



(12) 发明专利申请

(10) 申请公布号 CN 114817465 A

(43) 申请公布日 2022. 07. 29

(21) 申请号 202210394592.4

G06F 40/232 (2020.01)

(22) 申请日 2022.04.14

G06F 40/295 (2020.01)

G06F 40/30 (2020.01)

(71) 申请人 海信电子科技(武汉)有限公司

地址 430073 湖北省武汉市东湖新技术开发区软件园东路1号软件产业4.1期B2栋13层02号-2

(72) 发明人 胡胜元 曹晚霞

(74) 专利代理机构 北京弘权知识产权代理有限公司 11363

专利代理师 逯长明 许伟群

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/335 (2019.01)

G06F 16/36 (2019.01)

G06F 40/126 (2020.01)

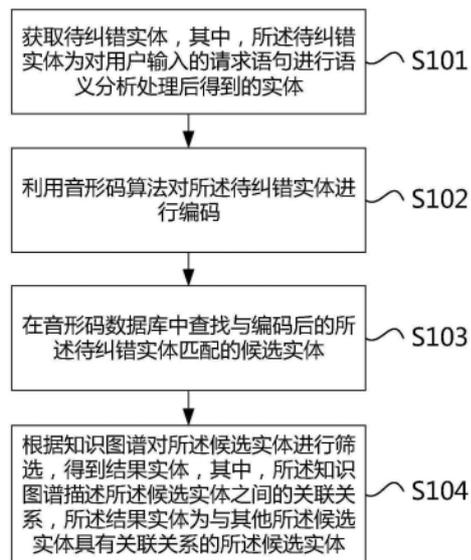
权利要求书2页 说明书10页 附图5页

(54) 发明名称

一种用于多语言语义理解的实体纠错方法及智能设备

(57) 摘要

本申请实施例提供一种用于多语言语义理解的实体纠错方法及智能设备,方法包括:获取待纠错实体后,利用音形码算法对待纠错实体进行编码。在音形码数据库中查找与编码后的待纠错实体匹配的候选实体。根据知识图谱对候选实体进行筛选,得到结果实体,知识图谱描述候选实体之间的关联关系,结果实体为与其他候选实体具有关联关系的候选实体。本申请提供的用于多语言语义理解的实体纠错方法及智能设备,针对多语言语义理解提供统一的框架,可以在缺少大规模训练数据的情况下,跨越不同语言的影响,实现对不同语言的文本进行实体纠错,从而提升语义理解和实体识别的准确率,提升多语言语音识别产品的性能,进而提升用户使用体验。



1. 一种用于多语言语义理解的实体纠错方法,其特征在于,包括:
获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体;
利用音形码算法对所述待纠错实体进行编码;
在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体;
根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述所述候选实体之间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。
2. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体,具体步骤为:
获取所述待纠错实体的语言类型;
根据所述待纠错实体的语言类型调用所述音形码数据库;
在所述音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体。
3. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体,具体步骤为:
获取所述待纠错实体的业务类型;
根据所述待纠错实体的业务类型调用所述音形码数据库;
在所述音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体。
4. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体,具体步骤为:
在根据编码后的所述待纠错实体从音形码数据库中召回的候选实体数量为零时,从所述待纠错实体中截取子字符串;
利用音形码算法对所述子字符串进行编码,以及在音形码数据库中查找与编码后的所述子字符串匹配的候选实体。
5. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体,具体步骤为:
在根据编码后的所述待纠错实体从音形码数据库中召回的候选实体数量大于数量阈值N时,计算所有所述候选实体与所述待纠错实体的编辑距离;
按照编辑距离对所有所述候选实体进行排序,以及将编辑距离按照由小到大的顺序排序的前N个所述候选实体确定为最终的所述候选实体。
6. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,实体编码后具有特征向量,第一实体的特征向量与第二实体的特征向量在编码空间的距离,与所述第一实体和所述第二实体的发音相似程度相匹配,其中,所述第一实体的语言类型和所述第二实体的语言类型不同。
7. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,根据知识图谱对所述候选实体进行筛选,得到结果实体,具体步骤为:
在所述知识图谱中存在与所述候选实体匹配的预存实体时,将所述预存实体确定为所述结果实体;
在所述知识图谱中不存在与所述候选实体匹配的预存实体,且在所述知识图谱中存在

与所述候选实体中的子字符匹配的预存实体,将所述预存实体确定为所述结果实体。

8. 根据权利要求1所述的用于多语言语义理解的实体纠错方法,其特征在于,所述音形码数据库为预设的常用指令集,所述候选实体为用于生成常用指令的实体。

9. 一种用于多语言语义理解的实体纠错的智能设备,其特征在于,包括:

待纠错实体获取单元,用于执行:获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体;

编码单元,用于执行:利用音形码算法对所述待纠错实体进行编码;

候选实体查找单元,用于执行:在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体;

筛选单元,用于执行:根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述所述候选实体之间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。

10. 根据权利要求9所述的用于多语言语义理解的实体纠错的智能设备,其特征在于,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体,具体步骤为:

获取所述待纠错实体的语言类型;

根据所述待纠错实体的语言类型调用所述音形码数据库;

在所述音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体。

一种用于多语言语义理解的实体纠错方法及智能设备

技术领域

[0001] 本申请涉及语音交互技术领域,尤其涉及一种用于多语言语义理解的实体纠错方法及智能设备。

背景技术

[0002] 随着智能语音交互技术的发展,语音交互功能逐渐成为智能终端产品的标准配置。用户可利用语音交互功能,实现语音控制智能终端产品,进行看视频、听音乐、查天气、电视控制等一系列操作。

[0003] 语音控制智能终端产品的过程通常是,语音识别模型将用户输入的语音识别为文本。之后语义理解模型对该文本进行词法句法和语义的分析,从而理解用户的意图。最后控制端根据理解结果控制智能终端产品进行相应的操作。

[0004] 在实际应用中,语音识别模型和语义理解模型都存在误差,两个阶段积累的误差直接影响整个系统的语音识别质量。因此,现有的语音识别系统都配置有纠错模型来对用户请求语句中的实体进行纠错,以提高整个语音识别系统的质量。目前多语言的纠错主要是对文本语法的纠错研究,通常采用大规模的文本数据进行深度学习模型训练。

[0005] 然而,由于多语言语音识别的实体通常是短文本,这些短文本实体对字符顺序不敏感,没有语法等语义信息,因此很难基于文本语法进行纠错。另外目前多语言语音识别使用的训练数据集和语音助手的应用场景不相符,而收集大量真实场景下的文本数据进行模型训练难度也较大。目前针对多语言语音识别的产品并不能很好的满足用户的使用需求。因此,亟需一种在缺少大规模训练数据的情况下,可以跨越不同语言影响,多语言通用的实时的实体纠错方法。

发明内容

[0006] 本申请提供了一种用于多语言语义理解的实体纠错方法及智能设备,用于解决由于多语言语音识别的实体通常是短文本,这些短文本实体对字符顺序不敏感,没有语法等语义信息,因此很难基于语法进行纠错。另外目前多语言语音识别使用的训练数据集和语音助手的应用场景不相符,而收集大量真实场景下的文本数据进行模型训练难度也较大的问题。

[0007] 第一方面,本申请实施例提供一种用于多语言语义理解的实体纠错方法,该方法包括:获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体;

[0008] 利用音形码算法对所述待纠错实体进行编码;

[0009] 在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体;

[0010] 根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述所述候选实体之间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。

[0011] 第二方面,本申请实施例提供一种用于多语言语义理解的实体纠错的智能设备,该智能设备包括:

[0012] 待纠错实体获取单元,用于执行:获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体;

[0013] 编码单元,用于执行:利用音形码算法对所述待纠错实体进行编码;

[0014] 候选实体查找单元,用于执行:在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体;

[0015] 筛选单元,用于执行:根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述所述候选实体之间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。

[0016] 本申请提供的技术方案包括以下有益效果:获取待纠错实体后,利用音形码算法对待纠错实体进行编码。在音形码数据库中查找与编码后的待纠错实体匹配的候选实体。根据知识图谱对候选实体进行筛选,得到结果实体。其中,知识图谱描述候选实体之间的关联关系,结果实体为与其他候选实体具有关联关系的候选实体。本申请提供的用于多语言语义理解的实体纠错方法及智能设备,针对多语言语义理解提供统一的框架,可以在缺少大规模训练数据的情况下,跨越不同语言的影响,实现对不同语言的文本进行实体纠错,从而提升语义理解和实体识别的准确率,提升多语言语音识别产品的性能,进而提升用户使用体验。

附图说明

[0017] 为了更清楚地说明本申请的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,对于本领域普通技术人员而言,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0018] 图1示例性示出了根据一些实施例的语音交互原理的示意图;

[0019] 图2示例性示出了根据一些实施例的实体纠错模块的框架示意图;

[0020] 图3示例性示出了根据一些实施例的用于多语言语义理解的实体纠错方法的流程示意图;

[0021] 图4示例性示出了根据一些实施例的中文音形码编码流程示意图;

[0022] 图5示例性示出了根据一些实施例的实体纠错方法具体示例框架示意图;

[0023] 图6示例性示出了根据一些实施例的又一种实体纠错方法流程示意图;

[0024] 图7示例性示出了根据一些实施例的又一种实体纠错方法流程示意图;

[0025] 图8示例性示出了根据一些实施例的多语言音形码编码算法示意图;

[0026] 图9示例性示出了根据一些实施例的又一种实体纠错方法流程示意图;

[0027] 图10示例性示出了根据一些实施例的用于多语言语义理解的实体纠错的智能设备框架示意图;

[0028] 图11示例性示出了根据一些实施例的多语言语音助手应用流程图。

具体实施方式

[0029] 为使本申请的目的和实施方式更加清楚,下面将结合本申请示例性实施例中的附

图,对本申请示例性实施方式进行清楚、完整地描述,显然,描述的示例性实施例仅是本申请一部分实施例,而不是全部的实施例。

[0030] 需要说明的是,本申请中对于术语的简要说明,仅是为了方便理解接下来描述的实施方式,而不是意图限定本申请的实施方式。除非另有说明,这些术语应当按照其普通和通常的含义理解。

[0031] 术语“包括”和“具有”以及他们的任何变形,意图在于覆盖但不排他的包含,例如,包含了一系列组件的产品或设备不必限于清楚地列出的所有组件,而是可包括没有清楚地列出的或对于这些产品或设备固有的其它组件。

[0032] 术语“模块”是指任何已知或后来开发的硬件、软件、固件、人工智能、模糊逻辑或硬件或/和软件代码的组合,能够执行与该元素相关的功能。

[0033] 为清楚说明本申请的实施例,下面结合图1对本申请实施例提供的一种语音识别网络架构进行描述。

[0034] 参见图1,图1为本申请实施例提供的一种语音识别网络架构示意图。图1中,智能设备用于接收输入的信息以及输出对该信息的处理结果。语音识别服务设备为部署有语音识别服务的电子设备,语义服务设备为部署有语义服务的电子设备,业务服务设备为部署有业务服务的电子设备。这里的电子设备可包括服务器、计算机等,这里的语音识别服务、语义服务(也可称为语义引擎)和业务服务为可部署在电子设备上的web服务,其中,语音识别服务用于将音频识别为文本,语义服务用于对文本进行语义解析,业务服务用于提供具体的服务如天气查询服务、音乐查询服务等。在一个实施例中,图1所示架构中可存在部署有不同业务服务的多个实体服务设备,也可以一个或多个实体服务设备中集合一项或多项功能服务。

[0035] 一些实施例中,下面对基于图1所示架构处理输入智能设备的信息的过程进行举例描述,以输入智能设备的信息为通过语音输入的查询语句为例,上述过程可包括如下三个过程:

[0036] 语音识别:智能设备可在接收到通过语音输入的查询语句后,将该查询语句的音频上传至语音识别服务设备,以由语音识别服务设备通过语音识别服务将该音频识别为文本后返回至智能设备。在一个实施例中,将查询语句的音频上传至语音识别服务设备前,智能设备可对查询语句的音频进行去噪处理,这里的去噪处理可包括去除回声和环境噪声等步骤。

[0037] 语义理解:智能设备将语音识别服务识别出的查询语句的文本上传至语义服务设备,以由语义服务设备通过语义服务对该文本进行语义解析,得到文本的业务领域、意图等。

[0038] 语义响应:语义服务设备根据对查询语句的文本的语义解析结果,向相应的业务服务设备下发查询指令以获取业务服务给出的查询结果。智能设备可从语义服务设备获取该查询结果并输出。作为一个实施例,语义服务设备还可将对查询语句的语义解析结果发送至智能设备,以由智能设备输出该语义解析结果中的反馈语句。

[0039] 需要说明的是,图1所示架构只是一种示例,并非对本申请保护范围的限定。本申请实施例中,也可采用其他架构来实现类似功能,例如:三个过程全部或部分可以由智能设备来完成,在此不做赘述。

[0040] 在一些实施例中,图1所示的智能设备可为显示设备,如智能电视,语音识别服务设备的功能可由显示设备上设置的声音采集器和控制器配合实现,语义服务设备和业务服务设备的功能可由显示设备的控制器实现,或者由显示设备的服务器来实现。

[0041] 在一些实施例中,智能设备支持语音交互功能,语音交互功能可作为智能终端产品的标准配置。用户可利用语音交互功能,实现语音控制智能终端产品,进行看视频、听音乐、查天气、电视控制等一系列操作。

[0042] 语音控制智能终端产品的过程通常是,语音识别模块将用户输入的语音识别为文本。之后语义分析模块对该文本进行词法句法和语义的分析,从而理解用户的意图。最后控制端根据理解结果控制智能终端产品进行相应的操作。

[0043] 在实际应用中,语音识别模型和语义理解模型都存在误差,两个阶段积累的误差直接影响整个系统的语音识别质量。因此,在一些实施例中,语音识别系统都配置有纠错模型来对用户请求语句中的实体进行纠错,以提高整个语音识别系统的质量。

[0044] 例如,中文下的用户请求“我想看吴某某的电影XX”,其中吴某某为实体,语音系统的纠错模块可能会纠正“吴某某”,从而更好触达用户实际意图,提升用户使用体验。需要说明的是,这里的实体指命名实体,命名实体是具有特定意义的实体,主要包括人名、地名、机构名、专有名词等。

[0045] 在一些实施例中,纠错模型主要基于两种方法:基于规则的方法和基于机器学习的方法。基于规则的方法指的是根据语言发音规则,结合用户使用习惯,对同音字纠错、模糊音纠错等。基于机器学习的方法指的是通过构建多个决策模型,利用大量的“错误-纠正后”的文本对模型进行训练,最后由决策模型给出统一的纠错结果。

[0046] 对于多语言的纠错方法,无法简单复用中文下的纠错方法。主要存在以下问题:中文是象形文字,而多语言中的大部分语言为表音文字,表音文字的字符含义逻辑与中文存在根本区别;不同语言的发音习惯不同,无法有效构建基于规则的纠正候选集;需要一个统一的框架对不同语言的识别提供复用功能。

[0047] 在一些实施例中,多语言的纠错主要是对文本语法的纠错研究,通常采用大规模的文本数据进行深度学习模型训练。然而,由于多语言语音识别的实体通常是短文本,这些短文本实体对字符顺序不敏感,没有语法等语义信息,因此很难基于语法进行纠错。另外多语言语音识别使用的训练数据集和语音助手的应用场景不相符,而收集大量真实场景下的文本数据进行模型训练难度也较大。针对多语言语音识别的产品并不能很好的满足用户的使用需求。因此,亟需一种在缺少大规模训练数据的情况下,可以跨越不同语言影响,多语言通用的实时的实体纠错方法。

[0048] 为了解决上述问题,本申请提供一种用于多语言语义理解的实体纠错方法,该方法针对多语言语义理解提供统一的框架,可以在缺少大规模训练数据的情况下,跨越不同语言的影响,实现对不同语言的文本进行实体纠错,从而提升语义理解和实体识别的准确率,提升多语言语音识别产品的性能,进而提升用户使用体验。

[0049] 在阐述本申请方法流程之前,对本申请涉及的技术术语进行解释:

[0050] 多语言语义理解模型采用大规模多语言的语料数据的预训练模型LaBSE(多语言嵌入向量模型)来对用请求文本进行编码分析,对其进行意图判断和实体识别。比如如果用户的请求为“search for XX by Tom”,经过多语言语义解析模型后,输出的语义解析结

果:"intent":"video.search","actor":"Tom","title":"XX",其中的intent为判断出的意图为搜索,actor和title均为实体。

[0051] Metaphone算法是一种根据文本英文发音方法进行编码的音形码算法,经过近年来的改进完善到Metaphone3。该算法主要是针对英语下的发音规则的音形码编码算法,其主要目的是将发音相似的文本编码成相同的键值。利用英语下的单词的发音规则,按照元音字母和辅音字母区分,将辅音字母按照预先设定的规则对输入文本进行编码。比如文本"volume up"中的辅音字母为"vlmp",因此整个文本编码后为"FLMP"。

[0052] Metaphone算法虽然是针对英语进行开发的,但是其基本思想是基于语言发音规则来的,该算法可以扩展到其它和英语相似的表音文字中。比如巴西葡萄牙语(pt-br)版的Metaphone算法已经成为巴西当地城市的数据库的解决方案,GitHub(一种面向开源及私有软件项目的托管平台)上也已经有学者开源了基于法语、西班牙语、俄语等Metaphone算法。

[0053] 本申请主要是基于多语言语音助手中的实体纠错模块,对语义理解模型识别出的实体进行纠错处理。如图2所示的实体纠错模块的框架示意图,实体纠错模块的输入为多语言语义理解模型对用户请求语句的解析结果,解析结果包含意图和实体。本申请的实体纠错模块主要是针对输入的实体进行纠错。实体纠错模块又主要包括音形码召回子模块和知识图谱检索子模块。基于图2所示的实体纠错模块,如图3所示的用于多语言语义理解的实体纠错方法的流程示意图,该方法包括以下步骤:

[0054] 步骤S101,获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体。

[0055] 语音文本为对用户输入的语音信号解析得到的。具体的,用户在终端设备接收信号的距离范围内输入语音信号。终端设备可以通过麦克风采集用户输入的语音信号,之后从语音信号中识别出语音文本。本申请实施例可由语音识别服务器识别出语音文本。由语义服务器对语音文本进行语义分析处理。需要说明的是,语义服务器对语音文本进行语义分析处理,可以是采用前文所提及的多语言语义理解模型,解析语音文本得到意图和实体。对语音文本进行语义分析处理的具体过程可以采用目前已有的技术,本申请对此不作详细的阐述。

[0056] 步骤S102,利用音形码算法对所述待纠错实体进行编码。

[0057] 该步骤主要利用音形码算法对解析出的实体进行编码,解析出的实体即为待纠错的实体。智能语音助手的输入为用户语音,因此发生文本错误的地方只存在于语音识别为文本的错误,因此文本的发音是一个重要的纠错依据。本申请基于现有的Metaphone算法对待纠错实体进行编码。如前文所述Metaphone是一种针对英文下的音形码算法,其通过英文发音规则对文本进行编码,将发音相似的文本编码成相同音形码。Metaphone算法是针对英语设计的,也可以扩展到类似的表音文字的语言(如英语、西班牙语、俄语等)。因此,本申请可以利用针对不同语言的Metaphone算法对待纠错实体进行编码。

[0058] 除表音文字可以使用Metaphone算法外,也可以基于发音规则开发象形文字的语言的音形码算法,比如中文下基于拼音规则来编码,日语下通过五十音进行编码。图4为中文音形码编码的示例,对于待编码的文本,先通过PyPinyin工具将文本转换为对应的拼音,对拼音中的声母和韵母分别进行编码,编码过程中考虑到拼音发音的相似性,比如声母n和声母l,韵母"an"和"ang"可以编码成相同的编码。

[0059] 步骤S103,在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体。

[0060] 本申请的音形码数据库中事先存储有大量的文本实体,音形码数据库中的实体也是经过相同的编码过程生成的,因此音形码数据库中的实体也是编码形式。该步骤将待纠错实体的编码形式与音形码数据库中的编码形式数据进行匹配,从而查找与待纠错实体匹配的候选实体。需要说明的是,这里的待纠错实体与候选实体匹配可以是待纠错实体的编码与候选实体的编码相同,或者待纠错实体的编码包含候选实体的编码,对于匹配形式本申请不作限制。

[0061] 步骤S104,根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述各种实体间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。

[0062] 该步骤主要为了得到最终的纠错结果,具体的,将待纠错实体和候选实体放入到知识图谱中查询。音形码数据库中的实体基本上和知识图谱中的实体相同。知识图谱指用于增强搜索引擎功能的知识库,旨在描述真实世界中存在的各种实体或概念及其关系,其构成一张巨大的语义网络图,节点表示实体或概念,边则由属性或关系构成。知识图谱中的所有数据可以采用RDF(Resource Description Framework,资源描述框架)进行表达和存储。

[0063] 举例说明,以[实体,属性,属性值]、[实体1,关系,实体2]等三元组形式表示实体之间的复杂关系。例如,[XX(电影名称),主演,吴某某],[于某某,出生地,北京]等。需要说明的是,本申请实施例中知识图谱还可以以其他形式表示和存储,本申请实施例对此不作具体限定。

[0064] 对候选实体进行筛选后得到结果实体,最后可以根据结果实体输出有效意图。具体的,将知识图谱检索输出的结果封装成统一的格式,方便下游终端执行命令。部分意图的实体还需要转换成统一的格式,比如电视的一些固定设置项需要转换成电视语言;不同语言的时间格式转化成终端能够解析的数字格式,如英文下的2hours转化成{h:2,m:0,s:0}。

[0065] 针对步骤S101至步骤S014的实体纠错方法,通过如图5所示具体的示例进行阐述:

[0066] 首先获取待纠错实体[Title:Barnaby]和[Actor:David]。利用音形码算法对上述两个待纠错实体进行编码,分别得到编码[Title:FTPRNPRJ]和[Actor:TFTKRLN]。在音形码数据路中查找候选实体,具体的在音形码数据库中召回编码相似的实体[Title:Barnaby,Bridge,Barrage]和[Actor:David]。此时召回三个Title实体和一个Actor实体。之后根据知识图谱对召回的所有实体进行筛选。

[0067] 具体的,知识图谱是描述真实世界中存在的各种实体或概念及其关系的知识库,演员David主演的电影为Barrage,在知识图谱中的表现形式为[David,main character(主演),Barrage]。因此,此时经过知识图谱的筛选,得到结果实体[Title:Barrage]和[Actor:David]。将待纠错实体[Title:Barnaby]纠正为结果实体[Title:Barrage]。最后根据得到的结果实体,输出有效意图:搜索演员David主演的电影Barrage。

[0068] 本申请提供的用于多语言语义理解的实体纠错方法,针对多语言语义理解提供统一的框架,可以在缺少大规模训练数据的情况下,跨越不同语言的影响,实现对不同语言的文本进行实体纠错,从而提升语义理解和实体识别的准确率,提升多语言语音识别产品的性能,进而提升用户使用体验。

[0069] 在一些实施例中,可以事先针对不同的语言设置相应的音形码数据库。在通过音形码数据库查找与待纠错实体匹配的候选实体之前,先确定待纠错实体的语言类型,之后再在相应的音形码数据库中进行查找。例如,待纠错实体spider,首先确定该待纠错实体的语言类型为英文。因此可以在英文音形码数据库中查找与该待纠错实体匹配的候选实体。这样区分不同的语言,在根据音形码数据库召回候选实体时,能够提升查找效率。

[0070] 在一些实施例中,可以预先针对不同的业务设置相应的音形码数据库。在通过音形码数据库查找与待纠错实体匹配的候选实体之前,先确定待纠错实体的业务类型,之后再在相应的音形码数据库中进行查找。在具体的业务需求中往往只有部分类型的实体需要纠错,比如在电视语音助手业务下,只需要对视频名、人名和频道名进行纠错。

[0071] 如图6所示的示例,在构建与实际业务相关的音形码数据库时,利用单语言的音形码对收集到的不同类型的实体分别进行编码,将编码后的结果作为key值,原实体的文本作为value值存放在音形码数据库S中。其中示例的,Schannel表示与频道相关的音形码数据库,Stitle表示与视频名称相关的音形码数据库,Sactor表示与人名相关的音形码数据库。

[0072] 对于待纠错实体text,按照如下方式召回发音相似的候选实体E:

[0073]
$$E_{\text{phoner ic}} = S(F_{\text{phonetic}}(\text{text}))$$

[0074] 例如,待纠错的实体text为Barnaby时,其实体类型为title,利用音形码算法F编码后的key值为FTPRNPRJ后,在数据库Stitle中进行召回,Stitle(FTPRNPRJ)能够召回出存放在知识库中的三个实体 {Barnaby, Barrage, Bridge}, 利用该召回后的三个实体作为后续纠错的候选集。这样,根据不同的业务类型划分不同的音形码数据库,在根据音形码数据库召回候选实体时,能够提升查找效率。

[0075] 在多语言识别系统的实际应用中,上述方式召回的候选实体可能会存在数量过多或数量为零的情况,需要对召回的纠错候选实体进行后处理来提高系统的质量。对于大型的多语言识别系统,收集的候选实体数量是千万级别的,在音形码数据库中召回于待纠错实体发音相似的候选实体,可能会召回数量巨大的候选实体。过多的候选实体会给后续的纠错带来计算查询上的负担,浪费计算资源和时间。因此需要对召回的候选实体进行数量上的清洗。

[0076] 具体的,如果从音形码数据库中召回的候选实体数量大于数量阈值N,则采用基于编辑距离的排序方法对候选实体进行清洗。编辑距离(Minimum Edit Distance, MED)是指一个文本变换成另一个文本所需要的最小次数,用来衡量两个文本的差异度。当召回的候选实体中的数量超过了预先设置的阈值N时,计算候选实体中的每个实体与待纠错实体之间的编辑距离,按照编辑距离进行排序,选取TopN个编辑距离较小的候选实体进行后续纠错。

[0077] 在一些实施例中,如果从音形码数据库中召回的候选实体数量为零时,则从待纠错实体中截取子字符串。之后利用音形码算法对子字符串进行编码,最后在音形码数据库中查找与编码后的子字符串匹配的候选实体。

[0078] 示例性的,多语言语义理解模型可能会将search for film Red Dog中的title实体识别为film Red Dog。由于待纠错实体编码时会含有film的编码,因此会导致该待纠错实体在音形码数据库中召回的候选实体数量为零。此时可以截取得待纠错实体film Red Dog的子字符串Red Dog。对子字符串"Red Dog"编码后去数据库S中检索召回。这种通过子字符

串编码去数据库中再检索时采用递归的方式实现,找出能够匹配出结果的最长子字符串,以其候选结果作为最终音形码召回结果。

[0079] 上述实施例均为一条用户请求语句中只包含一种语言的情形,在实际应用中还可能存在一条用户请求语句中包含多种语言的情形。这种情形主要有两种原因:不同语言的用户存在一定区域相关性。例如中文和日语的用户存在地域相关性,马来语用户中因为华侨较多,也会在请求中携带部分中文;英语使用的广泛性,和媒体资源的丰富性,导致很多用户会在语音助手语言不为英语时,输入携带英语的业务请求。

[0080] 例如,法语用户会有这样的请求rechercher le (法语) film spider (英语)。现有的语音识别模型(Automatic Speech Recognition,ASR)在面对这种多语言混杂的情况时,准确率会有下降,会概率性将非本国语言的发音误识别成本国语言的文本。比如马来语用户的请求语音为“buka沙发”会被ASR模型误识别为文本“buka soufa”,单语言的音形码召回无法适用于该类场景,多语言语音助手系统在面对这类问题的准确率也较低。

[0081] 为了解决上述问题,本实施例基于上述实施例对实体纠错方法进行改进。具体的,当待纠错实体编码后在音形码数据库中未召回有效实体(可生成有效意图的实体)时,采用候选的其他语言的音形码算法进行二次召回。这里采用的确定候选语言的方式有两种:基于文本的字符确定文本可能属于的语言类型;基于先验知识推测该语言用户可能会使用的其它语言。

[0082] 示例性的,如图7所示的多语言音形码召回流程示意图。当法语用户请求语句为rechercher le film man时,识别的命名实体文本为man。实体文本man经过法语的音形码召回后的候选集E为[mania]。召回的实体文本mania不能生成有效意图。则采用其它语言的音形码算法F进行二次召回。

[0083] 首先基于实体文本字符分析man可能属于的语言为英语。同时基于法语用户的语言习惯,基于先验知识推测法语的邻近语言为英语、西班牙语、意大利语和葡萄牙语等语言(西班牙语、意大利语、葡萄牙语、英语等与法语属于同一个语系,并且地域上也存在相关性)。综合两方面的信息得出二次召回可采用的语言为英语。采用英语的音形码算法Fen(图7中的Ffr表示法语的音形码算法,Sfr表示法语的音形码数据库,Sen表示英语的音形码数据库)进行二次召回后得到的候选实体E为[man]。此时待可根据召回的实体文本man生成有效意图(例如搜索电影man)。

[0084] 上述实施例的改进方法根本上还是组合不同语言的音形码算法的方式,但是ASR模型在面对多语言混合请求时,其输出文本的错误是具有多样性和不确定性的,因此上述实施例的改进方法无法适用所有ASR识别错误的情形。

[0085] 为了解决上述问题,进一步改进的方法是在设计单语言音形码算法时考虑多语言的发音问题,即将发音相近的不同语言的文本经过各自的语言编码后的编码相近。例如,英文下编码“soufa”与中文下的“沙发”编码成相近的编码。在候选实体召回时,采用相似度计算召回其他语言下发音相近的候选实体。可以收集不同语言下发音相似的大批量文本数据对语义理解模型进行训练,语义理解模型可以是LABSE模型。

[0086] 编码模型的网络结构可以采用Transformer模型,采用对比学习的方式使得不同语言下发音相似的文本对经过模型编码后,两者的特征向量在编码空间距离较近,而不同语言下发音差距较大的文本对的特征向量,在编码空间距离较远。例如,英语下的第一实体

的特征向量和中文的第二实体特征向量存在于编码空间中。如果第一实体和第二实体的发音比较相似,则两者的特征向量在编码空间的距离较近。如果第一实体和第二实体的发音相差较大,则两者的特征向量在编码空间的距离较远。

[0087] 示例性的深度学习网络如图8所示,英文下的实体文本“sofa”与中文下的实体文本“沙发”发音相似,在经过各自的编码算法进行编码时,可以共享编码参数,使得得到的编码的特征向量在编码空间中距离较近。在查找候选实体时,如果筛选的候选实体“sofa”不能输出有效意图,则可以直接将候选实体“sofa”替换成距离较近的候选实体“沙发”。这样不仅可以进一步提升语义理解和实体识别的准确率,还能够缩短系统计算时间,提升语义理解的效率。

[0088] 在一些实施例中,在根据知识图谱对候选实体进行筛选时,如果知识图谱中存在与候选实体匹配的预存实体,则将预存实体确定为结果实体。如果知识图谱中不存在与候选实体匹配的预存实体,但是存在与候选实体中的子字符匹配的预存实体,则将预存实体确定为结果实体。需要说明的是,预存实体与候选实体匹配可以是完全相同,也可以是具有包含关系,本申请对此不作限制。

[0089] 示例性的,如果候选实体为film man,在知识图谱中不存在与候选实体匹配的预存实体,但是存在于子字符man匹配的预存实体,则将该预存实体man确定为最终的结果实体。

[0090] 在一些实施例中,还可以查询候选实体之间的关联关系对不合适候选实体进行剔除。示例性的,英文请求“search for man by Tom”经过语义理解模型后会输出两个类型的实体:电影名为man;演员名为Tom。这种类型的实体经过音形码编码后可能会召回多个候选电影名和演员名。例如,man可能会召回多个与man相关,但是与演员Tom不相关的电影。经过知识图谱查询,筛选出同时和实体man与实体Tom关联的候选实体,即只保留既与man相关,又与Tom相关的实体。

[0091] 在一些实施例中,如果经过上述实施例中的筛选过程,仍然存在多个结果实体,这样将会产生多个有效意图,不利于响应用户请求。因此可以采用TF-IDF (Term Frequency-Inverse Document Frequency,词频-逆文件频率)对多个结果实体进行评分排序,选取评分最高的作为最终的结果实体。具体的可以参考实体的其他知识属性,例如电影的评分,电影的年份和热度,年份较新,热度较高的实体评分较高。这样不仅能够有效的响应用户请求,而且能够使得最终响应结果更大程度上符合用户的预期。

[0092] 在一些实施例中,可以针对语音助手的常用指令构建待匹配的指令集,当用户的请求符合预设条件(例如长度小于预设长度)时,对用户的原始请求文本直接进行音形码编码,然后与待匹配的指令集进行匹配。如果有匹配结果,则直接将预存的语义结果给下游任务。此时构建的待匹配指令集即为音形码数据库。需要说明的是,这里的匹配可以是两者完全相同,也可以是两者之间具有包含关系,本申请对此不作限制。

[0093] 示例性的,如图9所示的具体实施例,当用户在语言为英文时的请求为“pose”,按照多语言语音助手的经验来看,用户的请求应该为“pause(暂停)”,但是语音转文本识别错误,识别为“pose”。此时多语言语义理解模型无法有效确定出pose的意图,无法下达给终端命令。利用音形码算法对“pose”进行编码后得到编码“PS”,之后在待匹配指令集中检索,得到编码同样为“PS”的“pause”。之后可以在知识图谱中进一步筛选,本实施例由于是在待匹

配指令集中查找的候选实体,因此也可以不在知识图谱中筛选,直接将“pause”作为结果实体。进一步在指令集中检索得到与“pause”关联的指令“control.pause”,并将该指令下发至终端执行。这样不仅可以进一步提升语义理解和实体识别的准确率,还能够缩短系统计算时间,提升语义理解的效率。

[0094] 本申请实施例提供一种用于多语言语义理解的实体纠错的智能设备,用于执行图2对应的实施例,如图10所示,本申请提供的智能设备至少包括:

[0095] 待纠错实体获取单元U1001,用于执行:获取待纠错实体,其中,所述待纠错实体为对用户输入的请求语句进行语义分析处理后得到的实体;

[0096] 编码单元U1002,用于执行:利用音形码算法对所述待纠错实体进行编码;

[0097] 候选实体查找单元U1003,用于执行:在音形码数据库中查找与编码后的所述待纠错实体匹配的候选实体;

[0098] 筛选单元U1004,用于执行:根据知识图谱对所述候选实体进行筛选,得到结果实体,其中,所述知识图谱描述所述候选实体之间的关联关系,所述结果实体为与其他所述候选实体具有关联关系的所述候选实体。

[0099] 基于上述实施例的实体纠错方法,本申请还提供一种用于多语言语义理解的多语言语音助手,如图11所示的多语言语音助手应用流程图,其中虚线方框中的内容为本申请提供的实体纠错方法。语音转文本模型接收用户请求语音后,将用户请求语音转换为文本,之后输入多语言语义理解模型,输出意图,该意图包括实体。

[0100] 如果输出的意图需要进行实体纠错,利用音形码算法对实体进行编码,并在音形码数据路中召回候选实体。之后经过知识图谱查询,得到结果实体。最后将知识图谱检索输出封装成统一的格式,方便下游终端执行命令。

[0101] 如果输出的意图不需要进行实体纠错,将实体文本与待匹配的指令集中的文本进行音形码匹配。如果实体文本与待匹配的指令集中的文本音形码匹配,则将待匹配的指令集中的文本封装成统一格式之后输出。如果实体文本与待匹配的指令集中的文本音形码不匹配,则将实体文本封装成统一格式之后输出。

[0102] 以上已经描述的包括本发明的实现的示例。为了描述要求保护主题的目的,当然不可能描述组件或方法的每一个可设想组合,但是要意识到,本创新的许多另外组合和置换是可能的。相应地,要求保护主题打算包含全部这类改变、修改和变化,其落入所附权利要求的精神和范围之内。此外,包括“摘要”中所述的事物的本申请的所示实现的以上描述并不是要详细列举或者将所公开实现局限于所公开的精确形式。虽然本申请中为了说明目的而描述具体实现和示例,但是如相关领域的技术人员能够认识的,被认为在这类实现和示例的范围之内的各种修改是可能的。

[0103] 此外,词语“示例”或“示范”在本申请中用来表示“用作示例、实例或说明”。本申请中描述为“示范”的任何方面或设计不一定被理解为相对其它方面或设计是优选或有利的。相反,词语“示例”或“示范”的使用打算以具体方式呈现概念。

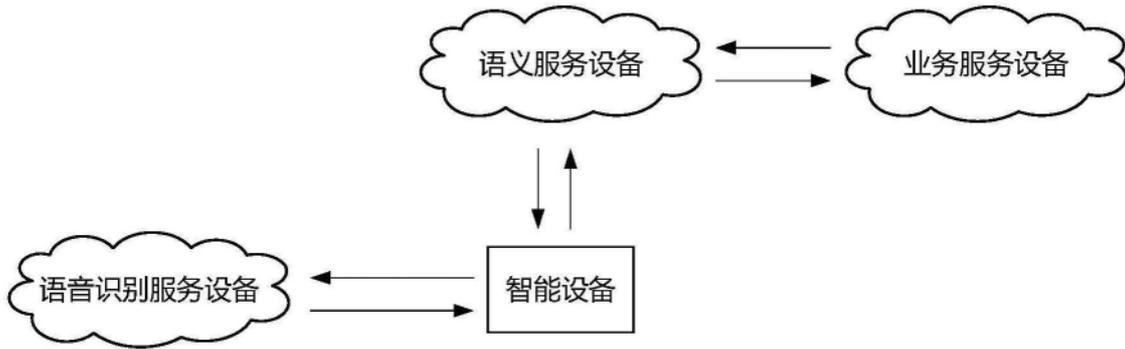


图1

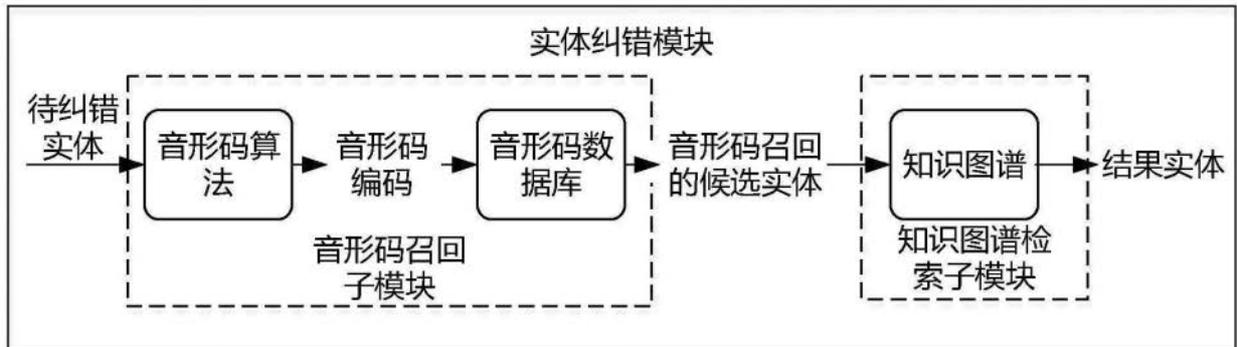


图2

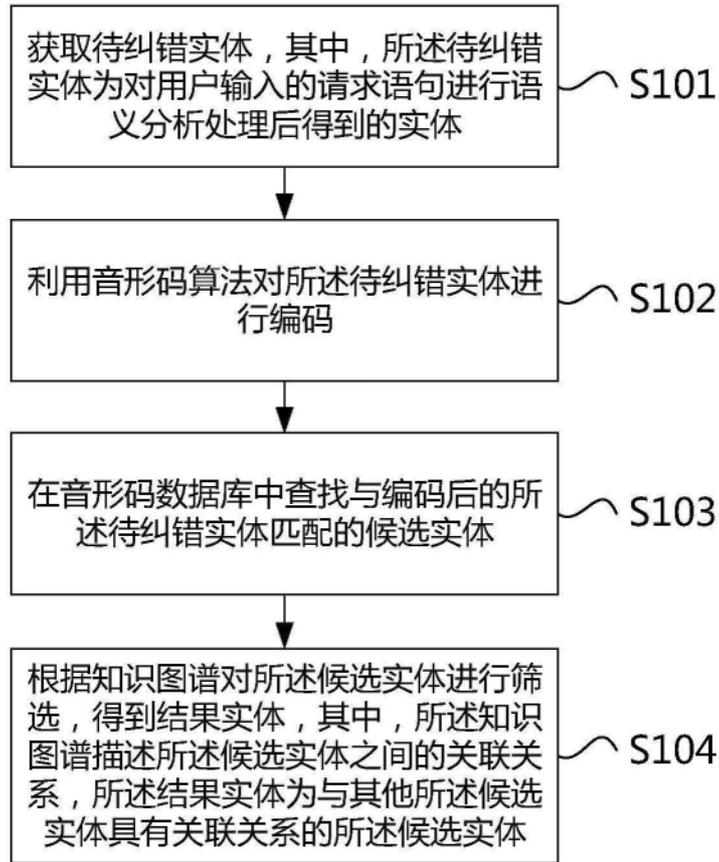


图3

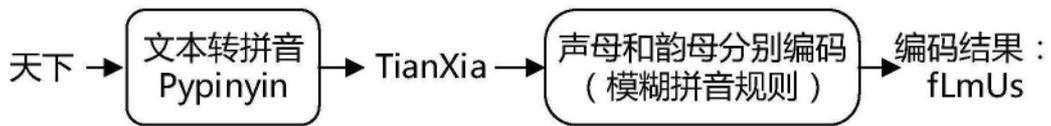


图4

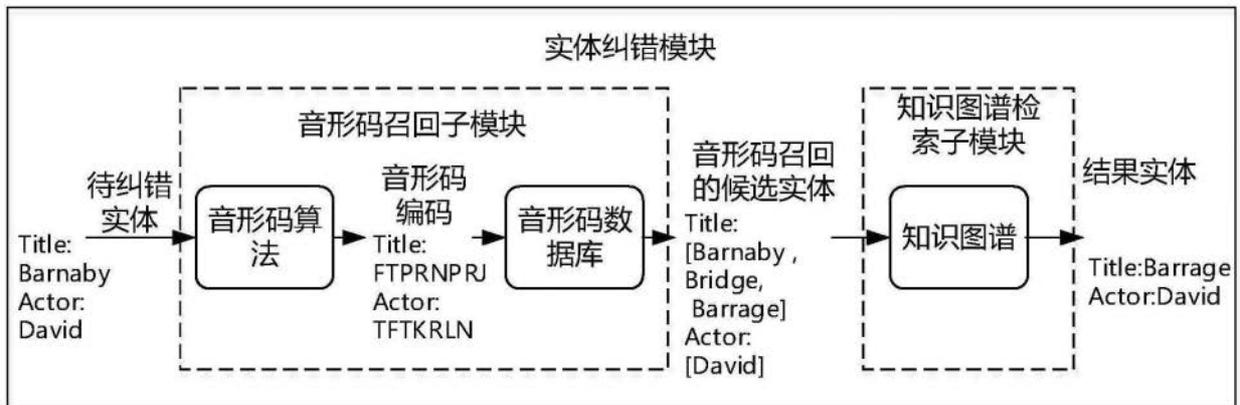


图5

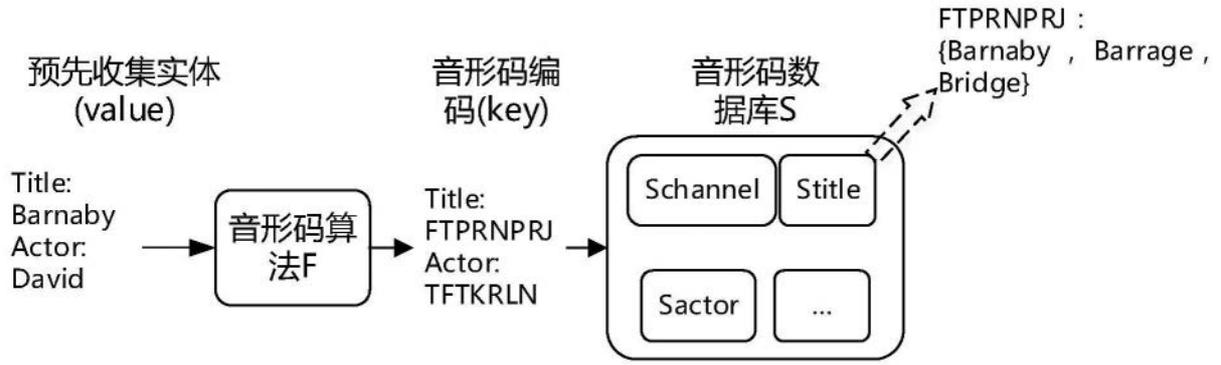


图6

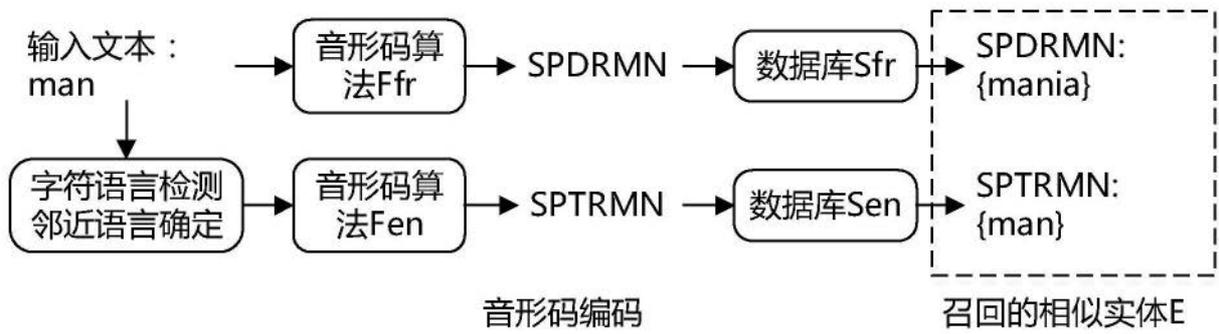


图7

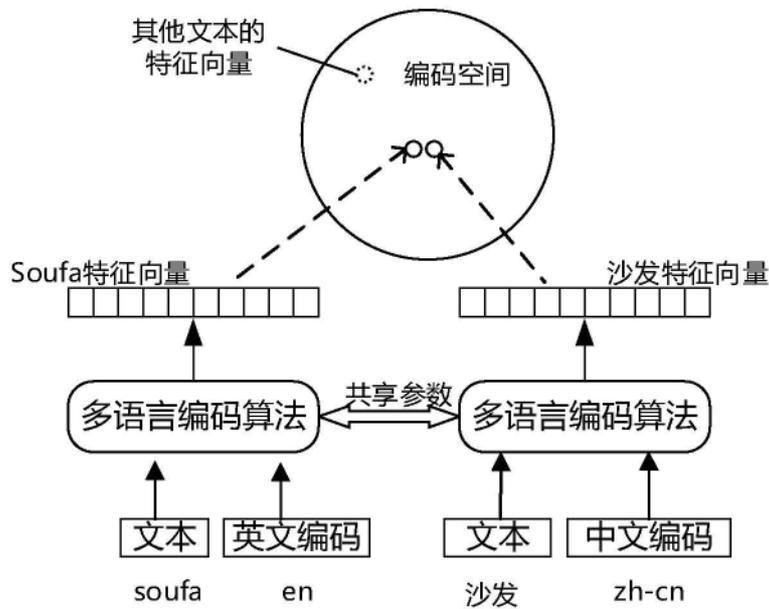


图8

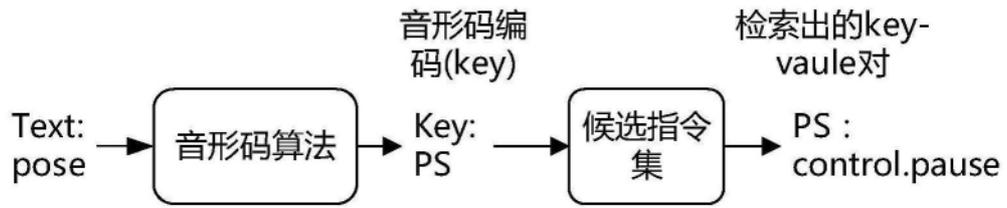


图9

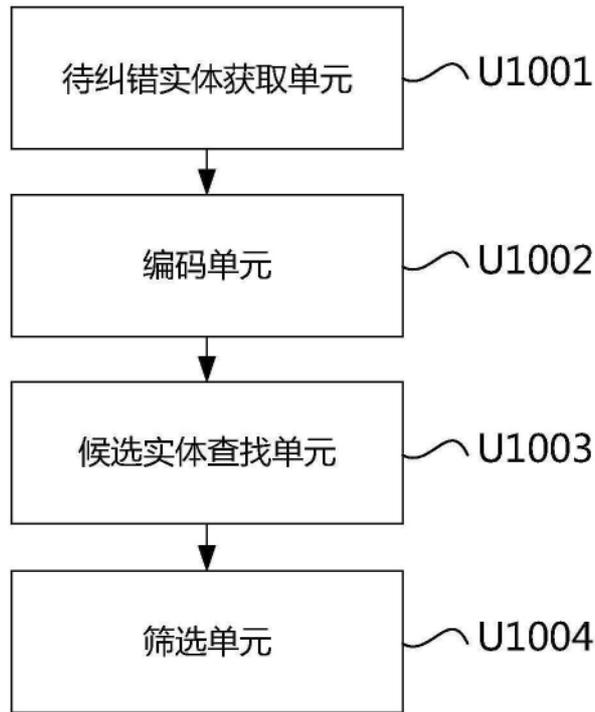


图10

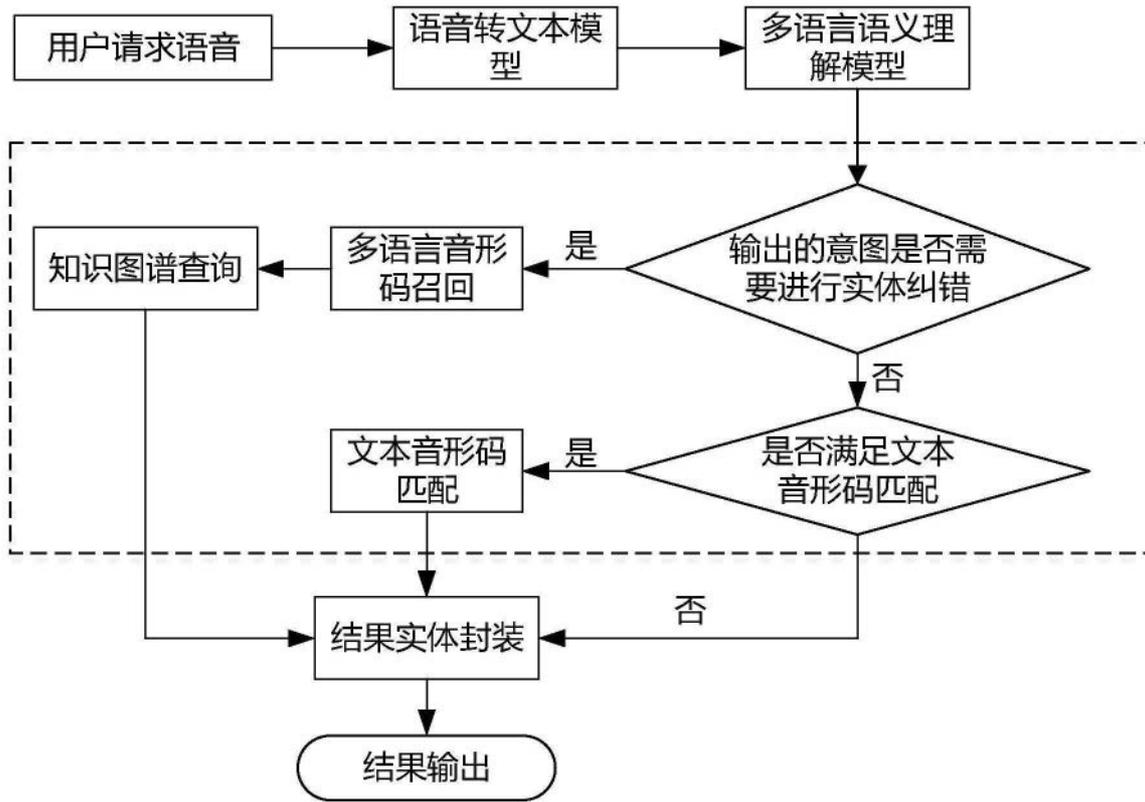


图11