



(12) 发明专利申请

(10) 申请公布号 CN 114187894 A

(43) 申请公布日 2022. 03. 15

(21) 申请号 202111506575.7

G10L 15/26 (2006.01)

(22) 申请日 2021.12.10

G06N 3/08 (2006.01)

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601

(72) 发明人 冯明超 陈蒙 乐雨泉 赵宇明
王泽勋

(74) 专利代理机构 北京路浩知识产权代理有限
公司 11002

代理人 毛宏宝

(51) Int. Cl.

G10L 15/02 (2006.01)

G10L 15/16 (2006.01)

G10L 15/22 (2006.01)

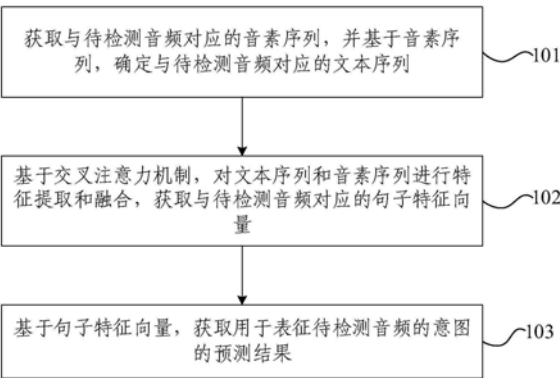
权利要求书2页 说明书10页 附图7页

(54) 发明名称

一种意图识别方法、装置及其相关设备

(57) 摘要

本公开提供了一种意图识别方法、装置及其相关设备,该方法包括:获取与待检测音频对应的音素序列,并基于所述音素序列,确定与待检测音频对应的文本序列;基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与待检测音频对应的句子特征向量;基于所述句子特征向量,获取用于表征待检测音频的意图的预测结果。该方法可以提升口语意图识别的准确率。



1. 一种意图识别方法,其特征在于,该方法包括:

获取与待检测音频对应的音素序列,并基于所述音素序列,确定与所述待检测音频对应的文本序列;

基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量;

基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果。

2. 根据权利要求1所述的意图识别方法,其特征在于,所述基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果具体包括:

输入所述句子特征向量至全连接神经网络进行分类预测,获取所述预测结果。

3. 根据权利要求2所述的意图识别方法,其特征在于,所述基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量包括:

分别对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量;

根据文本序列权重参数和音素序列权重参数,对所述文本序列特征向量和所述音素序列特征向量进行拼接,获取所述句子特征向量。

4. 根据权利要求3所述的意图识别方法,其特征在于,所述对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量包括:

采用双向神经网络对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量。

5. 根据权利要求3所述的意图识别方法,其特征在于,所述文本序列权重参数和所述音素序列权重参数通过对多个预先存储的音频样本进行训练获取,具体包括:

对样本文本序列和样本音素序列进行编码,获取与所述样本文本序列对应的样本文本序列特征向量,以及与所述样本音素序列对应的样本音素序列特征向量;其中,所述样本文本序列和所述样本音素序列通过对所述音频样本进行语音识别得到;

计算所述样本文本序列特征向量中每个文本的嵌入向量与所述样本音素序列特征向量的余弦相似度,得到每个文本的权重,通过归一化指数函数,归一化所有文本对应的权重,得到所述文本序列权重参数;

根据所述文本序列权重参数得到所述音素序列权重参数。

6. 根据权利要求5所述的意图识别方法,其特征在于,所述文本序列权重参数和所述音素序列权重参数通过如下方式进行调整,具体包括:

确定由人工听取的所述音频样本转写的真实文本;

基于交叉熵损失函数,计算表征所述音频样本的意图的预测结果与所述真实文本的差异;

基于所述差异,调整所述文本序列权重参数和所述音素序列权重参数。

7. 一种意图识别装置,其特征在于,包括:

语音识别模块,用于获取与待检测音频对应的音素序列,并基于所述音素序列,确定与所述待检测音频对应的文本序列;

交叉注意力模块,用于基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量;

全连接神经网络分类模块,用于基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果。

8.根据权利要求7所述的意图识别装置,其特征在于,所述全连接神经网络分类模块包括:

输入所述句子特征向量至全连接神经网络进行分类预测,获取所述预测结果。

9.根据权利要求8所述的意图识别装置,其特征在于,所述交叉注意力模块具体包括:

编码单元,分别对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量;

拼接单元,根据文本序列权重参数和所述音素序列权重参数,对所述文本序列特征向量和所述音素序列特征向量进行拼接,获取所述句子特征向量。

10.根据权利要求9所述的意图识别装置,其特征在于,所述编码单元具体包括:

采用双向神经网络对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量。

11.一种电子设备,包括存储器、处理器及存储在存储器上并可在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至6任一项所述意图识别方法的步骤。

12.一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在于,该计算机程序被处理器执行时实现如权利要求1至6任一项所述意图识别方法的步骤。

一种意图识别方法、装置及其相关设备

技术领域

[0001] 本公开涉及人工智能技术领域,尤其涉及一种意图识别方法、装置及其相关设备。

背景技术

[0002] 随着深度学习和自然语言处理技术的发展,许多重复繁琐的问答都移交给智能语音对话系统完成。智能语音对话系统包含语音识别、口语理解、对话管理、对话生成以及语音合成五大模块。口语理解模块主要由领域识别、意图识别、槽位填充三个任务组成,用于对话系统的下一步响应动作的意图识别的现有方法为利用文本层面的特征信息实现意图识别,现有方法在口语意图识别任务下准确率较低。如何提升口语意图识别的准确率亟待解决。

发明内容

[0003] 针对现有技术存在的问题,本公开实施例提供一种意图识别方法、装置及其相关设备,实现提升口语意图识别的准确率。

[0004] 第一方面,本公开提供一种意图识别方法,包括:获取与待检测音频对应的音素序列,并基于所述音素序列,确定与所述待检测音频对应的文本序列;基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量;基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果。

[0005] 根据本公开提供的意图识别方法,所述基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果具体包括:输入所述句子特征向量至全连接神经网络进行分类预测,获取所述预测结果。

[0006] 根据本公开提供的意图识别方法,所述基于交叉注意力机制,对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量包括:分别对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量;根据文本序列权重参数和音素序列权重参数,对所述文本序列特征向量和所述音素序列特征向量进行拼接,获取所述句子特征向量。

[0007] 根据本公开提供的意图识别方法,所述对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量包括:采用双向神经网络对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量。

[0008] 根据本公开提供的意图识别方法,所述文本序列权重参数和所述音素序列权重参数通过对多个预先存储的音频样本进行训练,具体包括:对样本文本序列和样本音素序列进行编码,获取与所述样本文本序列对应的所述样本文本序列特征向量,以及与所述样本音素序列对应的所述样本音素序列特征向量;其中,所述样本文本序列和所述样本音素序列通过对所述音频样本进行语音识别得到;计算所述样本文本序列特征向量中每个文本的

嵌入向量与所述样本音素序列特征向量的余弦相似度,得到每个文本的权重,通过归一化指数函数,归一化所有文本对应的权重,得到所述文本序列权重参数;根据所述文本序列权重参数得到所述音素序列权重参数。

[0009] 根据本公开提供的意图识别方法,所述文本序列权重参数和所述音素序列权重参数通过如下方式进行调整,具体包括:确定由人工听取的所述音频样本转写的真实文本;基于交叉熵损失函数,计算表征所述音频样本的意图的预测结果与所述真实文本的差异;基于所述差异,调整所述文本序列权重参数和所述音素序列权重参数。

[0010] 第二方面,本公开提供了一种意图识别装置,包括:语音识别模块、交叉注意力模块、全连接神经网络分类模块。其中,语音识别模块,用于获取与待检测音频对应的音素序列,并基于所述音素序列,确定与所述待检测音频对应的文本序列;交叉注意力模块,用于对所述文本序列和所述音素序列进行特征提取和融合,获取与所述待检测音频对应的句子特征向量;全连接神经网络分类模块,用于基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果。

[0011] 根据本公开提供的意图识别装置,所述全连接神经网络分类模块包括:输入所述句子特征向量至全连接神经网络进行分类预测,获取所述预测结果。

[0012] 根据本公开提供的意图识别装置,所述交叉注意力模块具体包括:编码单元和拼接单元。其中,编码单元,分别对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量;拼接单元,根据文本序列权重参数和所述音素序列权重参数,对所述文本序列特征向量和所述音素序列特征向量进行拼接,获取所述句子特征向量。

[0013] 根据本公开提供的意图识别装置,所述编码单元具体包括:采用双向神经网络对所述文本序列和所述音素序列进行编码,获取与所述文本序列对应的文本序列特征向量,以及与所述音素序列对应的音素序列特征向量。

[0014] 本公开还提供一种电子设备,包括存储器、处理器及存储在存储器上并可在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如上述任一种所述意图识别方法的步骤。

[0015] 本公开还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一种所述意图识别方法的步骤。

[0016] 本公开提供了一种意图识别方法、装置及其相关设备,通过获取与待检测音频对应的音素序列,并基于所述音素序列,确定与所述待检测音频对应的文本序列;对所述文本序列和所述音素序列进行特征提取和融合,而非仅利用文本序列的特征信息,获取到的句子特征向量将字的特征与音素的特征进行无缝融合,充分利用了语音识别输出的文本信息和音素信息,提升噪声文本下的句子特征向量表征,在面对发音声调错误和近音错误的时候,能根据句子文字序列和音素序列重建真实文本对应的句子特征向量表征;基于所述句子特征向量,获取用于表征所述待检测音频的意图的预测结果,预测结果的准确率提高,起到了提升口语意图识别的准确率。

附图说明

[0017] 为了更清楚地说明本公开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公开的一些实施例,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。

[0018] 图1是本公开提供的意图识别方法的流程示意图之一;

[0019] 图2是本公开提供的意图识别方法中,对文本序列和音素序列进行特征提取和融合,获取与待检测音频对应的句子特征向量的流程示意图之一;

[0020] 图3是本公开提供的意图识别方法中,对文本序列权重参数和音素序列权重参数进行训练的流程示意图之一;

[0021] 图4是本公开提供的意图识别方法的具体实施例的交叉注意力模型内部流程示意图;

[0022] 图5是本公开提供的意图识别方法的具体实施例流程示意图;

[0023] 图6是本公开提供的意图识别装置的结构示意图;

[0024] 图7是本公开提供的意图识别装置中的交叉注意力模块应用结构示意图;

[0025] 图8是本公开提供的意图识别装置中的交叉注意力模块训练结构示意图;

[0026] 图9是本公开提供的电子设备的结构示意图。

具体实施方式

[0027] 为使本公开实施例的目的、技术方案和优点更加清楚,下面将结合本公开实施例中的附图,对本公开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公开实施例一部分实施例,而不是全部的实施例。基于本公开实施例中的实施例,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,都属于本公开实施例保护的范围。

[0028] 下面结合图1描述本公开实施例的意图识别方法的各种实施例。

[0029] 参照图1,本公开实施例提供的意图识别方法,包括:

[0030] 步骤S101、获取与待检测音频对应的音素序列,并基于音素序列,确定与待检测音频对应的文本序列;

[0031] 步骤S102、基于交叉注意力机制,对文本序列和音素序列进行特征提取和融合,获取与待检测音频对应的句子特征向量;

[0032] 步骤S103、基于句子特征向量,获取用于表征待检测音频的意图的预测结果。

[0033] 本公开实施例提供的意图识别方法,通过获取与待检测音频对应的音素序列,并基于音素序列,确定与待检测音频对应的文本序列;对文本序列和音素序列进行特征提取和融合,而非仅利用文本序列的特征信息,获取到的句子特征向量将字的特征与音素的特征进行无缝融合,充分利用了语音识别输出的文本信息和音素信息,提升噪声文本下的句子特征向量表征,在面对发音声调错误和近音错误的时候,能根据句子文字序列和音素序列重建真实文本对应的句子特征向量表征;基于句子特征向量,获取用于表征待检测音频的意图的预测结果,预测结果的准确率提高,起到了提升口语意图识别的准确率。

[0034] 下面对图1所示的实施例的各个步骤做进一步说明。

[0035] 步骤S101、获取与待检测音频对应的音素序列,并基于音素序列,确定与待检测音频对应的文本序列;

[0036] 该步骤中,音素为根据语音的自然属性划分出来的最小语音单位,依据音节里的发音动作来分析,一个动作构成一个音素。音素分为元音与辅音两大类。如汉语音节啊(a)只有一个音素,爱(ai)有两个音素,代(dai)有三个音素等。

[0037] 获取与待检测音频对应的音素序列可以通过语音识别模块中的声学模型接收音频信号进行特征提取,经过转换得到。在声学模型给出音素序列之后,从候选的文字序列中找出概率最大的字符串序列作为识别top1文本,作为文本序列。

[0038] 步骤S102、基于交叉注意力机制,对文本序列和音素序列进行特征提取和融合,获取与待检测音频对应的句子特征向量;

[0039] 步骤S103、基于句子特征向量,获取用于表征待检测音频的意图的预测结果。

[0040] 该步骤中,待检测音频的意图的预测结果可以为分类结果,可以根据需要自己设置。

[0041] 可选地,步骤S103具体包括:输入句子特征向量至全连接神经网络进行分类预测,获取预测结果。

[0042] 全连接神经网络(FFNN)将信息从前面反馈到后面(分别是输入和输出)。神经网络通常被描述为具有多个层,其中每个层由并行的输入、隐藏或输出神经元组成。单独的一层不会有连接,通常相邻的两层是全连接的(一层的每个神经元连接到另一层的每个神经元)。最简单实用的网络有两个输入神经元和一个输出神经元,可用来建模逻辑门。全连接神经网络起到映射句子特征向量的作用,得到用户可以读懂的预测结果。

[0043] 人们通常通过反向传播算法来训练全连接神经网络,给网络一对数据集(输入数据集+期望的输出数据集),这叫做监督学习,而不是只给它输入,让网络来填补空白的非监督学习。反向传播的错误通常是输入和输出之间的差异的一些变体。假设网络有足够多的隐藏神经元,理论上它总是可以模拟输入和输出之间的关系。实际上,它们的使用非常有限,但是它们通常与其他网络结合在一起形成新的网络。

[0044] 可选地,参照图2,步骤S102包括:

[0045] 步骤S2011、分别对文本序列和音素序列进行编码,获取与文本序列对应的文本序列特征向量,以及与音素序列对应的音素序列特征向量;

[0046] 对于步骤S2011,编码是信息从一种形式或格式转换为另一种形式的过程。用预先规定的方法将文字、数字或其它对象编成数码,或将信息、数据转换成规定的电脉冲信号。这里的编码是将文本序列转换为特征向量或者可以叫做特征向量矩阵。编码的方式有多种,ASCII码,ISO-8859-1,GB18030,还可以选择神经网络进行编码,如双向神经网络。

[0047] 步骤S2012、根据文本序列权重参数和音素序列权重参数,对文本序列特征向量和音素序列特征向量进行拼接,获取句子特征向量。

[0048] 对于步骤S2012,权重参数是指某一因素或指标相对于某一事物的重要程度,体现的不仅仅是某一因素或指标所占的百分比,强调的是因素或指标的相对重要程度,倾向于贡献度或重要性。通过文本序列和音素序列不同的权重参数,应用文本序列权重参数对文本序列特征向量和音素序列权重参数对音素序列特征向量进行加权并求和,得到信息互补的加权后的文字特征向量和音素序列特征向量,对其进行拼接得到句子特征向量。

[0049] 可选地,步骤S2011包括:

[0050] S20111、采用双向神经网络对文本序列和音素序列进行编码,获取与文本序列对

应的文本序列特征向量,以及与音素序列对应的音素序列特征向量。

[0051] 具体来说,双向神经网络由前向神经网络和后向神经网络构成,包含了前向与后向的所有信息。双向神经网络的参数可以随机设定或系统自动生成。

[0052] 可选地,参照图3,文本序列权重参数和音素序列权重参数通过对多个预先存储的音频样本进行训练,具体包括:

[0053] 步骤S401、对样本文本序列和样本音素序列进行编码,获取与样本文本序列对应的样本文本序列特征向量,以及与样本音素序列对应的样本音素序列特征向量;

[0054] 步骤S401中的编码,是信息从一种形式或格式转换为另一种形式的过程。用预先规定的方法将文字、数字或其它对象编成数码,或将信息、数据转换成规定的电脉冲信号。这里的编码是将文本序列转换为特征向量或者可以叫做特征向量矩阵。编码的方式有多种,ASCII码,ISO-8859-1,GB18030,还可以选择神经网络进行编码,如双向神经网络。但步骤S401中的编码方式应该和步骤S2011相同。

[0055] 步骤S401中的样本文本序列和样本音素序列都可以通过对音频样本进行语音识别得到,多个音频样本为预先存储的多个音频,音频可为从互联网数据库中下载的,也可以是从实际使用中通话音频,还可以是指定文本进行人工阅读得到的音频录音。

[0056] 步骤S402、计算样本文本序列特征向量中每个文本的嵌入向量与样本音素序列特征向量的余弦相似度,得到每个文本的权重,通过归一化指数函数,归一化所有文本对应的权重,得到文本序列权重参数;

[0057] 具体来说,权重可通过划分多个层次指标进行判断和计算,常用的方法包括层次分析法、模糊法、模糊层次分析法和专家评价法等。

[0058] 归一化指数函数是二分类函数sigmoid在多分类上的推广,目的是将多分类的结果以概率的形式展现出来。归一化指数函数可以确保各个文本的权重的概率之和等于1。我们只需要将转换后的结果进行归一化处理。方法就是将转化后的结果除以所有转化后结果之和,可以理解为转化后结果占总数的百分比。这样就得到近似的概率。

[0059] 样本文本序列和样本音素序列可以通过对音频样本进行语音识别得到,音频样本可以为预先采集的网络音频,也可以为人工录制的指定文本音频,这里不做限制。对于样本文本序列的标签可以为人工听取的音频样本得出的结论。

[0060] 以一个具体实施例为例来讲解步骤S402,通过获取到的样本文本序列特征向量和样本音素序列特征向量,计算样本文本序列特征向量每个字 W_i 的嵌入向量与样本音素序列特征向量 P 的余弦相似度,得到字 W_i 的权重,通过softmax函数,归一化所有字对应的权重 $weight_i$ 。

[0061] 步骤S403、根据文本序列权重参数得到音素序列权重参数。

[0062] 由于音素序列权重参数和文本序列权重参数存在对应关系,通过文本序列参数进行运算可以获得音素序列权重参数,一般为互补关系。

[0063] 步骤S404、应用文本序列权重参数和音素序列权重参数加权求和文字序列和音素序列的特征向量矩阵,拼接文字和音素对应的特征向量接入全连接神经网络,得到表征音频样本的意图的预测结果;

[0064] 步骤S405、确定由人工听取的音频样本转写的真实文本;

[0065] 步骤S406、基于交叉熵损失函数,计算表征音频样本的意图的预测结果与真实文

本的差异；

[0066] 步骤S407、基于差异，调整文本序列权重参数和音素序列权重参数。

[0067] 由于初次训练计算的文本序列权重参数和音素序列权重参数与实际使用的文本序列权重参数和音素序列权重参数存在差异，为了能得到准确的预测结果，需要在训练的时候，不断调整文本序列权重参数和音素序列权重参数，得到合适的文本序列权重参数和音素序列权重参数。由此可见，通过本公开实施例进行调整，得到调整后的文本序列权重参数和音素序列权重参数，调整后的文本序列权重参数和音素序列权重参数更适用于实际应用，运用到实际应用中可以提升口语意图识别的准确率。

[0068] 以实际应用为例，在实际中，参照图4，交叉注意力模型训练包括在获取到训练数据后，首先采用两个不同的双向神经网络编码器分别对文字序列和音素序列进行编码，得到文字序列和音素序列的特征向量。根据交叉注意力机制在交叉注意力参数生成器计算文字序列和音素序列各自的权重参数，其中，文字序列的权重参数计算方式为对两个Bi-LSTM编码器输出的文字序列特征向量矩阵和音素序列特征向量矩阵，计算每个字 W_i 的嵌入向量与音素向量矩阵 P 的余弦相似度，得到字 W_i 的权重，通过softmax函数，归一化所有字对应的权重 $weight_i$ ，音素序列权重参数通过文字序列权重参数得到。应用权重参数加权求和文字序列和音素序列的特征向量矩阵，得到信息互补的文字向量特征和音素向量特征，拼接文字和音素对应的特征向量得到最终的句子向量表征接入全连接神经网络结合交叉熵损失函数，计算表征音频样本的意图的预测结果与真实文本的差异，返回交叉注意力模型继续训练权重参数至满足条件。

[0069] 交叉注意力模型需要的训练数据主要包括文本和音素两部分数据。其中文本数据由人工听取音频转写的真实文本和语音识别产出的识别top1文本组成。在音素数据中，真实文本的音素主要由音素对照表转换得到，对于多音字的情况，采用最常用的发音对应的音素。语音识别可以直接输出音素，即语音识别产出的识别top1文本的音素由语音识别直接输出。训练数据对应的意图标签采用人工听取音频标注的意图。

[0070] 以具体应用为例，参照图5，输入一段语音，对其进行语音识别，得到TOP1文本和对应音素，输入交叉注意力模型，得到文字序列和音素序列各自的权重参数，进行加权嵌入，得到加权文字序列权重参数的文本序列特征向量和加权音素序列权重参数的音素序列特征向量，进行拼接送入全连接网络，得到表征语音意图的预测结果。

[0071] 本公开实施例提供的意图识别装置进行描述，下文描述的意图识别装置与上文描述的意图识别方法可相互对应参照。

[0072] 参照图6，本公开实施例提供了一种意图识别装置，包括：语音识别模块601、交叉注意力模块602、全连接神经网络分类模块603。其中，语音识别模块601，用于获取与待检测音频对应的音素序列，并基于音素序列，确定与待检测音频对应的文本序列；交叉注意力模块602，用于基于交叉注意力机制，对文本序列和音素序列进行特征提取和融合，获取与待检测音频对应的句子特征向量；全连接神经网络分类模块603，用于基于句子特征向量，获取用于表征待检测音频的意图的预测结果。

[0073] 本公开实施例提供的意图识别装置，通过语音识别模块601获取与待检测音频对应的音素序列，并基于音素序列，确定与待检测音频对应的文本序列，通过交叉注意力模块602对文本序列和音素序列进行特征提取和融合，而非仅利用文本序列的特征信息，获取到

的句子特征向量将字的特征与音素的特征进行无缝融合,充分利用了语音识别输出的文本信息和音素信息,提升噪声文本下的句子特征向量表征,在面对发音声调错误和近音错误的时候,能根据句子文字序列和音素序列重建真实文本对应的句子特征向量表征,通过全连接神经网络分类模块603,基于句子特征向量,获取用于表征所述待检测音频的意图的预测结果,预测结果的准确率提高,起到了提升口语意图识别的准确率。

[0074] 下面对图6所示的实施例的各个模块做进一步说明。

[0075] 语音识别模块601中,音素为根据语音的自然属性划分出来的最小语音单位,依据音节里的发音动作来分析,一个动作构成一个音素。音素分为元音与辅音两大类。如汉语音节啊(a)只有一个音素,爱(ai)有两个音素,代(dai)有三个音素等。

[0076] 获取与待检测音频对应的音素序列可以通过语音识别模块中的声学模型接收音频信号进行特征提取,经过转换得到。在声学模型给出音素序列之后,从候选的文字序列中找出概率最大的字符串序列作为识别top1文本,作为文本序列。

[0077] 全连接神经网络分类模块603中,待检测音频的意图的预测结果可以为分类结果,可以根据需要自己设置。

[0078] 可选地,全连接神经网络分类模块603包括:输入句子特征向量至全连接神经网络进行分类预测,获取预测结果。

[0079] 全连接神经网络分类模块603中,全连接神经网络(FFNN)将信息从前面反馈到后面(分别是输入和输出)。神经网络通常被描述为具有多个层,其中每个层由并行的输入、隐藏或输出神经元组成。单独的一层不会有连接,通常相邻的两层是全连接的(一层的每个神经元连接到另一层的每个神经元)。最简单实用的网络有两个输入神经元和一个输出神经元,可用来建模逻辑门。全连接神经网络起到映射句子特征向量的作用,得到用户可以读懂的预测结果。

[0080] 人们通常通过反向传播算法来训练全连接神经网络,给网络一对数据集(输入数据集+期望的输出数据集),这叫做监督学习,而不是只给它输入,让网络来填补空白的非监督学习。反向传播的错误通常是输入和输出之间的差异的一些变体。假设网络有足够多的隐藏神经元,理论上它总是可以模拟输入和输出之间的关系。实际上,它们的使用非常有限,但是它们通常与其他网络结合在一起形成新的网络。

[0081] 可选地,参照图7交叉注意力模块602具体包括:编码单元6021和拼接单元6022。其中,编码单元,分别对文本序列和音素序列进行编码,获取与文本序列对应的文本序列特征向量,以及与音素序列对应的音素序列特征向量;拼接单元,根据文本序列权重参数和音素序列权重参数,对文本序列特征向量和音素序列特征向量进行拼接,获取句子特征向量。

[0082] 对于编码单元6021,编码是信息从一种形式或格式转换为另一种形式的过程。用预先规定的方法将文字、数字或其它对象编成数码,或将信息、数据转换成规定的电脉冲信号。这里的编码是将文本序列转换为特征向量或者可以叫做特征向量矩阵。编码的方式有多种,ASCII码,ISO-8859-1,GB18030,还可以选择神经网络进行编码,如双向神经网络。

[0083] 对于拼接单元6022,权重参数是指某一因素或指标相对于某一事物的重要程度,体现的不仅仅是某一因素或指标所占的百分比,强调的是因素或指标的相对重要程度,倾向于贡献度或重要性。通过文本序列和音素序列不同的权重参数,应用文本序列权重参数对文本序列特征向量和音素序列权重参数对音素序列特征向量进行加权并求和,得到信

息互补的加权后的文字特征向量和音素序列特征向量,对其进行拼接得到句子特征向量。

[0084] 可选地,编码单元6021具体包括:采用双向神经网络对文本序列和音素序列进行编码,获取与文本序列对应的文本序列特征向量,以及与音素序列对应的音素序列特征向量。

[0085] 具体来说,双向神经网络由前向神经网络和后向神经网络构成,包含了前向与后向的所有信息。双向神经网络的参数可以随机设定或系统自动生成。

[0086] 可选地,参照图8,文本序列权重参数和音素序列权重参数通过对多个预先存储的音频样本进行训练,具体包括:

[0087] 样本编码单元801,对样本文本序列和样本音素序列进行编码,获取与样本文本序列对应的样本文本序列特征向量,以及与样本音素序列对应的样本音素序列特征向量;

[0088] 样本编码单元801中的样本文本序列和样本音素序列都可以通过对音频样本进行语音识别得到。

[0089] 多个音频样本为预先存储的多个音频,音频可为从互联网数据库中下载的,也可以是从实际使用中通话音频,还可以是指定文本进行人工阅读得到的音频录音。

[0090] 文本序列参数计算单元802,计算样本文本序列特征向量中每个文本的嵌入向量与样本音素序列特征向量的余弦相似度,得到每个文本的权重,通过归一化指数函数,归一化所有文本对应的权重,得到文本序列权重参数;

[0091] 具体来说,权重可通过划分多个层次指标进行判断和计算,常用的方法包括层次分析法、模糊法、模糊层次分析法和专家评价法等。

[0092] 归一化指数函数是二分类函数sigmoid在多分类上的推广,目的是将多分类的结果以概率的形式展现出来。归一化指数函数可以确保各个文本的权重的概率之和等于1。我们只需要将转换后的结果进行归一化处理。方法就是将转化后的结果除以所有转化后结果之和,可以理解为转化后结果占总数的百分比。这样就得到近似的概率。

[0093] 样本文本序列和样本音素序列可以通过对音频样本进行语音识别得到,音频样本可以为预先采集的网络音频,也可以为人工录制的指定文本音频,这里不做限制。对于样本文本序列的标签可以为人工听取的音频样本得出的结论。

[0094] 以一个具体实施例为例来讲解文本序列参数计算单元802,文本序列参数计算单元802通过获取到的样本文本序列特征向量和样本音素序列特征向量,计算样本文本序列特征向量每个字 W_i 的嵌入向量与样本音素序列特征向量 P 的余弦相似度,得到字 W_i 的权重,通过softmax函数,归一化所有字对应的权重weight i 。

[0095] 音素序列参数计算单元803,根据文本序列权重参数得到音素序列权重参数;

[0096] 由于音素序列权重参数和文本序列权重参数存在对应关系,通过文本序列参数进行运算可以获得音素序列权重参数,一般为互补关系。

[0097] 应用单元804,应用文本序列权重参数和音素序列权重参数加权求和文字序列和音素序列的特征向量矩阵,拼接文字和音素对应的特征向量接入全连接神经网络,得到表征音频样本的意图的预测结果;

[0098] 确定文本单元805,确定由人工听取的音频样本转写的真实文本;

[0099] 表征差异单元806,基于交叉熵损失函数,计算表征音频样本的意图的预测结果与真实文本的差异;

[0100] 调整单元807,基于差异,调整文本序列权重参数和音素序列权重参数。

[0101] 由于初次训练计算的文本序列权重参数和音素序列权重参数与实际使用的文本序列权重参数和音素序列权重参数存在差异,为了能得到准确的预测结果,需要在训练的时候,不断调整文本序列权重参数和音素序列权重参数,得到合适的文本序列权重参数和音素序列权重参数。由此可见,通过本公开实施例进行调整,得到调整后的文本序列权重参数和音素序列权重参数,调整后的文本序列权重参数和音素序列权重参数更适用于实际应用,运用到实际应用中可以提升口语意图识别的准确率。

[0102] 以实际应用为例,在实际中,交叉注意力模块602是通过训练得到的,训练过程包括在获取到训练数据后,首先采用两个不同的双向神经网络编码器分别对文字序列和音素序列进行编码,得到文字序列和音素序列的特征向量。根据交叉注意力机制在交叉注意力参数生成器计算文字序列和音素序列各自的权重参数,其中,文字序列的权重参数计算方式为对两个Bi-LSTM编码器输出的文字序列特征向量矩阵和音素序列特征向量矩阵,计算每个字 W_i 的嵌入向量与音素向量矩阵 P 的余弦相似度,得到字 W_i 的权重,通过softmax函数,归一化所有字对应的权重 $weight_i$,音素序列权重参数通过文字序列权重参数得到。应用权重参数加权求和文字序列和音素序列的特征向量矩阵,得到信息互补的文字向量特征和音素向量特征,拼接文字和音素对应的特征向量得到最终的句子向量表征接入全连接神经网络结合交叉熵损失函数,计算表征音频样本的意图的预测结果与真实文本的差异,返回交叉注意力模块602继续训练权重参数至满足条件。

[0103] 交叉注意力模块602需要的训练数据主要包括文本和音素两部分数据。其中文本数据由人工听取音频转写的真实文本和语音识别产出的识别top1文本组成。在音素数据中,真实文本的音素主要由音素对照表转换得到,对于多音字的情况,采用最常用的发音对应的音素。语音识别可以直接输出音素,即语音识别产出的识别top1文本的音素由语音识别直接输出。训练数据对应的意图标签采用人工听取音频标注的意图。

[0104] 以装置具体应用实施例为例,输入一段语音,在语音模块601中对其进行语音识别,得到TOP1文本和对应音素,输入交叉注意力模块602,得到文字序列和音素序列各自的权重参数,进行加权嵌入,得到加权文字序列权重参数的文本序列特征向量和加权音素序列权重参数的音素序列特征向量,进行拼接送入全连接神经网络分类模块603,得到表征语音意图的预测结果。

[0105] 图9示例了一种电子设备的实体结构示意图,如图9所示,该电子设备可以包括:处理器(processor)910、通信接口(Communications Interface)920、存储器(memory)930和通信总线940,其中,处理器910,通信接口920,存储器930通过通信总线940完成相互间的通信。处理器910可以调用存储器930中的逻辑指令,以执行上述各实施例提供的意图识别方法。

[0106] 此外,上述的存储器930中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本公开实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案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一个存储介质中,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,服务器,或者网络设备等)执行本公开各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器

(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种上述各方法可以存储程序代码的介质。

[0107] 另一方面,本公开还提供一种计算机程序产品,所述计算机程序产品包括存储的非暂态计算机可读存储介质上的计算机程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,计算机能够执行上述各实施例所提供的意图识别方法。

[0108] 又一方面,本公开还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各实施例提供的意图识别方法。

[0109] 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。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0110] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述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,该计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、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,服务器,或者网络设备等)执行各个实施例或者实施例的某些部分所述的方法。

[0111] 最后应说明的是:以上实施例仅用以说明本公开的技术方案,而非对其限制;尽管参照前述实施例对本公开进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本公开各实施例技术方案的精神和范围。

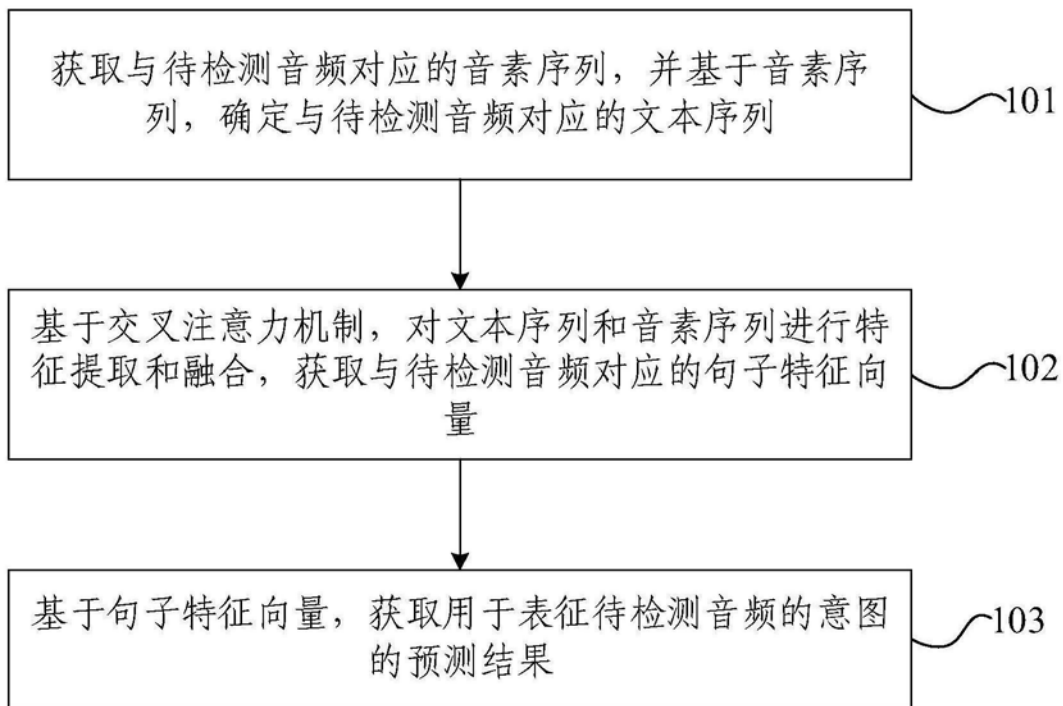


图1

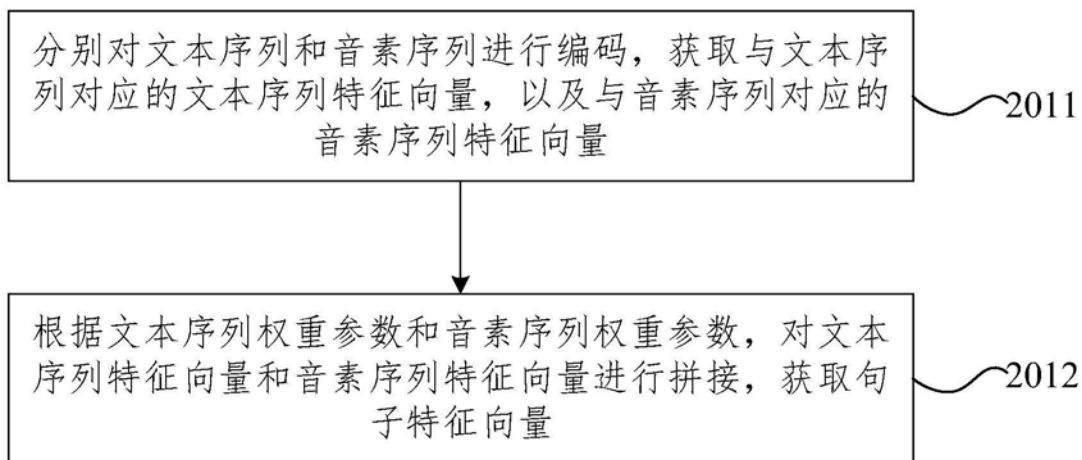


图2

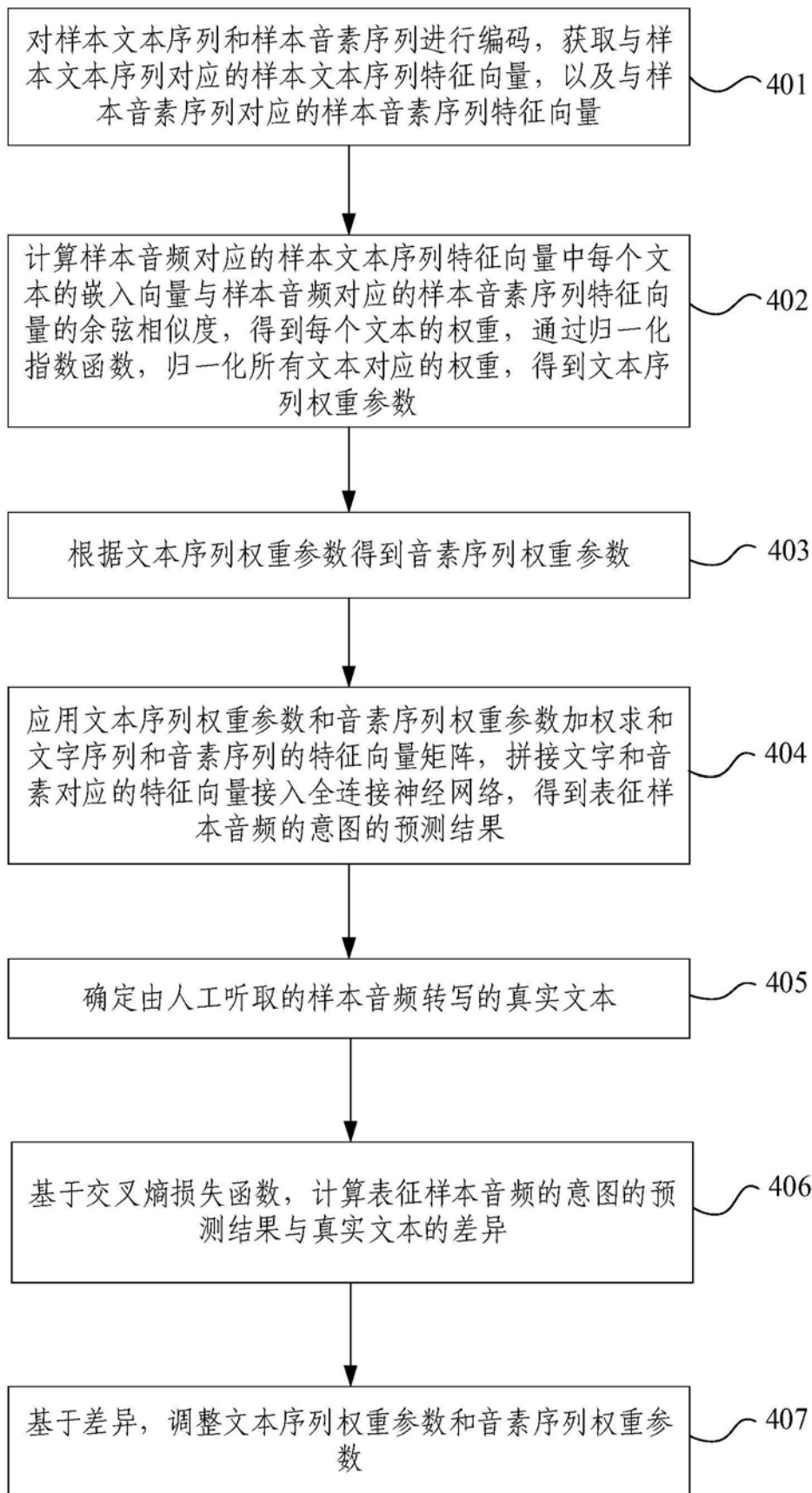


图3

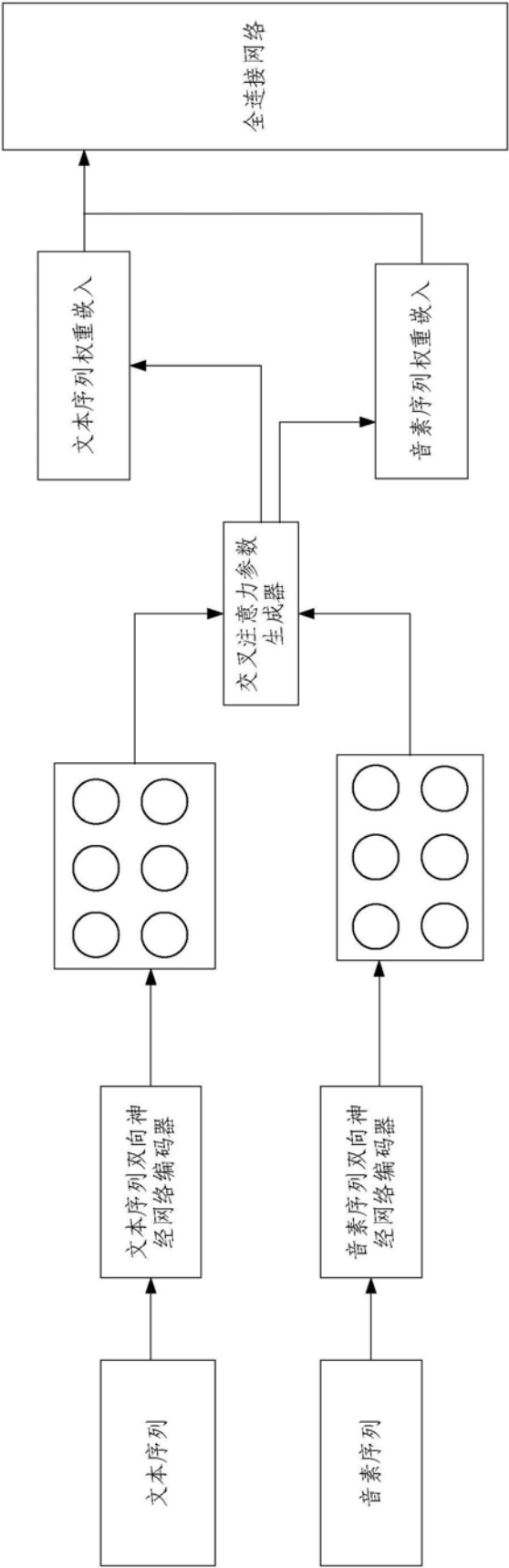


图4

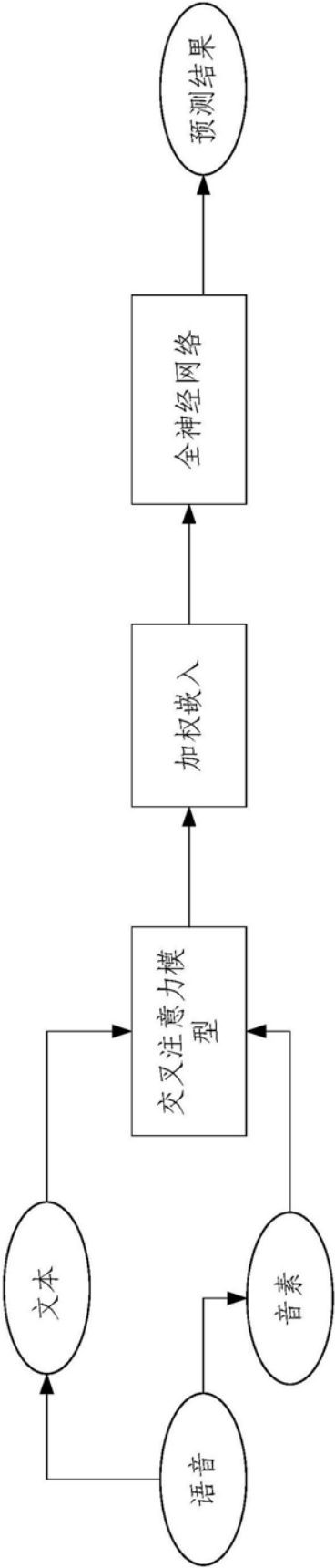


图5

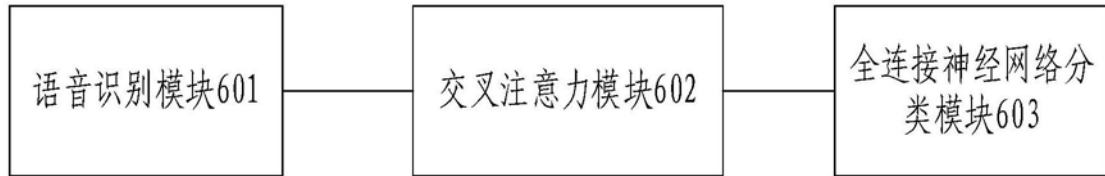


图6



图7



图8

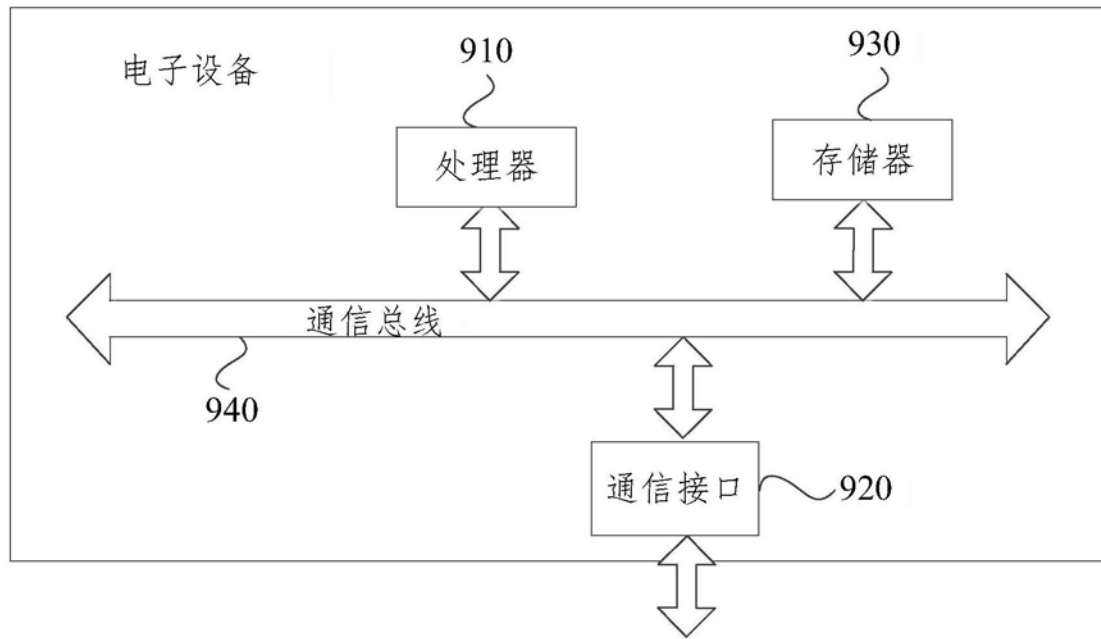


图9