(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118840996 A
(43)申请公布日 2024.10.25

(21)申请号 202410846834.8
(22)申请日 2024.06.27
(71)申请人 合肥智能语音创新发展有限公司
地址 230088 安徽省合肥市高新区习友路3333号中国(合肥)国际智能语音产业园A区2号科研楼1501室
(72)发明人 李沛霖 朱荣华 蔡明琦 方昕 吴江照 高建清
(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 侯珊
(51)Int.Cl.
G10L 13/08(2013.01)
G10L 19/00(2013.01)
G10L 25/30(2013.01)
G06F 40/284(2020.01)

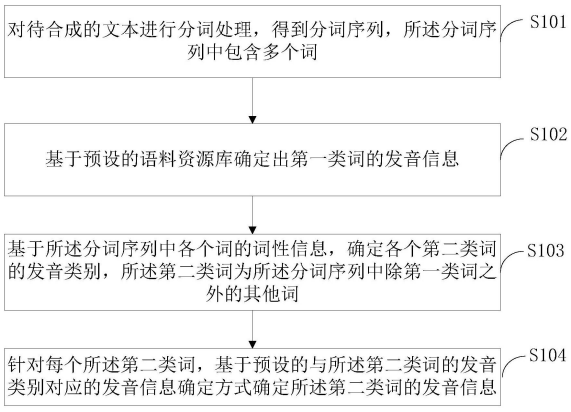
权利要求书3页 说明书14页 附图3页

(54)发明名称

一种发音预测方法及相关装置

(57)摘要

本申请公开了一种发音预测方法及相关装置,首先对待合成的文本进行分词处理,得到分词序列,对于分词序列中的第一类词,将基于预设的语料资源库确定出第一类词的发音信息,对于分词序列中除第一类词之外的第二类词,基于分词序列中各个词的词性信息,确定其发音类别,基于与其发音类别对应的发音信息确定方式确定其发音信息。在本申请中,结合语料资源库以及预设的各发音类别对应的发音信息确定方式,能够涵盖各种情况下的发音信息确定,因此,能够准确确定出待合成文本中各词的发音信息,进而能够提升语音合成的效果。



CN 118840996 A

1.一种发音预测方法,其特征在于,包括:

对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词;

基于预设的语料资源库确定出第一类词的发音信息,

基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,所述第二类词为所述分词序列中除第一类词之外的其他词;

针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

2.根据权利要求1所述的方法,其特征在于,所述基于预设的语料资源库确定出第一类词的发音信息,包括:

循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的发音信息,得到该词的发音信息查找结果,该词的发音信息查找结果用于指示所述语料资源库中与该词对应的发音信息;

基于各词的发音信息查找结果确定所述第一类词以及所述第一类词的发音信息;所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词,从所述语料资源库确定的发音信息为所述第一类词的发音信息。

3.根据权利要求2所述的方法,其特征在于,所述基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,包括:

循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的词性信息,得到该词的词性信息查找结果,该词的词性信息查找结果用于指示所述语料资源库中与该词对应的词性信息;

基于各词的所述发音信息查找结果,生成发音信息序列以及发音信息注意力掩码序列;

基于各词的所述词性信息查找结果,生成词性信息序列以及词性信息注意力掩码序列;

基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别。

4.根据权利要求3所述的方法,其特征在于,所述基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别,包括:

将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别;

其中,所述词性及发音类别预测模型是利用训练用分词序列、训练用发音信息序列、训练用发音信息注意力掩码序列、训练用词性信息序列以及训练用词性信息注意力掩码序列为训练样本,以所述训练用发音信息序列对应的发音类别标签序列以及所述训练用词性信息序列对应的词性信息标签序列为样本标签,以所述词性及发音类别预测模型输出的词性预测结果趋近于所述词性信息标签序列以及输出的发音类别预测结果趋近于所述发音类别标签序列为训练目标训练得到的。

5.根据权利要求4所述的方法,其特征在于,所述将所述分词序列、所述发音信息序列、

所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别,包括:

对所述分词序列进行特征提取处理,得到分词特征向量序列;

对所述发音信息序列进行特征提取处理,得到发音信息特征向量序列;

对所述词性信息序列进行特征提取处理,得到词性信息特征向量序列;

将所述分词特征向量序列、所述发音信息特征向量序列、所述词性信息特征向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列,进行特征融合处理,得到融合特征向量序列;

将所述融合特征向量序列输入所述词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果;

从所述分词序列中各个词的发音类别预测结果中,确定各个所述第二类词的发音类别。

6.根据权利要求1所述的方法,其特征在于,所述基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息,包括:

获取预设的各发音类别对应的发音预测模块;

从各发音预测模块中,确定目标发音预测模块,所述目标发音预测模块为所述各发音预测模块中与所述第二类词的发音类别对应的发音预测模块;

将所述文本或所述第二类词提供给所述目标发音预测模块,得到所述目标发音预测模块输出的所述第二类词的发音信息。

7.一种发音预测装置,其特征在于,包括:

分词处理单元,用于对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词;

第一发音信息确定单元,用于基于预设的语料资源库确定出第一类词的发音信息;

发音类别确定单元,用于基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,所述第二类词为所述分词序列中除第一类词之外的其他词;

第二发音信息确定单元,用于针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

8.一种计算机程序产品,其特征在于,包括计算机可读指令,当所述计算机可读指令在电子设备上运行时,使得所述电子设备实现如权利要求1至6中任意一项所述的发音预测方法。

9.一种电子设备,其特征在于,包括至少一个处理器和与所述处理器连接的存储器,其中:

所述存储器用于存储计算机程序;

所述处理器用于执行所述计算机程序,以使所述电子设备能够实现如权利要求1至6中任意一项所述的发音预测方法。

10.一种计算机存储介质,其特征在于,所述计算机存储介质承载有一个或多个计算机程序,当所述一个或多个计算机程序被电子设备执行时,能够使所述电子设备实现如权利

要求1至6中任意一项所述的发音预测方法。

一种发音预测方法及相关装置

技术领域

[0001] 本申请涉及语音合成技术领域,尤其涉及一种发音预测方法及相关装置。

背景技术

[0002] 目前的语音合成系统,首先会对待合成的文本进行分词,得到一个个独立的单词,再对各单词进行发音预测,确定各单词的发音信息,最后根据各单词的发音信息进行合成。其中,各单词的发音信息准确度对最终合成的效果起着至关重要的作用。

[0003] 现有的发音预测方案一般采用检索语料资源库的方式,即预先构建语料资源库,语料资源库中收录有多个词的发音信息。在对分词后的单词进行发音预测时,可以直接从语料资源库中确定单词的发音信息。但是,分词后的某些词,语料资源库中收录的发音信息可能有多种,这种情况下,将无法准确确定这些词的发音信息。另外,还有些词,语料库中可能并未收录,这种情况下,现有的发音预测方案会采用其他手段(如,跳过本词不发音或填充静音段、采用按字母读、使用C45决策树等机器学习方法进行发音预测中的任一手段)对这些词的发音信息进行预测,但是这些词的形式多样,采用的手段可能无法准确确定出某些形式的词的发音信息。

[0004] 因此,如何提供一种发音预测方法,以能够准确确定出待合成文本中各词的发音信息,进而提升语音合成的效果,成为本领域技术人员亟待解决的技术问题。

发明内容

[0005] 鉴于上述问题,本申请提供了一种发音预测方法及相关装置,以实现准确确定出待合成文本中各词的发音信息,进而提升语音合成的效果的目的。具体方案如下:

[0006] 本申请第一方面提供一种发音预测方法,包括:

[0007] 对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词;

[0008] 基于预设的语料资源库确定出第一类词的发音信息;

[0009] 基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,所述第二类词为所述分词序列中除第一类词之外的其他词;

[0010] 针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

[0011] 在一种可能的实现中,所述基于预设的语料资源库确定出第一类词的发音信息,包括:

[0012] 循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的发音信息,得到该词的发音信息查找结果,该词的发音信息查找结果用于指示所述语料资源库中与该词对应的发音信息;

[0013] 基于各词的发音信息查找结果确定所述第一类词以及所述第一类词的发音信息;所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词,从所述语料资源库确定的发音信息为所述第一类词的发音信息。

[0014] 在一种可能的实现中,所述基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,包括:

[0015] 循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的词性信息,得到该词的词性信息查找结果,该词的词性信息查找结果用于指示所述语料资源库中与该词对应的词性信息;

[0016] 基于各词的所述发音信息查找结果,生成发音信息序列以及发音信息注意力掩码序列;

[0017] 基于各词的所述词性信息查找结果,生成词性信息序列以及词性信息注意力掩码序列;

[0018] 基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别。

[0019] 在一种可能的实现中,所述基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别,包括:

[0020] 将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别;

[0021] 其中,所述词性及发音类别预测模型是利用训练用分词序列、训练用发音信息序列、训练用发音信息注意力掩码序列、训练用词性信息序列以及训练用词性信息注意力掩码序列为训练样本,以所述训练用发音信息序列对应的发音类别标签序列以及所述训练用词性信息序列对应的词性信息标签序列为样本标签,以所述词性及发音类别预测模型输出的词性预测结果趋近于所述词性信息标签序列以及输出的发音类别预测结果趋近于所述发音类别标签序列为训练目标训练得到的。

[0022] 在一种可能的实现中,所述将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别,包括:

[0023] 对所述分词序列进行特征提取处理,得到分词特征向量序列;

[0024] 对所述发音信息序列进行特征提取处理,得到发音信息特征向量序列;

[0025] 对所述词性信息序列进行特征提取处理,得到词性信息特征向量序列;

[0026] 将所述分词特征向量序列、所述发音信息特征向量序列、所述词性信息特征向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列,进行特征融合处理,得到融合特征向量序列;

[0027] 将所述融合特征向量序列输入所述词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果;

[0028] 从所述分词序列中各个词的发音类别预测结果中,确定各个所述第二类词的发音类别。

[0029] 在一种可能的实现中,所述基于预设的与所述第二类词的发音类别对应的发音信

息确定方式确定所述第二类词的发音信息,包括:

[0030] 获取预设的各发音类别对应的发音预测模块;

[0031] 从各发音预测模块中,确定目标发音预测模块,所述目标发音预测模块为所述各发音预测模块中与所述第二类词的发音类别对应的发音预测模块;

[0032] 将所述文本或所述第二类词提供给所述目标发音预测模块,得到所述目标发音预测模块输出的所述第二类词的发音信息。

[0033] 本申请第二方面提供一种发音预测装置,包括:

[0034] 分词处理单元,用于对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词;

[0035] 第一发音信息确定单元,用于基于预设的语料资源库确定出第一类词的发音信息;

[0036] 发音类别确定单元,用于基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,所述第二类词为所述分词序列中除第一类词之外的其他词;

[0037] 第二发音信息确定单元,用于针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

[0038] 在一种可能的实现中,所述第一发音信息确定单元,具体用于:

[0039] 循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的发音信息,得到该词的发音信息查找结果,该词的发音信息查找结果用于指示所述语料资源库中与该词对应的发音信息;

[0040] 基于各词的发音信息查找结果确定所述第一类词以及所述第一类词的发音信息;所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词,从所述语料资源库确定的发音信息为所述第一类词的发音信息。

[0041] 在一种可能的实现中,所述发音类别确定单元,包括:

[0042] 词性信息查找结果确定单元,用于循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的词性信息,得到该词的词性信息查找结果,该词的词性信息查找结果用于指示所述语料资源库中与该词对应的词性信息;

[0043] 发音信息序列以及发音信息注意力掩码序列生成单元,用于基于各词的所述发音信息查找结果,生成发音信息序列以及发音信息注意力掩码序列;

[0044] 词性信息序列以及词性信息注意力掩码序列生成单元,用于基于各词的所述词性信息查找结果,生成词性信息序列以及词性信息注意力掩码序列;

[0045] 发音类别确定子单元,用于基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别。

[0046] 在一种可能的实现中,所述发音类别确定子单元,具体用于:

[0047] 将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别;

[0048] 其中,所述词性及发音类别预测模型是利用训练用分词序列、训练用发音信息序列、训练用发音信息注意力掩码序列、训练用词性信息序列以及训练用词性信息注意力掩

码序列为训练样本,以所述训练用发音信息序列对应的发音类别标签序列以及所述训练用词性信息序列对应的词性信息标签序列为样本标签,以所述词性及发音类别预测模型输出的词性预测结果趋近于所述词性信息标签序列以及输出的发音类别预测结果趋近于所述发音类别标签序列为训练目标训练得到的。

[0049] 在一种可能的实现中,所述发音类别确定子单元,包括:

[0050] 第一特征提取处理单元,用于对所述分词序列进行特征提取处理,得到分词特征向量序列;

[0051] 第二特征提取处理单元,用于对所述发音信息序列进行特征提取处理,得到发音信息特征向量序列;

[0052] 第三特征提取处理单元,用于对所述词性信息序列进行特征提取处理,得到词性信息特征向量序列;

[0053] 特征融合处理单元,用于将所述分词特征向量序列、所述发音信息特征向量序列、所述词性信息特征向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列,进行特征融合处理,得到融合特征向量序列;

[0054] 词性预测结果和发音类别预测结果确定单元,用于将所述融合特征向量序列输入所述词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果;

[0055] 确定子单元,用于从所述分词序列中各个词的发音类别预测结果中,确定各个所述第二类词的发音类别。

[0056] 在一种可能的实现中,所述第二发音信息确定单元,具体用于:

[0057] 获取预设的各发音类别对应的发音预测模块;

[0058] 从各发音预测模块中,确定目标发音预测模块,所述目标发音预测模块为所述各发音预测模块中与所述第二类词的发音类别对应的发音预测模块;

[0059] 将所述文本或所述第二类词提供给所述目标发音预测模块,得到所述目标发音预测模块输出的所述第二类词的发音信息。

[0060] 本申请第三方面提供一种计算机程序产品,包括计算机可读指令,当所述计算机可读指令在电子设备上运行时,使得所述电子设备实现上述第一方面或第一方面任一实现方式的发音预测方法。

[0061] 本申请第四方面提供一种电子设备,包括至少一个处理器和与所述处理器连接的存储器,其中:

[0062] 所述存储器用于存储计算机程序;

[0063] 所述处理器用于执行所述计算机程序,以使所述电子设备能够实现上述第一方面或第一方面任一实现方式的发音预测方法。

[0064] 本申请第五方面提供一种计算机存储介质,所述存储介质承载有一个或多个计算机程序,当所述一个或多个计算机程序被电子设备执行时,能够使所述电子设备上述第一方面或第一方面任一实现方式的发音预测方法。

[0065] 借由上述技术方案,本申请提供的一种发音预测方法及相关装置,首先对待合成的文本进行分词处理,得到分词序列,对于分词序列中的第一类词,将基于预设的语料资源库确定出第一类词的发音信息,对于分词序列中除第一类词之外的第二类词,基于分词序

列中各个词的词性信息,确定其发音类别,基于与其发音类别对应的发音信息确定方式确定其发音信息。在本申请中,结合语料资源库以及预设的各发音类别对应的发音信息确定方式,能够涵盖各种情况下的发音信息确定,因此,能够准确确定出待合成文本中各词的发音信息,进而能够提升语音合成的效果。

附图说明

[0066] 结合附图并参考以下具体实施方式,本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中,相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的,原件和元素不一定按照比例绘制。

[0067] 图1为本申请提供的一种发音预测方法的流程示意图;

[0068] 图2为本申请提供的一种日语前向分词算法示例示意图;

[0069] 图3为本申请提供的一种日语前向分词算法得到的分词序列示意图;

[0070] 图4为本申请提供的一种词性及发音类别预测模型的原理示意图;

[0071] 图5为本申请提供的一种发音预测装置的结构示意图;

[0072] 图6为本申请提供的一种电子设备的结构示意图。

具体实施方式

[0073] 下面结合本申请实施例中的附图对本申请实施例进行描述。本申请的实施方式部分使用的术语仅用于对本申请的具体实施例进行解释,而非旨在限定本申请。

[0074] 下面结合附图,对本申请的实施例进行描述。本领域普通技术人员可知,随着技术的发展和新场景的出现,本申请实施例提供的技术方案对于类似的技术问题,同样适用。

[0075] 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换,这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它单元。

[0076] 目前的语音合成系统,首先会对待合成的文本进行分词,得到一个个独立的单词,再对各单词进行发音预测,确定各单词的发音信息,最后根据各单词的发音信息进行合成。其中,各单词的发音信息准确度对最终合成的效果起着至关重要的作用。

[0077] 现有的发音预测方案一般采用检索语料资源库的方式,即预先构建语料资源库,语料资源库中收录有多个词的发音信息。在对分词后的单词进行发音预测时,可以直接从语料资源库中确定单词的发音信息。但是,针对某个词,语料资源库中收录的该词的发音信息可能有多种,这种情况下,将无法准确确定该词的发音信息。

[0078] 另外,随着互联网的发展和全球化趋势,各种文化现象、热点事件层出不穷,也带来了各种新词、热词、流行语的涌现。如新冠疫情席卷全球后,“covid”一词的出现;又如大模型的火热带来的“ChatGPT”一词频繁出现在各种社交网络上。这些新词、热词、流行语往往无法及时收录到语料资源库中,成为未登录(Out OfVocabulary,OOV)词。但是,这些新词、热词、流行语在语音合成技术领域有着广泛的应用,比如在一些新闻、自媒体应用中,往

往会用到语音合成来发布包含各种新词、热词、流行语的内容。这种情况下,如果分词后的单词是OOV词,则无法直接从语料资源库中确定该单词的发音信息。

[0079] 目前可采用如下几种方案中的任意一种确定该未登录词的发音信息:跳过本词不发音或填充静音段;采用按字母读;使用C45决策树等机器学习方法进行发音预测。但是,跳过本词不发音或填充静音段的方案会使得语音合成结果出现漏词、词义突兀、语音不连贯自然等问题;采用按字母读的方案会使得语音合成结果中一些词(如covid,iflytek等)的发音不地道;采用C45决策树等机器学习方法进行发音预测的方案,仅考虑词面字母组合与发音音素之间的映射关系,当出现按字母发音的词(如,NBA、ChatGPT等)时,会出现拼读等奇怪发音。可见,上述方案无法适应形式多样的未登录词,会导致某些形式的未登录词的发音信息无法准确确定出。

[0080] 为了解决上述问题,本申请实施例提供了一种发音预测方法。本申请的发音预测方法,执行主体可以是手机、计算机、服务器或者服务器集群等电子设备或专门设计的智能设备,也可以是设置在该电子设备或智能设备中的发音预测装置,该发音预测装置可以通过软件、硬件或两者的结合来实现。

[0081] 下面结合附图对本申请实施例的发音预测方法进行详细的介绍。

[0082] 参照图1,图1为本申请实施例提供的一种发音预测方法的流程示意图,如图1所示,本申请实施例提供的一种发音预测方法,可以包括如下步骤:

[0083] S101、对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词。

[0084] 在本申请中,作为一种可实施方式,待合成的文本可以为初始文本,具体的,初始文本可以是包含任意语种的文字、数字、符号、缩略语等信息的文本。其中,可以通过直接输入、图文识别或数据库导入等任意方式获取初始文本。考虑到初始文本可能会存在不规范的问题,则在本申请中,可以对所述初始文本进行规整处理,得到规整后的文本作为待合成的文本。

[0085] 在本申请中,可以采用正则化的方式对初始文本进行规整处理,正则化主要包含两方面的内容:①按照语音合成系统的协议标准,将不规范字符进行规整,包括但不限于全角字符转换为对应的半角字符;数学符号、标点、阿拉伯语等文字的元音符号的合并及转义等。②基于发音语种判断,由文本转写系统完成特殊符号到本语种字符的撰写,如在英语语种合成场景下:初始文本“The price ofthe shirt is$5.”会被转写为“The price of the shirt is five dollars.”。

[0086] 在本申请中,作为一种可实施方式,可以采用基于规则资源库匹配、动态规划的前后向分词算法、CRF(Conditional Radom Field,条件随机场)的分词代价预测等任一分词方法对所述文本进行分词处理,得到分词序列。对此,本申请不进行任何限定。

[0087] 为便于理解,在本申请中给出如下分词处理示例:

[0088] 示例一:以英语、西班牙语等印欧语系为代表的空格分词语种,如“The price of the shirt is five dollars.”进行分词处理后,得到的分词序列可以为:[“The”,“price”,“of”,“the”,“shirt”,“is”,“five”,“dollars”];

[0089] 示例二:以中文、日语等汉藏语系为代表的非空格分词语种,如“南京市长江大桥”进行分词处理后,得到的分词序列可以为[“南京”,“市”,“长江”,“大桥”]。

[0090] 示例三:参照图2,图2为本申请实施例提供的一种日语前向分词算法示例,基于该算法,得到的分词序列如图3所示。

[0091] S102、基于预设的语料资源库确定出第一类词的发音信息;

[0092] 在本申请中,所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词,从所述语料资源库确定的发音信息为所述第一类词的发音信息。

[0093] 在本申请中,预设的语料资源库中包括多个词,以及每个词的具体信息,一个词的具体信息包括该词的发音信息以及该词的词性信息,其中,发音信息可以为发音音素重音,词性信息可以为词性缩写。

[0094] 在本申请中,词性缩写表如下:

[0095]

noun	pron	adj	adv	verb	num
名词	代词	形容词	副词	动词	数词
art	prep	conj	interj	aux	…
冠词	介词	连词	感叹词	助动词	…

[0096] S103、基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别,所述第二类词为所述分词序列中除第一类词之外的其他词。

[0097] 在本申请中,发音类别包括但不限于多音词、衍生词、复合词、字母缩写、单词+字母缩写、驼峰连词。

[0098] S104、针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

[0099] 在本申请中,可以预设各发音类别对应的发音信息确定方式,各发音类别对应的发音信息确定方式可以采用规则、算法或模型实现。

[0100] 需要说明的是,经过上述步骤可以确定出分词序列中各个词的发音信息,将其经过声学模型提取声学特征,再将声学特征传入声码器即可输出语音合成音频。

[0101] 本实施例公开的一种发音预测方法,首先对待合成的文本进行分词处理,得到分词序列,对于分词序列中能够基于预设语料资源库确定出唯一发音信息的第一类词,将从语料资源库确定的发音信息为第一类词的发音信息,对于分词序列中除第一类词之外的第二类词,基于分词序列中各个词的词性信息,确定其发音类别,基于与其发音类别对应的发音信息确定方式确定其发音信息。在本申请中,结合语料资源库以及预设的各发音类别对应的发音信息确定方式,能够涵盖各种情况下的发音信息确定,因此,能够准确确定出待合成文本中各词的发音信息,进而能够提升语音合成的效果。

[0102] 在本申请的另一个实施例中,对基于预设的语料资源库确定出第一类词的发音信息的具体实现方式进行说明,该方式可以包括如下步骤:

[0103] S201:循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的发音信息,得到该词的发音信息查找结果,该词的发音信息查找结果用于指示所述语料资源库中与该词对应的发音信息。

[0104] 需要说明的是,针对所述分词序列中的每一个词,其发音信息查找结果中可能只包含一个发音信息,也可能包含多个发音信息,还可能包含0个发音信息,如果发音信息查找结果中包含0个发音信息,说明语料资源库中未收录该词。如果发音信息查找结果中包含多个发音信息,说明语料资源库中虽然有收录该词,但是记录的该词的发音信息有歧义,需

要进一步确定。

[0105] S202:基于各词的发音信息查找结果确定所述第一类词以及所述第一类词的发音信息;所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词,从所述语料资源库确定的发音信息为所述第一类词的发音信息。

[0106] 在本申请的另一个实施例中,对基于所述分词序列中各个词的词性信息,确定各个第二类词的发音类别的具体实现方式进行说明,该方式可以包括如下步骤:

[0107] S301:循环遍历所述分词序列中的每一个词,在所述语料资源库中查找是否存在与该词对应的词性信息,得到该词的词性信息查找结果,该词的词性信息查找结果用于指示所述语料资源库中与该词对应的词性信息。

[0108] 需要说明的是,针对所述分词序列中的每一个词,其词性信息查找结果中可能只包含一个词性信息,也可能包含多个词性信息,还可能包含0个词性信息,如果词性信息查找结果中包含0个词性信息,说明语料资源库中未收录该词。如果词性信息查找结果中包含多个词性信息,说明语料资源库中虽然有收录该词,但是记录的该词的词性信息有歧义,需要进一步确定。

[0109] S302:基于各词的所述发音信息查找结果,生成发音信息序列以及发音信息注意力掩码序列。

[0110] 在本申请中,所述发音信息序列中包括所述分词序列中各个词的发音信息标注结果;其中,针对某个词,如果其发音信息查找结果中只包含一个发音信息,或者包含多个发音信息,则该词的发音信息标注结果可以为其发音信息查找结果,不同发音信息之间可以采用预设分隔符(如“/”)隔开,如果其发音信息查找结果中包含0个发音信息,则该词的发音信息标注结果可以为预设标识(比如,OOV)。所述发音信息注意力掩码序列用于指示所述分词序列中各个词的发音类别预测权重,其中,所述分词序列中第一类词的发音类别预测权重可以为0,第二类词的发音类别预测权重可以为1。

[0111] S303:基于各词的所述词性信息查找结果,生成词性信息序列以及词性信息注意力掩码序列;所述词性信息序列中包括所述分词序列中各个词的词性信息。

[0112] 在本申请中,所述词性信息序列中包括所述分词序列中各个词的词性信息标注结果;其中,针对某个词,如果其词性信息查找结果中只包含一个词性信息,或者包含多个词性信息,则该词的词性信息标注结果可以为其词性信息查找结果,不同词性信息之间可以采用预设分隔符(如“/”)隔开,如果其词性信息查找结果中包含0个词性信息,则该词的词性信息标注结果可以为预设标识(比如,OOV)。所述词性信息注意力掩码序列用于指示所述分词序列中各个词的词性信息预测权重,其中,词性信息查找结果中只包含一个词性信息的词,其词性信息预测权重可以为0,词性信息查找结果中包含多个词性信息的词一级包含0个词性信息的词,其词性信息预测权重可以为1。

[0113] 为便于理解,以待合成的文本为英语文本“I have read a book about chatGPT”为例,对其进行分词处理,得到的分词序列为[“I”,“have”,“read”,“a”,“book”,“about”,“chatGPT”],发音信息序列可以为[“((ai)1)”,“((hh ac vv)1)”,“((rr ii dd)1)/((rr ae dd)1)”,“((ax)1)/((ei)1)”,“((bb oc kg)1)”,“((ax)0)((bb ao td)1)”,“OOV”],发音信息注意力掩码序列可以为[0,0,1,1,0,1,1];词性信息序列可以为[“noun”,“aux/verb”,“verb”,“noun/art”,“noun”,“prep”,“OOV”],词性信息注意力掩码序列可以为[0,

1,0,1,0,0,1]。

[0114] 在本示例中,[“have”]一词的词性信息有2种可能:①作为完成时态的助动词②动词“拥有”。[“a”]一词的发音信息有2种可能:①作为冠词的发音((ax)1)②作为名词的字母发音((ei)1)。在语料资源库中没有查找到与[“chatGPT”]一词对应的发音信息以及词性信息,则[“chatGPT”]一词在发音信息序列中的标注为“OOV”,在词性信息序列中的标注也为“OOV”。

[0115] S304:基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,确定各个所述第二类词的发音类别。

[0116] 在本申请中,可以将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别;其中,所述词性及发音类别预测模型是利用训练用分词序列、训练用发音信息序列、训练用发音信息注意力掩码序列、训练用词性信息序列以及训练用词性信息注意力掩码序列为训练样本,以所述训练用发音信息序列对应的发音类别标签序列以及所述训练用词性信息序列对应的词性信息标签序列为样本标签,以所述词性及发音类别预测模型输出的词性预测结果趋近于所述词性信息标签序列以及输出的发音类别预测结果趋近于所述发音类别标签序列为训练目标训练得到的。

[0117] 在本申请中,词性及发音类别预测模型可以包括但不限于全连接层、卷积神经网络、循环神经网络、长短期记忆神经网络、残差神经网络、注意力机制等深度学习模型。

[0118] 具体的,所述将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列,以及,所述词性信息注意力掩码序列,输入词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别,可以包括如下步骤:

[0119] S401:对所述分词序列进行特征提取处理,得到分词特征向量序列。

[0120] 在本申请中,可以采用词嵌入(Word Embedding)处理的方式,对所述分词序列进行特征提取处理,得到分词特征向量序列。

[0121] S402:对所述发音信息序列进行特征提取处理,得到发音信息特征向量序列。

[0122] 在本申请中,可以采用词嵌入(Word Embedding)处理的方式,对所述发音信息序列进行特征提取处理,得到发音信息特征向量序列。

[0123] S403:对所述词性信息序列进行特征提取处理,得到词性信息特征向量序列。

[0124] 在本申请中,可以采用词嵌入(Word Embedding)处理的方式,对所述词性信息序列进行特征提取处理,得到词性信息特征向量序列。

[0125] S404:将所述分词特征向量序列、所述发音信息特征向量序列、所述词性信息特征向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列,进行特征融合处理,得到融合特征向量序列。

[0126] 为便于理解,仍然以待合成的文本为英语文本“I have read a book about chatGPT”为例,对其进行分词处理,得到的分词序列为[“I”,“have”,“read”,“a”,“book”,“about”,“chatGPT”],发音信息序列可以为[“((ai)1)”,“((hh ac vv)1)”,“((rr ii dd)

1)/((rr ae dd)1)”,“((ax)1)/((ei)1)”,“((bb oc kg)1)”,“((ax)0)((bb ao td)1)”,“00V”],发音信息注意力掩码序列可以为[0,0,1,1,0,0,1];词性信息序列可以为[“noun”,“aux/verb”,“verb”,“noun/art”,“noun”,“prep”,“00V”],词性信息注意力掩码序列可以为[0,1,0,1,0,0,1]。则融合特征向量序列可以如下:

[0127]

I	[-5.0421,-4.8561，……]
have	[-0.3704,-3.1329，……]
read	[-4.6420,-3.5719，……]
a	[-1.0441,-0.6443，……]
book	[-3.3375,-5.1706，……]
about	[-5.4136,-5.3791，……]

[0128] chatGPT [-7.0625,-5.3896，……],其中,第一列可以为分词特征向量序列,第二列可以为发音信息特征向量序列,后边省略号的部分可以包含多列,分别对应所述词性信息特征信息向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列。

[0129] S405:将所述融合特征向量序列输入所述词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果。

[0130] 在本申请中,词性及发音类别预测模型在得到所述融合特征向量之后,即可预测得到所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果。比如,预测出[“have”]一词词性为助动词aux,[“a”]一词词性为冠词art,[“chatGPT”]一词词性为名词noun,[“read”]一词为多音词发音类别,[“chatGPT”]一词为单词+字母缩写发音类别。为便于理解,参照图4,图4为本申请实施例提供的一种词性及发音类别预测模型的原理示意图。

[0131] S406:从所述分词序列中各个词的发音类别预测结果中,确定各个所述第二类词的发音类别。

[0132] 在本申请的另一个实施例中,对步骤S104针对每个所述第二类词,基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息的具体实现方式进行说明,该方式可以包括如下步骤:

[0133] S501:获取预设的各发音类别对应的发音预测模块。

[0134] 前述内容中指出,在本申请中,发音类别包括但不限于多音词、衍生词、复合词、字母缩写、单词+字母缩写、驼峰连词。则在本申请中,预设的各发音类别对应的发音预测模块包括但不限于多音词发音预测模块、衍生词发音预测模块、复合词发音预测模块、字母缩写发音预测模块、单词+字母缩写发音预测模块、驼峰连词发音预测模块。

[0135] 前述内容中指出,各发音类别对应的发音信息确定方式可以采用规则、算法或模型实现,则各发音预测模块可以采用规则、算法或模型实现。比如,多音词发音预测模块可以采用模型实现,单词+字母缩写发音预测模块可以采用动态规划算法逐步遍历词典树,前向匹配当前最长单词,确定切分位置,将字母部分按字母缩写发音,将单词部分查找语料资源库确定发音信息。

[0136] 驼峰连词发音预测模块由前向最大匹配算法加规则调整并限制栈调用深度,最后将词划分为多项,逐项查语料资源库确定发音信息即可。

[0137] 衍生词发音预测模块可以基于规则的方式确定衍生词的发音信息,规则具体为衍生词从词缀上分为前缀和后缀:①前缀如in/compitable②后缀如look/ing。相应的,前后缀匹配时存在词根形变情况,如:①双写t/d/p等情况,eg:sit/ting、pad/ding②辅音结尾去e/y/变y为i情况,eg:googl(e)/ing、happ/iness、ly/ing③……

[0138] 英混中拼音词,可以使用动态规划算法递归分词查找,确定发音信息。

[0139] 需要说明的是,各发音类别对应的发音预测模块不限语种。

[0140] 如阿拉伯语的元音恢复模块:阿拉伯语的书写中是不含元音的,仅有辅音。基于Transformer的发音预测,配合规则库资源,可将阿语文本中的元音重新恢复,从而获得正确的发音信息。如الصوت بـ كـ تم الملاحة تـ وجـ يه افـ تح中,阿语字符上方的字符即为恢复出的元音信息。

[0141] 又如韩语的jamo发音预测模块:韩语中的单词发音是由前jamo和后jamo及中间韩语文字共同预测获得。如“나는 영문 문장을 한 편을 읽었다”可得分词向量[“나는”, “영문”,“문장을”, “한”, “편을”, “읽었다”],对每个词进行jamo预测,可分为3种:①仅有前jamo②仅有后jamo③同时存在前后jamo,基于Transformer的模型预测及词典规则匹配,可完成韩语单词的正确发音。

[0142] S502:从各发音预测模块中,确定目标发音预测模块,所述目标发音预测模块为所述各发音预测模块中与所述第二类词的发音类别对应的发音预测模块;

[0143] S503:将所述文本或所述第二类词提供给所述目标发音预测模块,得到所述目标发音预测模块输出的所述第二类词的发音信息。

[0144] 如上述示例中的read一词,则可以将上述融合特征向量序列输入多音词发音预测模型,多音词发音预测模型即可预测到read的两个发音预测值向量,经过softMax的归一化可得[0.19652,0.80348],对应得到read一词的发音信息为((rr ae dd)1)。

[0145] 再如上述示例中的chatGPT一词,则可以将chatGPT一词输入单词+字母缩写发音预测模块,由动态规划算法逐步遍历词典树,前向匹配当前最长单词,可得切分位置为chat/GPT,后三位为按字母缩写发音,依次填充发音得该词发音信息为((chac td)1)((jhii)1)((pb ii)1)((td ii)1)。

[0146] 再如驼峰连词“NoPainsNoGains”合驼峰命名法单词,则可以将其输入驼峰连词发音预测模块,驼峰连词发音预测模块由前向最大匹配算法加规则调整并限制栈调用深度,最后将词划分为No/Pains/No/Gains,逐项查语料库资源将发音信息填充即可。

[0147] 以上介绍了本申请实施例提供的一种发音预测方法,以下将介绍执行上述的发音预测方法的装置。

[0148] 请参阅图5,图5为本申请实施例提供的一种发音预测装置的结构示意图。如图5所示,该发音预测装置,包括:

[0149] 分词处理单元11,用于对待合成的文本进行分词处理,得到分词序列,所述分词序列中包含多个词;

[0150] 第一发音信息确定单元12,用于基于预设的语料资源库确定出第一类词的发音信

息；

[0151] 发音类别确定单元13，用于基于所述分词序列中各个词的词性信息，确定各个第二类词的发音类别，所述第二类词为所述分词序列中除第一类词之外的其他词；

[0152] 第二发音信息确定单元14，用于针对每个所述第二类词，基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息。

[0153] 在一种可能的实现中，所述第一发音信息确定单元，具体用于：

[0154] 循环遍历所述分词序列中的每一个词，在所述语料资源库中查找是否存在与该词对应的发音信息，得到该词的发音信息查找结果，该词的发音信息查找结果用于指示所述语料资源库中与该词对应的发音信息；

[0155] 基于各词的发音信息查找结果确定所述第一类词以及所述第一类词的发音信息，所述第一类词为所述分词序列中能够基于所述语料资源库确定出唯一发音信息的词，从所述语料资源库确定的发音信息为所述第一类词的发音信息。

[0156] 在一种可能的实现中，所述发音类别确定单元，包括：

[0157] 词性信息查找结果确定单元，用于循环遍历所述分词序列中的每一个词，在所述语料资源库中查找是否存在与该词对应的词性信息，得到该词的词性信息查找结果，该词的词性信息查找结果用于指示所述语料资源库中与该词对应的词性信息；

[0158] 发音信息序列以及发音信息注意力掩码序列生成单元，用于基于各词的所述发音信息查找结果，生成发音信息序列以及发音信息注意力掩码序列；

[0159] 词性信息序列以及词性信息注意力掩码序列生成单元，用于基于各词的所述词性信息查找结果，生成词性信息序列以及词性信息注意力掩码序列；

[0160] 发音类别确定子单元，用于基于所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列，以及，所述词性信息注意力掩码序列，确定各个所述第二类词的发音类别。

[0161] 在一种可能的实现中，所述发音类别确定子单元，具体用于：

[0162] 将所述分词序列、所述发音信息序列、所述发音信息注意力掩码序列、所述词性信息序列，以及，所述词性信息注意力掩码序列，输入词性及发音类别预测模型，得到所述词性及发音类别预测模型输出的各个所述第二类词的发音类别；

[0163] 其中，所述词性及发音类别预测模型是利用训练用分词序列、训练用发音信息序列、训练用发音信息注意力掩码序列、训练用词性信息序列以及训练用词性信息注意力掩码序列为训练样本，以所述训练用发音信息序列对应的发音类别标签序列以及所述训练用词性信息序列对应的词性信息标签序列为样本标签，以所述词性及发音类别预测模型输出的词性预测结果趋近于所述词性信息标签序列以及输出的发音类别预测结果趋近于所述发音类别标签序列为训练目标训练得到的。

[0164] 在一种可能的实现中，所述发音类别确定子单元，包括：

[0165] 第一特征提取处理单元，用于对所述分词序列进行特征提取处理，得到分词特征向量序列；

[0166] 第二特征提取处理单元，用于对所述发音信息序列进行特征提取处理，得到发音信息特征向量序列；

[0167] 第三特征提取处理单元，用于对所述词性信息序列进行特征提取处理，得到词性

信息特征向量序列;

[0168] 特征融合处理单元,用于将所述分词特征向量序列、所述发音信息特征向量序列、所述词性信息特征向量序列、所述发音信息注意力掩码序列以及所述词性信息注意力掩码序列,进行特征融合处理,得到融合特征向量序列;

[0169] 词性预测结果和发音类别预测结果确定单元,用于将所述融合特征向量序列输入所述词性及发音类别预测模型,得到所述词性及发音类别预测模型输出的所述分词序列中各个词的词性预测结果和所述分词序列中各个词的发音类别预测结果;

[0170] 确定子单元,用于从所述分词序列中各个词的发音类别预测结果中,确定各个所述第二类词的发音类别。

[0171] 在一种可能的实现中,所述第二发音信息确定单元,具体用于:

[0172] 获取预设的各发音类别对应的发音预测模块;

[0173] 从各发音预测模块中,确定目标发音预测模块,所述目标发音预测模块为所述各发音预测模块中与所述第二类词的发音类别对应的发音预测模块;

[0174] 将所述文本或所述第二类词提供给所述目标发音预测模块,得到所述目标发音预测模块输出的所述第二类词的发音信息。

[0175] 本申请实施例中还提供一种电子设备。参考图6所示,其示出了适于用来实现本申请实施例中的电子设备的结构示意图。本申请实施例中的电子设备可以包括但不限于诸如移动电话、笔记本电脑、PDA(个人数字助理)、PAD(平板电脑)、台式计算机等等的固定终端。图6示出的电子设备仅仅是一个示例,不应对本申请实施例的功能和使用范围带来任何限制。

[0176] 如图6所示,该电子设备可以包括处理装置(例如中央处理器、图形处理器等)601,其可以根据存储在只读存储器(ROM)602中的程序或者从存储装置608加载到随机存取存储器(RAM)603中的程序而执行各种适当的动作和处理。在电子设备通电的状态下,RAM 603中还存储有电子设备操作所需的各种程序和数据。处理装置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。

[0177] 通常,以下装置可以连接至I/O接口605:包括例如触摸屏、触摸板、键盘、鼠标、摄像头麦克风加速度计陀螺仪等的输入装置606;包括例如液晶显示器(LCD)扬声器振动器等的输出装置607;包括例如内存卡、硬盘等的存储装置608;以及通信装置609。通信装置609可以允许电子设备与其他设备进行无线或有线通信以交换数据。虽然图6示出了具有各种装置的电子设备,但是应理解的是,并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

[0178] 本申请实施例中还提供一种包括计算机程序产品,包括计算机可读指令,当计算机可读指令在电子设备上运行时,使得电子设备实现本申请实施例提供的任一种发音预测方法。

[0179] 本申请实施例中还提供一种计算机可读存储介质,该存储介质承载有一个或多个计算机程序,当一个或多个计算机程序被电子设备执行时,能够使电子设备实现本申请实施例提供的任一种发音预测方法。

[0180] 另外需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也

可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外,本申请提供的装置实施例附图中,模块之间的连接关系表示它们之间具有通信连接,具体可以实现为一条或多条通信总线或信号线。

[0181] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现,当然也可以通过专用硬件包括专用集成电路、专用CPU、专用存储器、专用元器件等来实现。一般情况下,凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现,而且,用来实现同一功能的具体硬件结构也可以是多种多样的,例如模拟电路、数字电路或专用电路等。但是,对本申请而言更多情况下软件程序实现是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在可读取的存储介质中,如计算机的软盘、U盘、移动硬盘、ROM、RAM、磁碟或者光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,训练设备,或者网络设备等)执行本申请各个实施例所述的方法。

[0182] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。

[0183] 所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时,全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一计算机可读存储介质传输,例如,所述计算机指令可以从一个网站站点、计算机、训练设备或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、训练设备或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的训练设备、数据中心等数据存储设备。所述可用介质可以是磁性介质,(例如,软盘、硬盘、磁带)、光介质(例如,DVD)、或者半导体介质(例如固态硬盘(Solid State Disk,SSD))等。

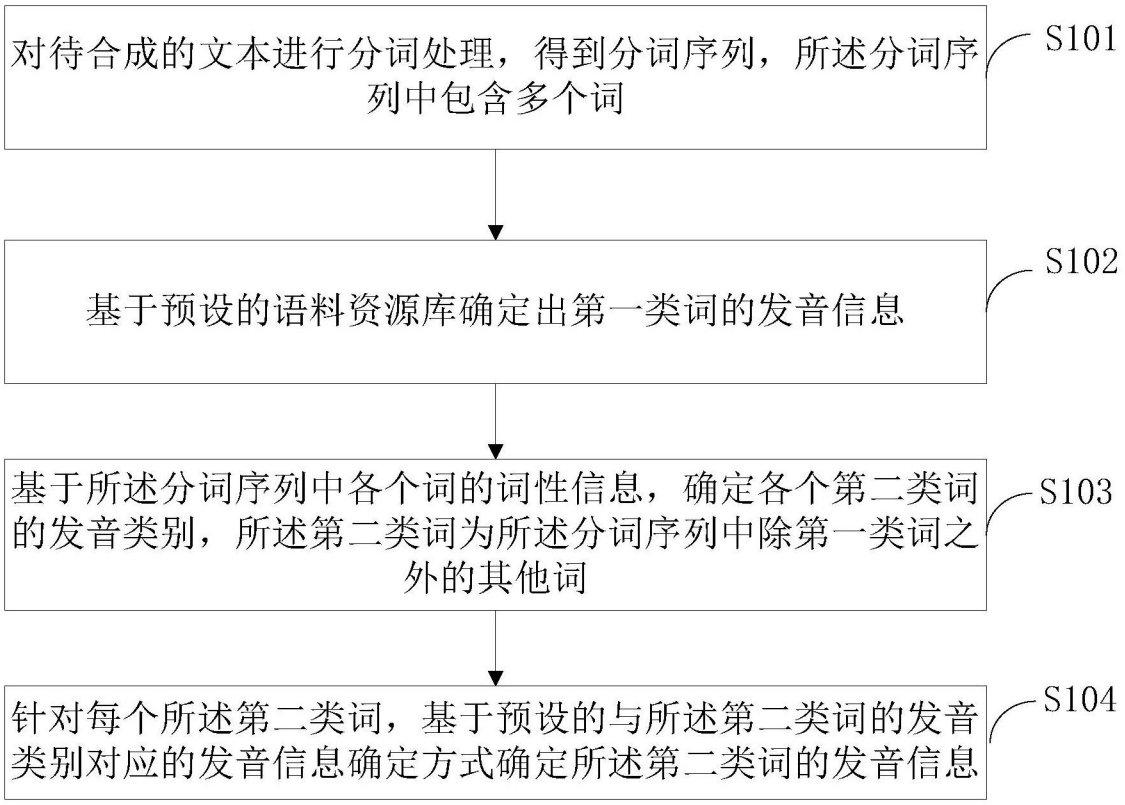

对待合成的文本进行分词处理，得到分词序列，所述分词序列中包含多个词
S101
基于预设的语料资源库确定出第一类词的发音信息
S102
基于所述分词序列中各个词的词性信息，确定各个第二类词的发音类别，所述第二类词为所述分词序列中除第一类词之外的其他词
S103
针对每个所述第二类词，基于预设的与所述第二类词的发音类别对应的发音信息确定方式确定所述第二类词的发音信息
S104

图1

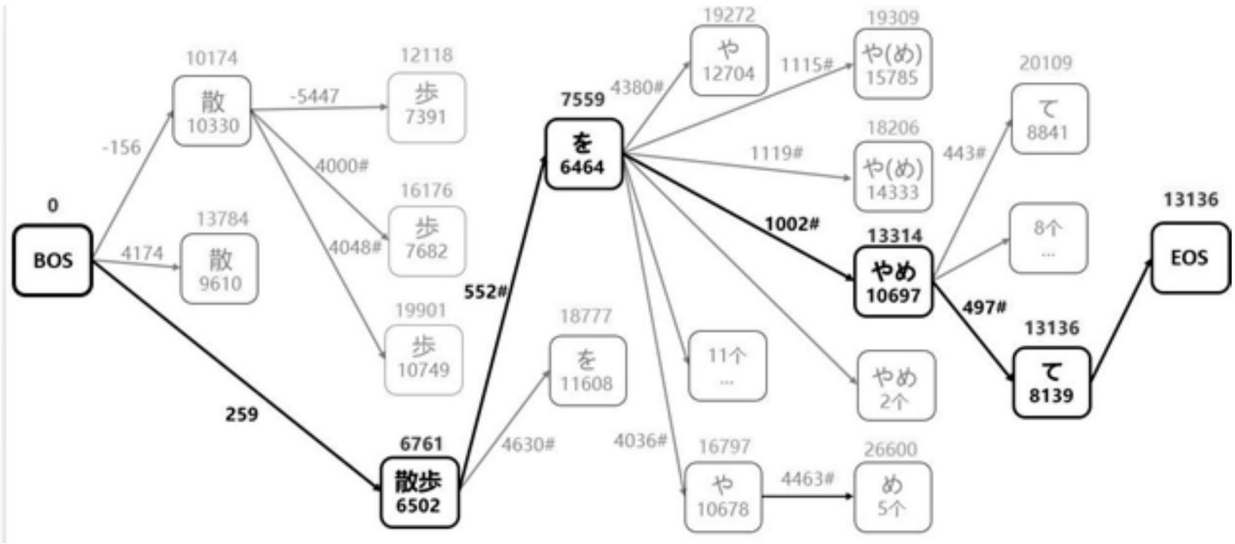

0
BOS
-156
10174
散
10330
4174
13784
散
9610
259
-5447
12118
歩
7391
4000#
16176
歩
7682
4048#
19901
歩
10749
552#
6761
散歩
6502
7559
を
6464
4630#
18777
を
11608
4380#
19272
や
12704
1115#
19309
や(め)
15785
1119#
18206
や(め)
14333
1002#
13314
やめ
10697
11个
...
やめ
2个
4036#
16797
や
10678
4463#
26600
め
5个
443#
20109
て
8841
8个
...
497#
13136
て
8139
13136
EOS

图2


散歩をやめて
散歩 / を / やめ / て

图3

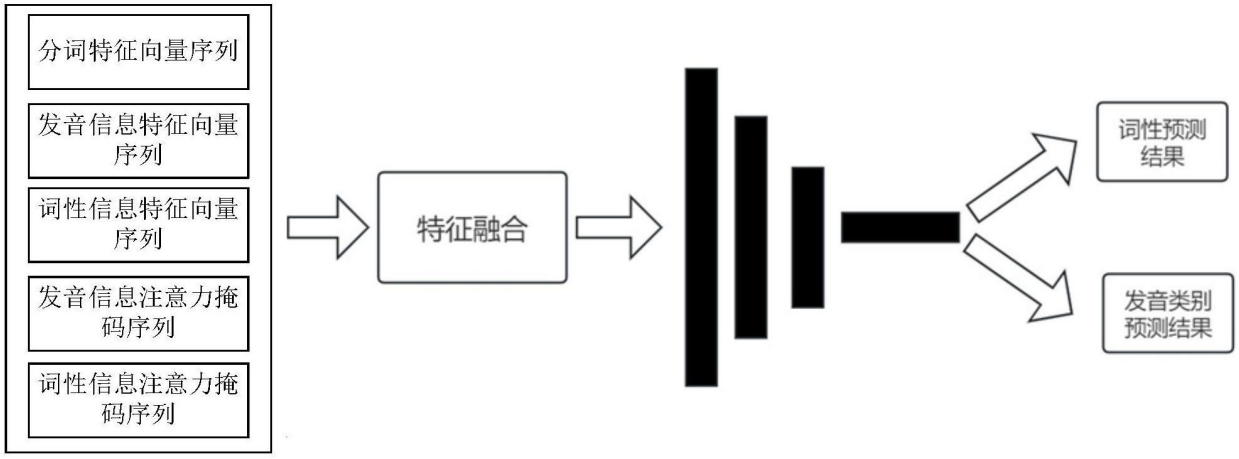

分词特征向量序列
发音信息特征向量序列
词性信息特征向量序列
发音信息注意力掩码序列
词性信息注意力掩码序列
特征融合
词性预测结果
发音类别预测结果

图4

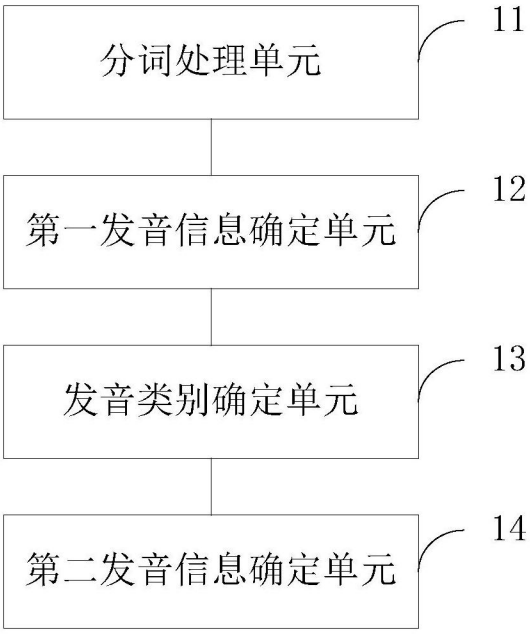

分词处理单元
11
第一发音信息确定单元
12
发音类别确定单元
13
第二发音信息确定单元
14

图5

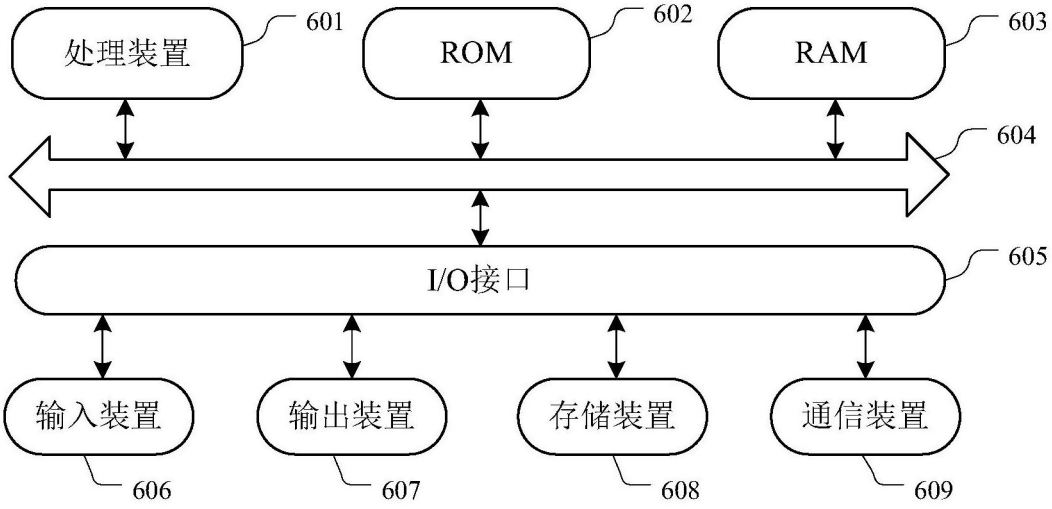

处理装置
601
ROM
602
RAM
603
604
I/O接口
605
输入装置
606
输出装置
607
存储装置
608
通信装置
609

图6