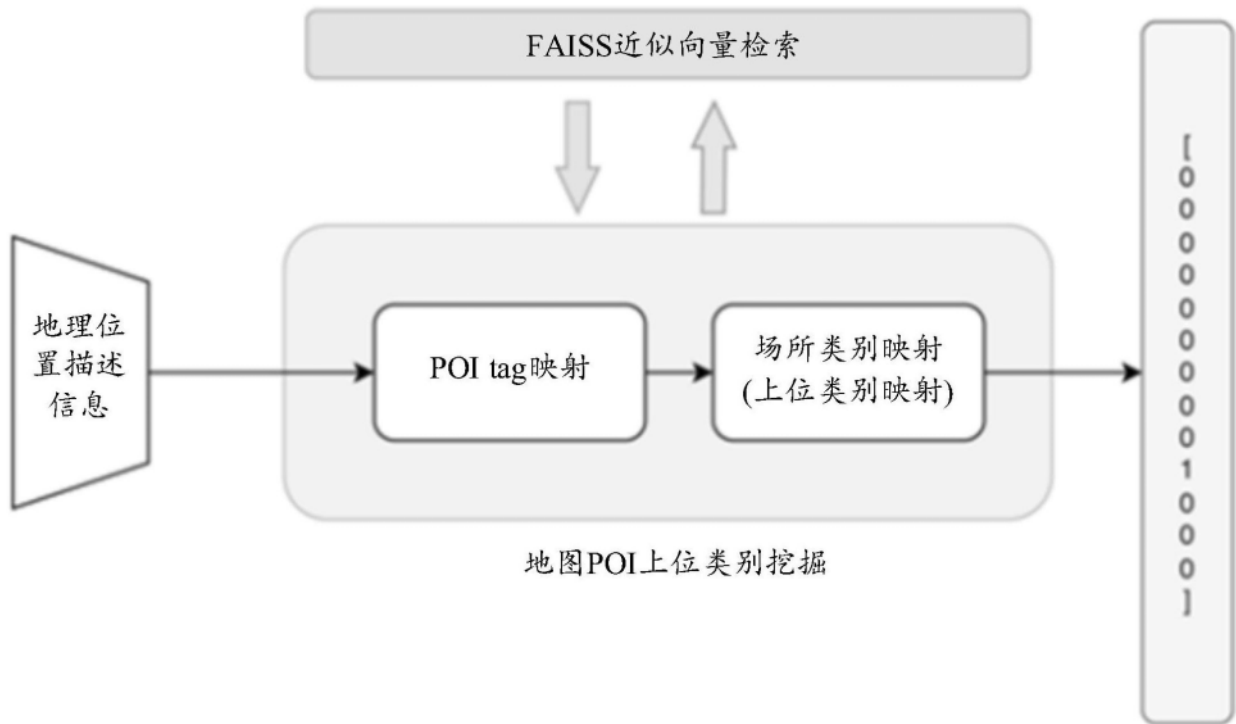


图10

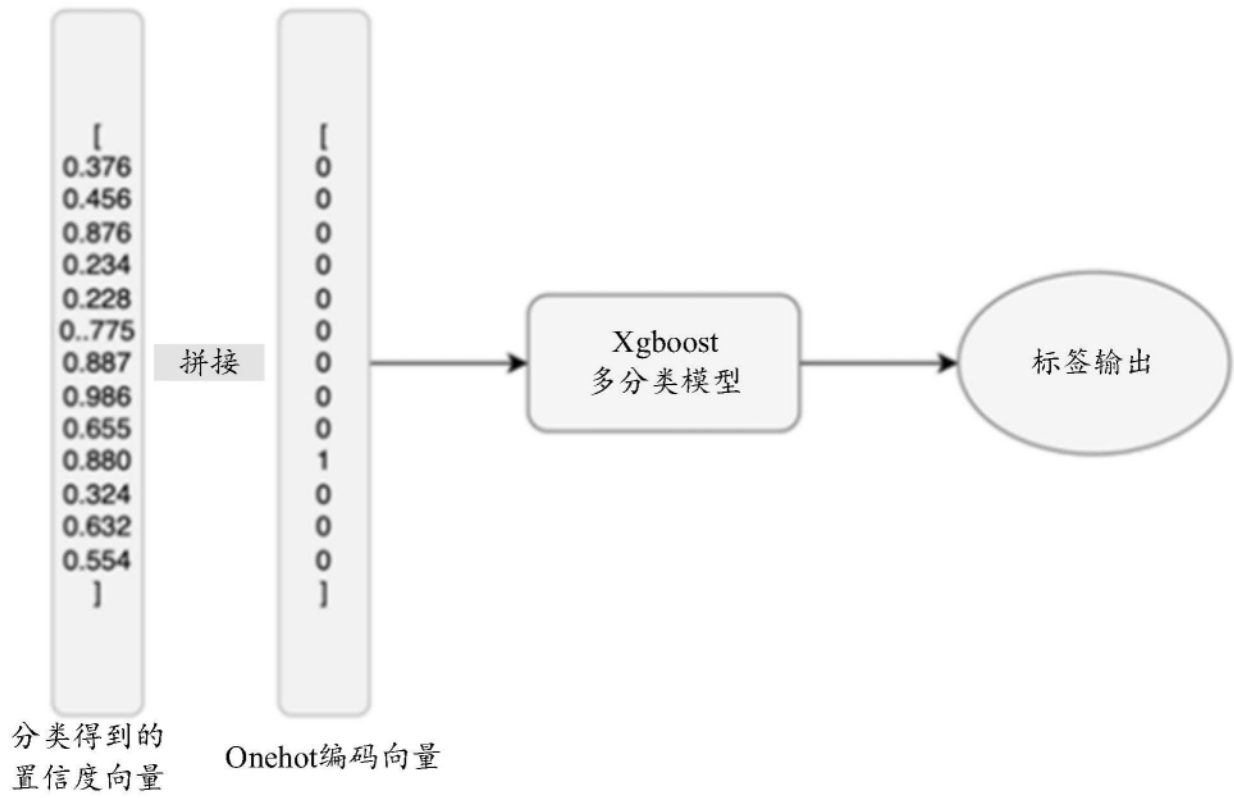


图11

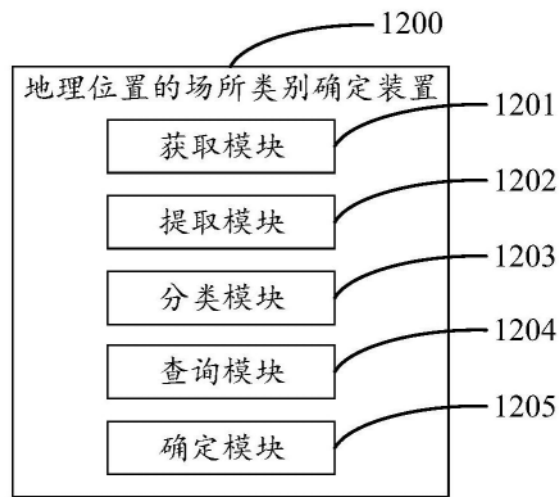


图12

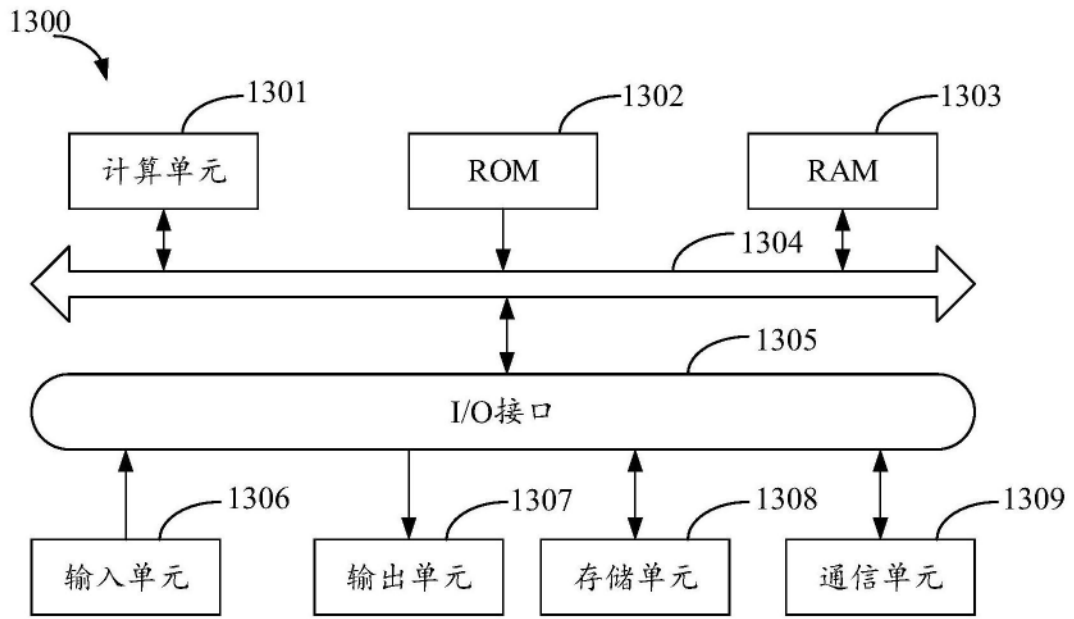


图13