



(12) 发明专利申请

(10) 申请公布号 CN 117743543 A

(43) 申请公布日 2024. 03. 22

(21) 申请号 202311764829.4

(22) 申请日 2023.12.20

(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  
百度大厦二层

(72) 发明人 陈禹燊 韩光耀 苏磊 岳洪达

王艺 安云静 常琳

(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 金星玉

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G06F 40/35 (2020.01)

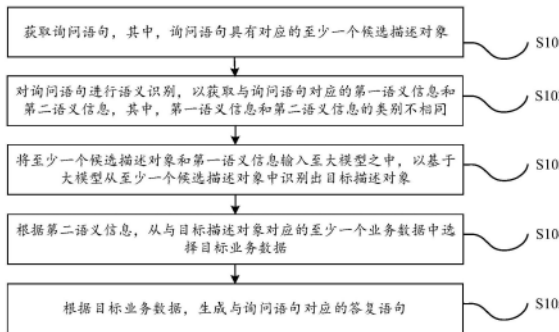
权利要求书3页 说明书16页 附图5页

(54) 发明名称

基于大语言模型的语句生成方法、装置及电子设备

(57) 摘要

本公开提供了一种基于大语言模型的语句生成方法、装置及电子设备,涉及计算机技术领域,具体为深度学习、自然语言处理、大语言模型等人工智能技术领域,具体实施方案为:获取询问语句,其中,询问语句具有对应的至少一个候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句。



1. 一种语句生成方法,包括:

获取询问语句,其中,所述询问语句具有对应的至少一个候选描述对象;

对所述询问语句进行语义识别,以获取与所述询问语句对应的第一语义信息和第二语义信息,其中,所述第一语义信息和所述第二语义信息的类别不相同;

将所述至少一个候选描述对象和所述第一语义信息输入至大模型之中,以基于所述大模型从所述至少一个候选描述对象中识别出目标描述对象;

根据所述第二语义信息,从与所述目标描述对象对应的至少一个业务数据中选择目标业务数据;

根据所述目标业务数据,生成与所述询问语句对应的答复语句。

2. 根据权利要求1所述的方法,其中,所述第一语义信息是地理区域维度的语义信息,所述第二语义信息是与所述候选描述对象相关的统计频率维度的语义信息,和/或与所述候选描述对象相关的统计时间维度的语义信息。

3. 根据权利要求1所述的方法,其中,所述至少一个候选描述对象,是基于以下方式确定得到:

获取所述询问语句对应的第一语义向量和第一关键词;

从预设检索库中确定与所述第一语义向量匹配的第一描述对象,及与所述第一关键词匹配的第二描述对象;

根据所述第一描述对象和所述第二描述对象,确定所述候选描述对象。

4. 根据权利要求3所述的方法,其中,所述预设检索库是基于以下方式构建得到:

获取业务数据库,其中,所述业务数据库中包括多条初始业务数据;

确定所述初始业务数据的业务描述对象,以及与所述业务描述对象对应的数据记录类型和数据记录频率;

分别基于所述业务描述对象、所述数据记录类型和所述数据记录频率对所述多条初始业务数据进行聚合处理,以得到聚合业务数据,其中,所述聚合业务数据包含所述与所述目标描述对象对应的至少一个业务数据;

基于语义向量模型对所述业务描述对象进行向量化处理,以得到第二语义向量;

基于所述聚合业务数据和所述第二语义向量,构建所述预设检索库。

5. 根据权利要求4所述的方法,其中,所述基于语义向量模型对所述业务描述对象进行向量化处理,以得到第二语义向量,包括:

基于单语言语义向量模型对所述业务描述对象进行向量化处理,得到所述第二语义向量;和/或

基于多语言语义向量模型对所述业务描述对象进行向量化处理,得到所述第二语义向量。

6. 根据权利要求3所述的方法,其中,所述获取所述询问语句对应的第一语义向量,包括:

基于单语言语义向量模型对所述询问语句进行向量化处理,得到所述第一语义向量;和/或

基于多语言语义向量模型对所述询问语句进行向量化处理,得到所述第一语义向量。

7. 根据权利要求3所述的方法,所述预设检索库包括:多个第二语义向量,其中,所述从

预设检索库中确定与所述第一语义向量匹配的第一描述对象,包括:

确定所述第一语义向量与所述预设检索库中所述第二语义向量的初始相似度;

根据所述初始相似度对多个所述第二语义向量进行排序;

根据排序在前的第一数量个所述第二语义向量,构建与所述第一语义向量对应的候选向量集合;

根据所述候选向量集合及所述候选向量集合的数量,确定所述第一描述对象。

8. 根据权利要求7所述的方法,其中,所述候选向量集合的数量是一个;其中,所述根据所述候选向量集合及所述候选向量集合的数量,确定所述第一描述对象,包括:

根据所述候选向量集合中所述第二语义向量的排序次序,对相应所述第二语义向量的所述初始相似度进行加权处理,以得到与所述第二语义向量对应的目标相似度;

根据所述目标相似度,确定所述第一描述对象。

9. 根据权利要求7所述的方法,其中,所述候选向量集合的数量是多个,不同所述候选向量集合分别是基于不同语义向量模型对所述询问语句进行向量化处理所得;其中,所述根据所述候选向量集合及所述候选向量集合的数量,确定所述第一描述对象,包括:

根据所述候选向量集合中所述第二语义向量的排序次序,对相应所述第二语义向量的所述初始相似度进行加权处理,以得到与所述第二语义向量对应的参考相似度;

确定与所述第二语义向量对应的多个所述参考相似度的和值作为目标相似度;

根据所述目标相似度,确定所述第一描述对象。

10. 根据权利要求8或9所述的方法,其中,所述预设检索库包括聚合业务数据,所述聚合业务数据包含多个业务描述对象,所述业务描述对象与所述第二语义向量相关联;其中,所述根据所述目标相似度,确定所述第一描述对象,包括:

根据所述目标相似度对多个所述第二语义向量进行排序;

将排序在前的第二数量个所述第二语义向量对应的所述业务描述对象作为所述第一描述对象。

11. 根据权利要求3所述的方法,其中,所述根据所述第一描述对象和所述第二描述对象,确定所述候选描述对象,包括:

基于预设评分模型确定所述第一描述对象的第一评分值,和所述第二描述对象的第二评分值,其中,所述第一评分值用于描述对应所述第一描述对象与所述询问语句的匹配程度,所述第二评分值用于描述对应所述第二描述对象与所述询问语句的匹配程度;

根据所述第一评分值和所述第二评分值,对所述第一描述对象和所述第二描述对象进行排序;

将排序在前的第三数量个所述第一描述对象和/或所述第二描述对象作为所述候选描述对象。

12. 根据权利要求11所述的方法,其中,所述预设评分模型,是基于以下方式训练得到:

获取样本数据,其中,所述样本数据包括:样本询问语句,以及与所述样本询问语句对应的多个样本答复语句;

确定所述询问语句对应的业务场景;

根据所述业务场景对所述样本答复语句进行标注处理,以从所述多个样本答复语句中确定正样本答复语句和负样本答复语句;

基于所述样本询问语句、所述正样本答复语句和所述负样本答复语句对初始评分模型进行训练,以得到所述预设评分模型。

13. 一种语句生成装置,包括:

获取模块,用于获取询问语句,其中,所述询问语句具有对应的至少一个候选描述对象;

第一识别模块,用于对所述询问语句进行语义识别,以获取与所述询问语句对应的第一语义信息和第二语义信息,其中,所述第一语义信息和所述第二语义信息的类别不相同;

第二识别模块,用于将所述至少一个候选描述对象和所述第一语义信息输入至大模型之中,以基于所述大模型从所述至少一个候选描述对象中识别出目标描述对象;

第三识别模块,用于根据所述第二语义信息,从与所述目标描述对象对应的至少一个业务数据中选择目标业务数据;

生成模块,用于根据所述目标业务数据,生成与所述询问语句对应的答复语句。

14. 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1-12中任一项所述的方法。

15. 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1-12中任一项所述的方法。

16. 一种计算机程序产品,包括计算机指令,所述计算机指令被处理器执行时实现权利要求1-12中任一项所述方法的步骤。

## 基于大语言模型的语句生成方法、装置及电子设备

### 技术领域

[0001] 本公开涉及计算机技术领域,具体为深度学习、自然语言处理、大语言模型等人工智能技术领域,具体涉及一种基于大语言模型的语句生成方法、装置及电子设备。

### 背景技术

[0002] 自然语言处理(Natural Language Processing,NLP)是计算机科学领域与人工智能领域中的一个重要方向,是指用计算机对自然语言的形、音、义等信息进行处理,即对字、词、句、篇章的输入、输出、识别、分析、理解、生成等的操作和加工,它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理的具体表现形式包括机器翻译、文本摘要、文本分类、文本校对、信息抽取、智能问答等。在实现智能问答时,如何准确地生成与询问语句相适配的答复语句成为重点关注的问题。

### 发明内容

- [0003] 本公开提供了一种基于大语言模型的语句生成方法、装置、电子设备及存储介质。
- [0004] 根据本公开的第一方面,提供了一种语句生成方法,包括:
- [0005] 获取询问语句,其中,所述询问语句具有对应的至少一个候选描述对象;
- [0006] 对所述询问语句进行语义识别,以获取与所述询问语句对应的第一语义信息和第二语义信息,其中,所述第一语义信息和所述第二语义信息的类别不相同;
- [0007] 将所述至少一个候选描述对象和所述第一语义信息输入至大模型之中,以基于所述大模型从所述至少一个候选描述对象中识别出目标描述对象;
- [0008] 根据所述第二语义信息,从与所述目标描述对象对应的至少一个业务数据中选择目标业务数据;
- [0009] 根据所述目标业务数据,生成与所述询问语句对应的答复语句。
- [0010] 根据本公开的第二方面,提供了一种语句生成装置,包括:
- [0011] 获取模块,用于获取询问语句,其中,所述询问语句具有对应的至少一个候选描述对象;
- [0012] 第一识别模块,用于对所述询问语句进行语义识别,以获取与所述询问语句对应的第一语义信息和第二语义信息,其中,所述第一语义信息和所述第二语义信息的类别不相同;
- [0013] 第二识别模块,用于将所述至少一个候选描述对象和所述第一语义信息输入至大模型之中,以基于所述大模型从所述至少一个候选描述对象中识别出目标描述对象;
- [0014] 第三识别模块,用于根据所述第二语义信息,从与所述目标描述对象对应的至少一个业务数据中选择目标业务数据;
- [0015] 生成模块,用于根据所述目标业务数据,生成与所述询问语句对应的答复语句。
- [0016] 根据本公开的第三方面,提供了一种电子设备,包括:
- [0017] 至少一个处理器;以及

[0018] 与至少一个处理器通信连接的存储器;其中,

[0019] 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如第一方面的语句生成方法。

[0020] 根据本公开第四方面,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,计算机指令用于使计算机执行如第一方面的语句生成方法。

[0021] 根据本公开的第五方面,提供了一种计算机程序产品,包括计算机指令,计算机指令在被处理器执行时实现如第一方面的语句生成方法的步骤。

[0022] 本公开提供的语句生成方法、装置、电子设备及存储介质,存在如下有益效果:

[0023] 本公开实施例中,通过获取询问语句,其中,询问语句具有对应的至少一个候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句,由此,可以识别询问语句不同类型的语义信息,从而基于第一语义信息和大模型从至少一个候选描述对象中识别出目标描述对象,基于第二语义信息从与目标描述对象对应的至少一个业务数据中选择目标业务数据,并目标业务数据生成答复语句,从而有效提升所得答复语句与询问语句的适配性,提升答复效果。

[0024] 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

[0025] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0026] 图1是根据本公开一实施例提供的一种语句生成方法的流程示意图;

[0027] 图1a是根据本公开提出的基于大模型识别询问语句对应第一语义信息的示意图;

[0028] 图2是根据本公开另一实施例提供的一种语句生成方法的流程示意图;

[0029] 图3是根据本公开又一实施例提供的一种语句生成方法的流程示意图;

[0030] 图4是根据本公开提出的智能投研指标问答算法流程图;

[0031] 图5是根据本公开一实施例提供的一种语句生成装置的结构示意图;

[0032] 图6是用来实现本公开实施例的语句生成方法的电子设备的框图。

## 具体实施方式

[0033] 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。

[0034] 本公开实施例涉及深度学习、自然语言处理、大语言模型等人工智能技术领域。

[0035] 人工智能(Artificial Intelligence),英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

[0036] 深度学习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。深度学习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。

[0037] 自然语言处理(Natural Language Processing,NLP)是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。

[0038] 大语言模型是指基于大规模语料库训练得到的具有大量参数和复杂结构的机器学习模型。这些模型可以应用于处理大规模的数据和复杂的问题。

[0039] 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,均符合相关法律法规的规定,且不违背公序良俗。

[0040] 下面参考附图描述本公开实施例的语句生成方法、装置、电子设备及存储介质。

[0041] 其中,需要说明的是,本实施例的语句生成方法的执行主体为语句生成装置,该装置可以由软件和/或硬件的方式实现,该装置可以配置在电子设备中,电子设备可以包括但不限于终端、服务器端等。

[0042] 图1是根据本公开一实施例提供的一种语句生成方法的流程示意图。

[0043] 如图1所示,该语句生成方法包括:

[0044] S101:获取询问语句,其中,询问语句具有对应的至少一个候选描述对象。

[0045] 其中,询问语句,可以是指用户对本公开实施例的执行主体所提出的用于询问指定信息的语句。

[0046] 其中,描述对象,可以是指语句对应描述的对象主体,例如可以是指标数据中的核心指标。举例而言,在“汽车和汽车底盘进口数量:累计值:月”指标数据中,“汽车和汽车底盘进口数量”就是对应的描述对象,即核心指标。而候选描述对象,则是指可能作为询问语句对应主体的描述对象。

[0047] 本公开实施例中,在获取询问语句对应的候选描述对象时,可以是基于预训练的描述对象识别模型,或者,还可以是基于第三方描述对象识别装置,对此不做限制。

[0048] S102:对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同。

[0049] 其中,语义信息,可以被用于描述对应语句的含义、意图和关系等。而第一语义信息和第二语义信息,则是针对询问语句进行语义识别所获取的不同类型的语义信息。

[0050] 即是说,本公开实施例中可以分别针对不同维度对询问语句进行语义识别,以获取第一语义信息和第二语义信息,从而为后续确定目标描述对象和目标业务数据提供可靠的判断依据。

[0051] 可以理解的是,本公开实施例中,第一语义信息和第二语义信息对应的指示内容可以根据应用场景进行灵活配置,对此不做限制。

[0052] 可选的,一些实施例中,第一语义信息是地理区域维度的语义信息,第二语义信息是与候选描述对象相关的统计频率维度的语义信息,和/或与候选描述对象相关的统计时间维度的语义信息。由此,第一语义信息可以在地理区域维度为筛选目标描述对象提供准确的筛选依据,而第二语义信息则可以在统计频率维度和/或统计时间维度为目标业务数据的选择过程中提供准确的取值范围。

[0053] 本公开实施例中,当第二语义信息是与候选描述对象相关的统计频率维度的语义信息时,可以根据第二语义信息,过滤掉其他无用统计频率的业务数据,提升所得目标业务数据的精确率,优化用户体验。

[0054] 举例而言,本公开实施例中可以基于大模型识别出与询问语句对应的第一语义信息和第二语义信息,当第一语义信息是地理区域维度的语义信息时,如图1a所示,图1a是根据本公开提出的基于大模型识别询问语句对应第一语义信息的示意图。其中,该第一语义信息可以包括行政区级别(如“国家”,“省级行政区”,“地级行政区”,“县级行政区”,“乡级行政区”等),以及地区范围(如对应地域粒度的部分地区,或者,对应地域粒度的地区列表),当询问语句中不涉及地理区域粒度或地理区域范围时,第一语义信息中对应内容为空,如下:

[0055] (1) 行政区级别:[“国家”,“省级行政区”,“地级行政区”,“县级行政区”,“乡级行政区”,null],其中null表示query中没有明确提到地理位置;

[0056] (2) 地区范围:[null,“all”,[“xxx”]]。其中null表示没有提到具体的地区;“all”表示提到了该行政区级别的所有地区;[“xxx”]是具体地区的列表。

[0057] 对第一语义信息进行识别答复时,举例如下

[0058] Query:2022年12月A所在哪个国家的会员数最多?回答:{"行政区级别":"国家",“范围”:"all"}

[0059] Query:今年a城市和b城市二手房销售均价的变化情况如何?回答:{"行政区级别":"地级行政区",“范围”:[“a城市”,“b城市”]}

[0060] 该第一语义信息可以为知识,添加到大模型的提示词prompt中作为知识增强。

[0061] 本公开实施例中在基于第一语义信息进行知识增强时,可以从“行政区级别”和“地区范围”的维度分别进行知识增强,以有效解决最终召回中带地理位置业务数据的漏召、误召的情况。

[0062] 本公开实施例中,统计频率维度的语义信息例如可以包括:年、季、月、旬、周、日等。

[0063] 对于召回的同一个指标,它可能会有多个统计频率,如:[“股票成交笔数:科创板:A所:当期值:年”,“股票成交笔数:科创板:A所:当期值:月”,“股票成交笔数:科创板:A所:当期值:周”,“股票成交笔数:科创板:A所:当期值:日”],当query是:“2021年7月1日A所科创板股票的成交数是多少?”此时需要拿到日频的指标:“股票成交笔数:科创板:A所:当期值:日”,而不是其他统计频率的指标。

[0064] 本公开实施例中,可以使用大模型,拼接query输入特定prompt,完成统计频率的分析,将query中的统计频率分为7个类别:年、季、月、旬、周、日、无。

[0065] 【few-shot示例】

[0066] 1、query:2022年全年国内生产总值(GDP)是多少?统计频率:年

[0067] 2、query:2022年Q2全国城镇调查失业率是多少?统计频率:季

[0068] 3、query:2019年最后一个月全国房地产开发投资累计同比是多少?统计频率:月

[0069] 完成分类后,对于召回的所有指标,可以进一步过滤,如:query的统计频率分类为“日”,但是某个召回的指标统计频率为“年”,该指标会被过滤掉。query的统计频率分类为“无”,那么所有的统计频率都可以接受,所有召回的指标会全部保留。



[0070] 本公开实施例中,询问语句query中提到的日期是用户用自然语言询问的日期,比如“B公司最近一个Q的净利润表现如何?”这样的query,在数据库中,“B公司净利润:当期值:季”这个指标收录的数据是从“2005-09-30”~“2023-09-30”的净利润数据,本发明的统计范围识别是将“最近一个Q”映射到“2023-07-01”~“2023-09-30”,并从指标库中拿到第三季度期间产生的净利润数值。

[0071] 具体实施方式如下:

[0072] 1) prompt中有动态变化的日期部分,如周的相对值,年的相对值,通过计算当前系统的日期current\_date,然后通过日期回溯函数找到对应的起始日期;

[0073] 2) prompt中添加few-shot,比如:对于2007年的Q3,是指:2007年07月01日~2007年09月30日;2022年初-2022年末是指:2022年01月01日~2022年12月31日...;

[0074] 3) 以上动态的few-shot和query结合,作为输入送入大模型推理,可以得到日期范围的推理结果;

[0075] 4) 对推理结果使用正则表达式提取开始日期和结束日期,并分别结构化为xxxx-xx-xx的格式;

[0076] 5) 使用格式化的日期范围可以从指标数据库中匹配得到对应日期范围内的指标数值。

[0077] S103:将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象。

[0078] 其中,目标描述对象,是指基于大模型所确定的适用于询问语句的描述对象。

[0079] 举例而言,本公开实施例中,可以将至少一个候选描述对象和第一语义信息输入至大模型之中,以得到大模型输出的至少一个参考描述对象,如果至少一个候选描述对象中包括该参考描述对象,则将该参考描述对象作为目标描述对象,由此,可以有效保证所得目标描述对象的准确性。

[0080] S104:根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据。

[0081] 其中,目标业务数据,是指根据第二语义信息对标描述对象的多个业务数据进行筛选后所得到的业务数据。

[0082] 可以理解的是,目标描述对象可能具有不同维度的业务数据,而询问语句则可能会指定获取对应维度的数据,因此,本公开实施例中可以根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,从而有效减少所得目标业务数据中的冗余数据。

[0083] 本公开实施例中,当根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,可以从统计频率维度和统计时间维度准确地获取对应的目标数据。

[0084] S105:根据目标业务数据,生成与询问语句对应的答复语句。

[0085] 其中,答复语句,是指用于答复上述询问语句所询问内容的语句。

[0086] 本公开实施例中,在根据目标业务数据,生成与询问语句对应的答复语句时,可以是将目标业务数据转换为适用于大模型处理的数据格式(例如markdown格式),然后将其输入至大模型中,以得到对应的答复语句。

[0087] 本公开实施例中,通过获取询问语句,其中,询问语句具有对应的至少一个候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句,由此,可以识别询问语句不同类型的语义信息,从而基于第一语义信息和大模型从至少一个候选描述对象中识别出目标描述对象,基于第二语义信息从与目标描述对象对应的至少一个业务数据中选择目标业务数据,并目标业务数据生成答复语句,从而有效提升所得答复语句与询问语句的适配性,提升答复效果。

[0088] 图2是根据本公开又一实施例提供的一种语句生成方法的流程示意图;

[0089] 如图2所示,该语句生成方法包括:

[0090] S201:获取询问语句。

[0091] S201的描述说明具体可以参见上述实施例,在此不再赘述。

[0092] S202:获取询问语句对应的第一语义向量和第一关键词。

[0093] 其中,第一语义向量,是指对询问语句进行语义向量化处理后所得到的向量。

[0094] 其中,第一关键词,是指对询问语句进行关键词提取之后所得到的关键词。举例而言,本公开实施例中可以对询问语句进行分词处理,以确定多个候选词,然后确定多个候选词的词频,将词频最高值对应的候选词作为第一关键词,或者,还可以基于任一任意可能的方法确定第一关键词,对此不做限制。

[0095] 可选的,一些实施例中在获取询问语句对应的第一语义向量时,可以是基于单语言语义向量模型对询问语句进行向量化处理,得到第一语义向量,和/或基于多语言语义向量模型对询问语句进行向量化处理,得到第一语义向量,由此,可以在询问语句的向量化处理过程中灵活使用单语言语义向量模型和/或多语言语义向量模型进行向量化处理,使第一语义向量的获取过程适用于询问语句个性化的文本语言场景,以保证所得第一语义向量的指示效果。

[0096] 其中,语义向量模型,是一种自然语言处理技术,主要通过将单词或短语映射到高维向量空间中,来实现文本的语义理解和处理。

[0097] 其中,单语言语义向量模型,可以是指针对单种语言文本进行向量化的语义向量化模型,例如可以是指针对汉语、英语等语言类型配置的语义向量模型。

[0098] 其中,多语言语义向量模型,可以是指针对多种语言混合文本进行向量化的语义向量化模型,例如可以是指针对汉语和英语等混合语言文本所配置的语义向量模型。

[0099] 本公开实施例中,单语言语义向量模型和多语言语义向量模型的数量可以根据应用场景进行灵活配置,以适用于个性化的应用场景,得到不同数量的第一语义向量。

[0100] 可以理解的是,每个独立预训练模型的语料、模型结构等都会有差异,对于同样是1024维large模型的稠密向量,还是768维base模型的稠密向量,不同的模型会有语义空间上的差异,因此本发明可以采用多种不同的模型,从而从不同语义方向提升召回率。当加入多语言向量化模型时,对于中英混合指标,如:“cpi:食品:环比:月”、“PIRM:建筑材料类:当期同比:月”这些指标,可以提升召回率。

[0101] 举例而言,本公开实施例中可以使用多个独立的Embedding模型对询问语句进行多路向量化。

[0102] S203:从预设检索库中确定与第一语义向量匹配的第一描述对象,及与第一关键词匹配的第二描述对象。

[0103] 其中,预设检索库,可以是预先配置的用于对第一语义向量和第一关键词进行检索的检索库。

[0104] 其中,第一描述对象,可以是指基于第一语义向量和预设检索库进行向量化检索后所得到的描述对象。而第二描述对象,则可以是指基于第一关键词和预设检索库进行关键词检索后所得到的描述对象。例如可以构建ES(ElasticSearch)检索库作为预设检索库,并基于ES检索库的term检索,召回分数最高的多个业务描述对象作为第二描述对象。

[0105] 即是说,本公开实施例中,可以分别基于向量化检索和关键词检索从预设检索库中确定第一描述对象和第二描述对象,从而为后续确定候选描述对象提供可靠的数据支持。

[0106] 可以理解的是,基于单种途径所获取的描述对象可能存在较大误差,而本公开实施例中分别从预设检索库中确定与第一语义向量匹配的第一描述对象,及与第一关键词匹配的第二描述对象,可以有效提升候选描述对象确定过程的鲁棒性。

[0107] 可选的,一些实施例中,可以获取业务数据库,其中,业务数据库中包括多条初始业务数据,确定初始业务数据的业务描述对象,以及与业务描述对象对应的数据记录类型和数据记录频率,分别基于业务描述对象、数据记录类型和数据记录频率对多条初始业务数据进行聚合处理,以得到聚合业务数据,其中,聚合业务数据包含与目标描述对象对应的至少一个业务数据,基于语义向量模型对业务描述对象进行向量化处理,以得到第二语义向量,基于聚合业务数据和第二语义向量,构建预设检索库,由此,可以基于业务数据库为答复语句的生成过程提供可靠的业务数据支持,业务数据库中可能包含多个不同的业务描述对象,每个业务描述对象又可能存在不同数据记录类型和数据记录频率的业务数据,因此可能存在较多的冗余数据,因此,可以分别基于业务描述对象、数据记录类型和数据记录频率对多条初始业务数据进行聚合处理,以有效减少所得预设检索库中的冗余数据,提升检索效率。

[0108] 其中,业务数据库,可以是包含业务描述对象各种业务数据的数据库,例如可以是金融数据指标库,如恒生聚源的指标库(Hermes Index)。

[0109] 举例而言,本公开实施例中在构建预设检索库时,可以是基于如下步骤:

[0110] (1) 核心指标(即上述业务描述对象)聚合

[0111] 对于指标数据“汽车和汽车底盘进口数量:累计值:月”,“汽车和汽车底盘进口数量”就是它的核心指标,可以对此清洗去重、聚合。

[0112] (2) 统计方式(即上述数据记录类型)聚合

[0113] 常见统计方式有:当期值、累计值、当期同比、累计同比、环比、环比增减量等。

[0114] 对于指标数据“汽车和汽车底盘进口数量:累计值:月”、“汽车和汽车底盘进口数量:当期值:月”,“累计值”和“当期值”就是它的统计方式,可以对此聚合,作为“统计方式”字段。

[0115] (3) 统计频率(即上述数据记录频率)聚合

[0116] 常见统计频率有:年、季、月、旬、周、日

[0117] 对于指标数据“汽车和汽车底盘进口数量:累计值:日”、“汽车和汽车底盘进口数量:累计值:月”,“日”和“月”就是它的统计频率,可以对此聚合,作为“统计频率字段”。

[0118] (4)Embedding构建指标ES(ElasticSearch)检索库(即上述预设检索库)

[0119] 对于从(1)到(3)聚合得到的指标库,具有指标ID、核心指标、统计方式、统计频率、指标数据等字段,可以对核心指标使用7个独立的Embedding模型,进行多路向量化,总计是6路中文向量化模型,1路多语言向量化模型,兼顾中文指标和中英文混合指标的向量化,然后构建ES检索库。

[0120] 上述步骤中聚合的目的是让核心指标、统计方式以及统计频率区分开,对于一个query,核心指标使用向量和关键词召回,统计方式以及统计频率分别通过意图识别来召回,一方面可以缩减核心指标的数量,让系统检索效率更高;另一方面统计方式和统计频率一般比较固定,可以枚举,让这两者从指标中剥离出来,使用专有模型进行意图判定,可以提升整体指标问答的准确率;对于指标问答,采用上述拆分式检索方式优于整体检索;相比于使用研报等材料挖掘指标的方法,本发明直接可以使用现有专业金融指标数据库进行处理,可以较大幅度地提升数据的丰富度和准确性。

[0121] 可选的,一些实施例中,在基于语义向量模型对业务描述对象进行向量化处理,以得到第二语义向量时,可以是基于单语言语义向量模型对业务描述对象进行向量化处理,得到第二语义向量,和/或基于多语言语义向量模型对业务描述对象进行向量化处理,得到第二语义向量,由此,可以在业务描述对象的向量化处理过程中灵活使用单语言语义向量模型和/或多语言语义向量模型进行向量化处理,以保证向量化处理过程适配于不同语言类型的业务描述对象,保证所得第二语义向量对业务描述对象的指示效果。

[0122] 可以理解的是,本公开实施例中可以基于一个或多个语义向量模型对业务描述对象进行向量化处理,以得到对应一个或多个第二语义向量,对此不做限制。

[0123] S204:根据第一描述对象和第二描述对象,确定候选描述对象。

[0124] 本公开实施例中,在根据第一描述对象和第二描述对象,确定候选描述对象时,可以是基于第三方描述对象确定装置从第一描述对象和第二描述对象中确定候选描述对象,或者,还可以基于其他任意可能的方法根据第一描述对象和第二描述对象,确定候选描述对象,对此不做限制。

[0125] 即是说,本公开实施例中,可以获取询问语句对应的第一语义向量和第一关键词,从预设检索库中确定与第一语义向量匹配的第一描述对象,及与第一关键词匹配的第二描述对象,根据第一描述对象和第二描述对象,确定候选描述对象,由此,可以基于第一语义向量在预设检索库中进行向量化检索后得到第一描述对象,基于第一关键词在预设检索库中进行关键词检索后得到第二描述对象,由于第一描述对象和第二描述对象分别是基于不同的检索逻辑得到的,当结合第一描述对象和第二描述对象确定候选描述对象时,可以有效降低单个检索逻辑所产生的误差影响,保证所得候选描述对象的可靠性。

[0126] 可选的,一些实施例中,在根据第一描述对象和第二描述对象,确定候选描述对象时,可以是基于预设评分模型确定第一描述对象的第一评分值,和第二描述对象的第二评分值,其中,第一评分值用于描述对应第一描述对象与询问语句的匹配程度,第二评分值用于描述对应第二描述对象与询问语句的匹配程度,根据第一评分值和第二评分值,对第一

描述对象和第二描述对象进行排序,将排序在前的第三数量个第一描述对象和/或第二描述对象作为候选描述对象,由于第一描述对象和第二描述对象分别是基于不同的检索逻辑得到的,因此第一描述对象和第二描述对象对应初始的评分值不能直接用于排序对比,当基于预设评分模型确定第一描述对象的第一评分值,和第二描述对象的第二评分值时,可以实现对第一描述对象和第二描述对象的重打分,从而为第一描述对象和第二描述对象的排序过程提供参考信息,并将排序在前的第三数量个第一描述对象和/或第二描述对象作为候选描述对象,以保证所得候选描述对象的准确性。

[0127] 其中,预设评分模型,是指预先训练的用于对第一描述对象和第二描述对象进行再次评分的模型。例如可以采用bge-rank-base模型作为重新打分的预设评分模型。

[0128] 可选的,一些实施例中,可以获取样本数据,其中,样本数据包括:样本询问语句,以及与样本询问语句对应的多个样本答复语句,确定询问语句对应的业务场景,根据业务场景对样本答复语句进行标注处理,以从多个样本答复语句中确定正样本答复语句和负样本答复语句,基于样本询问语句、正样本答复语句和负样本答复语句对初始评分模型进行训练,以得到预设评分模型,由此,可以结合询问语句对应的业务场景对样本答复语句进行标注处理,从而保证所得模型训练数据与询问语句对应业务场景的适配性,然后基于样本询问语句、正样本答复语句和负样本答复语句对初始评分模型进行训练,可以有效提升所得预设评分模型输出的评分值对描述对象与询问语句之间匹配程度的指示准确性。

[0129] 本发明中在根据业务场景对样本答复语句进行标注处理时,可以在业务场景为宏观经济分析场景时,标注的pos样例对应选取偏宏观的指标;在业务场景为财务分析场景时,标注的pos样例会选取偏公司财务、收入、利润以及负债等方面的指标;在业务场景为市场分析场景时,标注的pos样例会选取股票市场上的各种指标,如股价、成交量、市盈率等,用于分析市场趋势和估值;在业务场景为风险分析场景时;标注的pos样例会选取能衡量投资风险的指标,如波动率、贝塔系数等,有助于投资者评估投资组合的风险水平。

[0130] 举例而言,本公开实施例中,预设评分模型的训练和推理流程包括如下步骤:

[0131] 1) 对于query (即上述样本询问语句) 和候选指标列表 (即上述多个样本答复语句), 转换成一个query和一个指标列表的文本对: “query ||| 指标”, 然后标签定为1、0;

[0132] 2) 使用doccano标注工具,对于pos标注为1,neg标注为0;

[0133] 3) 根据标注结果构造目标数据结构,该目标数据结构由query、pos指标和neg指标构成;

[0134] 4) 调整学习参数,使用目标数据结构中的数据,对预训练模型进行Fine-tuning,评估并发布推理模型,使用python FastAPI框架封装推理服务;

[0135] 5) 部署到服务端,加载模型并使用GPU推理。在线使用时,对于一个query以及输入的多个召回指标,可以给出query对于每个召回指标的打分。

[0136] S205:对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同。

[0137] S206:将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象。

[0138] S207:根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据。

[0139] S208:根据目标业务数据,生成与询问语句对应的答复语句。

[0140] S205-S208的描述说明具体可以参见上述实施例,在此不再赘述。

[0141] 本公开实施例中,通过获取询问语句,获取询问语句对应的第一语义向量和第一关键词,从预设检索库中确定与第一语义向量匹配的第一描述对象,及与第一关键词匹配的第二描述对象,根据第一描述对象和第二描述对象,确定候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句。由此,可以基于第一语义向量在预设检索库中进行向量化检索后得到第一描述对象,基于第一关键词在预设检索库中进行关键词检索后得到第二描述对象,由于第一描述对象和第二描述对象分别是基于不同的检索逻辑得到的,当结合第一描述对象和第二描述对象确定候选描述对象时,可以有效降低单个检索逻辑所产生的误差影响,保证所得候选描述对象的可靠性。

[0142] 图3是根据本公开又一实施例提供的一种语句生成方法的流程示意图;

[0143] 如图3所示,该语句生成方法包括:

[0144] S301:获取询问语句,并获取询问语句对应的第一语义向量和第一关键词。

[0145] S302:从预设检索库中确定与第一关键词匹配的第二描述对象。

[0146] S301和S302的描述说明具体可以参见上述实施例,在此不再赘述。

[0147] S303:确定第一语义向量与预设检索库中第二语义向量的初始相似度。

[0148] 其中,初始相似度,是指第一语义向量与第二语义向量之间的相似度。

[0149] 可以理解的是,语义向量之间的相似度与对应语句描述对象之间的相似度通常呈正相关,由此,本公开实施例中当确定第一语义向量与预设检索库中第二语义向量的初始相似度时,可以为第一语义向量的向量化检索过程提供可靠的参考信息。

[0150] S304:根据初始相似度对多个第二语义向量进行排序。

[0151] 即是说,本公开实施例中在确定第一语义向量与预设检索库中第二语义向量的初始相似度之后,可以基于初始相似度对多个第二语义向量进行相似度从高到低排序,以便于确定与第一语义向量相似度较高的第二语义向量。

[0152] S305:根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合。

[0153] 其中,候选向量集合,是指由初始相似度排序在前的第一数量个第二语义向量所构建的向量集合。

[0154] 本公开实施例中,当根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合时,可以为第一描述对象的确定过程提供可靠的候选对象。

[0155] S306:根据候选向量集合及候选向量集合的数量,确定第一描述对象。

[0156] 可以理解的是,本公开实施例中可能会基于一个或多个语义向量模型对询问语句进行向量化处理,以得到一个或多个第一语义向量,因此,候选向量集合的数量也可能是一个或多个,而候选向量集合的不同数量可能会影响第一描述对象的确定过程,由此,需要结合候选向量集合及候选向量集合的数量,确定第一描述对象。

[0157] 即是说,本公开实施例中在获取询问语句对应的第一语义向量之后,可以确定第一语义向量与预设检索库中第二语义向量的初始相似度,根据初始相似度对多个第二语义向量进行排序,根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合,根据候选向量集合及候选向量集合的数量,确定第一描述对象,由此,可以确定第一语义向量与预设检索库中第二语义向量的初始相似度,并根据初始相似度对多个第二语义向量进行排序,根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合,从而基于初始相似度的排序实现对预设检索库中多个第二语义向量的粗筛选,而后结合候选向量集合及候选向量集合的数量,灵活采用对应的识别方法从多个业务描述对象中确定第一描述对象,可以有效提升所得第一描述对象与询问语句的适配性。

[0158] 可选的,一些实施例中,候选向量集合的数量是一个,则在根据候选向量集合及候选向量集合的数量,确定第一描述对象时,可以是根据候选向量集合中第二语义向量的排序次序,对相应第二语义向量的初始相似度进行加权处理,以得到与第二语义向量对应的目标相似度,根据目标相似度,确定第一描述对象,由此,在候选向量集合的数量是一个时,可以结合候选向量集合中第二语义向量的排序次序实现对初始相似度的调整,以有效提升所得目标相似度的指示准确性,从而根据目标相似度确定可靠的第一描述对象。

[0159] 举例而言,本公开实施例中在根据候选向量集合中第二语义向量的排序次序,对相应第二语义向量的初始相似度进行加权处理时,可以是根据候选向量集合中第二语义向量的排序次序确定对应的加权值,当第二语义向量的排序次序越靠前时,对应加权值越大,然后计算加权值与初始相似度的乘积值作为目标相似度。例如,当排序次序值为N时,对应的加权值可以为 $1/N$ 。

[0160] 可选的,一些实施例中,候选向量集合的数量是多个,不同候选向量集合分别是基于不同语义向量模型对询问语句进行向量化处理所得,则在根据候选向量集合及候选向量集合的数量,确定第一描述对象时,可以是根据候选向量集合中第二语义向量的排序次序,对相应第二语义向量的初始相似度进行加权处理,以得到与第二语义向量对应的参考相似度,确定与第二语义向量对应的多个参考相似度的和值作为目标相似度,根据目标相似度,确定第一描述对象,由此,可以在候选向量集合的数量是多个时,结合候选向量集合中第二语义向量的排序次序实现对初始相似度的调整,以得到参考相似度,而后确定与第二语义向量对应的多个参考相似度的和值作为目标相似度,以实现对多个候选向量集合中第二语义向量的去重处理,保证所得目标相似度的指示效果,从而基于目标相似度准确识别出第一描述对象。

[0161] 可以理解的是,同一个第二语义向量可能同时与多个第一语义向量具有较高的参考相似度,此时确定与第二语义向量对应的多个参考相似度的和值作为目标相似度,可以实现对多个候选向量集合中第二语义向量的去重处理。

[0162] 可选的,一些实施例中,预设检索库包括聚合业务数据,聚合业务数据包含多个业务描述对象,业务描述对象与第二语义向量相关联,则在根据目标相似度,确定第一描述对象时,可以是根据目标相似度对多个第二语义向量进行排序,将排序在前的第二数量个第二语义向量对应的业务描述对象作为第一描述对象,由此,可以根据目标相似度对多个第二语义向量进行排序,以确定多个第二语义向量对应的目标相似度大小关系,而后选择排

序在前的第二数量个第二语义向量对应的业务描述对象作为第一描述对象,以筛选出目标相似度较高的第二数量个第二语义向量,并将对应业务描述对象作为第一描述对象,以保证所得第一描述对象的适配性。

[0163] S307:根据第一描述对象和第二描述对象,确定候选描述对象。

[0164] S308:对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同。

[0165] S309:将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象。

[0166] S310:根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据。

[0167] S311:根据目标业务数据,生成与询问语句对应的答复语句。

[0168] S307-S311的描述说明具体可以参见上述实施例,在此不再赘述。

[0169] 本公开实施例中,通过获取询问语句,并获取询问语句对应的第一语义向量和第一关键词,从预设检索库中确定与第一关键词匹配的第二描述对象,确定第一语义向量与预设检索库中第二语义向量的初始相似度,根据初始相似度对多个第二语义向量进行排序,根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合,根据候选向量集合及候选向量集合的数量,确定第一描述对象,根据第一描述对象和第二描述对象,确定候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句。由此,可以确定第一语义向量与预设检索库中第二语义向量的初始相似度,并根据初始相似度对多个第二语义向量进行排序,根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合,从而基于初始相似度的排序实现对预设检索库中多个第二语义向量的粗筛选,而后结合候选向量集合及候选向量集合的数量,灵活采用对应的识别方法从多个业务描述对象中确定第一描述对象,可以有效提升所得第一描述对象与询问语句的适配性。

[0170] 举例而言,当本公开应用于智能投资研究指标问答场景中时,如图4所示,图4是根据本公开提出的智能投研指标问答算法流程图,其中,包括离线指标入库和在线指标问答两个大模块。图中处理逻辑可以参见上述实施例的技术方案,在此不在赘述。

[0171] 图5是根据本公开一实施例提供的一种语句生成装置的结构示意图;

[0172] 如图5所示,该语句生成装置50,包括:

[0173] 获取模块501,用于获取询问语句,其中,询问语句具有对应的至少一个候选描述对象;

[0174] 第一识别模块502,用于对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同;

[0175] 第二识别模块503,用于将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象;

[0176] 第三识别模块504,用于根据第二语义信息,从与目标描述对象对应的至少一个业



务数据中选择目标业务数据;

[0177] 生成模块505,用于根据目标业务数据,生成与询问语句对应的答复语句。

[0178] 在本公开的一些实施例中,其中,第一语义信息是地理区域维度的语义信息,第二语义信息是与候选描述对象相关的统计频率维度的语义信息,和/或与候选描述对象相关的统计时间维度的语义信息。

[0179] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0180] 获取询问语句对应的第一语义向量和第一关键词;

[0181] 从预设检索库中确定与第一语义向量匹配的第一描述对象,及与第一关键词匹配的第二描述对象;

[0182] 根据第一描述对象和第二描述对象,确定候选描述对象。

[0183] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0184] 获取业务数据库,其中,业务数据库中包括多条初始业务数据;

[0185] 确定初始业务数据的业务描述对象,以及与业务描述对象对应的数据记录类型和数据记录频率;

[0186] 分别基于业务描述对象、数据记录类型和数据记录频率对多条初始业务数据进行聚合处理,以得到聚合业务数据,其中,聚合业务数据包含与目标描述对象对应的至少一个业务数据;

[0187] 基于语义向量模型对业务描述对象进行向量化处理,以得到第二语义向量;

[0188] 基于聚合业务数据和第二语义向量,构建预设检索库。

[0189] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0190] 基于单语言语义向量模型对业务描述对象进行向量化处理,得到第二语义向量;和/或

[0191] 基于多语言语义向量模型对业务描述对象进行向量化处理,得到第二语义向量。

[0192] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0193] 基于单语言语义向量模型对询问语句进行向量化处理,得到第一语义向量;和/或

[0194] 基于多语言语义向量模型对询问语句进行向量化处理,得到第一语义向量。

[0195] 在本公开的一些实施例中,预设检索库包括:多个第二语义向量,其中,获取模块501,还用于:

[0196] 确定第一语义向量与预设检索库中第二语义向量的初始相似度;

[0197] 根据初始相似度对多个第二语义向量进行排序;

[0198] 根据排序在前的第一数量个第二语义向量,构建与第一语义向量对应的候选向量集合;

[0199] 根据候选向量集合及候选向量集合的数量,确定第一描述对象。

[0200] 在本公开的一些实施例中,其中,候选向量集合的数量是一个;其中,获取模块501,还用于:

[0201] 根据候选向量集合中第二语义向量的排序次序,对相应第二语义向量的初始相似度进行加权处理,以得到与第二语义向量对应的目标相似度;

[0202] 根据目标相似度,确定第一描述对象。

[0203] 在本公开的一些实施例中,其中,候选向量集合的数量是多个,不同候选向量集合

分别是基于不同语义向量模型对询问语句进行向量化处理所得;其中,获取模块501,还用于:

[0204] 根据候选向量集合中第二语义向量的排序次序,对相应第二语义向量的初始相似度进行加权处理,以得到与第二语义向量对应的参考相似度;

[0205] 确定与第二语义向量对应的多个参考相似度的和值作为目标相似度;

[0206] 根据目标相似度,确定第一描述对象。

[0207] 在本公开的一些实施例中,其中,预设检索库包括聚合业务数据,聚合业务数据包含多个业务描述对象,业务描述对象与第二语义向量相关联;其中,获取模块501,还用于:

[0208] 根据目标相似度对多个第二语义向量进行排序;

[0209] 将排序在前的第二数量个第二语义向量对应的业务描述对象作为第一描述对象。

[0210] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0211] 基于预设评分模型确定第一描述对象的第一评分值,和第二描述对象的第二评分值,其中,第一评分值用于描述对应第一描述对象与询问语句的匹配程度,第二评分值用于描述对应第二描述对象与询问语句的匹配程度;

[0212] 根据第一评分值和第二评分值,对第一描述对象和第二描述对象进行排序;

[0213] 将排序在前的第三数量个第一描述对象和/或第二描述对象作为候选描述对象。

[0214] 在本公开的一些实施例中,其中,获取模块501,还用于:

[0215] 获取样本数据,其中,样本数据包括:样本询问语句,以及与样本询问语句对应的多个样本答复语句;

[0216] 确定询问语句对应的业务场景;

[0217] 根据业务场景对样本答复语句进行标注处理,以从多个样本答复语句中确定正样本答复语句和负样本答复语句;

[0218] 基于样本询问语句、正样本答复语句和负样本答复语句对初始评分模型进行训练,以得到预设评分模型。

[0219] 需要说明的是,前述对语句生成方法的解释说明也适用于本实施例的语句生成装置,此处不再赘述。

[0220] 本公开实施例中,通过获取询问语句,其中,询问语句具有对应的至少一个候选描述对象,对询问语句进行语义识别,以获取与询问语句对应的第一语义信息和第二语义信息,其中,第一语义信息和第二语义信息的类别不相同,将至少一个候选描述对象和第一语义信息输入至大模型之中,以基于大模型从至少一个候选描述对象中识别出目标描述对象,根据第二语义信息,从与目标描述对象对应的至少一个业务数据中选择目标业务数据,根据目标业务数据,生成与询问语句对应的答复语句,由此,可以识别询问语句不同类型的语义信息,从而基于第一语义信息和大模型从至少一个候选描述对象中识别出目标描述对象,基于第二语义信息从与目标描述对象对应的至少一个业务数据中选择目标业务数据,并目标业务数据生成答复语句,从而有效提升所得答复语句与询问语句的适配性,提升答复效果。

[0221] 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。

[0222] 图6示出了可以用来实施本公开的实施例的示例电子设备600的示意性框图。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。

[0223] 如图6所示,设备600包括计算单元601,其可以根据存储在只读存储器 (ROM) 602中的计算机程序或者从存储单元608加载到随机访问存储器 (RAM) 603中的计算机程序,来执行各种适当的动作和处理。在RAM 603中,还可存储设备600操作所需的各种程序和数据。计算单元601、ROM 602以及RAM 603通过总线604彼此相连。输入/输出 (I/O) 接口605也连接至总线604。

[0224] 设备600中的多个部件连接至I/O接口605,包括:输入单元606,例如键盘、鼠标等;输出单元607,例如各种类型的显示器、扬声器等;存储单元608,例如磁盘、光盘等;以及通信单元609,例如网卡、调制解调器、无线通信收发机等。通信单元609允许设备600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。

[0225] 计算单元601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601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601执行上文所描述的各个方法和处理,例如语句生成方法。例如,在一些实施例中,语句生成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608。在一些实施例中,计算机程序的部分或者全部可以经由ROM 602和/或通信单元609而被载入和/或安装到设备600上。当计算机程序加载到RAM 603并由计算单元601执行时,可以执行上文描述的语句生成方法的一个或多个步骤。备选地,在其他实施例中,计算单元601可以通过其他任何适当的方式 (例如,借助于固件) 而被配置为执行语句生成方法。

[0226] 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、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。

[0227] 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。

[0228] 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。

[0229] 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。

[0230] 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)、互联网及区块链网络。

[0231] 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",或简称"VPS")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。

[0232] 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,本发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。

[0233] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本公开的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。在本公开的描述中,所使用的词语“如果”及“若”可以被解释成为“在……时”或“当……时”或“响应于确定”或“在……情况下”。

[0234] 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。

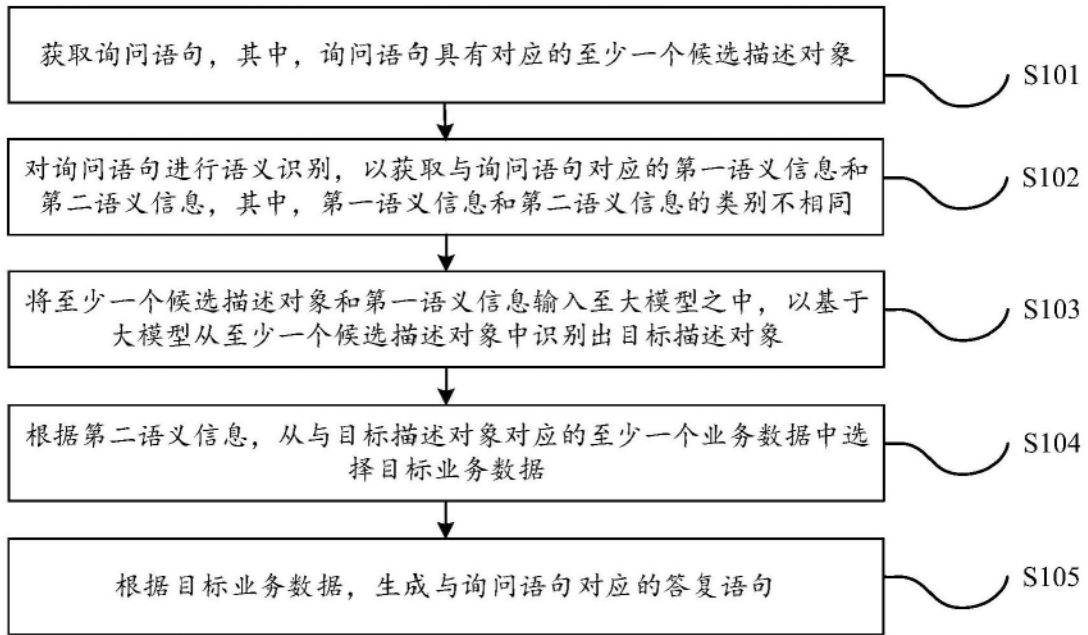


图1

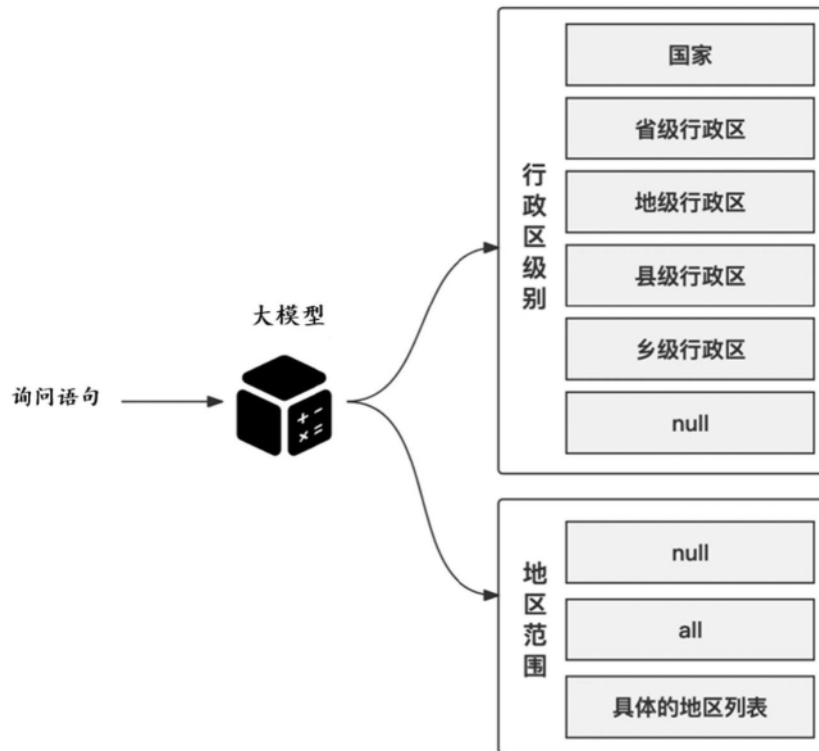


图1a

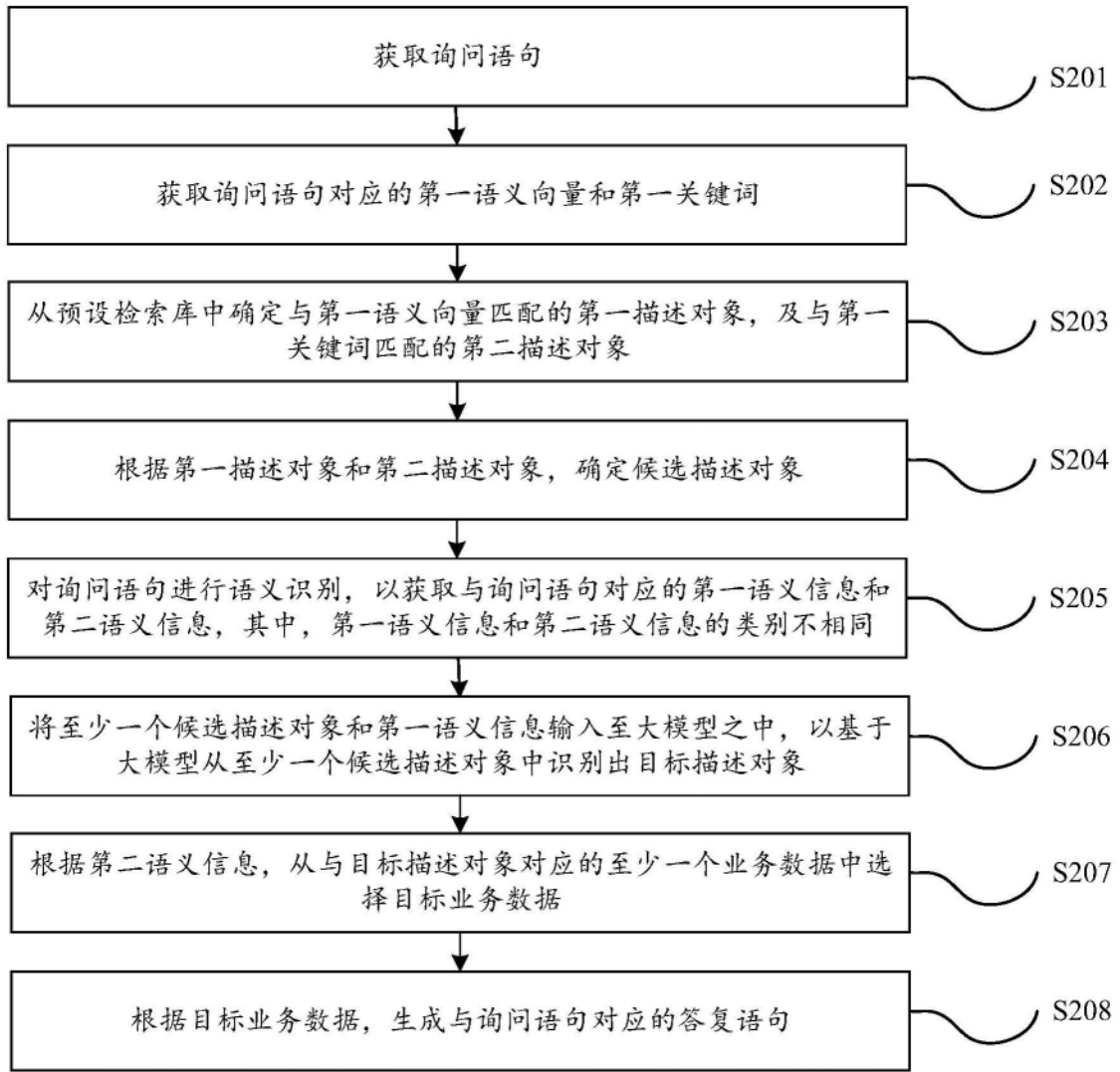


图2

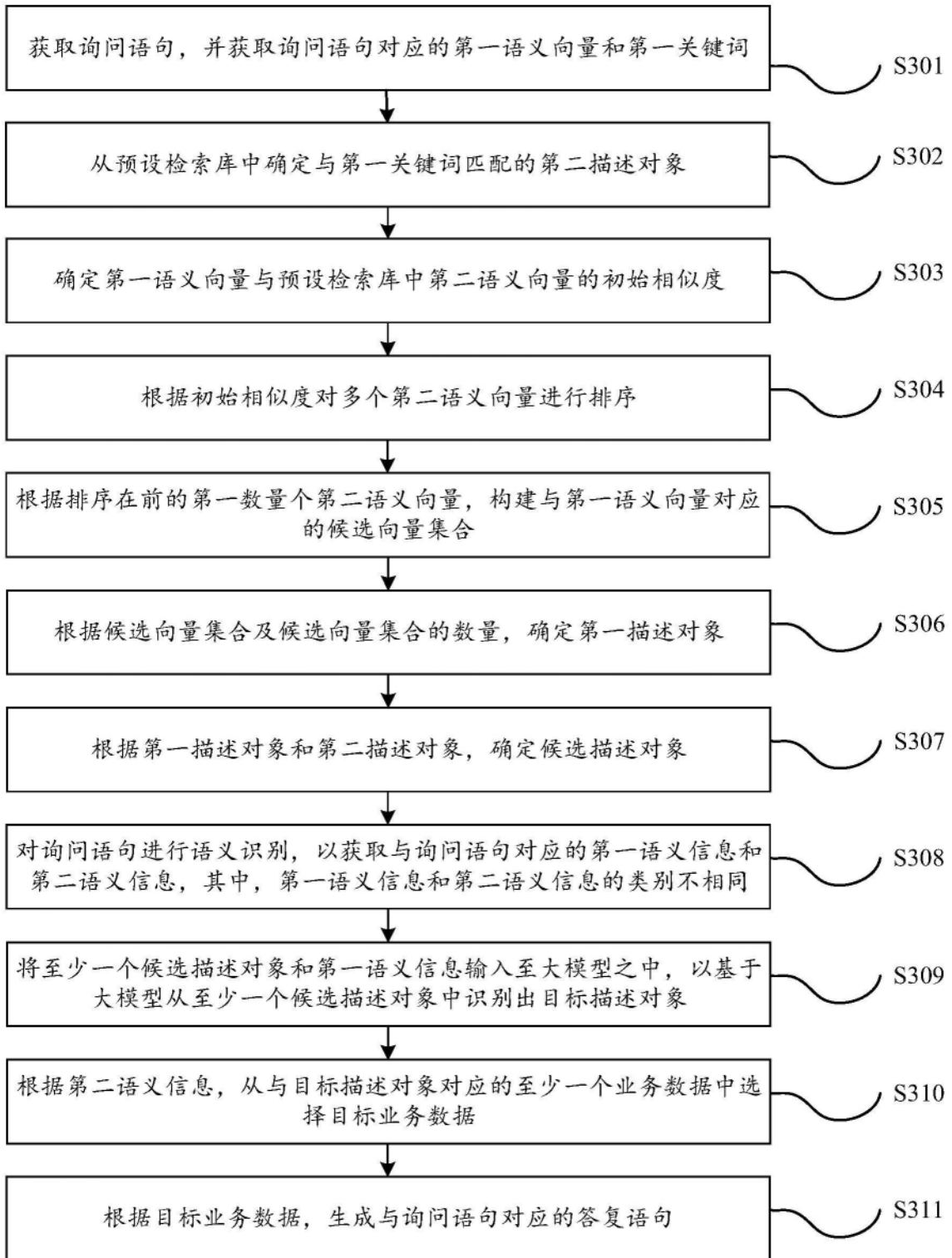


图3

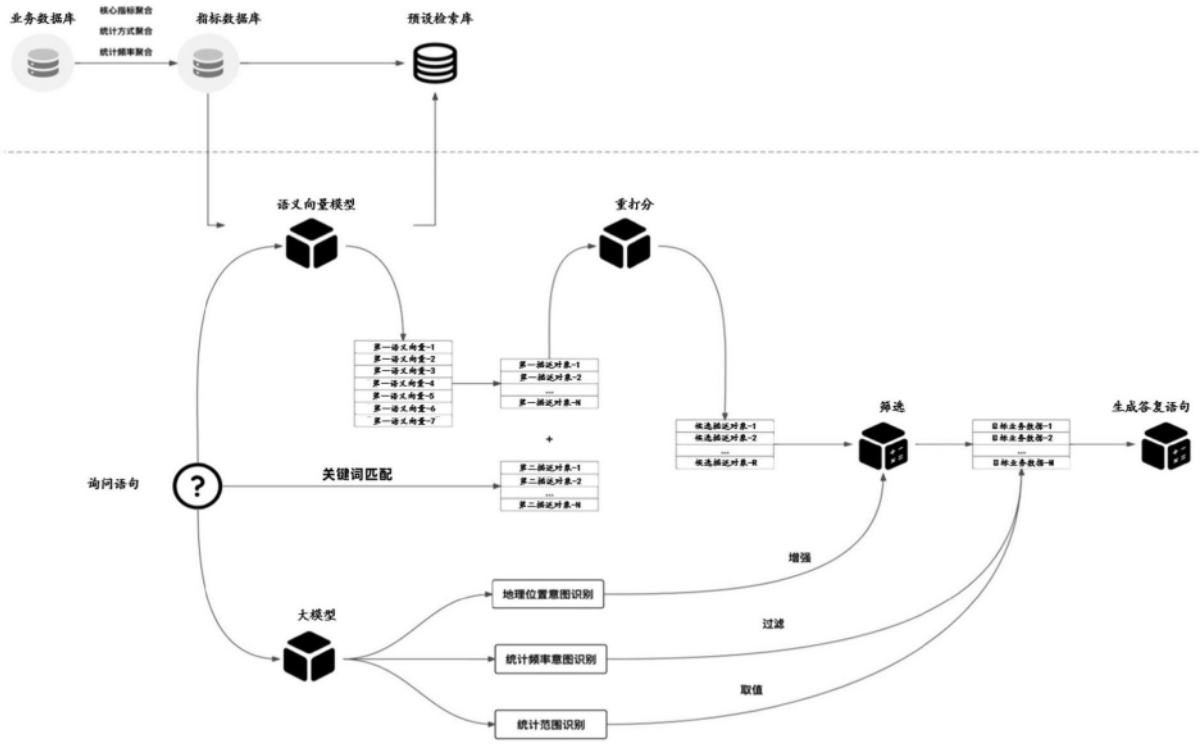


图4



图5



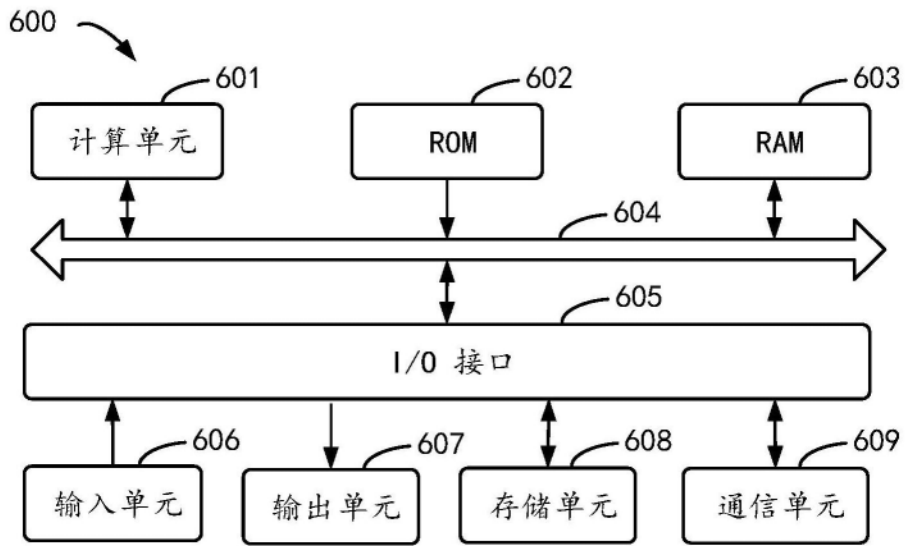


图6