



(21) 申请号 202311745175.0

(22) 申请日 2023.12.18

(71) 申请人 科大讯飞股份有限公司

地址 230088 安徽省合肥市高新开发区望江西路666号

(72) 发明人 王士进 韩凯 吴奎 金海

盛志超 刘聪 胡国平

(74) 专利代理机构 深圳市威世博知识产权代理

事务所(普通合伙) 44280

专利代理师 何倚雯

(51) Int. Cl.

G10L 15/26 (2006.01)

G10L 15/06 (2013.01)

G10L 15/16 (2006.01)

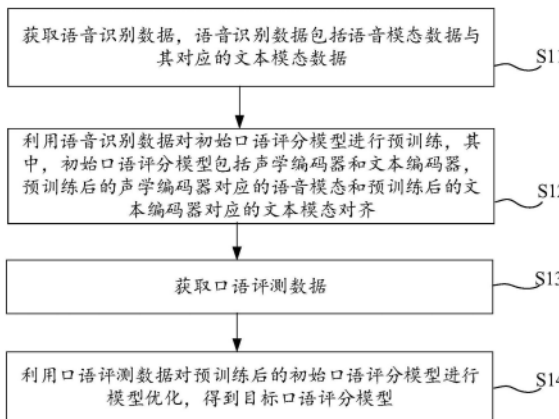
权利要求书2页 说明书8页 附图5页

(54) 发明名称

口语评分模型的训练方法、口语评分方法以及相关设备

(57) 摘要

本申请公开了一种口语评分模型的训练方法,该方法包括获取语音识别数据,所述语音识别数据包括语音模态数据与其对应的文本模态数据;利用所述语音识别数据对初始口语评分模型进行预训练,其中,所述初始口语评分模型包括声学编码器和文本编码器,预训练后的所述声学编码器对应的语音模态和预训练后的所述文本编码器对应的文本模态对齐;获取口语评测数据;利用所述口语评测数据对预训练后的所述初始口语评分模型进行模型优化,得到目标口语评分模型。本申请还公开了口语评分方法以及相关设备。本申请提高了自动口语评估的效率与准确度。



1. 一种口语评分模型的训练方法,其特征在于,包括:
 - 获取语音识别数据,所述语音识别数据包括语音模态数据与其对应的文本模态数据;
 - 利用所述语音识别数据对初始口语评分模型进行预训练,其中,所述初始口语评分模型包括声学编码器和文本编码器,预训练后的所述声学编码器对应的语音模态和预训练后的所述文本编码器对应的文本模态对齐;
 - 获取口语评测数据;
 - 利用所述口语评测数据对预训练后的所述初始口语评分模型进行模型优化,得到目标口语评分模型。
2. 根据权利要求1所述的方法,其特征在于,所述基于所述语音识别数据对初始口语评分模型进行预训练,包括:
 - 初始化所述声学编码器与所述文本编码器;以及
 - 利用损失函数,对所述初始口语评分模型进行训练。
3. 根据权利要求2所述的方法,其特征在于,所述损失函数包括语音识别损失函数,所述初始化所述声学编码器,包括:
 - 利用预设编码器初始化所述声学编码器;
 - 在所述声学编码器的预设位置连接所述语音识别损失函数,以学习文本模态的表征能力。
4. 根据权利要求2所述的方法,其特征在于,所述文本编码器包括语音文本对比编码模块和语音文本匹配编码模块,其中,所述语音文本对比编码模块和所述语音文本匹配编码模块分别包括至少一transformer子层,所述语音文本对比编码模块和所述语音文本匹配编码模块共用模型参数;
 - 所述初始化所述文本编码器,包括:
 - 在所述语音文本匹配编码模块中每一transformer子层的自注意层与前馈层之间增加交叉注意层,以输入所述声学编码器的输出信息。
5. 根据权利要求2所述的方法,其特征在于,所述损失函数包括语音文本对比损失函数,所述利用损失函数对所述初始口语评分模型进行训练,包括:
 - 计算每个语音向量与其匹配的文本向量之间的距离语音到文本的损失,以得到所述语音模态数据到与其对应的文本模态数据的损失;
 - 计算每个文本向量与其匹配的语音向量之间的距离,以得到所述文本模态数据到与其对应的语音模态数据的损失;
 - 最小化所述语音模态数据到与其对应的文本模态数据的损失与所述文本模态数据到与其对应的语音模态数据的损失之和,以实现所述初始口语评分模型进行训练。
6. 根据权利要求2所述的方法,其特征在于,所述损失函数包括语音文本匹配损失函数,所述利用损失函数对所述初始口语评分模型进行训练,包括:
 - 利用语音文本匹配损失函数,对所述初始口语评分模型的输出进行二分类,以确定所述语音模态数据与所述文本模态数据是否相匹配。
7. 根据权利要求1所述的方法,其特征在于,所述口语评测数据包括语音作答数据与对应的试题文本数据,所述利用所述口语评测数据对预训练后的所述初始口语评分模型进行模型优化,包括:

将所述口语评测数据输入到预训练后的所述初始口语评分模型；
基于预训练后的所述初始口语评分模型的输出，利用均方误差进行口语分数预测。

8. 一种口语评分方法，其特征在于，包括：

获取口语测试数据；

将所述口语测试数据输入口语评分模型，以输出对应的口语评分；

其中，所述口语评分模型是基于权利要求1至7任一项所述的口语评分模型的训练方法得到的。

9. 一种电子设备，其特征在于，包括相互耦接的存储器和处理器，所述处理器用于执行所述存储器中存储的程序指令，以实现权利要求1至7任一项所述的口语评分模型的训练方法，或以实现权利要求8所述的口语评分方法。

10. 一种非易失性计算机可读存储介质，其上存储有程序指令，其特征在于，所述程序指令被处理器执行时实现权利要求1至7任一项所述的口语评分模型的训练方法，或者实现权利要求8所述的口语评分方法。

口语评分模型的训练方法、口语评分方法以及相关设备

技术领域

[0001] 本申请的所公开实施例涉及人工智能技术领域,且更具体而言,涉及一种口语评分模型的训练方法、口语评分方法以及相关设备。

背景技术

[0002] 计算机辅助发音训练(Computer-Assisted Pronunciation Training, CAPT)已经成为非母语(L2)说话者学习不同外语语言的有效工具。自动语音评估(Automatic speech Assessment)作为CAPT重要的组成部分,在帮助自主语言学习者提高口语水平方面发挥着举足轻重的作用。自动语音评估任务也由最初的朗读句子、单词等受限任务中扩展到更多半开放测试任务中,如话题讨论、听后复述、看图说话、听后问答等。在这些任务中期待学生的回答中包含题目给的提示信息。传统的开放自动口语评估方法依赖于从ASR转录文本中提取声音特征(例如发音准确性和流畅性)或文本特征(例如语法和内容)作为回归器或者分类器的输入来对口语作答进行评分。近些年,随着ASR效果的不断提升,自动语音评估问题逐渐变为在转写结果上进行,并通过自然语言处理(NLP)技术来解决评分问题。无论是基于声学特征还是文本特征的语音评估方案,都高度依赖于ASR的准确性,然而ASR错误会产生级联误差,极大地影响自动口语评估的性能。

发明内容

[0003] 根据本申请的实施例,本申请提出一种口语评分模型的训练方法、口语评分方法以及相关设备,以解决相关问题。

[0004] 本申请的第一方面公开了口语评分模型的训练方法,包括:获取语音识别数据,所述语音识别数据包括语音模态数据与其对应的文本模态数据;利用所述语音识别数据对初始口语评分模型进行预训练,其中,所述初始口语评分模型包括声学编码器和文本编码器,预训练后的所述声学编码器对应的语音模态和预训练后的所述文本编码器对应的文本模态对齐;获取口语评测数据;利用所述口语评测数据对预训练后的所述初始口语评分模型进行模型优化,得到目标口语评分模型。

[0005] 在一些实施例中,所述基于所述语音识别数据对初始口语评分模型进行预训练,包括:初始化所述声学编码器与所述文本编码器;以及利用损失函数,对所述初始口语评分模型进行训练。

[0006] 在一些实施例中,所述损失函数包括语音识别损失函数,所述初始化所述声学编码器,包括:利用预设编码器初始化所述声学编码器;在所述声学编码器的预设位置连接所述语音识别损失函数,以学习文本模态的表征能力。

[0007] 在一些实施例中,所述文本编码器包括语音文本对比编码模块和语音文本匹配编码模块,其中,所述语音文本对比编码模块和所述语音文本匹配编码模块分别包括至少一transformer子层,所述语音文本对比编码模块和所述语音文本匹配编码模块共用模型参数;所述初始化所述文本编码器,包括:在所述语音文本匹配编码模块中每一transformer

子层的自注意层与前馈层之间增加交叉注意层,以输入所述声学编码器的输出信息。

[0008] 在一些实施例中,所述损失函数包括语音文本对比损失函数,所述利用损失函数对所述初始口语评分模型进行训练,包括:计算每个语音向量与其匹配的文本向量之间的距离语音到文本的损失,以得到所述语音模态数据到与其对应的文本模态数据的损失;计算每个文本向量与其匹配的语音向量之间的距离,以得到所述文本模态数据到与其对应的语音模态数据的损失;最小化所述语音模态数据到与其对应的文本模态数据的损失与所述文本模态数据到与其对应的语音模态数据的损失之和,以实现所述初始口语评分模型进行训练。

[0009] 在一些实施例中,所述损失函数包括语音文本匹配损失函数,所述利用损失函数对所述初始口语评分模型进行训练,包括:利用语音文本匹配损失函数,对所述文本编码器的输出进行二分类,以确定所述语音模态数据与所述文本模态数据是否相匹配。

[0010] 在一些实施例中,所述口语评测数据包括语音作答数据与对应的试题文本数据,所述利用所述口语评测数据对预训练后的所述初始口语评分模型进行模型优化,包括:将所述口语评测数据输入到预训练后的所述初始口语评分模型;基于预训练后的所述初始口语评分模型的输出,利用均方误差进行口语分数预测。

[0011] 本申请第二方面公开了一种口语评分方法,包括:获取口语测试数据;将所述口语测试数据输入口语评分模型,以输出对应的口语评分;其中,所述口语评分模型是基于第一方面中所述的口语评分模型的训练方法得到的。

[0012] 本申请第三方面公开了一种电子设备,包括相互耦接的存储器和处理器,所述处理器用于执行所述存储器中存储的程序指令,以实现第一方面中所述的口语评分模型的训练方法,或以实现第二方面中所述的口语评分方法。

[0013] 本申请第四方面公开了一种非易失性计算机可读存储介质,其上存储有程序指令,所述程序指令被处理器执行时实现第一方面中所述的口语评分模型的训练方法,或以实现第二方面中所述的口语评分方法。

[0014] 本申请的有益效果有:利用语音识别数据对初始口语评分模型进行预训练,其中,初始口语评分模型包括声学编码器和文本编码器,预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐,进而利用口语评测数据对预训练后的初始口语评分模型进行模型优化,得到目标口语评分模型,通过将预训练使得语音模态与文本模态对齐,提高了自动口语评估的效率与准确度。

附图说明

[0015] 下面将结合附图及实施方式对本申请作进一步说明,附图中:

[0016] 图1是本申请实施例的口语评分模型的训练方法的流程示意图;

[0017] 图2是本申请一实施例的声学编码器的结构示意图;

[0018] 图3是本申请一实施例的文本编码器的结构示意图;

[0019] 图4是本申请一实施例的预训练模型的结构示意图;

[0020] 图5是本申请实施例的口语评分方法的流程示意图;

[0021] 图6是本申请实施例的电子设备的结构示意图;

[0022] 图7是本申请实施例的非易失性计算机可读存储介质的结构示意图。

具体实施方式

[0023] 在本申请中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0024] 本申请中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。另外,本文中字符“/”,一般表示前后关联对象是一种“或”的关系。此外,本文中的“多”表示两个或者多于两个。另外,本文中术语“至少一种”表示多种中的任意一种或多种中的至少两种的任意组合,例如,包括A、B、C中的至少一种,可以表示包括从A、B和C构成的集合中选择的任意一个或多个元素。另外,本申请中的术语“第一”、“第二”、“第三”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。

[0025] 为使本领域的技术人员更好地理解本申请的技术方案,下面结合附图和具体实施方式对本申请的技术方案做进一步详细描述。

[0026] 请参阅图1,图1是本申请实施例的口语评分模型的训练方法的流程示意图。该方法的执行主体可以是有计算功能的电子设备,例如,微型计算机、服务器,以及笔记本电脑、平板电脑等移动设备等。

[0027] 需注意的是,若有实质上相同的结果,本申请的方法并不以图1所示的流程顺序为限。

[0028] 在一些可能的实现方式中,该方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现,如图1所示,该方法可以包括如下步骤:

[0029] S11:获取语音识别数据,语音识别数据包括语音模态数据与其对应的文本模态数据。

[0030] 获取语音识别数据,语音识别数据包括语音模态数据与其对应的文本模态数据,例如获取大量语音识别数据,其中语音模态数据可以是一段语音,例如“今天天气真好”,则与其对应的文本模态数据,即文字“今天天气真好”。

[0031] S12:利用语音识别数据对初始口语评分模型进行预训练,其中,初始口语评分模型包括声学编码器和文本编码器,预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐。

[0032] 初始口语评分模型包括声学编码器,声学编码器采用的是conformer的网络结构,Conformer是用卷积增强的Transformer,声学编码器的Conformer Blocks模块包括多个Conformer子模块,每个Conformer子模块主要包括多头自注意力模块(Multi-Head SelfAttention)、卷积模块(Convolution)、前馈层(Feed Forward)和层标准化模块(Layer Norm),其中多头自注意力模块负责建模全局的上下文依赖,而卷积模块负责抓取局部特征之间的相关性。

[0033] 初始口语评分模型包括文本编码器,文本编码器包括多个transformer子模块,其中,transformer子模块包括自注意层(Self-Attention)和前馈网络(Feed Forward),或者transformer子模块包括自注意层(Self-Attention)、前馈网络(Feed Forward)与交叉注意层(Cross Attention)。

[0034] 利用语音识别数据对初始口语评分模型进行预训练,即利用获取的语音识别数据对初始口语评分模型进行预训练,得到预训练后的初始口语评分模型,其中预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐,即通过预训练使得初始口语评分模型中对应的语音模态与文本模态之间可以相互表示。

[0035] S13:获取口语评测数据。

[0036] 获取口语评测数据,即口语评测任务数据,例如口语问答对应的语音数据和文本数据,朗读评测对应的语音数据和文本数据,话题陈述对应的语音数据和文本数据等。

[0037] S14:利用口语评测数据对预训练后的初始口语评分模型进行模型优化,得到目标口语评分模型。

[0038] 利用口语评测数据对预训练后的初始口语评分模型进行模型优化,例如基于口语评测数据对预训练后的初始口语评分模型进行模型优化,以实现预测对应的口语评分,进而得到目标口语评分模型。

[0039] 在本实施例中,利用语音识别数据对初始口语评分模型进行预训练,其中,初始口语评分模型包括声学编码器和文本编码器,预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐,进而利用口语评测数据对预训练后的初始口语评分模型进行模型优化,得到目标口语评分模型,通过将预训练使得语音模态与文本模态对齐,提高了自动口语评估的性能,例如自动口语评估的效率与准确度。

[0040] 在一些实施例中,基于语音识别数据对初始口语评分模型进行预训练,包括:初始化声学编码器与文本编码器;以及利用损失函数,对初始口语评分模型进行训练。

[0041] 基于语音识别数据对初始口语评分模型进行预训练,以使得预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐,具体地,预训练包括初始化声学编码器与文本编码器,以及利用损失函数,对初始口语评分模型进行训练,即对初始口语评分模型的声学编码器与文本编码器进行初始化操作,并利用预设的损失函数进行模型预训练,损失函数可以用于量化模型预测与真实匹配之间的差异。

[0042] 在一些实施例中,损失函数包括语音识别损失函数,初始化声学编码器,包括:利用预设编码器初始化声学编码器;在声学编码器的预设位置连接语音识别损失函数,以学习文本模态的表征能力。

[0043] 初始口语评分模型包括声学编码器,如图2所示,图2是本申请一实施例的声学编码器的结构示意图,对输入的音频进行处理,例如以原始音频的10ms为一帧,经SpecAug模块与ConvolutionSubsampling模块进行下采样后变为每帧40ms,接着经过Linear和Dropout模块,得到40维Filter bank特征,进而输入到声学编码器的Conformer Blocks模块,Conformer Block模块包括N个Conformer子模块,如图2的右侧所示,每个Conformer子模块主要包括多头自注意力模块(Multi-Head Self Attention)、卷积模块(Convolution)、前馈层(Feed Forward)和层标准化模块(Layer Norm),其中多头自注意力模块负责建模全局的上下文依赖,而卷积模块负责抓取局部特征之间的相关性。

[0044] 可以理解地,声学编码器中,利用Transformer的自主注意力机制可以对音频的全局相关性进行建模,而卷积可以对抓取局部的相关性,进而可以实现用卷积来增强Transformer可以获取不同视野范围的上下文依赖。

[0045] 损失函数包括语音识别损失函数,即CTC(Connective)

TemporalClassification) 损失函数,通过在神经网络模型中引入CTC约束,使模型能够将输入序列映射到输出序列,同时不需要知道输入序列和输出序列之间的对应关系。

[0046] 进一步地,利用预设编码器初始化声学编码器,例如使用ASR (Automatic Speech Recognition) 的编码器来初始化声学编码器,其中预训练好的ASR编码器中已经包含了内容等信息,跟口语评测任务契合,且使用ASR编码器初始化可以加速多模态预训练收敛速度,节省成本。在声学编码器的预设位置连接语音识别损失函数,以学习文本模态的表征能力,例如,声学编码器具有12层conformer层,在预设位置连接语音识别损失函数,例如可以在第6层conformer层外接语音识别损失函数,即语音识别的CTC目标,使得口语评分模型在训练过程中仍然保留ASR的能力,而后6层conformer层具有学习适配文本模态的表征能力。需要注意的是,声学编码器需要分别提供帧级和句子级的表征同文本编码器进行交互,在第6层的隐层输出起始位置拼接上一个和隐层相同维度的可学习的矢量 h_0 。

[0047] 在一些实施例中,文本编码器包括语音文本对比编码模块和语音文本匹配编码模块,其中,语音文本对比编码模块和语音文本匹配编码模块分别包括至少一transformer子层,语音文本对比编码模块和语音文本匹配编码模块共用模型参数;初始化文本编码器,包括:在语音文本匹配编码模块中每一transformer子层的自注意层与前馈层之间增加交叉注意层,以输入声学编码器的输出信息。

[0048] 初始口语评分模型包括文本编码器,文本编码器中的模型结构与标准BERT模型一致,如图3所示,图3是本申请一实施例的文本编码器的结构示意图,文本编码器包括语音文本对比编码模块和语音文本匹配编码模块,文本编码器包括多个transformer子模块,即transformer子层,其中,语音文本对比编码模块和语音文本匹配编码模块共用模型参,初始化文本编码器,包括在语音文本匹配编码模块中每一transformer子层的自注意层与前馈层之间增加交叉注意层,以输入声学编码器的输出信息,即通过交叉注意层在本文编码器中注入声学信息,具体地,语音文本对比编码模块中的M个transformer子模块分别包括自注意层(Self-Attention)和前馈网络(Feed Forward),语音文本匹配编码模块中的M个transformer子模块分别包括自注意层(Self-Attention)、前馈网络(Feed Forward),以及在自注意层与前馈网络之间的交叉注意层(Cross Attention)。

[0049] 在一些实施例中,损失函数包括语音文本对比损失函数,利用损失函数对初始口语评分模型进行训练,包括:计算每个语音向量与其匹配的文本向量之间的距离语音到文本的损失,以得到语音模态数据到与其对应的文本模态数据的损失;计算每个文本向量与其匹配的语音向量之间的距离,以得到文本模态数据到与其对应的语音模态数据的损失;最小化语音模态数据到与其对应的文本模态数据的损失与文本模态数据到与其对应的语音模态数据的损失之和,以实现初始口语评分模型进行训练。

[0050] 损失函数包括语音文本对比损失函数,即ATC (Audio-TextComparison) 损失函数,ATC损失指语音-文本对比损失,通过正负语音-文本对的对比学习,以对齐语音和文本的潜在特征空间。可以理解地,在训练过程中试图最大化正确的语音-文本匹配的相似性分数,并最小化与负样本(即不匹配的语音-文本对)的相似性,例如学生作答通过声学编码器得到声学特征向量,学生作答语音对应的转写文本经过文本编码器得到文本特征向量,进而两个模态的特征向量的点积计算为相似性分数。

[0051] 具体地,计算每个语音向量与其匹配的文本向量之间的距离语音到文本的损失,

以得到语音模态数据到与其对应的文本模态数据的损失,即计算语音到文本的损失,其中批量处理内(batch)的语音和文本相对应为正例,不对应的为负例;计算每个文本向量与其匹配的语音向量之间的距离,以得到文本模态数据到与其对应的语音模态数据的损失,即计算文本到语音的损失,其中批量处理内(batch)的语音和文本相对应为正例,不对应的为负例。最小化语音模态数据到与其对应的文本模态数据的损失与文本模态数据到与其对应的语音模态数据的损失之和,即最小化语音文本对比中产生的损失,以实现初始口语评分模型进行训练。

[0052] 在一些实施例中,损失函数包括语音文本匹配损失函数,利用损失函数对初始口语评分模型进行训练,包括:利用语音文本匹配损失函数,对文本编码器的输出进行二分类,以确定语音模态数据与文本模态数据是否相匹配。

[0053] 损失函数包括语音文本匹配损失函数,即ATM(Audio-Textmatching)损失函数,利用语音文本匹配损失函数,对文本编码器的输出进行二分类,以确定语音模态数据与文本模态数据是否相匹配,即通过对语音-文本匹配性进行二分类,建模语音-文本多模态信息的相关性,例如利用文本编码器隐层对声学编码器的隐层进行交叉注意机制,并在文本编码器的[CLS]位置进行二分类,若语音数据与文本数据相对应,则目标值为1,若语音数据与文本数据不对应,则目标值为0。

[0054] 在一些实施例中,口语评测数据包括语音作答数据与对应的试题文本数据,利用口语评测数据对预训练后的初始口语评分模型进行模型优化,包括:将口语评测数据输入到预训练后的初始口语评分模型;基于预训练后的初始口语评分模型的输出,利用均方误差进行口语分数预测。

[0055] 口语评测数据包括语音作答数据与对应的试题文本数据,即口语问答,朗读评测,话题陈述等口语评测任务,例如学生的语音作答和文本信息两个模态的信息,其中文本信息包括口语问答题的问题和参考答案,朗读题的文本,话题陈述的话题和参考答案,评分规则等。

[0056] 进一步地,利用口语评测数据对预训练后的初始口语评分模型进行模型优化,例如,将口语评测数据输入到预训练后的初始口语评分模型中,进而基于预训练后的初始口语评分模型的输出,利用均方误差进行口语分数预测,即将文本信息经过文本编码器编码后对声学编码器的隐层进行交叉注意力机制,通过池化平均投影到一维,利用均方误差(MSE)来预测学生作答的分数,进而实现了多模态端到端评测,省去识别的过程并减少级联误差。

[0057] 为了便于理解,对本申请多模态预训练的过程进行举例说明,如图4所示,图4是本申请一实施例的预训练模型的结构示意图,初始口语评分模型包括声学编码器和文本编码器。声学编码器共有 $2N$ 层conformer子模块,可以在第 N 层数据外接语音识别的CTC损失函数,其中 N 可以为6。文本编码器包括语音文本对比编码模块和语音文本匹配编码模块,语音文本对比编码模块中的transformer子模块,包括自注意层(Self-Attention)与前馈网络(Feed Forward),且ATC损失函数连接声学编码器与文本编码器,例如ATC损失函数连接声学编码器与语音文本对比编码模块;语音文本匹配编码模块中的transformer子模块包括自注意层(Self-Attention)、前馈网络(Feed Forward),以及在自注意层与前馈网络之间的交叉注意层(Cross Attention);并使用ATM损失函数,以通过对语音-文本匹配性进行二

分类,建立语音-文本多模态信息的相关性。

[0058] 请参阅图5,图5是本申请实施例的口语评分方法的流程示意图,该方法可以应用于具有计算等功能的电子设备。需注意的是,若有实质上相同的结果,本申请的方法并不以图5所示的流程顺序为限。

[0059] 在一些可能的实现方式中,该方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现,如图5所示,该方法可以包括如下步骤:

[0060] S51:获取口语测试数据。

[0061] 口语评测数据包括语音作答数据与对应的试题文本数据,即口语问答,朗读评测,话题陈述等口语评测任务,例如学生的语音作答和文本信息两个模态的信息,其中文本信息包括口语问答题的问题和参考答案,朗读题的文本,话题陈述的话题和参考答案等。

[0062] S52:将口语测试数据输入口语评分模型,以输出对应的口语评分。

[0063] 将语音作答数据与对应的试题文本数据输入口语评分模型,以得到对应考生的口语评分,其中,口语评分模型是基于上述的口语评分模型的训练方法得到的,即获取语音识别数据,语音识别数据包括语音模态数据与其对应的文本模态数据,利用语音识别数据对初始口语评分模型进行预训练,其中,初始口语评分模型包括声学编码器和文本编码器,预训练后的声学编码器对应的语音模态和预训练后的文本编码器对应的文本模态对齐,获取口语评测数据,利用口语评测数据对预训练后的初始口语评分模型进行模型优化,得到目标口语评分模型,具体内容不再赘述。

[0064] 本领域技术人员可以理解,在具体实施方式的上述方法中,各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定,各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

[0065] 请参阅图6,图6是本申请实施例的电子设备的结构示意图。电子设备60包括相互耦接的存储器61和处理器62,处理器62用于执行存储器61中存储的程序指令,以实现上述的口语评分模型的训练方法实施例的步骤,或者以实现上述口语评分方法实施例的步骤。在一个具体的实施场景中,电子设备60可以包括但不限于:微型计算机、服务器,在此不做限定。

[0066] 具体而言,处理器62用于控制其自身以及存储器61以实现上述口语评分模型的训练方法实施例的步骤,或者以实现上述口语评分方法实施例的步骤。处理器62还可以称为CPU(Central Processing Unit,中央处理单元),处理器62可能是一种集成电路芯片,具有信号的处理能力。处理器62还可以是通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。另外,处理器62可以由集成电路芯片共同实现。

[0067] 请参阅图7,图7为本申请实施例的非易失性计算机可读存储介质的结构示意图。非易失性计算机可读存储介质70用于存储计算机程序701,计算机程序701在被处理器执行时,例如被上述图6实施例中的处理器62执行时,用于实现上述用于口语评分模型的训练方法实施例的步骤,或者以实现上述口语评分方法实施例的步骤。

[0068] 上文对各个实施例的描述倾向于强调各个实施例之间的不同之处,其相同或相似

之处可以互相参考,为了简洁,本文不再赘述。

[0069] 在本申请所提供的几个实施例中,应该理解到,所揭露的方法和相关设备,可以通过其它的方式实现。例如,以上所描述的相关设备实施方式仅仅是示意性的,例如,模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信断开连接可以通过一些接口,装置或单元的间接耦合或通信断开连接,可以是电性、机械或其它的形式。

[0070] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0071] 集成的单元如果以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或者该技术方案的全部或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)或处理器(processor)执行本申请各个实施方式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-OnlyMemory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0072] 所属领域的技术人员易知,可在保持本申请的教导内容的同时对装置及方法作出诸多修改及变动。因此,以上公开内容应被视为仅受随附权利要求书的范围的限制。

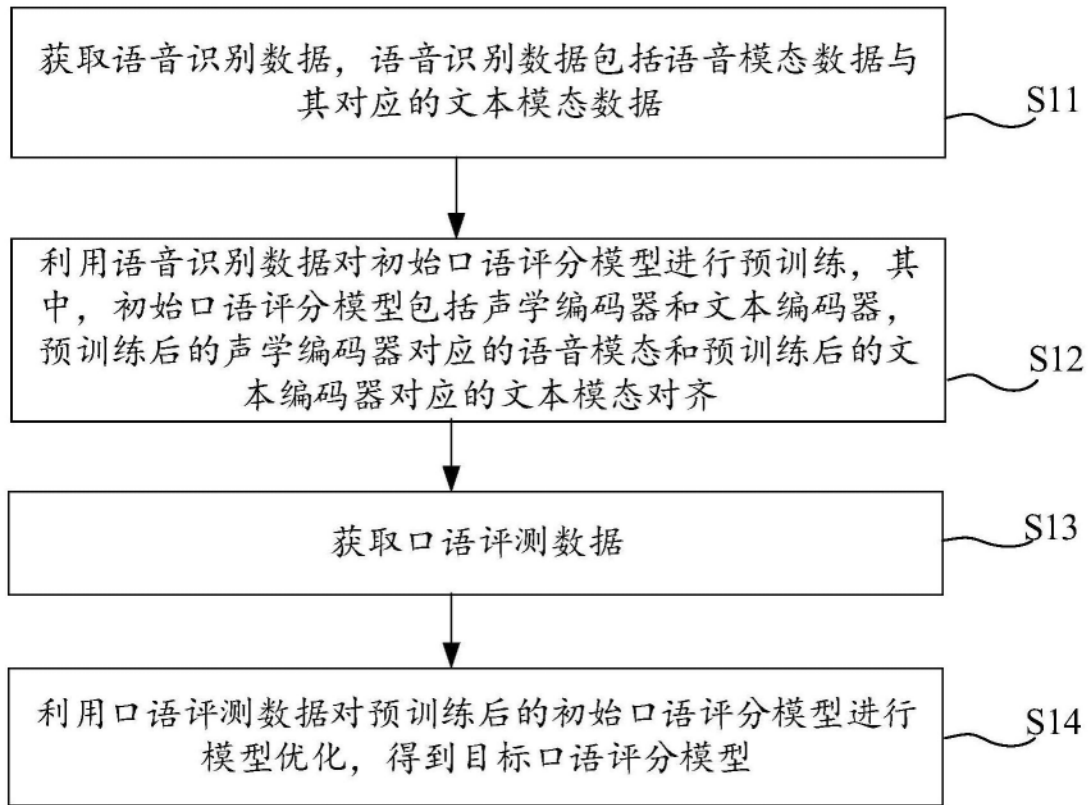


图1

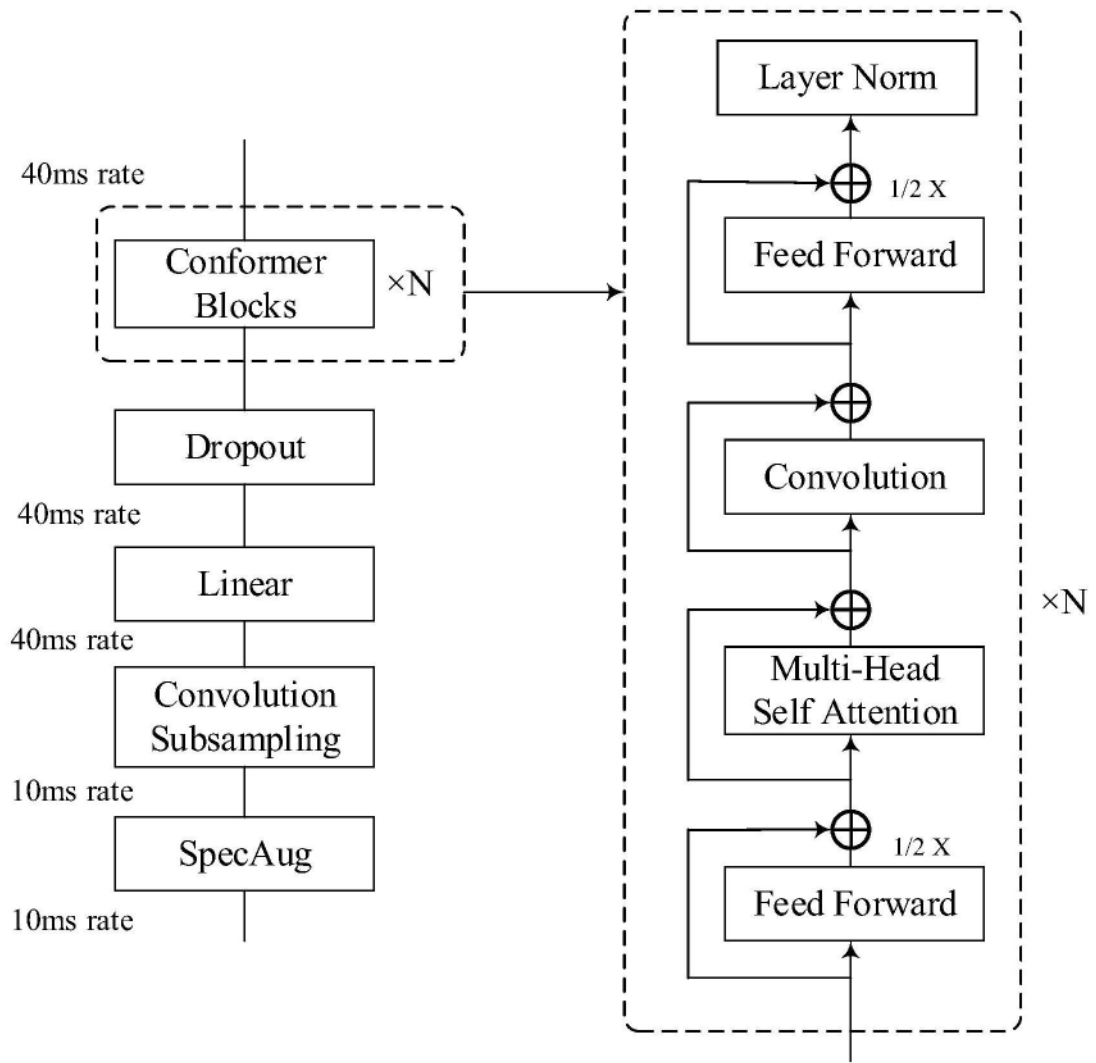


图2

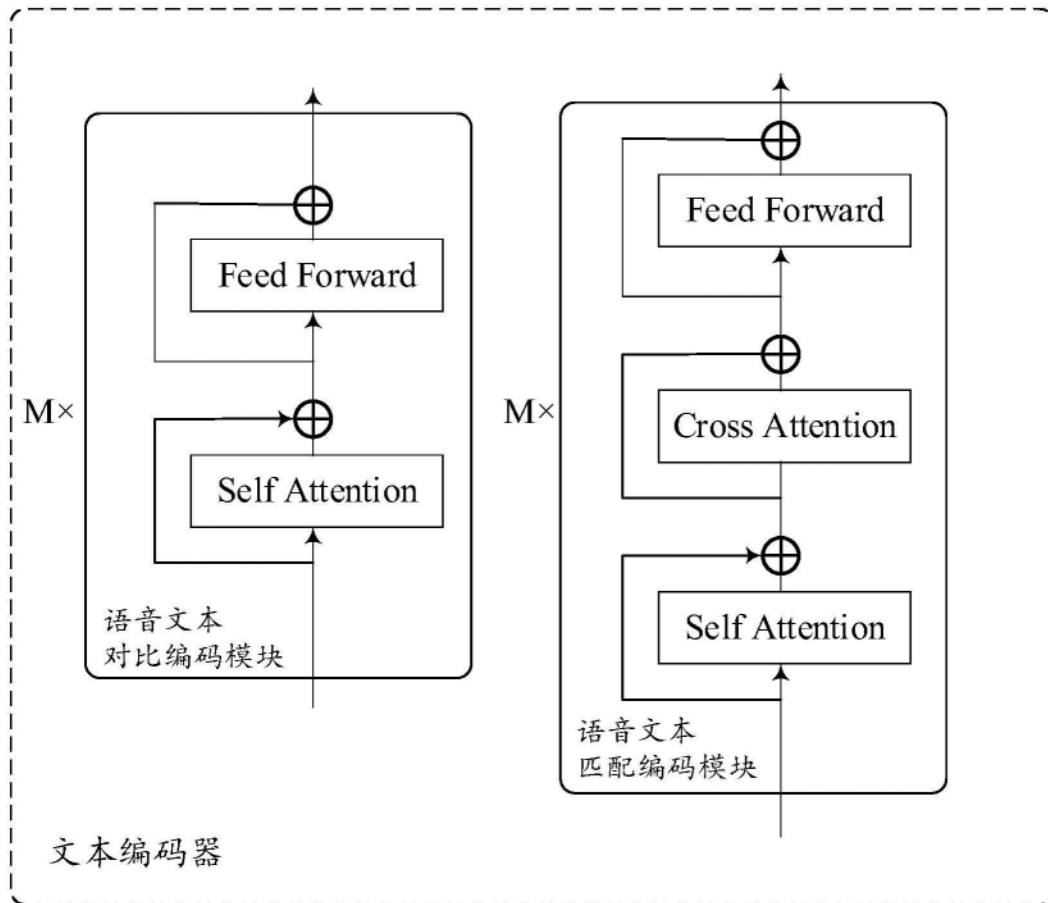


图3

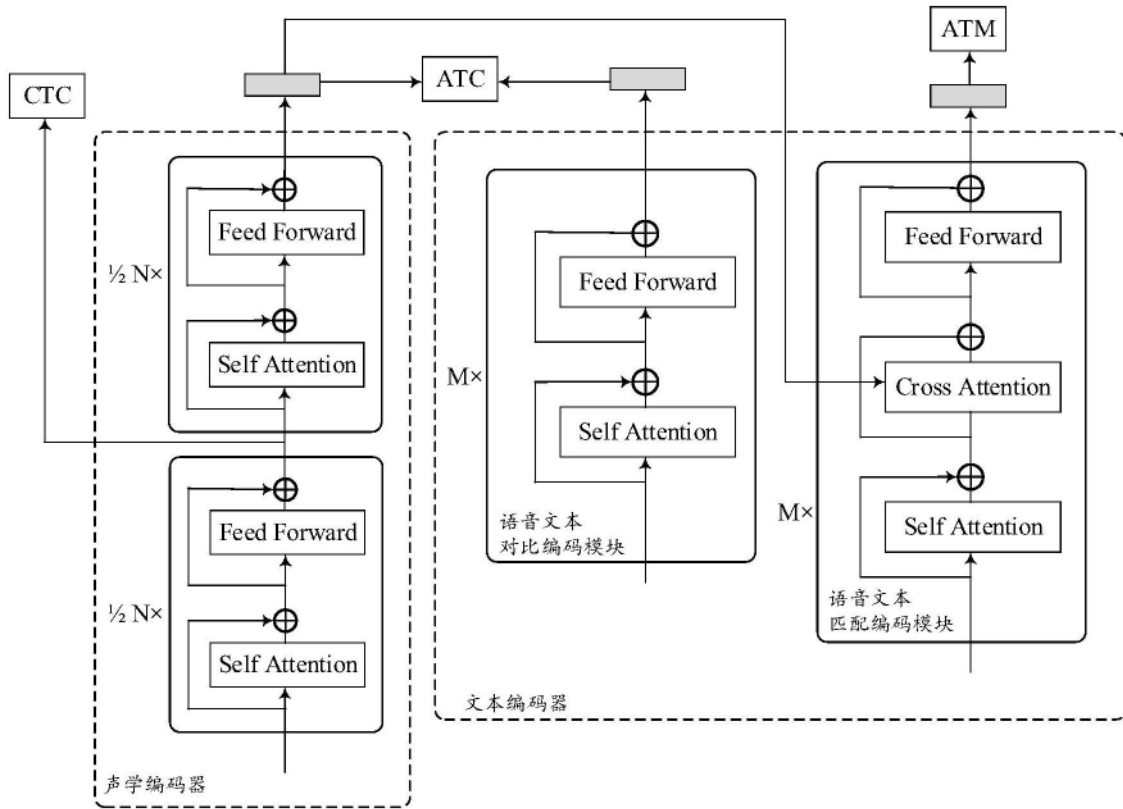


图4

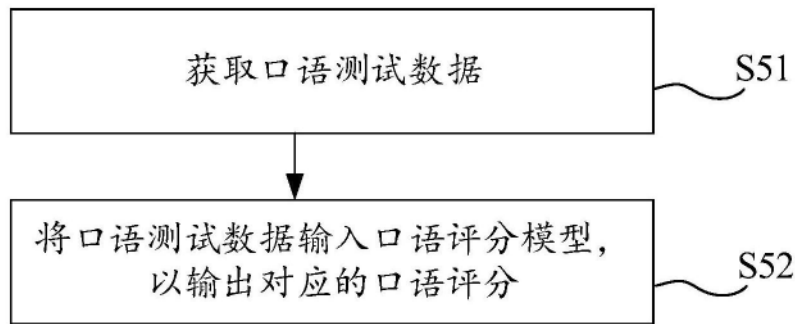


图5

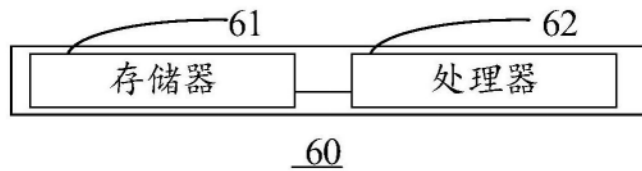


图6

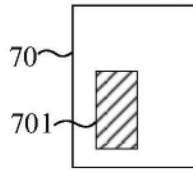


图7