

语音文本生成方法、语音文本生成模型的训练方法、装置

本申请要求于 2022 年 10 月 9 日提交的、申请号为 202211231004.1 的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开涉及人工智能技术领域和智能客服技术领域，更具体地，涉及一种语音文本生成方法、语音文本生成模型的训练方法、装置、设备、介质及程序产品。

背景技术

随着人工智能技术的发展，智能对话系统的应用场景越来越广泛，智能对话系统通过自动生成与业务相关的智能语音信息，或者根据用户的语音信息生成智能回复信息，从而实现利用智能对话系统自动与用户进行语音交互，以满足用户的相关需求。

在实现本公开构思的过程中，发明人发现相关技术中至少存在如下问题，智能对话系统通常采用具有标准书面句式的文本转换为智能语音信息，生成的智能语音信息较为生硬呆板，与人类正常对话中产生的语音信息差别较大。

发明内容

有鉴于此，本公开提供了一种语音文本生成方法、语音文本生成模型的训练方法、装置、设备、介质及程序产品。

本公开的一个方面提供了一种语音文本生成方法，包括：

对标准文本进行词性标注，得到词性标注结果；

根据语气词分布特征从上述词性标注结果中确定目标词性；

根据与上述目标词性对应的内容在上述标准文本中的位置确定预测插入位置；

根据上述预测插入位置在上述标准文本中插入目标语气词，得到目标口语文本；以

及

根据上述目标口语文本生成目标语音文本。

根据本公开的实施例，根据上述预测插入位置在上述标准文本中插入目标语气词，得到目标口语文本包括：

对上述标准文本的预测插入位置进行掩码，得到掩码标准文本；

将上述掩码标准文本输入至语音文本生成模型，以便上述语音文本生成模型在上述预测插入位置中的目标插入位置插入目标语气词，生成上述目标口语文本。

根据本公开的实施例，上述语音文本生成方法还包括：

获取初始语料集，其中，上述初始语料集包括根据口语语音语料生成的初始口语语料文本；

对上述初始口语语料文本进行词性标注，得到口语语料词性标注结果、口语语料语气词标注结果；

根据上述口语语料词性标注结果和上述口语语料语气词标注结果，确定上述语气词分布特征。

根据本公开的实施例，对标准文本进行词性标注，得到词性标注结果包括：

将上述标准文本输入至语义识别模型，得到上述词性标注结果；

其中，上述语义识别模型包括：

基于循环神经网络模型与条件随机场模型构建的第一语义识别模型；或者

基于依存句法分析构建的第二语义识别模型。

本公开的另一方面还提供了一种语音文本生成模型的训练方法，包括：

对训练样本集中的训练样本标准文本和与上述训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到上述训练样本标准文本的第一样本词性标注结果、上述训练样本口语文本的第二样本词性标注结果、上述训练样本口语文本的样本语气词标注结果；

根据样本语气词分布特征从上述第一样本词性标注结果中确定样本目标词性；

根据与上述样本目标词性对应的样本内容在上述训练样本标准文本中的位置确定样本预测插入位置；

对上述训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，上述训练样本掩码标准文本具有第一样本词性标注结果；

利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本生成模型，其中，上述目标训练集包括上述训练样本掩码标准文本、上述训练样本口语文本的第二样本词性标注结果、上述训练样本口语文本的样本语气词标注结果。

根据本公开的实施例，上述语音文本生成模型的训练方法还包括：

利用样本混淆词典中的样本混淆词分别更新第一样本集中的第一样本标准文本和与上述第一样本标准文本关联的第一样本口语文本，得到包含有第二样本标准文本和

二样本口语文本的第二样本集；

根据上述第一样本集与上述第二样本集构建上述训练样本集。

根据本公开的实施例，上述语音文本生成模型的训练方法还包括：

利用语音合成装置处理样本标准语料文本，得到样本语音语料；

对上述样本语音语料进行语音识别，得到样本混淆语料文本；

根据上述样本标准语料文本和上述样本混淆语料文本，构建上述样本混淆词典。

根据本公开的实施例，上述语音文本生成模型的训练方法还包括：

获取样本初始语料集，其中，上述样本初始语料集包括根据样本口语语音语料生成的样本初始口语语料文本；

对上述样本初始口语语料文本进行词性标注，得到样本口语语料词性标注结果、样本口语语料语气词标注结果；

根据上述样本口语语料词性标注结果和上述样本口语语料语气词标注结果，确定上述样本语气词分布特征。

本公开的另一方面还提供了一种语音文本生成装置，包括：

标注模块，用于对标准文本进行词性标注，得到词性标注结果；

第一确定模块，用于根据语气词分布特征从上述词性标注结果中确定目标词性；

第二确定模块，用于根据与上述目标词性对应的内容在上述标准文本中的位置确定预测插入位置；

插入模块，用于根据上述预测插入位置在上述标准文本中插入目标语气词，得到目标口语文本；以及

生成模块，用于根据上述目标口语文本生成目标语音文本。

本公开的另一方面还提供了一种语音文本生成模型的训练装置，包括：

样本标注模块，用于对训练样本集中的训练样本标准文本和与上述训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到上述训练样本标准文本的第一样本词性标注结果、上述训练样本口语文本的第二样本词性标注结果、上述训练样本口语文本的样本语气词标注结果；

样本第一确定模块，用于根据样本语气词分布特征从上述第一样本词性标注结果中确定样本目标词性；

样本第二确定模块，用于根据与上述样本目标词性对应的样本内容在上述训练样本标准文本中的位置确定样本预测插入位置；

样本掩码模块，用于对上述训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，上述训练样本掩码标准文本具有第一样本词性标注结果；

训练模块，用于利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本生成模型，其中，上述目标训练集包括上述训练样本掩码标准文本、上述训练样本口语文本的第二样本词性标注结果、上述训练样本口语文本的样本语气词标注结果。

本公开的另一个方面提供了一种电子设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序，

其中，当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器实现如上所述的方法。

本公开的另一方面提供了一种计算机可读存储介质，存储有计算机可执行指令，所述指令在被执行时用于实现如上所述的方法。

本公开的另一方面提供了一种计算机程序产品，所述计算机程序产品包括计算机可执行指令，所述指令在被执行时用于实现如上所述的方法。

根据本公开的实施例，在对标准文本进行词性标注，得到词性标注结果后，可以根据词性标注结果获得标准文本中每个标准词的词性，根据于此分布特征从词性标准结果中确定可以插入语气词的预测插入位置，根据该预测插入位置在标准文本中插入目标语气词，可以使得到的目标口语文本具有人类正常对话所具备的口语化特性，从而使根据目标口语文本生成的目标语音文本可以至少部分解决相关智能语音信息中较为生硬呆板，与人类对话差别较大的技术问题，使目标语音文本可以更加接近人类对话的语音信息中的口语化特征，使目标语音文本具有拟人化特点的技术效果，提升用户在进行语音交互过程中的使用体验。

附图说明

通过以下参照附图对本公开实施例的描述，本公开的上述以及其他目的、特征和优点将更为清楚，在附图中：

图1示意性示出了根据本公开实施例的可以应用语音文本生成方法、装置的示例性系统架构；

图2示意性示出了根据本公开实施例的语音文本生成方法的流程图；

图3示意性示出了根据本公开另一实施例的语音文本生成方法的流程图；

图4示意性示出了根据本公开实施例的根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本的流程图；

图5示意性示出了根据本公开实施例的语音文本生成方法的应用场景图；

图6示意性示出了根据本公开实施例的语音文本生成模型的训练方法的流程图；

图7示意性示出了根据本公开的实施例的语音文本生成装置的框图；

图8示意性示出了根据本公开的实施例的语音文本生成装置的框图；以及

图9示意性示出了根据本公开实施例的适于实现语音文本生成方法、语音文本生成模型的训练方法的电子设备的框图。

具体实施方式

以下，将参照附图来描述本公开的实施例。但是应该理解，这些描述只是示例性的，而并非要限制本公开的范围。在下面的详细描述中，为便于解释，阐述了许多具体的细节以提供对本公开实施例的全面理解。然而，明显地，一个或多个实施例在没有这些具体细节的情况下也可以被实施。此外，在以下说明中，省略了对公知结构和技术的描述，以避免不必要地混淆本公开的概念。

在此使用的术语仅仅是为了描述具体实施例，而并非意在限制本公开。在此使用的术语“包括”、“包含”等表明了所述特征、步骤、操作和/或部件的存在，但是并不排除存在或添加一个或多个其他特征、步骤、操作或部件。

在此使用的所有术语（包括技术和科学术语）具有本领域技术人员通常所理解的含义，除非另外定义。应注意，这里使用的术语应解释为具有与本说明书的上下文相一致的含义，而不应以理想化或过于刻板的方式来解释。

在使用类似于“A、B和C等中至少一个”这样的表述的情况下，一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释（例如，“具有A、B和C中至少一个的系统”应包括但不限于单独具有A、单独具有B、单独具有C、具有A和B、具有A和C、具有B和C、和/或具有A、B、C的系统等）。

智能对话系统产生的语音信息通常根据书面化的文本生成语音信息，忽略了真实场景中人与人交谈可能存在的语气词、犹豫词、重述词等。因此，基于语音合成装置或者由人工转译生成的语音信息虽然非常标准，但是十分生硬呆板，很容易让用户察觉到是在与机器对话，从而会降低用户的使用体验。

本公开的实施例提供了一种语音文本生成方法、语音文本生成模型的训练方法、装置、设备、介质及程序产品。该语音文本生成方法包括：对标准文本进行词性标注，得到词性标注结果；根据语气词分布特征从词性标注结果中确定目标词性；根据与目标词性对应的内容在标准文本中的位置确定预测插入位置；根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本；以及根据目标口语文本生成目标语音文本。

根据本公开的实施例，在对标准文本进行词性标注，得到词性标注结果后，可以根据词性标注结果获得标准文本中每个标准词的词性，根据于此分布特征从词性标准结果中确定可以插入语气词的预测插入位置，根据该预测插入位置在标准文本中插入目标语气词，可以使得到的目标口语文本具有人类正常对话所具备的口语化特性，从而使根据目标口语文本生成的目标语音文本可以至少部分解决相关智能语音信息中较为生硬呆板，与人类对话差别较大的技术问题，使目标语音文本可以更加接近人类对话的语音信息中的口语化特征，使目标语音文本具有拟人化特点，提升用户在进行语音交互过程中的使用体验。

在本公开的技术方案中，在获取或采集用户个人信息之前，均获取了用户的授权或同意。

在本公开的技术方案中，所涉及的用户个人信息的收集、存储、使用、加工、传输、提供、公开和应用等处理，均符合相关法律法规的规定，采取了必要保密措施，且不违背公序良俗。

图1示意性示出了根据本公开实施例的可以应用语音文本生成方法、装置的示例性系统架构。需要注意的是，图1所示仅为可以应用本公开实施例的系统架构的示例，以帮助本领域技术人员理解本公开的技术内容，但并不意味着本公开实施例不可以用于其他设备、系统、环境或场景。

如图1所示，根据该实施例的系统架构100可以包括终端设备101、102、103、网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型，例如有线和/或无线通信链路等等。

用户可以使用终端设备101、102、103通过网络104与服务器105交互，以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用，例如购物类应用、网页浏览器应用、搜索类应用、即时通信工具、邮箱客户端和/或社交平台软件等（仅为示例）。

终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备，包

括但不限于智能手机、平板电脑、膝上型便携计算机和台式计算机等等。

服务器 105 可以是提供各种服务的服务器，例如对用户利用终端设备 101、102、103 所浏览的网站提供支持的后台管理服务器（仅为示例）。后台管理服务器可以对接收到的用户请求等数据进行分析等处理，并将处理结果（例如根据用户请求获取或生成的网页、信息、或数据等）反馈给终端设备。

需要说明的是，本公开实施例所提供的语音文本生成方法一般可以由服务器 105 执行。相应地，本公开实施例所提供的语音文本生成装置一般可以设置于服务器 105 中。本公开实施例所提供的语音文本生成方法也可以由不同于服务器 105 且能够与终端设备 101、102、103 和/或服务器 105 通信的服务器或服务器集群执行。相应地，本公开实施例所提供的语音文本生成装置也可以设置于不同于服务器 105 且能够与终端设备 101、102、103 和/或服务器 105 通信的服务器或服务器集群中。或者，本公开实施例所提供的语音文本生成方法也可以由终端设备 101、102、或 103 执行，或者也可以由不同于终端设备 101、102、或 103 的其他终端设备执行。相应地，本公开实施例所提供的语音文本生成装置也可以设置于终端设备 101、102、或 103 中，或设置于不同于终端设备 101、102、或 103 的其他终端设备中。

例如，标准文本可以原本存储在终端设备 101、102、或 103 中的任意一个（例如，终端设备 101，但不限于此）之中，或者存储在外部存储设备上并可以导入到终端设备 101 中。然后，终端设备 101 可以在本地执行本公开实施例所提供的语音文本生成方法，或者将标准文本发送到其他终端设备、服务器、或服务器集群，并由接收该标准文本的其他终端设备、服务器、或服务器集群来执行本公开实施例所提供的语音文本生成方法。

应该理解，图 1 中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、网络和服务器的。

图 2 示意性示出了根据本公开实施例的语音文本生成方法的流程图。

如图 2 所示，该方法包括操作 S210~S250。

在操作 S210，对标准文本进行词性标注，得到词性标注结果。

根据本公开的实施例，标准文本可以包括用于书面沟通交流的文本，例如标准客服人员应答文本、邮件正文文本等。

需要说明的是，标准客服人员应答文本可以应用于智能问客服答装置中，该智能问客服答装置根据标准化的标准客服人员应答文本，生成对应的语音文本，从而实现与用户进行语音交互，但生成的语音文本通常较为生硬呆板，与人类对话语音差别过大，不

具备口语化特点。

根据本公开的实施例，对标注文本进行词性标注，得到的词性标注结果可以包括标准文本中进行分词后产生的标准词和标准词的词性特征，词性特征例如可以包括形容词词性、动词词性等。

需要说明的是，本公开的实施例针对词性标注的具体方法不做限定，例如可以采用基于神经网络构建的网络模型对标准文本进行词性标注，但不仅限于此，还可以采用基于统计算法构建的语义识别模型对标准文本进行词性标注，本公开的实施例对词性标注的具体技术手段不做限定，本领域技术人员可以根据实际情况进行选择。

在操作 S220，根据语气词分布特征从词性标注结果中确定目标词性。

根据本公开的实施例，语气词可以包括表示犹豫、疑惑等人类情感的词汇，例如“嗯”、“就是”、“比如”“类似的”等。但不仅限于此，还可以包括在人类语音对话的开始和/或结束会进行表达的礼貌用于类词，例如“如果有空的话”、“如果你喜欢”等。

根据本公开的实施例，可以根据大量的对话文本，统计对话文本中和语气词邻接的词的词性，确定语气词分布特征，从而可以根据语气词分别特征来预测标准文本中的目标词性。

在本公开的实施例中，目标词性可以包括动词词性和形容词词性。

在操作 S230，根据与目标词性对应的内容在标准文本中的位置确定预测插入位置。

根据本公开的实施例，在确定目标词性后，可以根据标准文本中的标准词的词性特征，确定预测插入位置，该预测插入位置可以是具有目标词性的标准词的邻接位置。

在操作 S240，根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本。

根据本公开的实施例，可以在预测插入位置有多个的情况下，可以在每个预测插入位置均插入相应地目标语气词，或者还可以在预测插入位置中的目标插入位置插入目标语气词。在标准文本中插入目标语气词后，可以使目标口语文本在不改变标准文本语义信息的情况下，增强口语化特性。

在操作 S250，根据目标口语文本生成目标语音文本。

根据本公开的实施例，目标语音文本可以是语音信息，可以利用相关语音合成设备将目标口语文本转化为目标语音文本，根据目标口语文本生成的目标语音文本。

根据本公开的实施例，在对标准文本进行词性标注，得到词性标注结果后，可以根据词性标注结果获得标准文本中每个标准词的词性，根据于此分布特征从词性标准结果中确定可以插入语气词的预测插入位置，根据该预测插入位置在标准文本中插入目标语

气词，可以使得到的目标口语文本具有人类正常对话所具备的口语化特性，从而使根据目标口语文本生成的目标语音文本可以至少部分解决相关智能语音信息中较为生硬呆板，与人类对话差别较大的技术问题，使目标语音文本可以更加接近人类对话的语音信息中的口语化特性，使目标语音文本具有拟人化特点，提升用户在进行语音交互过程中的使用体验。

图3示意性示出了根据本公开另一实施例的语音文本生成方法的流程图。

如图3所示，语音文本生成方法还可以包括操作S310~操作S330。

在操作S310，获取初始语料集，其中，初始语料集包括根据口语语音语料生成的初始口语语料文本。

在操作S320，对初始口语语料文本进行词性标注，得到口语语料词性标注结果、口语语料语气词标注结果。

在操作S330，根据口语语料词性标注结果和口语语料语气词标注结果，确定语气词分布特征。

根据本公开的实施例，初始口语语料文本例如可以包括根据人类真实对话场景中对话语音信息生成的语料文本，该语料文本记录有对话语音信息的文本，即初始口语语料文本中包含有人类习惯添加的口语语料语气词。通过对初始口语语料文本进行词性标注，可以得到初始口语语料文本中，口语语料词的词性，还可以得到口语语料语气词与各个口语语料词之间的位置关系，通过分析统计该位置关系，可以确定口语语料语气词的分布特征，即可以确定在口语语音信息中的语气词分布特征。

例如，语气词分布特征可以表征在具有动词词性的口语语料词之后位置的统计概率为0.9，在具有形容词词性的口语语料词之前位置的统计概率为0.8，通过统计该些位置的统计概率，确定语气词分布特征。

根据本公开的实施例，可以设定位置概率阈值，在并将统计概率阈值大于或等于位置概率阈值的位置作为语气词分布特征中的分布统计概率。

需要说明的是，本公开实施例中的“之前”表示与该词邻接，且位于该词位置之前，相应地，“之后”表示与该词邻接，且位于该词位置之后。

根据本公开的实施例，操作S210，对标准文本进行词性标注，得到词性标注结果可以包括如下操作。

将标准文本输入至语义识别模型，得到词性标注结果；其中，语义识别模型包括：基于循环神经网络模型与条件随机场模型构建的第一语义识别模型；或者基于依存句法

分析构建的第二语义识别模型。

根据本公开的实施例，第一语义识别模型例如可以是根据循环神经网络模型（RNN 模型）与条件随机场模型（CRF 模型）依序连接构建得到的，或者还可以是基于双向长短期记忆网络模型（Bi-LSTM 模型）和与条件随机场模型（CRF 模型）依序连接构建得到的。

应该理解的是，循环神经网络模型可以包括长短期记忆网络模型（LSTM 模型），（双向长短期记忆网络模型（Bi-LSTM 模型），本领域技术人员可以根据实际需求对第一语义识别模型的具体网络结构进行设计。

根据本公开的实施例，基于依存句法分析构建的第二语义识别模型例如可以包括 LTP（Language Technology Platform）语言处理系统等。

图 4 示意性示出了根据本公开实施例的根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本的流程图。

如图 4 所示，操作 S240，根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本包括操作 S410~S420。

在操作 S410，对标准文本的预测插入位置进行掩码，得到掩码标准文本。

在操作 S420，将掩码标准文本输入至语音文本生成模型，以便语音文本生成模型在预测插入位置中的目标插入位置插入目标语气词，生成目标口语文本。

根据本公开的实施例，语音文本生成模型可以是基于 BERT 模型构建得到的，例如可以包括 BERT-WWM 模型，掩码标准文本输入至 BERT-WWM 模型可以对预测插入位置的掩码进一步迭代预测，进而从预测插入位置中确定目标预测位置，并基于 BERT-WWM 模型的预测能力从语气词集中确定各个目标预测位置的目标语气词，实现生产目标口语文本。

需要说明的是，语音文本生成模型可以是经过相关训练方法训练后得到的，在语音文本生成模型是 BERT-WWM 模型的情况下，还可以基于 BERT-WWM 模型的预测能力，对掩码标准文本中的至少部分标准词替换为同义词、同音词，从而进一步提升目标口语文本的口语化特性。

图 5 示意性示出了根据本公开实施例的语音文本生成方法的应用场景图。

如图 5 所示，该应用场景中可以包括标准文本 510“请问您需要大尺寸的显示器么”，将标准文本 510 输入至语义识别模型 520，可以实现对标准文本进行词性标注，得到词性标注结果 530。

在本公开的实施例中，语义识别模型 520 可以是基于双向长短期记忆网络模型（Bi-LSTM 模型）和与条件随机场模型（CRF 模型）依序连接构建得到的。

词性标注结果 530 可以包括标准文本 510 中的标准词“请问”、“您”、“需要”、“大尺寸的”、“显示器”、“么”。还可以包括每个标准词各自的词性，其中“v”表示动词词性，“r”表示代词词性，“a”表示形容词词性，“n”表示名词词性，“c”表示语气词词性。

根据语气词分布特征，可以从所述词性标注结果 530 中确定目标词性为动词词性和形容词词性，并根据目标词性对应的内容，在标准文本 510 中的动词词性标准词“请问”之后，动词词性标准词“需要”之后，以及形容词词性标准词“大尺寸的”之前，确定预测插入位置，并对每个预测插入位置进行掩码，得到掩码标准文本 540。掩码标准文本 540 中，可以包括每个预测插入位置对应的掩码单元 541、542。

将掩码标准文本 540 输入至语音文本生成模型 550，语音文本生成模型可以将预测插入位置确定为目标预测插入位置，并将目标语气词“嗯”插入至目标预测插入位置对应的掩码单元 541，将目标语气词“这个”插入至目标预测插入位置对应的掩码单元 542，进而生成目标口语文本 560“请问嗯您需要这个大尺寸的显示器么”。从而可以使目标口语文本 560 具有接近人类口语化语音信息的口语化特性，根据目标口语文本 560 生成的目标语音文本可以具备口语化特性，至少部分避免生成的语音信息生硬呆板，减少与人类对话语音信息的差别。

图 6 示意性示出了根据本公开实施例的语音文本生成模型的训练方法的流程图。

如图 6 所示，该方法包括操作 S610~S650。

在操作 S610，对训练样本集中的训练样本标准文本和与训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到训练样本标准文本的第一样本词性标注结果、训练样本口语文本的第二样本词性标注结果、训练样本口语文本的样本语气词标注结果。

在操作 S620，根据样本语气词分布特征从第一样本词性标注结果中确定样本目标词性。

在操作 S630，根据与样本目标词性对应的样本内容在训练样本标准文本中的位置确定样本预测插入位置。

在操作 S640，对训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，训练样本掩码标准文本具有第一样本词性标注结果。

在操作 S650，利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本

生成模型，其中，目标训练集包括训练样本掩码标准文本、训练样本口语文本的第二样本词性标注结果、训练样本口语文本的样本语气词标注结果。

根据本公开的实施例，训练样本标准文本可以包括标准的书面化文本，训练样本口语文本可以包括样本用户对训练样本标准文本发音转述后生成的语音信息转化为的口语化文本，训练样本口语文本由于经过样本用户的语音转述后，在根据转述的语音生成，因此训练样本口语文本可以包含有样本语气词。至少部分克服了相关技术中采用标准文本训练语音文本生成模型，从而使训练得到的语音文本生成模型不能学习到口语对话表达中可能存在的语气词、犹豫词等语气词的特性。

根据本公开的实施例，可以将训练样本掩码标准文本和训练样本口语文本组成训练样本对，并根据训练样本对中训练样本掩码标准文本和训练样本口语文本的相似度确定相似度标签值，该相似度标签值可以用于迭代地调整初始语音文本生成模型中的权重参数，使生成的语音文本生成模型，可以预测样本语气词与训练样本标准文本中第一样本词性标注结果之间的位置关系，从而可以准确地根据预测插入位置确定目标插入位置，并从样本语气词中确定目标样本语气词。

需要说明的是，根据本公开实施例提供的语音文本生成模型的训练方法训练得到的语音文本生成模型，可以用于上述语音文本生成方法。

根据本公开的实施例，语音文本生成模型的训练方法还包括如下操作。

利用样本混淆词典中的样本混淆词分别更新第一样本集中的第一样本标准文本和与第一样本标准文本关联的第一样本口语文本，得到包含有第二样本标准文本和第二样本口语文本的第二样本集；根据第一样本集与第二样本集构建训练样本集。

根据本公开的实施例，样本混淆词典可以包括样本标准词和样本混淆词构成的样本混淆词对，通过样本混淆词典中的样本混淆词替换第一样本标准文本中的样本标准词，以及通过样本混淆词典中的样本混淆词替换第一样本口语文本中的样本标准词，可以分别得到大量的第二样本标准文本和第二样本口语文本，从而根据第一样本集与第二样本集构建得到训练样本集，可以扩充训练样本数据的数量，以增强训练样本集的训练能力。进一步地，利用包含有第一样本集与第二样本集的训练样本集训练初始语音文本生成模型，可以使初始语音文本生成模型充分学习到标准词与混淆词之间的相似关联关系，从而使训练后得到的语音文本生成模型自动将标准文本中的标准词替换为样本混淆词，从而进一步丰富目标口语文本的语义表达方式，使目标口语文本更贴近人类正常对话的口语化特性。

根据本公开的实施例，语音文本生成模型的训练方法还可以包括如下操作。

利用语音合成装置处理样本标准语料文本，得到样本语音语料；对样本语音语料进行语音识别，得到样本混淆语料文本；根据样本标准语料文本和样本混淆语料文本，构建样本混淆词典。

根据本公开的实施例，样本标准语料文本可以包括用于书面沟通交流的文本，例如标准客服人员应答文本、邮件正文文本等。样本语音语料可以包括语音合成装置自动识别样本标准语料文本后，生成的语音信息。利用语音识别（Automatic Speech Recognition，ASR）装置识别样本语音语料，可以得到识别后的样本混淆语料文本，由于语音识别装置的识别能力限制，因此可以将样本标准语料文本中的至少部分样本标准词识别为样本混淆词，从而使样本混淆语料文本包含有语音识别装置识别到的样本混淆词。根据样本标准语料文本和样本混淆语料文本的比对结果，可以将样本标准词和样本混淆词做成样本混淆词对，进而构建得到样本混淆词典。

根据本公开的实施例，可以在利用语音识别装置对样本语音语料进行语音识别后，通过确定语音识别装置输出的初始样本混淆语料文本的置信度信息，将置信度信息小于或等于预设置信度阈值的初始样本混淆语料文本确定为样本混淆语料文本，从而可以根据样本混淆语料文本中选择出容易被识别错误的样本混淆词，使构建得到的样本混淆词典的样本混淆词对更加准确地体现样本混淆词与样本标准词的关联特征。

根据本公开的实施例，语音文本生成模型的训练方法还可以包括如下操作。

获取样本初始语料集，其中，所述样本初始语料集包括根据样本口语语音语料生成的样本初始口语语料文本；对所述样本初始口语语料文本进行词性标注，得到样本口语语料词性标注结果、样本口语语料语气词标注结果；根据所述样本口语语料词性标注结果和所述样本口语语料语气词标注结果，确定所述样本语气词分布特征。

根据本公开的实施例，样本初始口语语料文本例如可以包括根据人类真实对话场景中对话语音信息生成的语料文本，该语料文本记录有对话语音信息的文本，即样本初始口语语料文本中包含有人类习惯添加的口语语料语气词。通过对样本初始口语语料文本进行词性标注，可以得到样本初始口语语料文本中，样本口语语料词的词性，还可以得到样本口语语料语气词与各个样本口语语料词之间的位置关系，通过分析统计该位置关系，可以确定样本口语语料语气词的分布特征，即可以确定在样本口语语音信息中的样本语气词分布特征。

需要说明的是，根据本公开的实施例提供的语音文本生成模型的训练方法训练得到

的语音文本生成模型，可以用于上述语音文本生成方法。

图7示意性示出了根据本公开的实施例的语音文本生成装置的框图。

如图7所示，语音文本生成装置700可以包括标注模块710、第一确定模块720、第二确定模块730、插入模块740和生成模块750。

标注模块710用于对标准文本进行词性标注，得到词性标注结果。

第一确定模块720用于根据语气词分布特征从词性标注结果中确定目标词性。

第二确定模块730用于根据与目标词性对应的内容在标准文本中的位置确定预测插入位置。

插入模块740用于根据预测插入位置在标准文本中插入目标语气词，得到目标口语文本。

生成模块750用于根据目标口语文本生成目标语音文本。

根据本公开的实施例，插入模块可以包括：掩码单元和生成单元。

掩码单元用于对标准文本的预测插入位置进行掩码，得到掩码标准文本。

生成单元用于将掩码标准文本输入至语音文本生成模型，以便语音文本生成模型在预测插入位置中的目标插入位置插入目标语气词，生成目标口语文本。

根据本公开的实施例，语音文本生成方法还可以包括：获取模块、语料标注模块和第三确定模块。

获取模块用于获取初始语料集，其中，初始语料集包括根据口语语音语料生成的初始口语语料文本。

语料标注模块用于对初始口语语料文本进行词性标注，得到口语语料词性标注结果、口语语料语气词标注结果。

第三确定模块用于根据口语语料词性标注结果和口语语料语气词标注结果，确定语气词分布特征。

根据本公开的实施例，标注模块可以包括标注单元。

标注单元用于将标准文本输入至语义识别模型，得到词性标注结果。

其中，语义识别模型包括：

基于循环神经网络模型与条件随机场模型构建的第一语义识别模型；或者基于依存句法分析构建的第二语义识别模型。

需要说明的是，本公开的实施例中语音文本生成装置部分与本公开的实施例中语音文本生成方法部分是相对应的，语音文本生成装置部分的描述具体参考语音文本生成方

法部分，在此不再赘述。

图8示意性示出了根据本公开的实施例的语音文本生成装置的框图。

如图8所示，语音文本生成装置800可以包括样本标注模块810、样本第一确定模块820、样本第二确定模块830、样本掩码模块840和训练模块850。

样本标注模块810用于对训练样本集中的训练样本标准文本和与训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到训练样本标准文本的第一样本词性标注结果、训练样本口语文本的第二样本词性标注结果、训练样本口语文本的样本语气词标注结果。

样本第一确定模块820用于根据样本语气词分布特征从第一样本词性标注结果中确定样本目标词性。

样本第二确定模块830用于根据与样本目标词性对应的样本内容在训练样本标准文本中的位置确定样本预测插入位置。

样本掩码模块840用于对训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，训练样本掩码标准文本具有第一样本词性标注结果。

训练模块850用于利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本生成模型，其中，目标训练集包括训练样本掩码标准文本、训练样本口语文本的第二样本词性标注结果、训练样本口语文本的样本语气词标注结果。

根据本公开的实施例，语音文本生成模型的训练装置还可以包括：样本更新模块和样本构建模块。

样本更新模块用于利用样本混淆词典中的样本混淆词分别更新第一样本集中的第一样本标准文本和与第一样本标准文本关联的第一样本口语文本，得到包含有第二样本标准文本和第二样本口语文本的第二样本集。

样本构建模块用于根据第一样本集与第二样本集构建训练样本集。

根据本公开的实施例，语音文本生成模型的训练装置还可以包括：语料处理模块、识别模块和混淆词典构建模块。

语料处理模块用于利用语音合成装置处理样本标准语料文本，得到样本语音语料。

识别模块用于对样本语音语料进行语音识别，得到样本混淆语料文本。

混淆词典构建模块用于根据样本标准语料文本和样本混淆语料文本，构建样本混淆词典。

根据本公开的实施例，语音文本生成模型的训练装置还可以包括：样本初始语料获

取模块、样本语料标注模块和样本第三确定模块。

样本初始语料获取模块用于获取样本初始语料集，其中，样本初始语料集包括根据样本口语语音语料生成的样本初始口语语料文本；

样本语料标注模块用于对样本初始口语语料文本进行词性标注，得到样本口语语料词性标注结果、样本口语语料语气词标注结果。

样本第三确定模块用于根据样本口语语料词性标注结果和样本口语语料语气词标注结果，确定样本语气词分布特征。

需要说明的是，本公开的实施例中语音文本生成模型的训练装置部分与本公开的实施例中语音文本生成模型的训练方法部分是相对应的，语音文本生成模型的训练装置部分的描述具体参考语音文本生成模型的训练方法部分，在此不再赘述。

根据本公开的实施例的模块、单元中的任意多个、或其中任意多个的至少部分功能可以在一个模块中实现。根据本公开实施例的模块、子模块、单元、子单元中的任意一个或多个可以被拆分成多个模块来实现。根据本公开实施例的模块、单元中的任意一个或多个可以至少被部分地实现为硬件电路，例如现场可编程门阵列（FPGA）、可编程逻辑阵列（PLA）、片上系统、基板上的系统、封装上的系统、专用集成电路（ASIC），或可以通过对电路进行集成或封装的任何其他的合理方式的硬件或固件来实现，或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者，根据本公开实施例的模块、单元中的一个或多个可以至少被部分地实现为计算机程序模块，当该计算机程序模块被运行时，可以执行相应的功能。

例如，标注模块 710、第一确定模块 720、第二确定模块 730、插入模块 740 和生成模块 750 中的任意多个可以合并在一个模块/单元中实现，或者其中的任意一个模块/单元可以被拆分成多个模块/单元。或者，这些模块/单元中的一个或多个模块/单元的至少部分功能可以与其他模块/单元/子单元的至少部分功能相结合，并在一个模块/单元中实现。根据本公开的实施例，标注模块 710、第一确定模块 720、第二确定模块 730、插入模块 740 和生成模块 750 中的至少一个可以至少被部分地实现为硬件电路，例如现场可编程门阵列（FPGA）、可编程逻辑阵列（PLA）、片上系统、基板上的系统、封装上的系统、专用集成电路（ASIC），或可以通过对电路进行集成或封装的任何其他的合理方式等硬件或固件来实现，或以软件、硬件以及固件三种实现方式中任意一种或以其中任意几种的适当组合来实现。或者，标注模块 710、第一确定模块 720、第二确定模块 730、插入模块 740 和生成模块 750 中的至少一个可以至少被部分地实现为计算机程

序模块, 当该计算机程序模块被运行时, 可以执行相应的功能。

图 9 示意性示出了根据本公开实施例的适于实现语音文本生成方法、语音文本生成模型的训练方法的电子设备的框图。图 9 示出的电子设备仅仅是一个示例, 不应对本公开实施例的功能和使用范围带来任何限制。

如图 9 所示, 根据本公开实施例的电子设备 900 包括处理器 901, 其可以根据存储在只读存储器 (ROM) 902 中的程序或者从存储部分 908 加载到随机访问存储器 (RAM) 903 中的程序而执行各种适当的动作和处理。处理器 901 例如可以包括通用微处理器 (例如 CPU)、指令集处理器和/或相关芯片组和/或专用微处理器 (例如, 专用集成电路 (ASIC)), 等等。处理器 901 还可以包括用于缓存用途的板载存储器。处理器 901 可以包括用于执行根据本公开实施例的方法流程的不同动作的单一处理单元或者是多个处理单元。

在 RAM 903 中, 存储有电子设备 900 操作所需的各种程序和数据。处理器 901、ROM 902 以及 RAM 903 通过总线 904 彼此相连。处理器 901 通过执行 ROM 902 和/或 RAM 903 中的程序来执行根据本公开实施例的方法流程的各种操作。需要注意, 所述程序也可以存储在除 ROM 902 和 RAM 903 以外的一个或多个存储器中。处理器 901 也可以通过执行存储在所述一个或多个存储器中的程序来执行根据本公开实施例的方法流程的各种操作。

根据本公开的实施例, 电子设备 900 还可以包括输入/输出 (I/O) 接口 905, 输入/输出 (I/O) 接口 905 也连接至总线 904。系统 900 还可以包括连接至 I/O 接口 905 的以下部件中的一项或多项: 包括键盘、鼠标等的输入部分 906; 包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分 907; 包括硬盘等的存储部分 908; 以及包括诸如 LAN 卡、调制解调器等网络接口卡的通信部分 909。通信部分 909 经由诸如因特网的网络执行通信处理。驱动器 910 也根据需要连接至 I/O 接口 905。可拆卸介质 911, 诸如磁盘、光盘、磁光盘、半导体存储器等等, 根据需要安装在驱动器 910 上, 以便于从其上读出的计算机程序根据需要被安装入存储部分 908。

根据本公开的实施例, 根据本公开实施例的方法流程可以被实现为计算机软件程序。例如, 本公开的实施例包括一种计算机程序产品, 其包括承载在计算机可读存储介质上的计算机程序, 该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中, 该计算机程序可以通过通信部分 909 从网络上被下载和安装, 和/或从可拆卸介质 911 被安装。在该计算机程序被处理器 901 执行时, 执行本公开实施例的系统中限

定的上述功能。根据本公开的实施例，上文描述的系统、设备、装置、模块、单元等可以通过计算机程序模块来实现。

本公开还提供了一种计算机可读存储介质，该计算机可读存储介质可以是上述实施例中描述的设备/装置/系统中所包含的；也可以是单独存在，而未装配入该设备/装置/系统中。上述计算机可读存储介质承载有一个或者多个程序，当上述一个或者多个程序被执行时，实现根据本公开实施例的方法。

根据本公开的实施例，计算机可读存储介质可以是非易失性的计算机可读存储介质。例如可以包括但不限于：便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

例如，根据本公开的实施例，计算机可读存储介质可以包括上文描述的 ROM 902 和/或 RAM 903 和/或 ROM 902 和 RAM 903 以外的一个或多个存储器。

本公开的实施例还包括一种计算机程序产品，其包括计算机程序，该计算机程序包含用于执行本公开实施例所提供的方法的程序代码，当计算机程序产品在电子设备上运行时，该程序代码用于使电子设备实现本公开实施例所提供的上述方法。

在该计算机程序被处理器 901 执行时，执行本公开实施例的系统/装置中限定的上述功能。根据本公开的实施例，上文描述的系统、装置、模块、单元等可以通过计算机程序模块来实现。

在一种实施例中，该计算机程序可以依托于光存储器件、磁存储器件等有形存储介质。在另一种实施例中，该计算机程序也可以在网络介质上以信号的形式进行传输、分发，并通过通信部分 909 被下载和安装，和/或从可拆卸介质 911 被安装。该计算机程序包含的程序代码可以用任何适当的网络介质传输，包括但不限于：无线、有线等等，或者上述的任意合适的组合。

根据本公开的实施例，可以以一种或多种程序设计语言的任意组合来编写用于执行本公开实施例提供的计算机程序的程序代码，具体地，可以利用高级过程和/或面向对象的编程语言、和/或汇编/机器语言来实施这些计算程序。程序设计语言包括但不限于诸如 Java, C++, python, “C” 语言或类似的程序设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、部分在远程计算设备上执行、或者完全

在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图或流程图中的每个方框、以及框图或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。本领域技术人员可以理解，本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合，即使这样的组合或结合没有明确记载于本公开中。特别地，在不脱离本公开精神和教导的情况下，本公开的各个实施例和/或权利要求中记载的特征可以进行多种组合和/或结合。所有这些组合和/或结合均落入本公开的范围。

以上对本公开的实施例进行了描述。但是，这些实施例仅仅是为了说明的目的，而并非为了限制本公开的范围。尽管在以上分别描述了各实施例，但是这并不意味着各个实施例中的措施不能有利地结合使用。本公开的范围由所附权利要求及其等同物限定。不脱离本公开的范围，本领域技术人员可以做出多种替代和修改，这些替代和修改都应落在本公开的范围之内。

权 利 要 求

1. 一种语音文本生成方法，包括：

对标准文本进行词性标注，得到词性标注结果；

根据语气词分布特征从所述词性标注结果中确定目标词性；

根据与所述目标词性对应的内容在所述标准文本中的位置确定预测插入位置；

根据所述预测插入位置在所述标准文本中插入目标语气词，得到目标口语文本；

以及

根据所述目标口语文本生成目标语音文本。

2. 根据权利要求1所述的语音文本生成方法，其中，根据所述预测插入位置在所述标准文本中插入目标语气词，得到目标口语文本包括：

对所述标准文本的预测插入位置进行掩码，得到掩码标准文本；

将所述掩码标准文本输入至语音文本生成模型，以便所述语音文本生成模型在所述预测插入位置中的目标插入位置插入目标语气词，生成所述目标口语文本。

3. 根据权利要求1所述的语音文本生成方法，还包括：

获取初始语料集，其中，所述初始语料集包括根据口语语音语料生成的初始口语语料文本；

对所述初始口语语料文本进行词性标注，得到口语语料词性标注结果、口语语料语气词标注结果；

根据所述口语语料词性标注结果和所述口语语料语气词标注结果，确定所述语气词分布特征。

4. 根据权利要求1所述的语音文本生成方法，对标准文本进行词性标注，得到词性标注结果包括：

将所述标准文本输入至语义识别模型，得到所述词性标注结果；

其中，所述语义识别模型包括：

基于循环神经网络模型与条件随机场模型构建的第一语义识别模型；或者

基于依存句法分析构建的第二语义识别模型。

5. 一种语音文本生成模型的训练方法，包括：

对训练样本集中的训练样本标准文本和与所述训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到所述训练样本标准文本的第一样本词性标注结果、所述训练样本口语文本的第二样本词性标注结果、所述训练样本口语文本的样本语气词标注结果；

根据样本语气词分布特征从所述第一样本词性标注结果中确定样本目标词性；

根据与所述样本目标词性对应的样本内容在所述训练样本标准文本中的位置确定样本预测插入位置；

对所述训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，所述训练样本掩码标准文本具有第一样本词性标注结果；

利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本生成模型，其中，所述目标训练集包括所述训练样本掩码标准文本、所述训练样本口语文本的第二样本词性标注结果、所述训练样本口语文本的样本语气词标注结果。

6. 根据权利要求5所述的训练方法，还包括：

利用样本混淆词典中的样本混淆词分别更新第一样本集中的第一样本标准文本和与所述第一样本标准文本关联的第一样本口语文本，得到包含有第二样本标准文本和第二样本口语文本的第二样本集；

根据所述第一样本集与所述第二样本集构建所述训练样本集。

7. 根据权利要求6所述的训练方法，还包括：

利用语音合成装置处理样本标准语料文本，得到样本语音语料；

对所述样本语音语料进行语音识别，得到样本混淆语料文本；

根据所述样本标准语料文本和所述样本混淆语料文本，构建所述样本混淆词典。

8. 根据权利要求5所述的训练方法，还包括：

获取样本初始语料集，其中，所述样本初始语料集包括根据样本口语语音语料生成的样本初始口语语料文本；

对所述样本初始口语语料文本进行词性标注，得到样本口语语料词性标注结果、样本口语语料语气词标注结果；

根据所述样本口语语料词性标注结果和所述样本口语语料语气词标注结果，确定所述样本语气词分布特征。

9. 一种语音文本生成装置，包括：

标注模块，用于对标准文本进行词性标注，得到词性标注结果；

第一确定模块，用于根据语气词分布特征从所述词性标注结果中确定目标词性；

第二确定模块，用于根据与所述目标词性对应的内容在所述标准文本中的位置确定预测插入位置；

插入模块，用于根据所述预测插入位置在所述标准文本中插入目标语气词，得到目标口语文本；以及

生成模块，用于根据所述目标口语文本生成目标语音文本。

10. 一种语音文本生成模型的训练装置，包括：

样本标注模块，用于对训练样本集中的训练样本标准文本和与所述训练样本标准文本关联的训练样本口语文本分别进行词性标注，得到所述训练样本标准文本的第一样本词性标注结果、所述训练样本口语文本的第二样本词性标注结果、所述训练样本口语文本的样本语气词标注结果；

样本第一确定模块，用于根据样本语气词分布特征从所述第一样本词性标注结果中确定样本目标词性；

样本第二确定模块，用于根据与所述样本目标词性对应的样本内容在所述训练样本标准文本中的位置确定样本预测插入位置；

样本掩码模块，用于对所述训练样本标准文本中的样本预测插入位置进行掩码，得到训练样本掩码标准文本，其中，所述训练样本掩码标准文本具有第一样本词性标注结果；

训练模块，用于利用目标训练集训练初始语音文本生成模型，得到训练后的语音文本生成模型，其中，所述目标训练集包括所述训练样本掩码标准文本、所述训练样本口语文本的第二样本词性标注结果、所述训练样本口语文本的样本语气词标注结果。

11. 一种电子设备，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序，

其中，当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器实现权利要求1至8中任一项所述的方法。

12. 一种计算机可读存储介质，其上存储有可执行指令，该指令被处理器执行时使处理器实现权利要求1至8中任一项所述的方法。

13. 一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现根据权利要求1至8中任一项所述的方法。