



(12) 发明专利申请

(10) 申请公布号 CN 116597809 A

(43) 申请公布日 2023. 08. 15

(21) 申请号 202310484396.0

(22) 申请日 2023.04.28

(71) 申请人 北京捷通华声科技股份有限公司
地址 100193 北京市海淀区东北旺西路8号
9号楼3层2区318

(72) 发明人 李睿端 陈明 李健 武卫东

(74) 专利代理机构 北京润泽恒知识产权代理有限公司 11319
专利代理师 莎日娜

(51) Int. Cl.
G10L 13/10 (2013.01)

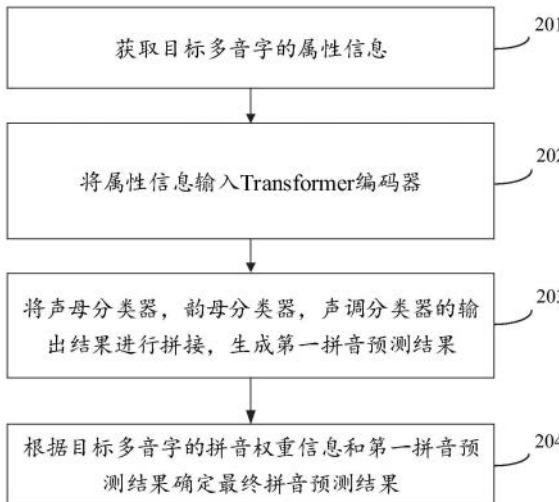
权利要求书2页 说明书10页 附图6页

(54) 发明名称

多音字消歧方法、装置、电子设备及可读存储介质

(57) 摘要

本发明实施例提供了一种多音字消歧方法、装置、电子设备及存储介质,包括:获取目标多音字包括掩码信息,分词信息、词性信息和语义信息的属性信息,将属性信息输入包括:声母分类器,韵母分类器,声调分类器的Transformer编码器后,将输出结果进行拼接,生成第一拼音预测结果,根据目标多音字的拼音权重信息和第一拼音预测结果确定最终拼音预测结果,本发明实施例通过将Transformer编码器拆解为三个分类器使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高多音字预测正确率,同时通过增加拼音权重信息,可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。



1. 一种多音字消歧方法,其特征在于,所述方法包括:
获取目标多音字的属性信息,所述属性信息包括所述目标多音字的掩码信息,分词信息、词性信息和语义信息;
将所述属性信息输入Transformer编码器,所述Transformer编码器包括:声母分类器,韵母分类器,声调分类器;
将所述声母分类器,所述韵母分类器,所述声调分类器的输出结果进行拼接,生成第一拼音预测结果;
根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。
2. 根据权利要求1所述的方法,其特征在于,所述获取目标多音字的属性信息之前,还包括:
预先设定包括目标数量第一多音字的多音字列表;
根据所述多音字列表获取目标数量的第二多音字;
将所述第二多音字进行拼音、分词和词性的标注,并进行校对。
3. 根据权利要求2所述的方法,其特征在于,所述将所述多音字进行拼音、分词和词性的标注,并进行校对之后,还包括:
若校对通过,则获取所述多音字列表的任一所述第一多音字对应的所述第二多音字的拼音标注信息;
根据所述拼音标注信息确认任一所述第一多音字的拼音权重信息。
4. 根据权利要求2所述的方法,其特征在于,所述将所述第二多音字进行拼音、分词和词性的标注,并进行校对之后,还包括:
获取待预测文本;
根据所述多音字列表对所述待预测文本进行判断,确定所述待预测文本的目标多音字。
5. 根据权利要求1所述的方法,其特征在于,所述将所述属性信息输入Transformer编码器之后,还包括:
通过所述声母分类器、所述属性信息确认所述目标多音字的声母预测结果;
通过所述韵母分类器、所述属性信息确认所述目标多音字的韵母预测结果;
通过所述声调分类器、所述属性信息确认所述目标多音字的声调预测结果。
6. 根据权利要求1所述的方法,其特征在于,所述根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果,包括:
获取所述目标多音字的拼音权重信息;
根据所述拼音权重信息计算所述第一拼音预测结果,以及所述目标多音字其他拼音信息的发音概率;
取所述发音概率中的最大值对应的拼音信息作为最终拼音预测结果。
7. 一种多音字消歧装置,其特征在于,所述装置包括:
第一获取模块,用于获取目标多音字的属性信息,所述属性信息包括所述目标多音字的掩码信息,分词信息、词性信息和语义信息;
第一输入模块,用于将所述属性信息输入Transformer编码器,所述Transformer编码

器包括:声母分类器,韵母分类器,声调分类器;

第一生成模块,用于将所述声母分类器,所述韵母分类器,所述声调分类器的输出结果进行拼接,生成第一拼音预测结果;

第一确定模块,用于根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。

8. 根据权利要求7所述的装置,其特征在于,所述多音字消歧装置还包括:

预先设定模块,用于预先设定包括目标数量第一多音字的多音字列表;

第二获取模块,用于根据所述多音字列表获取目标数量的第二多音字;

第一校对模块,用于将所述第二多音字进行拼音、分词和词性的标注,并进行校对。

9. 一种电子设备,其特征在于,包括:

处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1-6中任一所述的多音字消歧方法。

10. 一种可读存储介质,其特征在于,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行权利要求1-6中一个或多个所述的多音字消歧方法。

多音字消歧方法、装置、电子设备及可读存储介质

技术领域

[0001] 本发明属于语音处理技术领域,特别是涉及一种多音字消歧方法、装置、电子设备及可读存储介质。

背景技术

[0002] 语音合成(Text-to-speech,TTS)技术可以完成根据文本生成音频的任务。语音合成技术经历了多个发展阶段,但是无论是哪个阶段的合成技术,都离不开前端系统。前端系统一般包括语种判断模块、断句、分词、韵律预测、多音字消歧模块等。语音合成的前端系统中的多音字消歧模块对于整个合成效果起到非常关键的作用,因为如果多音字读错或者韵律停顿出现问题,会直接影响听者的感受,因此,优化语音合成前端系统的多音字消歧模块是亟须解决的问题。

[0003] 现有技术中多音字消歧可以通过词表加规则方式,或者通过模型方法来完成,前者结合通用词表及自定义词表,以及大量规则或专家知识来实现消歧,后者则通过常见的2071类的分类模型来完成,然而前者会存在维护费力以及规则和规则之间容易冲突或者相互影响的问题,后者对数据量要求很大,另外,样本数据大的话,会导致计算量变大,使得该方法存在一定应用上的限制;故现有技术中实现多音字消歧的方法存在多音字消歧效果差的问题。

发明内容

[0004] 本发明提供一种多音字消歧方法、装置、电子设备及可读存储介质,以便解决现有技术中实现多音字消歧的方法存在多音字消歧效果差的问题。

[0005] 为了解决上述技术问题,本发明是这样实现的:

[0006] 第一方面,本发明提供一种多音字消歧方法,所述方法包括:

[0007] 获取目标多音字的属性信息,所述属性信息包括所述目标多音字的掩码信息,分词信息、词性信息和语义信息;

[0008] 将所述属性信息输入Transformer编码器,所述Transformer编码器包括:声母分类器,韵母分类器,声调分类器;

[0009] 将所述声母分类器,所述韵母分类器,所述声调分类器的输出结果进行拼接,生成第一拼音预测结果;

[0010] 根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。

[0011] 进一步地,所述获取目标多音字的属性信息之前,还包括:

[0012] 预先设定包括目标数量第一多音字的多音字列表;

[0013] 根据所述多音字列表获取目标数量的第二多音字;

[0014] 将所述第二多音字进行拼音、分词和词性的标注,并进行校对。

[0015] 进一步地,所述将所述多音字进行拼音、分词和词性的标注,并进行校对之后,还

包括：

[0016] 若校对通过，则获取所述多音字列表的任一所述第一多音字对应的所述第二多音字的拼音标注信息；

[0017] 根据所述拼音标注信息确认任一所述第一多音字的拼音权重信息。

[0018] 进一步地，所述将所述第二多音字进行拼音、分词和词性的标注，并进行校对之后，还包括：

[0019] 获取待预测文本；

[0020] 根据所述多音字列表对所述待预测文本进行判断，确定所述待预测文本的目标多音字。

[0021] 进一步地，所述将所述属性信息输入Transformer编码器之后，还包括：

[0022] 通过所述声母分类器、所述属性信息确认所述目标多音字的声母预测结果；

[0023] 通过所述韵母分类器、所述属性信息确认所述目标多音字的韵母预测结果；

[0024] 通过所述声调分类器、所述属性信息确认所述目标多音字的声调预测结果。

[0025] 进一步地，所述根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果，包括：

[0026] 获取所述目标多音字的拼音权重信息；

[0027] 根据所述拼音权重信息计算所述第一拼音预测结果，以及所述目标多音字其他拼音信息的发音概率；

[0028] 取所述发音概率中的最大值对应的拼音信息作为最终拼音预测结果。

[0029] 第二方面，本发明提供一种多音字消歧装置，所述装置包括：

[0030] 第一获取模块，用于获取目标多音字的属性信息，所述属性信息包括所述目标多音字的掩码信息，分词信息、词性信息和语义信息；

[0031] 第一输入模块，用于将所述属性信息输入Transformer编码器，所述Transformer编码器包括：声母分类器，韵母分类器，声调分类器；

[0032] 第一生成模块，用于将所述声母分类器，所述韵母分类器，所述声调分类器的输出结果进行拼接，生成第一拼音预测结果；

[0033] 第一确定模块，用于根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。

[0034] 第三方面，本发明提供一种电子设备，包括：处理器、存储器以及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述程序时实现上述多音字消歧方法。

[0035] 第四方面，本发明提供一种可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行上述多音字消歧方法。

[0036] 在本发明实施例中通过获取目标多音字的属性信息，属性信息包括所述目标多音字的拼音信息、掩码信息，分词信息、词性信息和语义信息，使得在识别时获取更加准确的目标多音字的发音，将属性信息输入Transformer编码器，Transformer编码器包括：声母分类器，韵母分类器，声调分类器，将声母分类器，韵母分类器，声调分类器的输出结果进行拼接，生成第一拼音预测结果，通过将Transformer编码器拆解使得在数据量不多或者数据不平衡的情况下，声韵母模型可以得到充分训练，提高预测正确的概率，根据目标多音字的拼

音权重信息和第一拼音预测结果确定最终拼音预测结果,通过加入拼音权重信息,可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。本发明实施例通过将Transformer编码器拆解为三个分类器使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高多音字预测正确的概率,同时通过增加拼音权重信息(专家知识),可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。

附图说明

[0037] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0038] 图1是本发明实施例提供的一种包括多音字消歧方法的语音合成技术系统的架构示意图;

[0039] 图2是本发明实施例提供的一种多音字消歧方法的步骤流程图;

[0040] 图3是本发明实施例提供的另一种多音字消歧方法的步骤流程图;

[0041] 图4是图2所示的本发明实施例提供的一种多音字消歧方法的步骤204的步骤流程图;

[0042] 图5是本发明实施例提供的另一种多音字消歧方法的步骤流程图;

[0043] 图6是本发明实施例提供的一种多音字消歧装置的结构图;

[0044] 图7是本发明实施例提供的一种电子设备的结构图。

具体实施方式

[0045] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0046] 为解决上述现有技术中存在的问题,本发明的发明人对现有的语音合成技术进行深入了解,如图1所示的一种包括多音字消歧方法的语音合成技术系统的架构示意图,以中文为例,输入待预测文本后经过文本分析(包含断句、分词、文本正则化、命名实体识别等)后,进入韵律预测模块,接着标音模块会对这个输出结果进行音素标注(包含多音字消歧),得到的音素序列随后进行时长模型和声学模型的训练,训练后的结果过声码器,生成目标发言人的音频结果。

[0047] 其中,多音字消歧模块从本质上讲,就是一个分类任务,即从众多的拼音候选中“挑选”出最恰当的一个标注结果。基于此,本发明提供一种通过3个小的分类模型(声母分类器,韵母分类器,声调分类器),并加入专家知识(拼音权重信息)辅助提高预测准确率的多音字消歧方法。

[0048] 图2是本发明实施例提供的一种多音字消歧方法的步骤流程图,如图2所示,该方法可以包括:

[0049] 步骤201,获取目标多音字的属性信息。

[0050] 本发明实施例中的目标多音字的属性信息包括目标多音字的掩码信息,分词信息、词性信息和语义信息。其中,目标多音字的掩码信息是为了将目标多音字与非目标多音字进行区分,将非目标多音字在初始就Mask掉,以免后续非目标多音字的属性信息对预测结果造成影响,分词信息、词性信息和语义信息是为了帮助辨析发音,提高预测结果的准确性,其中分词信息可以使用自然语言处理模型进行编码和标注,词性信息可以通过预先设计好的映射表进行编码和标注,语义信息可以通过Bert预训练模型进行编码和标注,当然,本发明实施例还可以通过其他的深度学习模型或规则方法对分词信息、词性信息和语义信息进行编码和标注,本发明在此不作具体限定。

[0051] 示例的,输入待预测文本为“登录人行征信系统”,其中确定目标多音字为“行”,分词信息使用自然语言处理模型中常见的BIOS输入方式标注,即单字词标为4,词首标为1,词末标为3,词中标为2,因此待预测文本的分词信息为“1、3、1、3、1、3、1、3”,词性信息通过预先设计好的映射表进行编码和标注,如动词标记为5,名词标记为13,则待预测文本的词性信息为“5、5、13、13、13、13、13、13”,语义信息的编码则是根据预训练的语言模型获得:以目前常用的Bert为例,由于Bert预训练模型是由大量非监督文本训练所得,而且Bert对文本的编码是向量形式的,因此,文本中的语义信息可以得到很好地编码。

[0052] 步骤202,将属性信息输入Transformer编码器。

[0053] 本发明实施例中的Transformer编码器包括:声母分类器,韵母分类器,声调分类器,其中声母分类器包括27个分类、韵母分类器包括35个分类、声调分类器包括5个分类(阴平、阳平、上声、去声、轻声),将Transformer编码器拆解后可以使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练(同样的数据量下,三个小模型更容易得到充分训练),从而提高预测正确的概率。

[0054] 需要说明的是,属性信息输入到Transformer编码器后主要是根据属性信息从Transformer编码器的三个分类器中对目标多音字的声母、韵母和声调进行预测,得到第一拼音预测结果,具体地,将属性信息输入Transformer编码器之后,还包括:

[0055] 通过声母分类器、属性信息确认目标多音字的声母预测结果;

[0056] 通过韵母分类器、属性信息确认目标多音字的韵母预测结果;

[0057] 通过声调分类器、属性信息确认目标多音字的声调预测结果。

[0058] 示例的,目标多音字为“行”,将这一信息分别输入Transformer编码器包括的声母分类器,韵母分类器,声调分类器,通过目标多音字的属性信息从三个分类器中分别获取预测结果,即从声母分类器得到声母预测结果为“h”,从韵母分类器得到韵母预测结果为“ang”,从声调分类器得到声调预测结果为“2(阳平)”。

[0059] 步骤203,将声母分类器,韵母分类器,声调分类器的输出结果进行拼接,生成第一拼音预测结果。

[0060] 本发明实施例中的第一拼音预测结果是在全连接层上将声母分类器,韵母分类器,声调分类器分别得到的声母预测结果、韵母预测结果和声调预测结果,根据拼音的拼接规则将这三个预测结果进行拼接,得到的,需要说明的是,得到声母分类器,韵母分类器,声调分类器的输出结果后会这三个结果拼接后输入到全连接层,通过计算交叉熵,得到概率最高的标音结果作为第一拼音预测结果。

[0061] 示例的,根据上述步骤202的内容可知,当获取的声母预测结果为“h”、韵母预测结

果为“ang”和声调预测结果为“2(阳平)”,将这三个结果输入到深度学习后的全连接层,在全连接层中将三个部分的内容进行合理拼接得到的第一拼音预测结果为“hang2”。

[0062] 步骤204,根据目标多音字的拼音权重信息和第一拼音预测结果确定最终拼音预测结果。

[0063] 本发明实施例通过属性信息和声母分类器,韵母分类器,声调分类器得到第一拼音预测结果后,还会获取目标多音字的拼音权重信息,因为不同目标多音字的多个发音,在自然语言中出现的概率差异很大,示例的,以17w句多音字标注数据为例(数据来源,网络随机抓取)，“长”的两个发音的分布就相对平均(chang2:7613次;zhang3:5973),权重可以设定为0.5和0.5,而“正”的两个发音的比例就极度偏移(zheng1:20次;zhang4:8342次)此时权重设定为0.01和0.99,因此如果根据属性信息分析得到的第一拼音预测结果与实际不符合时,还可以通过拼音权重信息来提前限制好可能的多音字读音,使得整个模型训练更加准确。

[0064] 在本发明实施例中通过获取目标多音字的属性信息,属性信息包括所述目标多音字掩码信息,分词信息、词性信息和语义信息,使得在后续识别时获取更加准确的目标多音字的发音,将属性信息输入Transformer编码器,Transformer编码器包括:声母分类器,韵母分类器,声调分类器,将声母分类器,韵母分类器,声调分类器的输出结果进行拼接,生成第一拼音预测结果,通过将Transformer编码器拆解使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高了根据属性信息预测正确的概率,根据目标多音字的拼音权重信息和第一拼音预测结果确定最终拼音预测结果,通过加入拼音权重信息,可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。本发明实施例通过将Transformer编码器拆解为三个分类器使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高多音字预测正确的概率,同时通过增加拼音权重信息(专家知识),可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。

[0065] 图3是本发明实施例提供的另一种多音字消歧方法的步骤流程图,本实施例公开的多音字消歧方法其步骤与如图2所示的基本相同,区别在于,在步骤201之前,还可以包括:

[0066] 步骤301,预先设定包括目标数量第一多音字的多音字列表。

[0067] 本发明实施例中为了准确识别出待预测文本中的多音字会预先设定一组多音字列表,其中的第一多音字可以是“行、卡、为、还、长……”,目标数量可以是100,也可以是120,本发明在此不做具体限定。

[0068] 步骤302,根据多音字列表获取目标数量的第二多音字。

[0069] 本发明实施例中的目标数量可以是5000,也可以是8000,本发明在此不做具体限定,第二多音字与第一多音字对应,只是第二多音字是从网络数据中随机爬取的目标数量的句子中获取的,所以第二多音字会有重复。

[0070] 示例的,多音字列表中包括第一多音字“行”,从网络爬取5000条数据中,有500条句子里包括这个“行”,所以第二多音字的“行”就有500个。

[0071] 步骤303,将第二多音字进行拼音、分词和词性的标注,并进行校对。

[0072] 本发明实施例中在获取第二多音字的时候会获取一个文本,所以可以根据文本的含义先进行自动拼音、分词、词性的标注后再进行人工校对;也可爬取语音数据,识别转

写后进行校对。

[0073] 需要说明的是,分词信息和词性信息均可帮助辨析发音。以“登录人行征信系统”和“白雪给人行道盖了床白被子”为例,如果能将分词分对(“人行”和“人行道”),即可很容易标注正确的读音即人行读音选项是hang2,人行道读音选项是xing2(词的差异很大,消歧难度低)。再比如,“这间房间朝阳”和“朝阳产业发展之我见”,“朝阳”的读音仅靠分词无法辨音,而需辅以词性信息,即“房间朝阳”中“朝阳”是动词,所以读音选项是chao2,而“朝阳产业”中“朝阳”是形容词,所以读音选项是zhao1。其中,1指的是该字符读音的声调为阴平(第一声),2指的是该字符读音的声调为阳平(第二声),此外,3指的是该字符读音的声调为上声(第三声),4指的是该字符读音的声调为去声(第四声),5指的是该字符读音的声调为轻声。

[0074] 除此之外,在校对通过后,还会针对多音字列表的每一个多音字计算拼音权重信息,为了保证拼音权重信息的合理性,会使计算得到的拼音权重信息更接近自然语言的真实比例,示例的,以多音字列表的“长”为例,5000句多音字标注数据中,包括“长”的句子有500条,其中两个发音的分布为chang2:240次;zhang3:260),因为分布比较平均,所以权重可以设定为0.5和0.5。

[0075] 具体地,将多音字进行拼音、分词和词性的标注,并进行校对之后,还包括:

[0076] 若校对通过,则获取多音字列表的任一第一多音字对应的第二多音字的拼音标注信息;

[0077] 根据拼音标注信息确认任一第一多音字的拼音权重信息。

[0078] 需要说明的是,本发明实施例的设置的的多音字列表不仅可以设定多音字的拼音权重信息还可以确定待预测文本中的目标多音字,即将待预测文本的每一个字与多音字列表中的内容进行比对,当比对一致时,确认此时的汉字为目标多音字。

[0079] 具体地,将第二多音字进行拼音、分词和词性的标注,并进行校对之后,还包括:

[0080] 步骤304,获取待预测文本。

[0081] 步骤305,根据多音字列表对待预测文本进行判断,确定待预测文本的目标多音字。

[0082] 在本发明实施例中通过获取目标多音字的属性信息,属性信息包括所述目标多音字掩码信息,分词信息、词性信息和语义信息,使得在后续识别时获取更加准确的目标多音字的发音,将属性信息输入Transformer编码器,Transformer编码器包括:声母分类器,韵母分类器,声调分类器,将声母分类器,韵母分类器,声调分类器的输出结果进行拼接,生成第一拼音预测结果,通过将Transformer编码器拆解使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高了根据属性信息预测正确的概率,根据目标多音字的拼音权重信息和第一拼音预测结果确定最终拼音预测结果,通过加入拼音权重信息,可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。本发明实施例通过将Transformer编码器拆解为三个分类器使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高多音字预测正确的概率,同时通过增加拼音权重信息(专家知识),可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。同时通过预先设置的多音字列表可以确定待预测文本的目标多音字,从而使得后续在添加掩码信息时将目标多音字与非目标多音字进行区分,将非目标多音字在初始就Mask掉,以

免后续非目标多音字的属性信息对预测结果造成影响。

[0083] 图4是图2所示的本发明实施例提供的一种多音字消歧方法的步骤204的一种可行实现方案,具体包括:

[0084] 步骤401,获取目标多音字的拼音权重信息。

[0085] 本发明实施例中在通过属性信息和声母分类器,韵母分类器,声调分类器得到第一拼音预测结果后,还会获取目标多音字的拼音权重信息,因为不同目标多音字的多个发音,在自然语言中出现的概率差异很大,此外权重信息的计算也可考虑行业信息,例如银行领域应用的多音字,可以赋予“行hang2”更高的权重。

[0086] 步骤402,根据拼音权重信息计算第一拼音预测结果,以及目标多音字其他拼音信息的发音概率。

[0087] 本发明实施例将目标多音字的拼音权重输入到深度学习模型中得到目标多音字的第一拼音预测结果的权重占比,而后与目标多音字其他拼音信息的权重占比融合计算得到目标多音字所有拼音信息的发音概率。

[0088] 步骤403,取发音概率中的最大值对应的拼音信息作为最终拼音预测结果。

[0089] 本发明实施例中将通过属性信息和Transformer编码器的三个分类器得到第一预测结果,而后通过增加拼音权重信息(专家知识),将拼音权重信息和经过全连接层得到的第一预测结果一起经过softmax计算多元交叉熵,取概率最高的拼音为最终拼音预测结果,其中,拼音权重信息可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。

[0090] 图5是本发明实施例提供的另一种多音字消歧方法的步骤流程图,具体包括:

[0091] 本发明实施例在获取待预测文本后,获取掩码Mask信息,分词信息和词性信息,将这些信息输入预训练语言模型得到语义信息,将掩码Mask信息,分词信息,词性信息,语义信息统称为属性信息,将属性信息输入Transformer编码器的三个分类器:声母分类器、韵母分类器和声调分类器,得到输出结果后进行拼接发送到全连接层,通过计算交叉熵,得到概率最高的标音结果作为第一拼音预测结果,而后通过拼音权重信息和第一拼音预测结果进行最终多元交叉熵计算得到最终拼音预测结果。

[0092] 图6是本发明实施例提供的一种多音字消歧装置的结构图,该装置可以包括:

[0093] 第一获取模块501,用于获取目标多音字的属性信息,所述属性信息包括所述目标多音字的掩码信息,分词信息、词性信息和语义信息。

[0094] 第一输入模块502,用于将所述属性信息输入Transformer编码器,所述Transformer编码器包括:声母分类器,韵母分类器,声调分类器。

[0095] 第一生成模块503,用于将所述声母分类器,所述韵母分类器,所述声调分类器的输出结果进行拼接,生成第一拼音预测结果。

[0096] 第一确定模块504,用于根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。

[0097] 可选地,所述多音字消歧装置还包括:

[0098] 预先设定模块,用于预先设定包括目标数量第一多音字的多音字列表。

[0099] 第二获取模块,用于根据所述多音字列表获取目标数量的第二多音字。

[0100] 第一校对模块,用于将所述第二多音字进行拼音、分词和词性的标注,并进行校

对。

[0101] 第三获取模块,用于若校对通过,则获取所述多音字列表的任一所述第一多音字对应的所述第二多音字的拼音标注信息。

[0102] 第一确认模块,用于根据所述拼音标注信息确认任一所述第一多音字的拼音权重信息。

[0103] 第四获取模块,用于获取待预测文本。

[0104] 第二确定模块,用于根据所述多音字列表对所述待预测文本进行判断,确定所述待预测文本的目标多音字。

[0105] 第二确认模块,用于通过所述声母分类器、所述属性信息确认所述目标多音字的声母预测结果。

[0106] 第三确认模块,用于通过所述韵母分类器、所述属性信息确认所述目标多音字的韵母预测结果。

[0107] 第四确认模块,用于通过所述声调分类器、所述属性信息确认所述目标多音字的声调预测结果。

[0108] 可选地,第一确定模块504还包括:

[0109] 第一获取子模块,用于获取所述目标多音字的拼音权重信息。

[0110] 第一计算子模块,用于根据所述拼音权重信息计算所述第一拼音预测结果,以及所述目标多音字其他拼音信息的发音概率。

[0111] 最终拼音预测子模块,用于取所述发音概率中的最大值对应的拼音信息作为最终拼音预测结果。

[0112] 在本发明实施例中通过获取目标多音字的属性信息,属性信息包括所述目标多音字掩码信息,分词信息、词性信息和语义信息,使得在后续识别时获取更加准确的目标多音字的发音,将属性信息输入Transformer编码器,Transformer编码器包括:声母分类器,韵母分类器,声调分类器,将声母分类器,韵母分类器,声调分类器的输出结果进行拼接,生成第一拼音预测结果,通过将Transformer编码器拆解使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高了根据属性信息预测正确的概率,根据目标多音字的拼音权重信息和第一拼音预测结果确定最终拼音预测结果,通过加入拼音权重信息,可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。本发明实施例通过将Transformer编码器拆解为三个分类器使得在数据量不多或者数据不平衡的情况下,声韵母模型可以得到充分训练,提高多音字预测正确的概率,同时通过增加拼音权重信息(专家知识),可以提前限制好可能的多音字读音,使得多音字消歧的预测结果更加准确。同时通过预先设置的多音字列表可以确定待预测文本的目标多音字,从而使得后续在添加掩码信息时将目标多音字与非目标多音字进行区分,将非目标多音字在初始就Mask掉,以免后续非目标多音字的属性信息对预测结果造成影响。

[0113] 本发明还提供了一种电子设备,图6是本发明实施例提供的一种电子设备的结构框图,参见图6,包括处理器601、通信接口602、存储器603和通信总线604,其中,处理器601,通信接口602,存储器603通过通信总线604完成相互间的通信,

[0114] 存储器603,用于存放计算机程序;

[0115] 处理器601,用于执行存储器603上所存放的程序时,实现如下步骤:

[0116] 获取目标多音字的属性信息,所述属性信息包括所述目标多音字的掩码信息,分词信息、词性信息和语义信息;

[0117] 将所述属性信息输入Transformer编码器,所述Transformer编码器包括:声母分类器,韵母分类器,声调分类器;

[0118] 将所述声母分类器,所述韵母分类器,所述声调分类器的输出结果进行拼接,生成第一拼音预测结果;

[0119] 根据所述目标多音字的拼音权重信息和所述第一拼音预测结果确定最终拼音预测结果。

[0120] 本发明还提供了一种可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行前述实施例的多音字消歧方法。

[0121] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0122] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其他设备固有相关。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0123] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0124] 类似地,应当理解,为了精简本发明并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图,或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0125] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0126] 本发明的各个部件实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明的排序设备中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序。这样的实现本发明的程序可以存储在计算机可读介质上,或者可以

具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0127] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

[0128] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0129] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所做的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

[0130] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求的保护范围为准。

[0131] 需要说明的是,本申请实施例中获取各种数据相关过程,都是在遵照所在地国家相应的数据保护法规政策的前提下,并获得由相应装置所有者给予授权的情况下进行的。

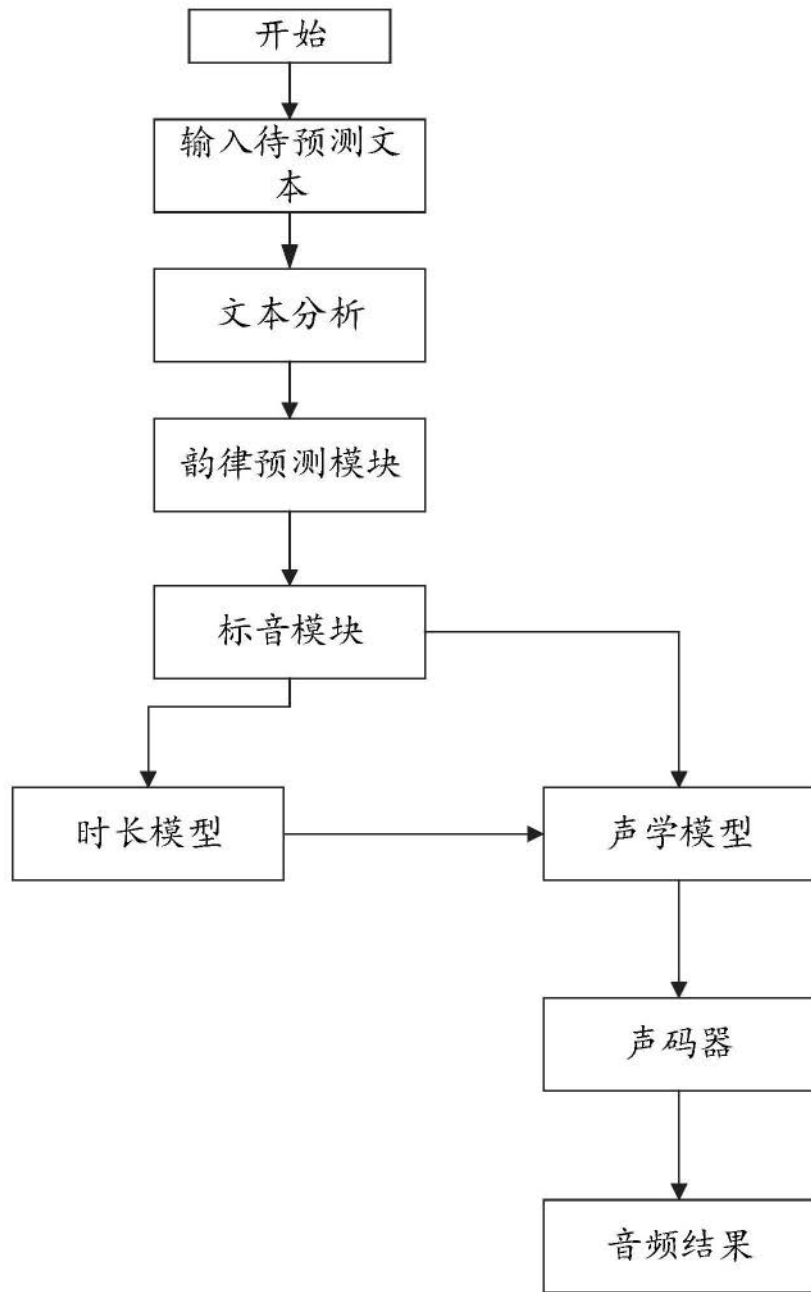


图1

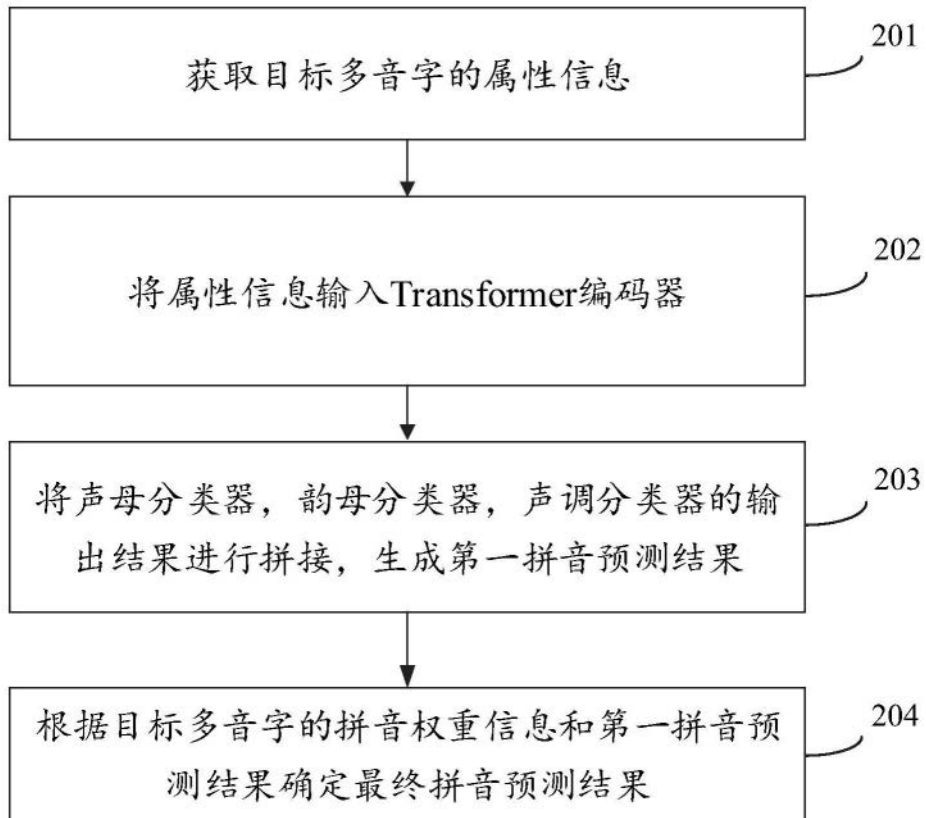


图2

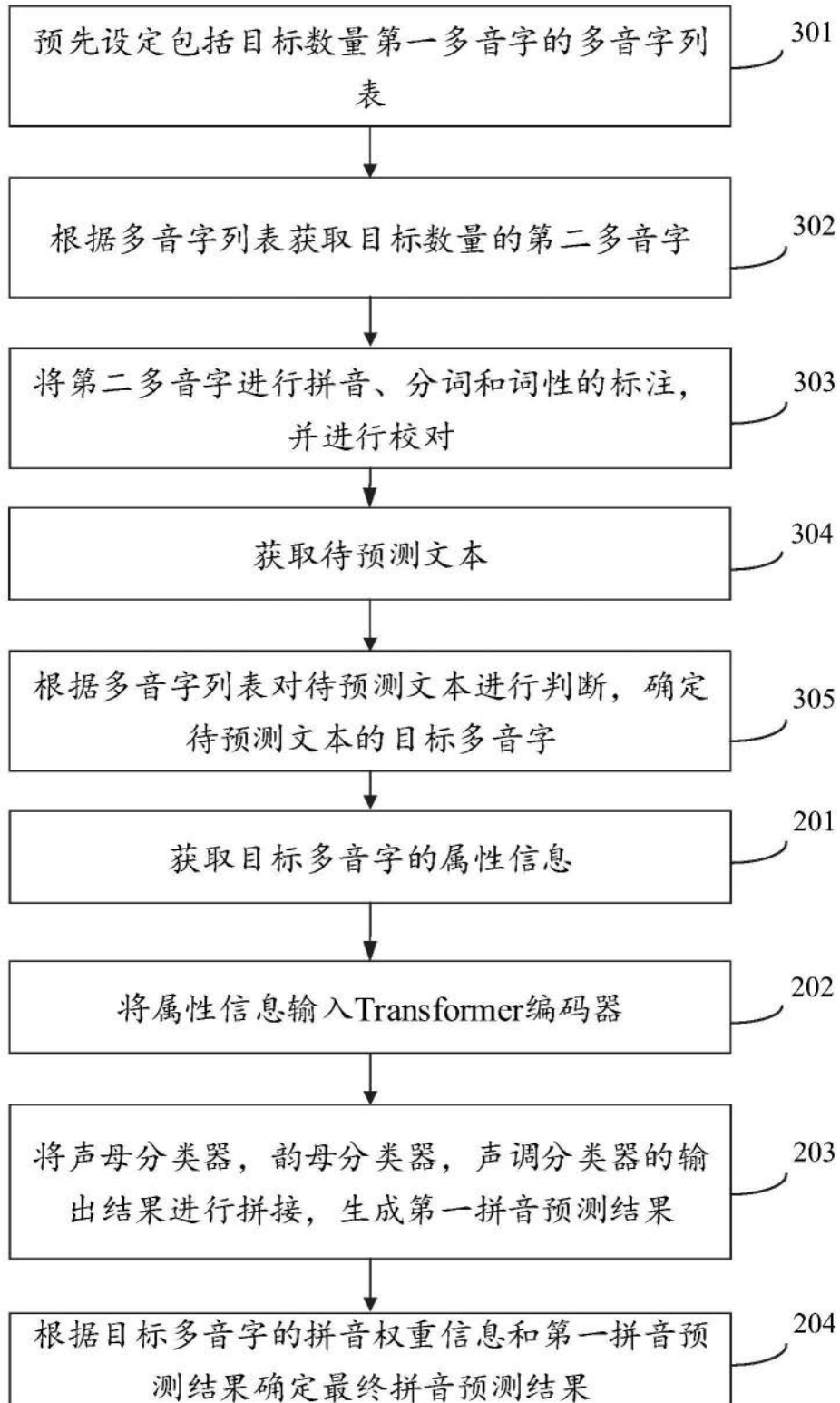


图3

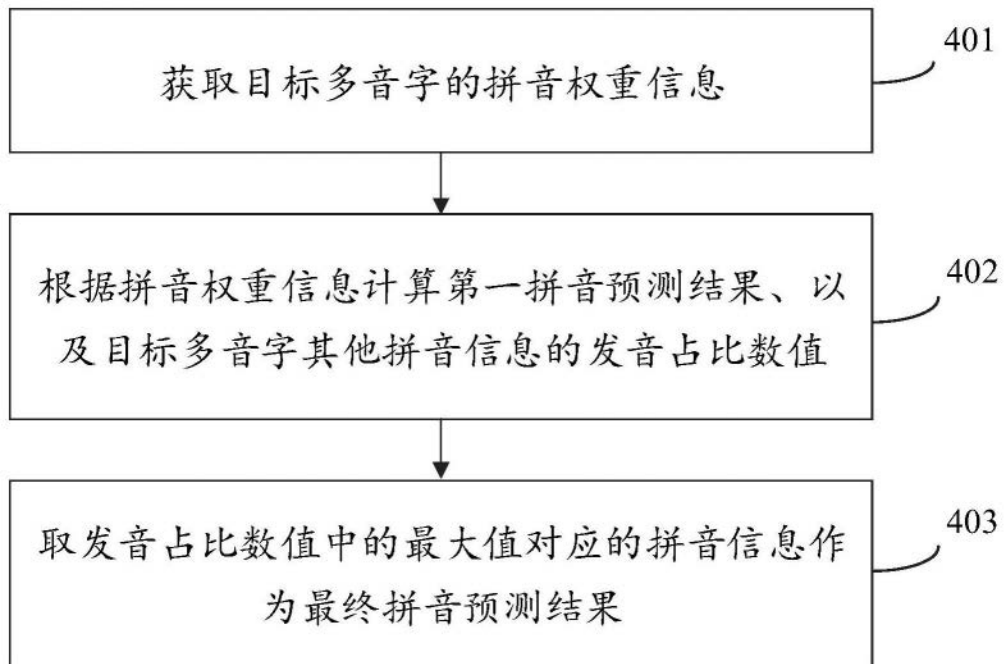


图4

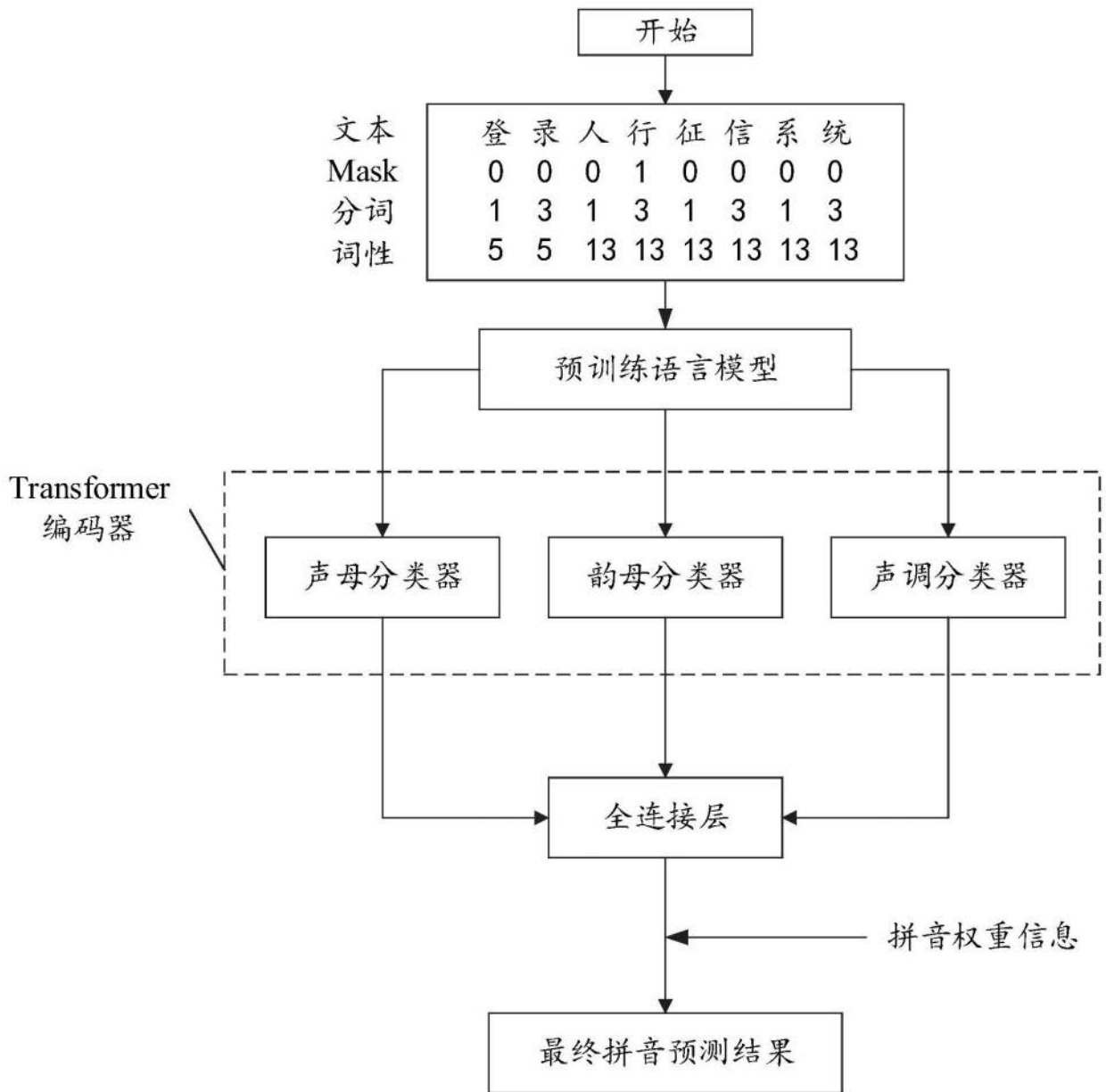


图5

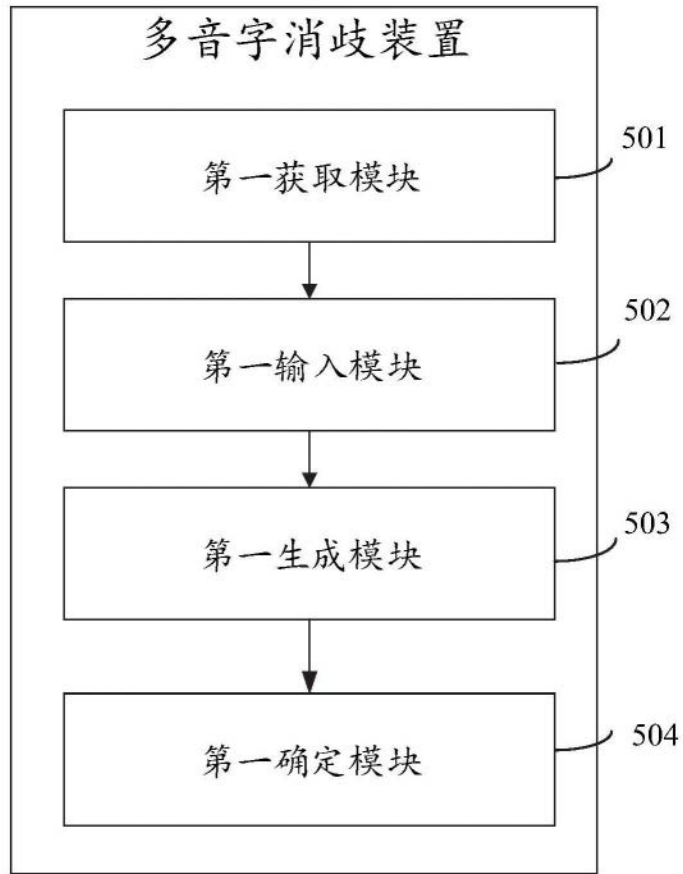


图6

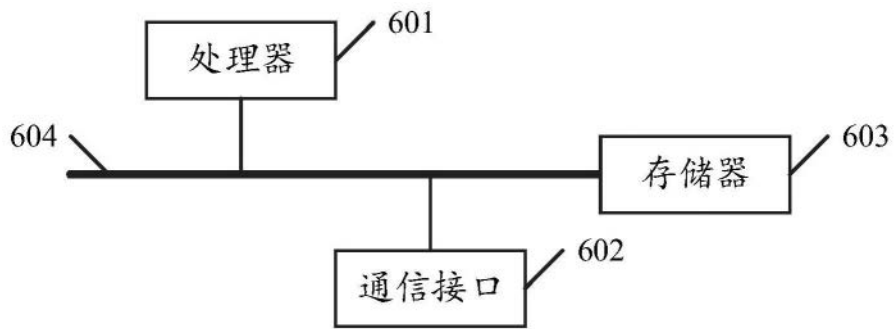


图7