



(12) 发明专利申请

(10) 申请公布号 CN 112925995 A

(43) 申请公布日 2021.06.08

(21) 申请号 202110199748.9

(22) 申请日 2021.02.22

(71) 申请人 北京百度网讯科技有限公司  
地址 100085 北京市海淀区上地十街10号  
百度大厦2层

(72) 发明人 黄际洲 孙一博 王海峰

(74) 专利代理机构 北京鸿德海业知识产权代理  
有限公司 11412  
代理人 田宏宾

(51) Int. Cl.

G06F 16/9537 (2019.01)

G06F 16/387 (2019.01)

G06F 16/23 (2019.01)

G06F 40/295 (2020.01)

G06F 40/30 (2020.01)

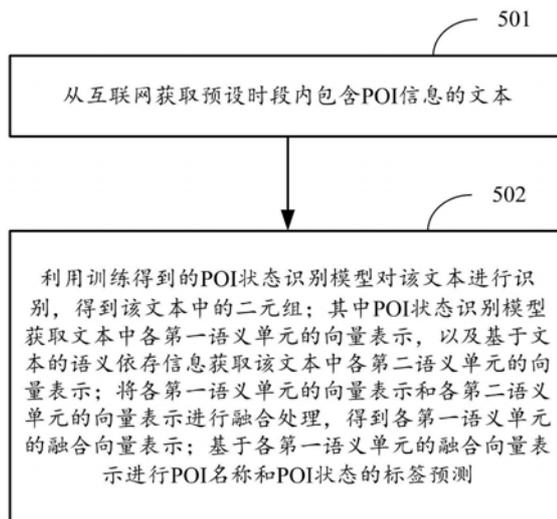
权利要求书4页 说明书13页 附图7页

(54) 发明名称

获取POI状态信息的方法及装置

(57) 摘要

本公开公开了一种获取POI状态信息的方法及装置,涉及人工智能技术领域下的大数据技术。具体实现方案为:从互联网获取预设时段内包含POI信息的文本;利用预先训练得到的POI状态识别模型对文本进行识别,得到文本中的二元组;其中POI状态识别模型获取文本中各第一语义单元的向量表示,以及基于文本的语义依存信息获取文本中各第二语义单元的向量表示;将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。本公开的技术方案能够节约人力成本,提高时效性和准确性。



1. 一种获取兴趣点POI状态信息的方法,包括:
  - 从互联网获取预设时段内包含POI信息的文本;
  - 利用预先训练得到的POI状态识别模型对所述文本进行识别,得到所述文本中的二元组,所述二元组包括POI名称和该POI名称对应的POI状态信息;
  - 其中所述POI状态识别模型获取所述文本中各第一语义单元的向量表示,以及基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。
2. 根据权利要求1所述的方法,其中,所述从互联网获取预设时段内包含POI信息的文本包括:
  - 从互联网获取预设时段内的文本;
  - 利用预先训练得到的POI识别器对获取到的文本进行POI信息识别;
  - 保留包含POI信息的文本。
3. 根据权利要求2所述的方法,其中,所述POI识别器采用以下方式训练得到:
  - 从POI数据库中抽取POI信息;
  - 利用抽取的POI信息在互联网中进行搜索;
  - 利用搜索得到的文本以及对应的POI信息训练标签预测模型,得到所述POI识别器。
4. 根据权利要求1所述的方法,其中,获取所述文本中各第一语义单元的向量表示包括:
  - 通过预训练语言模型对所述文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。
5. 根据权利要求1所述的方法,其中,基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示包括:
  - 对所述文本进行句法分析得到所述文本的语义依存树;
  - 获取所述语义依存树对应的有向图,所述有向图中的各节点为所述各第二语义单元;
  - 利用关系图卷积神经网络RGCN获取所述有向图中各节点的向量表示。
6. 根据权利要求1所述的方法,其中,将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示包括:
  - 若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示;或者,
  - 将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。
7. 根据权利要求1所述的方法,其中,所述基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测包括:
  - 利用条件随机场CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,所述标签指示对应第一语义单元是否为POI名称以及POI状态。
8. 根据权利要求1至7中任一项所述的方法,其中,所述第一语义单元为字,所述第二语义单元为词语。

9. 一种建立POI状态识别模型的方法,包括:

获取训练数据,所述训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果;

将所述文本作为POI状态识别模型的输入,将所述文本中所包含POI名称和POI状态的标签标注结果作为所述POI状态识别模型的目标预测结果,以训练所述POI状态识别模型;

其中所述POI状态识别模型获取所述文本中各第一语义单元的向量表示,以及基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

10. 根据权利要求9所述的方法,其中,所述获取训练数据包括:

获取种子文本以及对所述种子文本所包含POI名称和该POI名称对应的POI状态的标签标注结果;

利用POI数据库中与所述种子文本所包含POI名称同类型的POI名称,对所述种子文本中的POI名称进行替换,得到扩展文本;

由所述种子文本和所述扩展文本得到所述训练数据。

11. 根据权利要求10所述的方法,其中,所述获取种子文本包括:

从互联网获取包含POI名称的文本;

利用预先构建的词典从所述包含POI名称的文本中过滤掉不包含POI状态变化信息的文本,得到所述种子文本;其中所述词典包括表达POI状态变化信息的词。

12. 根据权利要求9所述的方法,其中,获取所述文本中各第一语义单元的向量表示包括:

通过预训练语言模型对所述文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。

13. 根据权利要求9所述的方法,其中,基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示包括:

对所述文本进行句法分析得到所述文本的语义依存树;

获取所述语义依存树对应的有向图,所述有向图中的各节点为所述各第二语义单元;

利用关系图卷积神经网络RGCN获取所述有向图中各节点的向量表示。

14. 根据权利要求9所述的方法,其中,将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示包括:

若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示;或者,

将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

15. 根据权利要求9所述的方法,其中,所述基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测包括:

利用条件随机场CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,所述标签指示是否为POI名称以及POI状态。

16. 一种获取POI状态信息的装置,包括:

文本获取单元,用于从互联网获取预设时段内包含POI信息的文本;

状态识别单元,用于利用预先训练得到的POI状态识别模型对所述文本进行识别,得到所述文本中的二元组,所述二元组包括POI名称和该POI名称对应的POI状态信息;

其中所述POI状态识别模型包括:

语义表示模块,用于获取所述文本中各第一语义单元的向量表示;

语义依存表示模块,用于基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;

表示融合模块,用于将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;

标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

17. 根据权利要求16所述的装置,其中,所述文本获取单元,具体用于从互联网获取预设时段内的文本;利用预先训练得到的POI识别器对获取到的文本进行POI信息识别;保留包含POI信息的文本。

18. 根据权利要求16所述的装置,其中,所述语义表示模块,具体用于使用预训练语言模型对所述文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。

19. 根据权利要求16所述的装置,其中,所述语义依存表示模块,具体用于对所述文本进行句法分析得到所述文本的语义依存树;获取所述语义依存树对应的有向图,所述有向图中的各节点为所述各第二语义单元;利用关系图卷积神经网络RGCN获取所述有向图中各节点的向量表示。

20. 根据权利要求16所述的装置,其中,所述表示融合模块,具体用于若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示;或者,将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

21. 根据权利要求16所述的装置,其中,所述标签预测模块,具体用于利用条件随机场CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,所述标签指示对应第一语义单元是否为POI名称以及POI状态。

22. 根据权利要求16至21中任一项所述的装置,其中,所述第一语义单元为字,所述第二语义单元为词语。

23. 一种建立POI状态识别模型的装置,包括:

获取单元,用于获取训练数据,所述训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果;

训练单元,用于将所述文本作为POI状态识别模型的输入,将所述文本中所包含POI名称和POI状态的标签标注结果作为所述POI状态识别模型的目标预测结果,以训练所述POI状态识别模型;

其中所述POI状态识别模型包括:

语义表示模块,用于获取所述文本中各第一语义单元的向量表示;

语义依存表示模块,用于基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;

表示融合模块,用于将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;

标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

24.根据权利要求23所述的装置,其中,所述获取单元,具体用于获取种子文本以及对所述种子文本所包含POI名称和该POI名称对应的POI状态的标签标注结果;利用POI数据库中与所述种子文本所包含POI名称同类型的POI名称,对所述种子文本中的POI名称进行替换,得到扩展文本;由所述种子文本和所述扩展文本得到所述训练数据。

25.根据权利要求23所述的装置,其中,所述语义表示模块,具体用于使用预训练语言模型对所述文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。

26.根据权利要求23所述的装置,其中,所述语义依存表示模块,具体用于对所述文本进行句法分析得到所述文本的语义依存树;获取所述语义依存树对应的有向图,所述有向图中的各节点为所述各第二语义单元;利用关系图卷积神经网络RGCN获取所述有向图中各节点的向量表示。

27.根据权利要求23所述的装置,其中,所述表示融合模块,具体用于若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示;或者,将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

28.根据权利要求23所述的装置,其中,所述标签预测模块,具体用于利用条件随机场CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,所述标签指示是否为POI名称以及POI状态。

29.一种电子设备,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1-15中任一项所述的方法。

30.一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权利要求1-15中任一项所述的方法。

31.一种计算机程序产品,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1-15中任一项所述的方法。

## 获取POI状态信息的方法及装置

### 技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及人工智能技术领域下的大数据技术。

### 背景技术

[0002] POI(Point of Interest)指的是兴趣点,是地理信息系统中的一个术语,泛指一切可以抽象为点的地理对象,一个POI可以是一栋房子、一个商铺、一个邮筒、一个公交站、一所学校、一个医院,等等。POI的主要用途是对事物或事件的位置进行描述,从而增强对事物或事件位置的描述能力和查询能力。

[0003] 在互联网地图类应用中,POI扮演了非常重要的角色。通过使用POI用户可以在地图中方便地找到感兴趣的地点以及到达该地点的路线。但由于受到城市规划以及其他因素的影响,随着时间的推移会有部分POI的信息发生变化,例如新增的POI、搬迁的POI、更名的POI、关闭的POI等等。如果不能及时对POI数据库进行更新,用户在使用地图搜索时可能找不到他们感兴趣的信息,甚至会对用户产生误导。例如用户导航到某个商场后,发现该商场已经搬迁或者停业,显然会给用户带来非常不好的体验。

[0004] 目前对POI状态信息的获取主要依靠人力,例如由工作人员到现场采集,或者由用户主动上报。但这些方式一方面浪费人力资源,成本较高;另一方面严重依赖人为的主动性,时效性和准确性难以保证。

### 发明内容

[0005] 本公开提供了一种获取POI状态信息的方法、装置、设备、计算机存储介质和计算机程序产品,以便于节约人力成本,提高时效性和准确性。

[0006] 根据本公开的第一方面,提供了一种获取POI状态信息的方法,包括:

[0007] 从互联网获取预设时段内包含POI信息的文本;

[0008] 利用预先训练得到的POI状态识别模型对所述文本进行识别,得到所述文本中的二元组,所述二元组包括POI名称和该POI名称对应的POI状态信息;

[0009] 其中所述POI状态识别模型获取所述文本中各第一语义单元的向量表示,以及基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0010] 根据本公开的第二方面,提供了一种建立POI状态识别模型的方法,包括:

[0011] 获取训练数据,所述训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果;

[0012] 将所述文本作为POI状态识别模型的输入,将所述文本中所包含POI名称和POI状态的标签标注结果作为所述POI状态识别模型的目标预测结果,以训练所述POI状态识别模型;

[0013] 其中所述POI状态识别模型获取所述文本中各第一语义单元的向量表示,以及基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测

[0014] 根据本公开的第三方面,提供了一种获取POI状态信息的装置,包括:

[0015] 文本获取单元,用于从互联网获取预设时段内包含POI信息的文本;

[0016] 状态识别单元,用于利用预先训练得到的POI状态识别模型对所述文本进行识别,得到所述文本中的二元组,所述二元组包括POI名称和该POI名称对应的POI状态信息;

[0017] 其中所述POI状态识别模型包括:

[0018] 语义表示模块,用于获取所述文本中各第一语义单元的向量表示;

[0019] 语义依存表示模块,用于基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;

[0020] 表示融合模块,用于将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;

[0021] 标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0022] 根据本公开的第四方面,提供了一种建立POI状态识别模型的装置,包括:

[0023] 获取单元,用于获取训练数据,所述训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果;

[0024] 训练单元,用于将所述文本作为POI状态识别模型的输入,将所述文本中所包含POI名称和POI状态的标签标注结果作为所述POI状态识别模型的目标预测结果,以训练所述POI状态识别模型;

[0025] 其中所述POI状态识别模型包括:

[0026] 语义表示模块,用于获取所述文本中各第一语义单元的向量表示;

[0027] 语义依存表示模块,用于基于所述文本的语义依存信息获取所述文本中各第二语义单元的向量表示;

[0028] 表示融合模块,用于将所述各第一语义单元的向量表示和所述各第二语义单元的向量表示进行融合处理,得到所述各第一语义单元的融合向量表示;

[0029] 标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0030] 根据本公开的第五方面,提供了一种电子设备,包括:

[0031] 至少一个处理器;以及

[0032] 与所述至少一个处理器通信连接的存储器;其中,

[0033] 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如上所述的方法。

[0034] 根据本公开的第六方面,提供了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如上所述的方法。

[0035] 根据本公开的第七方面,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如上所述的方法。

[0036] 由以上技术方案可以看出,本公开通过从包含POI信息的互联网文本中挖掘POI名称和POI名称对应的POI状态信息,充分利用了互联网对POI状态改变事件的及时反应,相比较人工确定发生状态改变的POI信息的方式,节约了人力成本,提高了时效性和准确性。

[0037] 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。本公开的其他特征将通过以下的说明书而变得容易理解。

[0038] 上述可选方式所具有的其他效果将在下文中结合具体实施例加以说明。

## 附图说明

[0039] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0040] 图1示出了可以应用本公开实施例的方法或装置的示例性系统架构;

[0041] 图2为本公开实施例提供的一种建立POI状态识别模型的方法流程图;

[0042] 图3为本公开实施例提供的获取训练数据的方法流程图;

[0043] 图4为本公开实施例提供的POI状态识别模型的示意图;

[0044] 图5为本公开实施例提供的获取POI状态信息的方法流程图;

[0045] 图6为本公开实施例提供的对文本进行标签预测的实例图;

[0046] 图7为本公开实施例提供的建立POI状态识别模型的装置结构图;

[0047] 图8为本公开实施例提供的获取POI状态信息的装置结构图;

[0048] 图9是用来实现本公开实施例的电子设备的框图。

## 具体实施方式

[0049] 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。

[0050] 图1示出了可以应用本公开实施例的方法或装置的示例性系统架构。如图1所示,该系统架构可以包括终端设备101和102,网络103和服务器104。网络103用以在终端设备101、102和服务器104之间提供通信链路的介质。网络103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。

[0051] 用户可以使用终端设备101和102通过网络103与服务器104交互。终端设备101和102上可以安装有各种应用,例如地图类应用、网页浏览器应用、通信类应用等。

[0052] 终端设备101和102可以是能够运行地图类应用的各种用户设备。包括但不限于智能手机、平板电脑、PC、智能电视等等。本公开所提供的获取兴趣点POI状态信息的装置以及建立POI状态识别模型的装置可以设置并运行于上述服务器104中,也可以运行于独立于服务器104的设备中。其可以实现成多个软件或软件模块(例如用来提供分布式服务),也可以实现成单个软件或软件模块,在此不做具体限定。服务器104可以与POI数据库105之间进行交互,具体地,服务器104可以从POI数据库105中获取数据,也可以将数据存储于POI数据库105中。POI数据库105中存储有包括POI信息的地图数据。

[0053] 例如,获取POI状态信息的装置设置并运行于上述服务器104中,服务器104采用本

公开实施例提供的方法进行获取POI的状态信息,然后利用获取的POI的状态信息更新POI数据库105。服务器104能够响应于终端设备101、102的查询请求,查询POI数据库105,并向终端设备101、102返回所查询POI的信息。

[0054] 服务器104可以是单一服务器,也可以是多个服务器构成的服务器群组。另外104除了以服务器的形式存在之外,也可以是具有较高计算性能的其他计算机系统或处理器。应该理解,图1中的终端设备、网络、服务器和数据库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、服务器和数据库。

[0055] 本公开基于互联网大数据所具有的丰富性和实时性,将互联网大数据作为数据源,从互联网文本中挖掘POI状态信息。核心思想在于,从互联网获取预设时段内包含POI信息的文本;利用预先训练得到的POI状态识别模型对所述文本进行识别,得到文本中的二元组,二元组包括POI名称和该POI名称对应的POI状态信息。其中,从本公开的整体方案上来看,涉及到两个部分,一个是POI状态识别模型的建立,另一个是利用POI状态识别模型对POI状态信息的获取。下面分别结合实施例对这两个部分进行详细描述。

[0056] 图2为本公开实施例提供的一种建立POI状态识别模型的方法流程图,如图2中所示,该方法可以包括以下步骤:

[0057] 在201中,获取训练数据,其中训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果。

[0058] 在202中,将文本作为POI状态识别模型的输入,将该文本中所包含POI名称和POI状态的标签标注结果作为POI状态识别模型的目标预测结果,以训练该POI状态识别模型。

[0059] 其中POI状态识别模型获取文本中各第一语义单元的向量表示,以及基于该文本的语义依存信息获取该文本中各第二语义单元的向量表示;将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0060] 下面结合实施例对上述步骤201即“获取训练数据”进行详细描述。

[0061] 在获取训练数据时,可以采用人工标注的方式对一些包含POI名称和POI状态的文本进行标注以作为训练数据。但这种方式因人力的限制只能产生少量数据,对模型训练的准确性会带来一定制约。在此,作为一种优选的实施方式,可以采用如图3中所示的方式来获取训练数据,具体包括:

[0062] 在301中,利用预先训练得到的POI识别器从互联网获取包含POI名称的文本。

[0063] 从网页上爬取互联网文本,为了确保文本中包含POI信息,可以训练一个POI识别器,并保留被POI识别器识别出包含POI信息的文本。

[0064] 其中预先训练POI识别器的过程为:从POI数据库中抽取出POI信息,其中抽取的POI信息可以是热门的POI信息,即可以为检索热度超过一定程度的POI信息,例如设定时长内的检索次数超过预设次数阈值的POI信息;然后利用抽取的POI信息在互联网中进行搜索,利用搜索得到的文本以及对应的POI信息训练标签预测模型,从而得到POI识别器。POI识别器能够在输入文本后,输出该文本中包含的POI信息。其中,POI信息可以包括POI名称。需要说明的是,在本公开中,POI名称指代能够标明和区分一个POI的信息,可以是中文名称、英文名称、名称缩写,等等。

[0065] 训练POI识别器过程中采用的标签预测模型可以是由双向LSTM(Long Short-Term

Memory,长短期记忆网络)和CRF(conditional random field,条件随机场)构成的结构。当然也可以采用其他类型的结构,在此不做一一列举和详细描述。

[0066] 在302中,利用预先构建的词典从包含POI名称的文本中过滤掉不包含POI状态变化信息的文本,得到种子文本。

[0067] 其中词典包括包含POI状态变化信息的词。为了确保文本中包含描述POI状态变化的信息,预先构建了一个词典,这个词典中的词语义与POI状态变化相关。其中POI状态变化可以包括新增、搬迁、更名、关闭等。

[0068] 经过上述301和302的处理后,得到的种子文本中均包括POI名称和POI状态变化信息。上述301和302是获取种子文本的一种优选实施方式,除了上述方式之外,也可以简单地采用人工构建文本方式来得到种子文本。

[0069] 在303中,获取对种子文本所包含POI名称和该POI名称对应的POI状态的标签标注结果。

[0070] 本步骤中获取的种子文本数量不多,可以采用人工标注的方式。然而,若数量不多的种子文本作为训练数据进行POI状态识别模型的训练,则可能会使得POI状态识别模型的准确度不够高。为了进一步提高训练效果,可以在此基础上利用POI数据库对种子文本进行数据增强。即在304中,利用POI数据库中种子文本所包含POI名称同类型的POI名称,对种子文本中的POI名称进行替换,得到扩展文本。

[0071] 具体地,可以采用一个分类模型对种子文本中包含的POI名称和POI数据库中的POI名称分别进行分类。分类得到的类别用以区分POI的特性,例如餐厅类、公司类、住宅类等等。

[0072] 举个例子,若一个种子文本为“A公司已搬迁,搬迁期间暂停办理业务”,其中种子文本中包含的POI名称“A公司”为公司类,则可以从POI数据库中找到同类型的POI名称“B公司”、“C公司”等等,分别利用这些名称对种子文本中的“A”公司进行替换后,得到扩展文本:“B公司已搬迁,搬迁期间暂停办理业务”、“C公司已搬迁,搬迁期间暂停办理业务”等等。

[0073] 其中分类模型可以预先利用POI数据库中大量的(POI名称,类型标签)对进行训练,分类模型的结构可以采用诸如双向LSTM和softmax构成的结构。

[0074] 在305中,由种子文本和扩展文本得到训练数据。

[0075] 经过对种子文本进行扩展后,可以得到大量的扩展文本,然后将这些种子文本和扩展文本共同构成训练数据。这些训练数据中包含文本以及对文本中包含的POI名称和该POI名称对应的POI状态的标签标注结果。

[0076] 下面结合实施例对上述步骤202即“将文本作为POI状态识别模型的输入,将该文本中所包含POI名称和POI状态的标签标注结果作为POI状态识别模型的目标预测结果,以训练该POI状态识别模型”进行详细描述。

[0077] 在本公开涉及的POI状态识别模型中,如图4中所示,主要包含四个部分:语义表示模块、语义依存表示模块、表示融合模块和标签预测模块。

[0078] 其中,语义表示模块用以获取文本中各第一语义单元的向量表示。在本公开中,第一语义单元可以是字,这在中文中比较常见。在英文中,第一语义单元可以是一个单词(word),也可以是更细粒度的,例如复合词中的每一个“字”。举个例子,对于可以将“yellowstone”作为一个第一语义单元,也可以将“yellow”和“stone”分别作为第一语义单

元。

[0079] 可以使用传统的word2vec模型、glove模型等实现对文本中各字的语义特征提取。但由于POI名称常常使用长尾词,采用传统的模型并不能很好地对各字的语义进行表达。因此,本公开中采用了一种优选的实施方式,由ERNIE(Enhanced Representation from Knowledge Integration,知识增强语义表示)、BERT(Bidirectional Encoder Representations from Transformers,Transformer的双向编码器表示)等预训练语言模型对文本对应的词序列进行语义特征的提取。由于ERNIE模型使用subword(子词)策略,即采用BPE(Byte-Pair Encoding,双字节编码)算法将POI名称拆解成更细粒度的字,从而高效地解决长尾词的问题。

[0080] 假设训练数据中某文本D对应的字序列表示为 $\{x_1, x_2, \dots, x_n\}$ ,其中,n为D中字的数量。 $\{x_1, x_2, \dots, x_n\}$ 作为ERNIE模型的输入,ERNIE模型提取语义特征后输出隐层向量序列 $\{h_1, h_2, \dots, h_n\}$ 。其中, $h_i$ 是字 $x_i$ 对应的向量表示。

[0081] 语义依存表示模块用以基于文本的语义依存信息获取该文本中各第二语义单元的向量表示。

[0082] 具体地,可以对文本进行句法分析得到文本的语义依存树;然后获取语义依存树对应的有向图,在有向图中各节点为各第二语义单元;利用RGCN(Relational Graph Convolutional Network,关系图卷积神经网络)获取有向图中各节点的向量表示。

[0083] 其中可以使用依存句法分析器对文本进行分析得到文本的语义依存树。依存句法分析是自然语言处理的核心技术之一,旨在通过分析句子中词语之间的依存关系确定句子的句法结构。通常包括对文本进行分词、词性标注、命名实体识别、语法分析等处理。依存句法分析器是目前较为成熟的工具,本公开中可以直接使用此类工具实现。

[0084] 本公开中的第二语义单元可以是词语。除此之外,也可以是短语、词语组合等。在此以词语为例。

[0085] 在语义依存树中,节点为文本中的词语,边为词语之间的语义关系。可以将语义依存树转化为有向图。在有向图中节点为文本中的词语,边体现词语之间的语义关系。

[0086] 然后有向图作为RGCN的输入,由RGCN输出有向图中各节点的向量表示。

[0087] 对于语义依存表示模块而言,针对D对应的词序列表示为 $\{y_1, y_2, \dots, y_m\}$ ,其中,m为D中词语的数量。经过语义依存表示模块后,得到各词语对应的向量表示分别为 $\{l_1, l_2, \dots, l_m\}$ 。

[0088] 表示融合模块用以将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示。

[0089] 其中,表示融合模块可以采用但不限于以下两种方式:

[0090] 第一种方式:若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后,得到该第一语义单元的融合向量。

[0091] 以第一语义单元是字、第二语义单元是词为例,假设词A是由字A1和字A2组成,则将字A1的向量表示与词A的向量表示进行拼接后得到字A1的融合向量表示,将A2的向量表示与词A的向量表示进行拼接后得到字A2的融合向量表示。

[0092] 第二种方式:若各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

[0093] 以第一语义单元是字、第二语义单元是词为例,将得到的所有词的向量表示对字A1的向量表示进行Attention(注意力)处理。由于第二语义单元是从有向图获取的,因此实际上就相当于使用整个有向图对字A1进行Attention,从而得到整个有向图关于该字A1的向量表示。即可以将字的向量表示作为query,将每个词的向量表示分别作为key,从而进行Attention。

[0094] 标签预测模块用以基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0095] 在此可以利用诸如CRF(条件随机场)等序列预测模型进行标签预测。对于包含n个字的文档D,  $D = \{x_1, x_2, \dots, x_n\}$ 。经过上述表示融合模块的融合后,得到各字的融合向量表示,即序列  $\{h'_1, h'_2, \dots, h'_n\}$ 。对于文档中的每个字  $x_i$ , 标签预测模块的任务是预测其标签  $t_i$ , 其中,  $t_i \in T^p$ 。在进行标签预测时,对于字  $x_i$  的标签指示该字  $x_i$  是否是POI名称以及POI状态,也就是说,一个标签会对POI名称和POI状态信息同时进行指示。例如,标签集合可以采用:

[0096]  $T^p = \{B-NEW, I-NEW, B-RELOCATE, I-RELOCATE, B-RENAME, I-RENAME, B-CLOSE, I-CLOSE, B-NONE, I-NONE, 0\}$

[0097] 其中,B、I和0指示该字是POI名称的开始、中间和非POI名称。

[0098] NEW、RELOCATE、RENAME、CLOSE和NONE分别指示POI状态为新增、更名、搬迁、关闭和未识别出POI状态。

[0099] 因此,标签B-NEW指示该字是POI名称的开始且POI状态为新增,标签I-NEW指示该字是POI名称的中间且POI状态为新增,其他标签类似。

[0100] 若该子模型输出的预测序列为  $T = \{t_1, t_2, \dots, t_n\}$ , 则可以得到这个序列的得分:

$$[0101] \quad s(D, T) = \sum_{i=1}^n p_i^{t_i} + \sum_{i=0}^n T_{t_i, t_{i+1}}$$

[0102] 其中,  $p_i^{t_i} = \text{soft max}(W_{t_i}^T h'_i)$ ,  $W_{t_i}$  为权重参数,是模型参数之一。 $W_{t_i} \in \mathfrak{R}^{\{2d \times c\}}$ , 其中,  $\mathfrak{R}$  代表实数域,  $d$  为  $h'_i$  的维度,  $c$  为模型输出标签的数量。 $T_{t_i, t_{i+1}}$  表示从  $t_i$  到  $t_{i+1}$  的转移概率矩阵。

[0103] 最终可以使用softmax层得到每个预测序列T的概率:

$$[0104] \quad p(T|D) = \frac{e^{s(D, T)}}{\sum_{\tilde{T} \in T_D} e^{s(D, \tilde{T})}}$$

[0105] 其中,  $T_D$  为所有的预测序列构成的集合。

[0106] 训练阶段采用的损失函数为:  $L_0 = \sum \log(p(T|D))$ 。

[0107] 训练过程中利用损失函数的值更新模型参数,包括语义表示模块、语义依存表示模块、表示融合模块和标签预测模块的参数值,直至达到训练结束条件。其中训练结束条件可以包括:损失函数的值小于或等于预设的损失函数阈值,迭代次数达到预设的次数阈值,等等。

[0108] 最终训练得到的POI状态识别模型能够对输入的文本进行识别,得到文本中的二元组,该二元组包括POI名称和POI名称对应的POI状态信息,表示为(POI名称,状态信息)。

[0109] 图5为本公开实施例提供的获取POI状态信息的方法流程图,该方法基于图2所示

流程所建立的POI状态识别模型。如图5中所示,该方法可以包括以下步骤:

[0110] 在501中,从互联网获取预设时段内包含POI信息的文本。

[0111] 为了保证POI状态信息的实时性,可以从互联网爬取最近一段时间的文本,然后利用预先训练得到的POI识别器对获取到的文本进行POI信息识别,保留包含POI信息的文本。

[0112] 关于POI识别器的相关记载可以参见上面实施例中对步骤301的相关记载,在此不做赘述。

[0113] 在502中,利用预先训练得到的POI状态识别模型对该文本进行识别,得到该文本中的二元组,即(POI名称,状态信息)。

[0114] 其中POI状态识别模型获取文本中各第一语义单元的向量表示,以及基于文本的语义依存信息获取该文本中各第二语义单元的向量表示;将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示;基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0115] 在步骤502中,可以使用图4中所示的POI状态识别模型对文本进行识别。

[0116] 其中,语义表示模块用以获取文本中各第一语义单元的向量表示。在本公开中,第一语义单元可以是字,这在中文中比较常见。在英文中,第一语义单元可以是一个单词(word),也可以是更细粒度的,例如复合词中的每一个“字”。

[0117] 可以使用传统的word2vec模型、glove模型等实现对文本中各字的语义特征提取。但由于POI名称常常使用长尾词,采用传统的模型并不能够很好地对各字的语义进行表达。因此,本公开中采用了一种优选的实施方式,由ERNIE、BERT等预训练语言模型对文本对应的词序列进行语义特征的提取。由于ERNIE模型使用subword策略,即采用BPE算法将POI名称拆解成更细粒度的字,从而高效地解决长尾词的问题。

[0118] 语义依存表示模块用以基于文本的语义依存信息获取该文本中各第二语义单元的向量表示。

[0119] 具体地,可以对文本进行句法分析得到文本的语义依存树;然后获取语义依存树对应的有向图,在有向图中各节点为各第二语义单元;利用RGCN获取有向图中各节点的向量表示。

[0120] 表示融合模块用以将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示。

[0121] 其中,表示融合模块可以采用但不限于以下两种方式:

[0122] 第一种方式:若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后,得到该第一语义单元的融合向量。

[0123] 第二种方式:若各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

[0124] 标签预测模块用以基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。在此可以利用诸如CRF(条件随机场)等序列预测模型进行标签预测。预测的标签会对POI名称和POI状态信息同时进行指示。

[0125] 上述POI状态识别模型中各模块的具体实现方式可以参见图2所示实施例中的相关记载,在此不做一一赘述。

[0126] 举个例子:

[0127] 假设输入文本为“继坪山图书馆开业之后,坪山艺术博物馆也在3月30日正式开业”。

[0128] 经过POI状态识别模型的预测之后,打上的标签如图6中所示。可以确定出二元组(坪山图书馆,新增)和(坪山艺术博物馆,新增)。

[0129] 在针对互联网文本挖掘出一系列二元组后,可以将挖掘出的二元组与POI数据库中记录的POI名称及其对应的POI状态信息进行比对,如果存在不一致,可以利用挖掘出的二元组对POI数据库中记录的POI名称及其对应的POI状态信息进行更新。

[0130] 其中,在进行更新时,可以依据挖掘出的二元组的出现次数对二元组的准确性进行衡量,例如认为出现次数大于预设次数阈值的二元组认为是准确的,可以利用准确的二元组对POI数据库进行更新。

[0131] 以上是对本公开所提供方法进行的详细描述,下面结合实施例对本公开所提供的装置进行详细描述。

[0132] 图7为本公开实施例提供的建立POI状态识别模型的装置结构图,如图7中所示,该装置可以包括:获取单元01和训练单元02,其中各组成单元的主要功能如下:

[0133] 获取单元01,用于获取训练数据,训练数据包括:文本以及对该文本中所包含POI名称和POI状态的标签标注结果。

[0134] 其中获取单元01可以获取种子文本以及对种子文本所包含POI名称和该POI名称对应的POI状态的标签标注结果;利用POI数据库中与种子文本所包含POI名称同类型的POI名称,对种子文本中的POI名称进行替换,得到扩展文本;由种子文本和扩展文本得到训练数据。

[0135] 其中获取单元01在获取种子文本时,可以从互联网获取包含POI名称的文本;利用预先构建的词典从所述包含POI名称的文本中过滤掉不包含POI状态变化信息的文本,得到所述种子文本;其中所述词典包括表达POI状态变化信息的词。

[0136] 训练单元02,用于将文本作为POI状态识别模型的输入,将文本中所包含POI名称和POI状态的标签标注结果作为POI状态识别模型的目标预测结果,以训练POI状态识别模型。

[0137] 其中POI状态识别模型如图4中所示,可以包括:

[0138] 语义表示模块,用于获取文本中各第一语义单元的向量表示。

[0139] 具体地,语义表示模块可以使用诸如ERNIE、BERT等预训练语言型对文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。

[0140] 语义依存表示模块,用于基于文本的语义依存信息获取文本中各第二语义单元的向量表示。

[0141] 作为一种优选的实施方式,语义依存表示模块可以对文本进行句法分析得到文本的语义依存树;获取语义依存树对应的有向图,有向图中的各节点为各第二语义单元;利用RGCN获取有向图中各节点的向量表示。

[0142] 表示融合模块,用于将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示。

[0143] 其中,表示融合模块可以采用但不限于以下两种方式:

[0144] 第一种方式:若第二语义单元包含第一语义单元,则将该第一语义单元的向量表

示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示。

[0145] 第二种方式:将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

[0146] 标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0147] 作为一种优选的实施方式,标签预测模块可以利用CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,标签指示是否为POI名称以及POI状态。

[0148] 图8为本公开实施例提供的获取POI状态信息的装置结构图,如图8中所示,该装置可以包括:文本获取单元11和状态识别单元12,还可以进一步包括更新单元13。其中各组成单元的主要功能如下:

[0149] 文本获取单元11,用于从互联网获取预设时段内包含POI信息的文本。

[0150] 作为其中一种实现方式,文本获取单元11可以从互联网获取预设时段内的文本;利用预先训练得到的POI识别器对获取到的文本进行POI信息识别;保留包含POI信息的文本。

[0151] 状态识别单元12,用于利用预先训练得到的POI状态识别模型对文本进行识别,得到文本中的二元组,二元组包括POI名称和该POI名称对应的POI状态信息。

[0152] 其中POI状态识别模型如图4中所示,可以包括:

[0153] 语义表示模块,用于获取文本中各第一语义单元的向量表示。

[0154] 具体地,语义表示模块可以使用诸如ERNIE、BERT等预训练语言型对文本中的各第一语义单元进行语义特征的提取,得到各第一语义单元的向量表示。

[0155] 语义依存表示模块,用于基于文本的语义依存信息获取文本中各第二语义单元的向量表示。

[0156] 作为一种优选的实施方式,语义依存表示模块可以对文本进行句法分析得到文本的语义依存树;获取语义依存树对应的有向图,有向图中的各节点为各第二语义单元;利用RGCN获取有向图中各节点的向量表示。

[0157] 表示融合模块,用于将各第一语义单元的向量表示和各第二语义单元的向量表示进行融合处理,得到各第一语义单元的融合向量表示。

[0158] 其中,表示融合模块可以采用但不限于以下两种方式:

[0159] 第一种方式:若第二语义单元包含第一语义单元,则将该第一语义单元的向量表示和该第二语义单元的向量表示进行拼接后得到该第一语义单元的融合向量表示。

[0160] 第二种方式:将各第二语义单元的向量表示对第一语义单元进行注意力处理,得到该第一语义单元的融合向量表示。

[0161] 标签预测模块,用于基于各第一语义单元的融合向量表示进行POI名称和POI状态的标签预测。

[0162] 作为一种优选的实施方式,标签预测模块可以利用CRF模型将各第一语义单元的融合向量进行映射,得到各第一语义单元的标签,标签指示是否为POI名称以及POI状态。

[0163] 更新单元13,用于利用得到的二元组更新POI数据库。

[0164] 在针对互联网文本挖掘出一系列二元组后,更新单元13可以将挖掘出的二元组与POI数据库中记录的POI名称及其对应的POI状态信息进行比对,如果存在不一致,可以利用

挖掘出的二元组对POI数据库中记录的POI名称及其对应的POI状态信息进行更新。

[0165] 其中,更新单元13在进行更新时,可以依据挖掘出的二元组的出现次数对二元组的准确性进行衡量,例如认为出现次数大于预设次数阈值的二元组认为是准确的,可以利用准确的二元组对POI数据库进行更新。

[0166] 下面列举具体的应用实例:

[0167] 从互联网中获取最近一个月内包含POI信息的文本,这些文本可能是数量较大的,在此举几个其中的例子:

[0168] 互联网文本1:上海大学科技园区与上海望源企业发展有限公司在宝山区罗森宝商务中心揭牌建立上大望源科技园;

[0169] 互联网文本2:之前火岩大峡谷在申报风景区时,已更名为乌龙山大峡谷;

[0170] 互联网文本3:嵩明县不动产登记中心于2019年6月28日搬迁至秀嵩街66号办公,搬迁期间暂停办理业务;

[0171] 互联网文本4:为了保护三江源的自然生态环境,年保玉则风景区已全面关闭。

[0172] 将获取的包含POI信息的互联网文本分别输入POI状态识别模型,该POI状态识别模型能够基于文本中词语的语义依存关系,更加准确地识别出其中的二元组。识别结果如下:

[0173] 互联网文本1对应的二元组:(上大望源科技园,新增);

[0174] 互联网文本2对应的二元组:(火岩大峡谷,更名);

[0175] 互联网文本3对应的二元组:(嵩明县不动产登记中心,搬迁)

[0176] 互联网文本4对应的二元组:(年保玉则风景区,关闭)。

[0177] 对于获取到的大量二元组,可以通过出现次数的方式来确定其准确性,例如认为某二元组被识别出的次数超过预设阈值,则认为是准确的,可以利用其更新更新POI数据库。也可以通过人工核验的方式,将获取到的二元组核验其准确性。

[0178] 经过上述过程就能够自动、及时地从互联网文档中挖掘出发生变化的POI状态信息,例如新增、更名、搬迁和关闭,并及时更新POI数据库,使得利用POI数据库提供的服务是基于准确的POI信息,提升用户体验。

[0179] 由以上实施例可以看出,本公开可以具备以下优点:

[0180] 1) 本公开通过从包含POI信息的互联网文本中挖掘POI名称和POI名称对应的POI状态信息,充分利用了互联网对POI状态改变事件的及时反应,相比较人工确定发生状态改变的POI信息的方式,节约了人力成本,提高了时效性和准确性。

[0181] 2) 本公开不仅能够识别出互联网文本中包含了POI状态变化信息,更能够识别出POI名称发生变化的具体POI状态信息。

[0182] 3) 本公开在建立的POI状态识别模型中融入了语义依存信息,从而帮助模型更加准确地识别文本中的POI名称以及理解各POI的状态,从而提高识别准确性。

[0183] 4) 本公开在训练POI状态识别模型时,利用POI数据库对种子文本进行数据扩展从而得到训练数据,从而扩大了训练文本的数量,以提高POI状态识别模型的准确度。

[0184] 5) 本公开的POI状态识别模型中采用ERNIE模型对输入文本对应的词序列进行语义特征的提取,从而高效地解决POI名称中长尾词的问题。

[0185] 6) 本公开利用挖掘得到的二元组更新POI数据库,使得后续基于POI数据库所提供

的服务更加准确,从而提高用户体验。

[0186] 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。

[0187] 如图9所示,是根据本公开实施例的获取POI状态信息或建立POI状态识别模型的方法的电子设备的框图。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、和其他适合的计算机。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、可穿戴设备和其他类似的计算装置。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。

[0188] 如图9所示,设备900包括计算单元901,其可以根据存储在只读存储器 (ROM) 902中的计算机程序或者从存储单元908加载到随机访问存储器 (RAM) 903中的计算机程序,来执行各种适当的动作和处理。在RAM 903中,还可存储设备900操作所需的各种程序和数据。计算单元901、ROM 902以及RAM 903通过总线904彼此相连。输入/输出 (I/O) 接口905也连接至总线904。

[0189] 设备900中的多个部件连接至I/O接口905,包括:输入单元906,例如键盘、鼠标等;输出单元907,例如各种类型的显示器、扬声器等;存储单元908,例如磁盘、光盘等;以及通信单元909,例如网卡、调制解调器、无线通信收发机等。通信单元909允许设备900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。

[0190] 计算单元901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901的一些示例包括但不限于中央处理单元 (CPU)、图形处理单元 (GPU)、各种专用的人工智能 (AI) 计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器 (DSP) 以及任何适当的处理器、控制器、微控制器等。计算单元901执行上文所描述的各个方法和处理,例如获取POI状态信息或建立POI状态识别模型的方法。例如,在一些实施例中,获取POI状态信息或建立POI状态识别模型的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908。

[0191] 在一些实施例中,计算机程序的部分或者全部可以经由ROM 802和/或通信单元909而被载入和/或安装到设备900上。当计算机程序加载到RAM 903并由计算单元901执行时,可以执行上文描述的获取POI状态信息或建立POI状态识别模型的方法的一个或多个步骤。备选地,在其他实施例中,计算单元901可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行获取POI状态信息或建立POI状态识别模型的方法。

[0192] 此处描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、负载可编程逻辑设备 (CPLD)、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。

[0193] 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30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服务器上执行。

[0194] 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、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0195] 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(阴极射线管)或者LCD(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。

[0196] 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:局域网(LAN)、广域网(WAN)和互联网。

[0197] 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。

[0198] 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公开的技术方案所期望的结果,本文在此不进行限制。

[0199] 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做的修改、等同替换和改进等,均应包含在本公开保护范围之内。

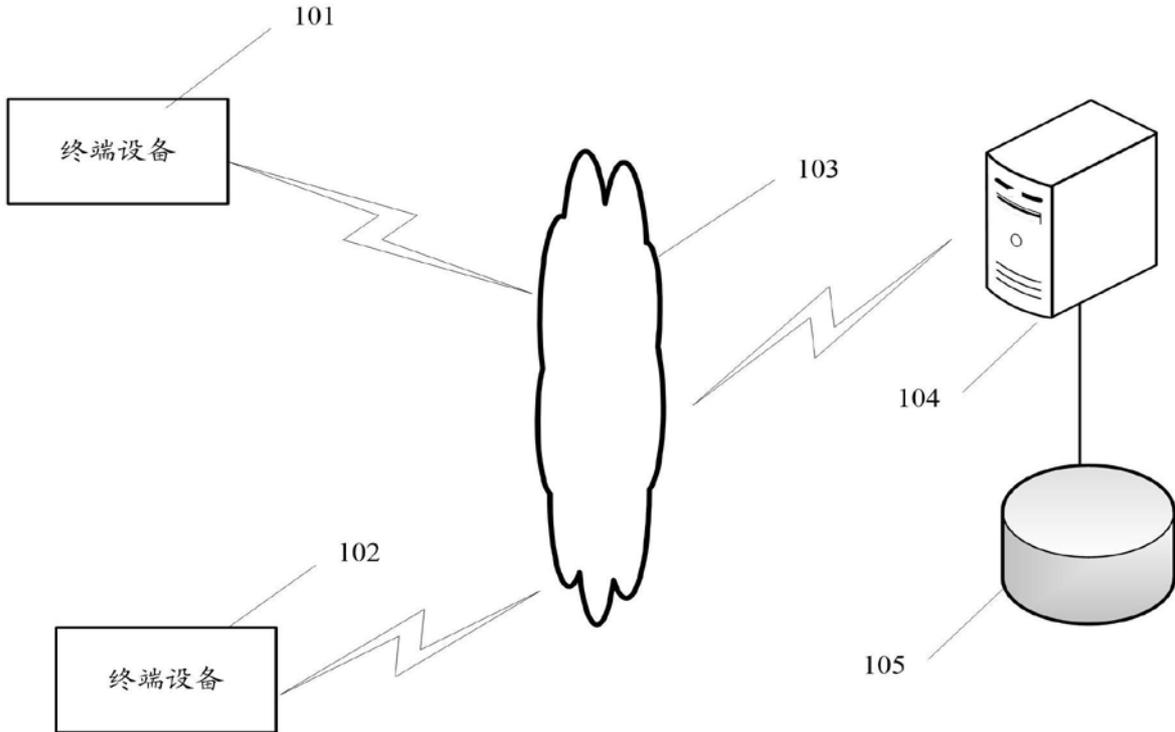


图1

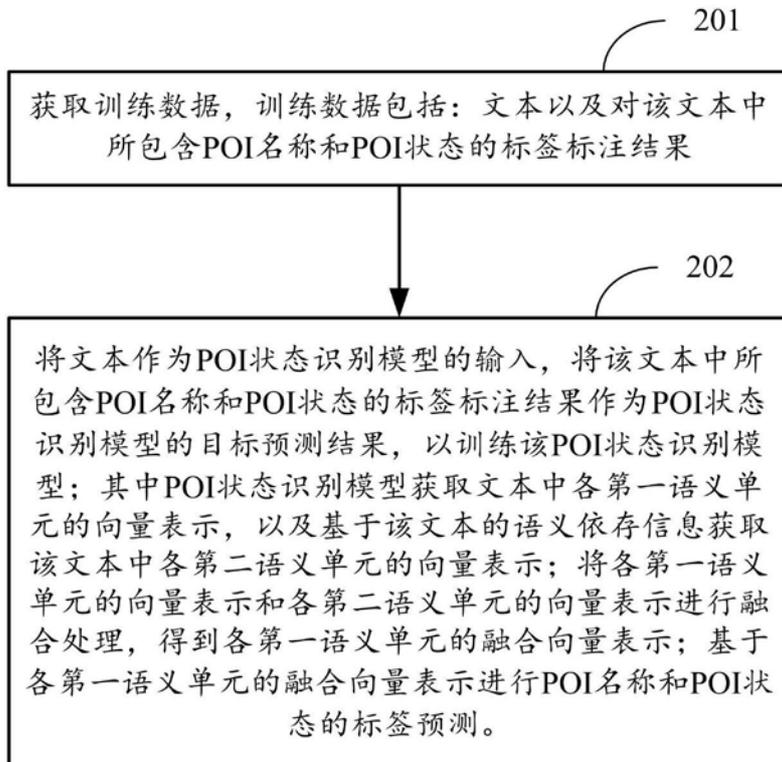


图2

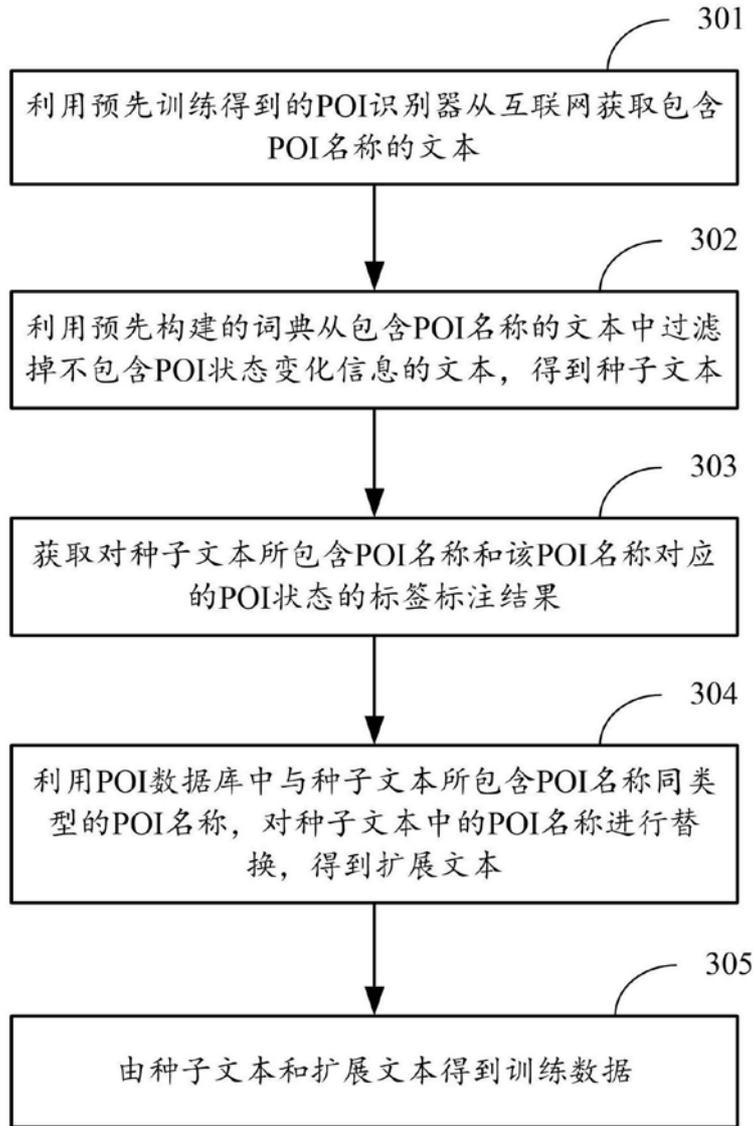


图3

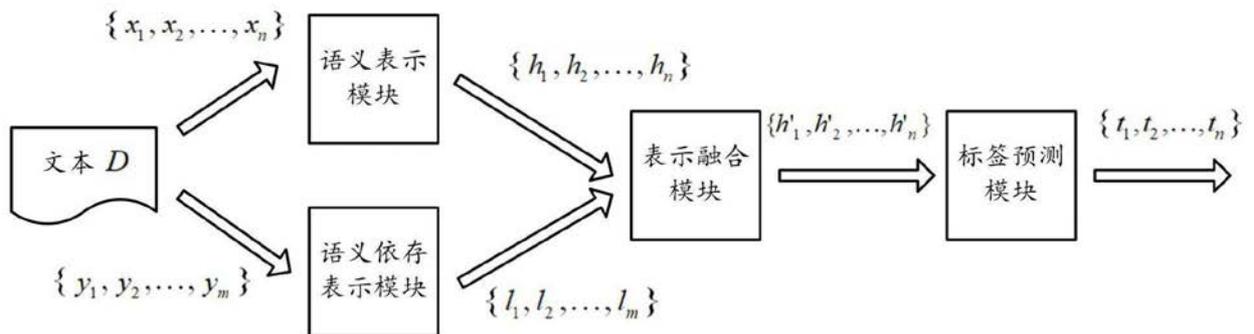


图4

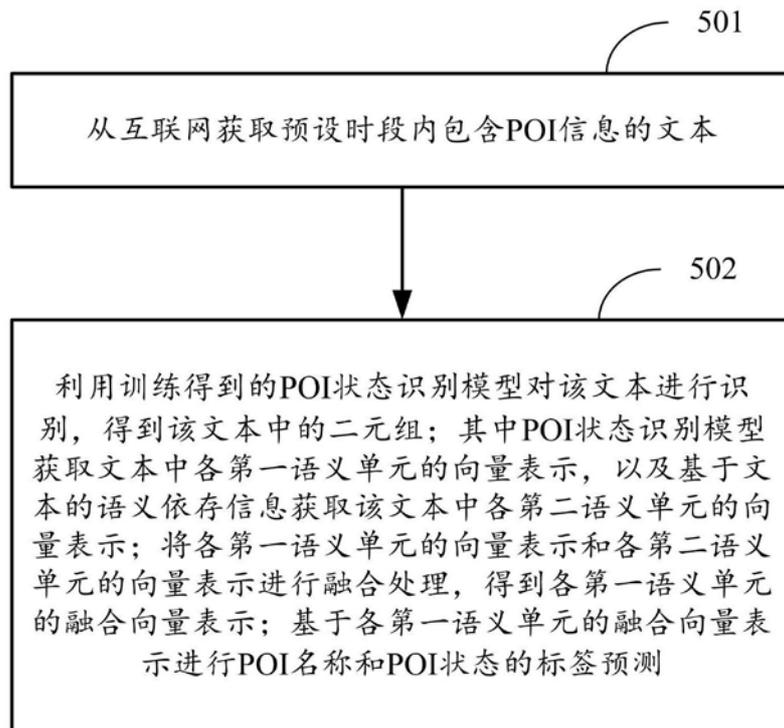


图5

输入文本	预测的标签
继	O
坪	B-NEW
山	I-NEW
图	I-NEW
书	I-NEW
馆	I-NEW
开	O
业	O
之	O
后	O
,	O
坪	B-NEW
山	I-NEW
艺	I-NEW
术	I-NEW
博	I-NEW
物	I-NEW
馆	I-NEW
也	O
在	O
3	O
月	O
30	O
日	O
正	O
式	O
开	O
业	O
,	O
其	O
在	O
坪	B-NONE
山	I-NONE
中	I-NONE
心	I-NONE
公	I-NONE
园	I-NONE
内	O

图6

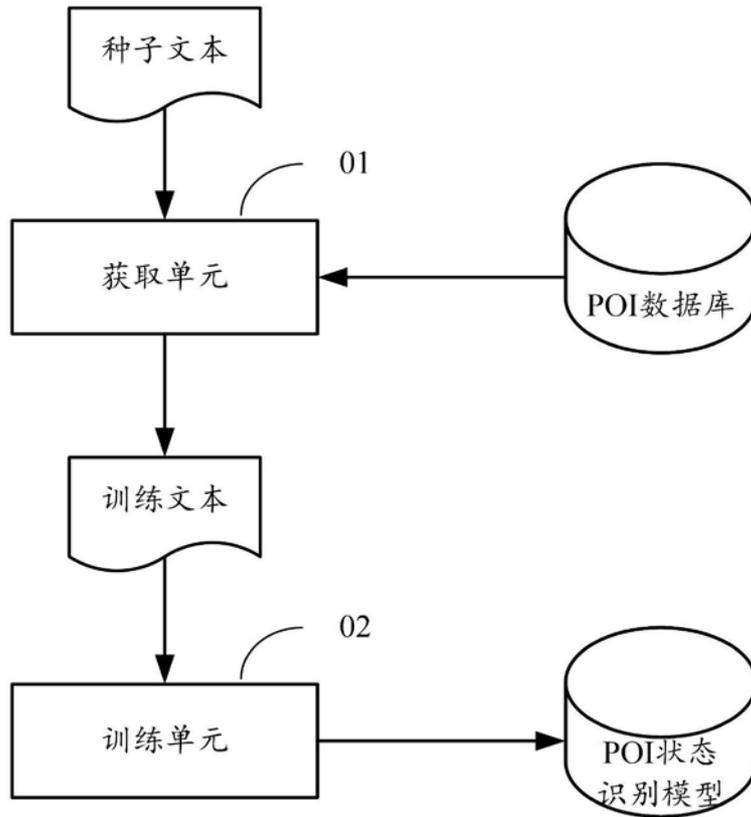


图7

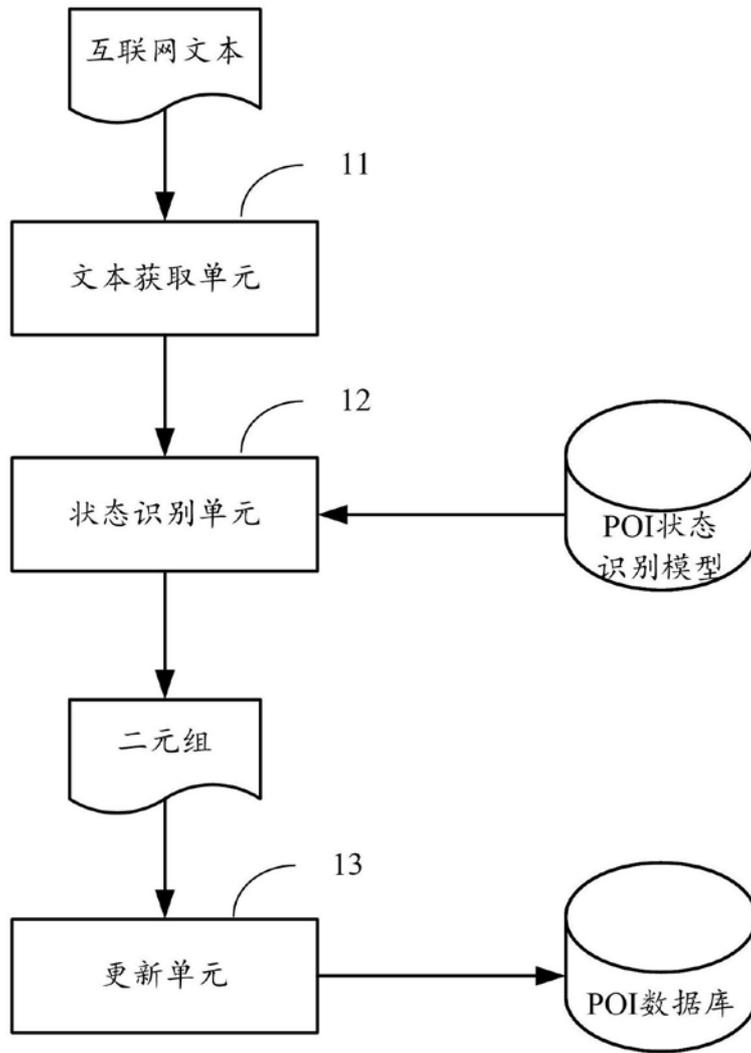


图8

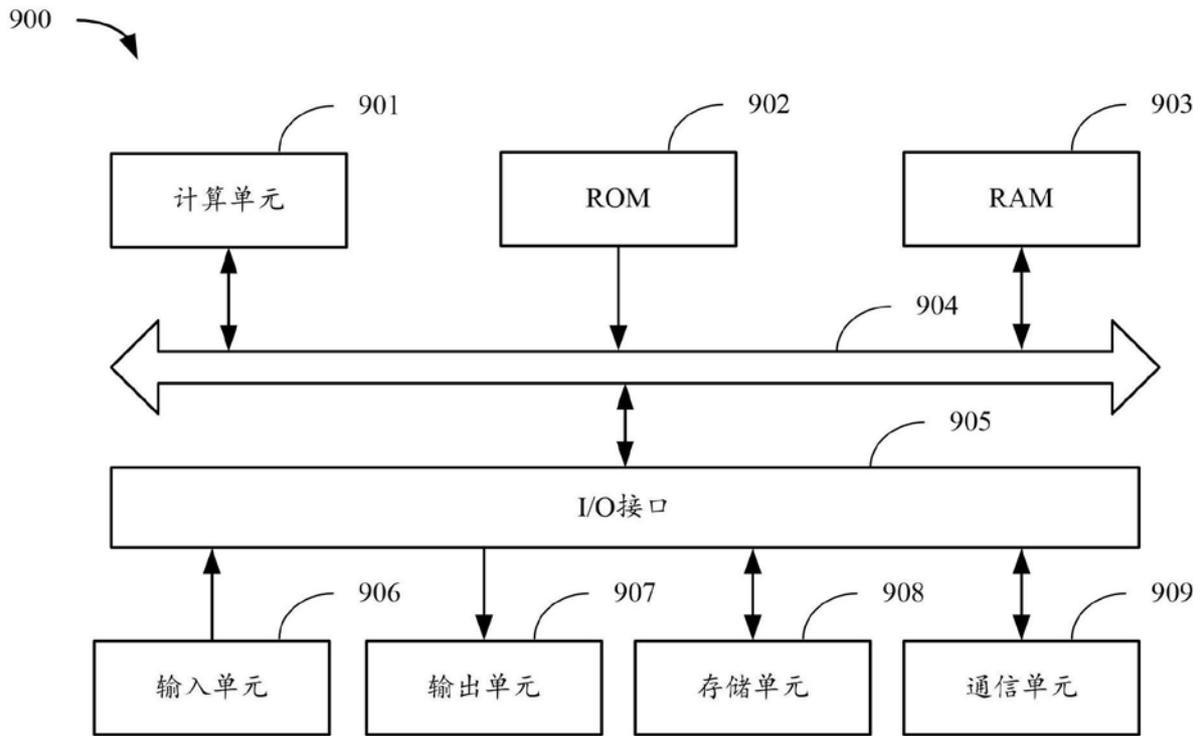


图9