语言处理模型训练方法、装置、电子设备和可读存储介质

相关申请的交叉引用

本申请要求于 2025 年 02 月 21 日提交中国专利局的申请号为 2025101964074、名称为“语言处理模型训练方法、装置、电子设备和可读存储介质”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

本公开涉及语音合成技术领域，具体而言，涉及一种语言处理模型训练方法、装置、电子设备和计算机可读存储介质。

背景技术

随着信息技术的快速发展，语音合成技术（Text-To-Speech, TTS）作为一种将文本转化为自然流畅语音的技术，已经成为人机交互的重要组成部分，直接影响用户的体验感。TTS 广泛应用于外呼播报、语音导航、有声读物等领域。

近年来，基于神经网络的语音合成大模型训练成为 TTS 领域的研究热点，然而，现有的语音合成大模型在处理风格信息（如韵律、情感等）时表现欠佳。

发明内容

本公开的目的在于提供一种语言处理模型训练方法、装置、电子设备和可读存储介质，以解决现有的语音合成大模型在处理风格信息时表现欠佳的问题。

第一方面，本公开提供一种语言处理模型训练方法，所述方法包括：

获得训练文本数据以及真实语音标记数据；

将所述训练文本数据以及所述真实语音标记数据输入至语言处理模型中的预测模块，获得所述预测模块输出的预测语音标记数据；

根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征；

根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数；

根据所述总损失函数对所述预测模块进行优化训练。

在可选的实施方式中，所述语言处理模型包括风格编码模块，所述根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征，包括：

通过所述风格编码模块对所述真实语音标记数据进行风格提取，获得所述真实风格特征；

通过所述风格编码模块对所述预测语音标记数据进行风格提取，获得所述预测风格特征。

在可选的实施方式中，所述根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数，包括：

根据所述真实语音标记数据以及所述预测语音标记数据，获得预测损失函数；

根据所述真实风格特征以及所述预测风格特征，获得风格损失函数；

根据所述预测损失函数以及所述风格损失函数，获得所述总损失函数。

在可选的实施方式中，所述根据所述真实语音标记数据以及所述预测语音标记数据，获得预测损失函数，包括：

计算所述真实语音标记数据与所述预测语音标记数据之间的分布距离，获得所述预测损失函数。

在可选的实施方式中，所述根据所述真实风格特征以及所述预测风格特征，获得风格损失函数，包括：

计算所述真实风格特征与所述预测风格特征之间的相似度，获得所述风格损失函数。

在可选的实施方式中，所述根据所述预测损失函数以及所述风格损失函数，获得所述总损失函数，包括：

对所述预测损失函数以及所述风格损失函数进行加权计算，获得所述总损失函数。

第二方面，本公开提供一种语言处理模型训练装置，所述装置包括：

预测语音标记数据获得模块，配置成获得训练文本数据以及真实语音标记数据；将所述训练文本数据以及所述真实语音标记数据输入至语言处理模型中的预测模块，获得所述预测模块输出的预测语音标记数据；

风格特征获得模块，配置成根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征；

损失函数获得模块，配置成根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数；

优化训练模块，配置成根据所述总损失函数对所述预测模块进行优化训练。

在可选的实施方式中，所述语言处理模型包括风格编码模块，所述风格特征获得模块还配置成通过所述风格编码模块对所述真实语音标记数据进行风格提取，获得所述真实风格特征；通过所述风格编码模块对所述预测语音标记数据进行风格提取，获得所述预测风格特征。

第三方面，本公开提供一种电子设备，包括处理器和存储器，所述存储器存储有能够被所述处理器执行的计算机程序，所述处理器可执行的计算机程序以实现前述实施方式任一所述的语言处理模型训练方法。

第四方面，本公开提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现如前述实施方式中任一项所述的语言处理模型训练方法。

本公开实施例提供的语言处理模型训练方法、装置、电子设备和可读存储介质，该方法包括：获取训练文本数据和真实语音标记数据并输入至语言处理模型中的预测模块，获得预测语音标记数据，分别提取真实和预测语音标记数据中的风格特征，再根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征构建总损失函数，再根据总损失函数对语言处理模型进行优化训练，将风格特征损失引入总损失函数，确保语言处理模型能够学习风格信息，使得语言处理模型更准确地复现语音风格，从而提高语言处理模型在处理语音风格信息的表现，并且语音处理模型是语音合成模型中的关键部分，进而能够提高语音合成大模型在处理语音风格信息的表现。

附图说明

为了更清楚地说明本公开实施例的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，应当理解，以下附图仅示出了本公开的某些实施例，因此不应被看作是对范围的限定，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相关的附图。

图 1 示出了本公开实施例提供的语言处理模型训练方法的一种流程示意图；

图 2 示出了本公开实施例提供的语言处理模型训练方法的另一种流程示意图；

图 3 示出了本公开实施例提供语音合成大模型的框架示意图；

图 4 示出了本公开实施例提供的语言处理模型训练方法的又一种流程示意图；

图 5 示出了本公开实施例提供的语言处理模型训练装置的模块框图；

图 6 示出了本公开实施例提供的电子设备的方框示意图。

图标：100-电子设备；110-存储器；120-处理器；130-通信模块；200-语言处理模型训练装置；210-预测语音标记数据获得模块；220-风格特征获得模块；230-损失函数获得模块；240-优化训练模块。

具体实施方式

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一部分实施例，而不是全部的实施例。通常在此处附图中描述和示出的本公开实施例的组件可以以各种不同的配置来布置和设计。

因此，以下对在附图中提供的本公开的实施例的详细描述并非旨在限制要求保护的本公开的范围，而是仅仅表示本公开的选定实施例。基于本公开中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

需要说明的是，术语“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、物品或者设备中还存在另外的相同要素。

随着信息技术的快速发展，语音合成技术（Text-To-Speech, TTS）作为一种将文本转化为自然流畅语音的技术，已经成为人机交互的重要组成部分，直接影响用户的体验感。TTS广泛应用于外呼播报、语音导航、有声读物等领域。

现有语音合成大模型主要模型有三部分：语言处理模型（Large Language Model，LLM）、声学特征预测模型、声码器模型。在第一阶段语言处理模型输出的预测语音标记数据中主要包括语义信息以及部分风格信息，现有的语言处理模型的损失函数主要关注的是语义信息，关注风格信息较少，因此语音合成大模型在进行语音合成时，不能较好地对语音风格进行复现，从而导致语音合成大模型在处理风格信息（如韵律、情感等）时表现欠佳。

针对上述问题，发明人提出了本公开实施例提供的语言处理模型训练方法、装置、服务器以及存储介质，通过引入风格特征损失，并根据总损失函数对语言处理模型进行优化训练，从而提高语言处理模型在处理语音风格信息的表现，进而提高语音合成大模型在处理语音风格信息的表现。其中，具体的语言处理模型训练方法在后续的实施例中进行详细的说明。

请参阅图 1，图 1 示出了本公开实施例提供的语言处理模型训练方法的一种流程示意图。该语言处理模型训练方法可以应用于电子设备，下面以电子设备为例，说明本实施例的具体流程。下面针对图 1 所示的流程进行详细的阐述，该语言处理模型训练方法具体可以包括以下步骤：

步骤 110：获得训练文本数据以及真实语音标记数据。

其中，训练文本数据是指经过整理和标注的文本句子，将训练文本数据输入至语言处理模型，语言处理模型可以学习如何根据训练文本数据以及真实语音标记数据生成相应的预测语音标记数据。真实语音标记数据是从参考语音中提取的离散标记。真实语音标记数据中包括参考语音的真实语义信息以及部分真实风格信息，真实语义信息为参考语音的真实的语义信息，真实风格信息为参考语音的真实的风格信息，风格信息例如可以包括语速、韵律、音调、情感等。参考语音的内容可以与训练文本数据的内容相同，也即，训练文本数据为参考语音的文本句子。

在一些实施方式中，电子设备可以获取参考语音，对参考语音进行预处理，例如，清洗处理、去噪处理以及统一化处理等，再对预处理后的参考语音进行特征提取，获得语音特征，并对每一帧语音特征进行标记，获得真实语音标记数据。

步骤 120：将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得预测模块输出的预测语音标记数据。

其中，预测模块可以为 LLM 模型。

在一些实施方式中，电子设备可以通过以下公式，获得语言处理模型中的预测模块输出的训练文本数据的预测语音标记数据，预测语音标记数据包括参考语音的预测语义信息以及预测风格信息，预测语义信息为预测出的参考语音的语义信息，预测风格信息为预测出的参考语音的风格信息。

$$speech_token_{predict} = LLM（speech_token_{truth}，text）$$

其中，$speech_token_{predict}$ 表示预测语音标记数据，LLM 表示 LMM 模型，$speech_token_{truth}$ 表示真实语音标记数据，text 表示训练文本数据。

步骤 130：根据真实语音标记数据获得真实风格特征，并根据预测语音标记数据获得预测风格特征。

在一些实施方式中，电子设备可以对真实语音标记数据进行风格提取，获得真实风格特征，电子设备还可以对预测语音标记数据进行风格提取，获得预测风格特征。

步骤 140：根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征，获得总损失函数。

在一些实施方式中，电子设备可以根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征构建总损失函数。

步骤 150：根据总损失函数对预测模块进行优化训练。

在一些实施方式中，预测模块可以根据总损失函数的反馈逐步调整内部参数，改善在生成预测语音标记数据和风格信息捕捉的能力，预测模块可以根据总损失函数不断进行迭代，直到总损失函数达到阈值，或者总损失函数趋于稳定。

本公开实施例提供的语言处理模型训练方法，获取训练文本数据和真实语音标记数据并输入至语言处理模型中的预测模块，获得预测语音标记数据，分别提取真实和预测语音标记数据中的风格特征，再根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征构建总损失函数，再根据总损失函数对预测模块进行优化训练，将风格特征损失引入总损失函数，确保语言处理模型能够学习风格信息，使得语言处理模型更准确地复现语音风格，从而提高语言处理模型在处理语音风格信息的表现，并且语音处理模型是语音合成模型中的关键部分，进而能够提高语音合成大模型在处理语音风格信息的表现。

为了确保语言处理模型能够学习风格信息，在语言处理模型中引入风格编码模块，语言处理模型包括风格编码模块，如图 2 所示，步骤 130 具体包括以下步骤：

步骤 131：通过风格编码模块对真实语音标记数据进行风格提取，获得真实风格特征。

在引入风格编码模块的语音合成大模型的架构如图 3 所示，将训练文本数据（text）以及真实语音标记数据（$speech_token_{truth}$）输入至语音合成大模型的第一阶段的语言处理模型（LLM），获得语言处理模型（LLM）输出的预测语音标记数据（$speech_token_{predict}$），通过风格编码模块（style encoder）对真实语音标记数据（$speech_token_{truth}$）进行风格提取，获得真实风格特征（style truth），通过风格编码模块（style encoder）对预测语音标记数据（$speech_token_{predict}$）进行风格提取，获得预测风格特征（style predict），将语言处理模型输出的预测语音标记数据（$speech_token_{predict}$）输入至语音合成大模型的第二阶段的声学特征预测模型（Token-to-mel），再将声学特征预测模型（Token-to-mel）输出的数据输入至声码器模型（Mel-to-wav）。

在一些实施方式中，风格编码模块可以利用线性预测编码（Linear Predictive Coding，LPC）技术对语音标记进行风格提取，获得风格特征。

需要说明的是，LPC 模型通过提取线性预测系数（LPC 系数）来表征语音标记数据的频谱特性，LPC 系数能够反映语音标记数据的共振峰位置和形状，从而间接表示语音的风格特征，例如语调、语速和音色等。

在一些实施方式中，电子设备可以通过以下公式，获得真实风格特征。

$$style_hidden_{truth}=Style_encoder(speech_token_{truth})$$

其中，$style_hidden_{truth}$ 表示真实风格特征，Style_encoder 表示为风格编码模块，$speech_token_{truth}$ 表示真实语音标记数据。

请继续参阅图 2，步骤 132：通过风格编码模块对预测语音标记数据进行风格提取，获得预测风格特征。

在一些实施方式中，电子设备可以通过以下公式，获得预测风格特征。

$$style_hidden_{predict}=Style_encoder(speech_token_{predict})$$

其中，style_hidden $_{predict}$ 表示预测风格特征，Style_encoder 表示为风格编码模块，$speech_token_{predict}$ 表预测语音标记数据。

在一些实施方式中，语言处理模型中包括的风格编码模块是已经训练好的模型，在预测模块训练过程中，风格编码模块作为监督语言处理模型训练的部分。

在另一些实施方式中，语言处理模型中包括的风格编码模块是未经过训练的，可以在预测模块训练过程中同时对风格编码模块中的参数进行调整。

可以理解的是，通过统一的风格编码模块分别对真实语音标记数据以及预测语音标记数据进行风格提取，可以保证真实风格特征和预测风格特征在同一标准下进行比较，从而提高语言处理模型在训练过程中风格损失函数计算的准确性，进而提高语言处理模型在处理语言风格的表现。

为了提高语言处理模型在处理语言风格的表现，引入风格损失函数，如图 4 所示，步骤 140 具体包括以下步骤：

步骤 141：根据真实语音标记数据以及预测语音标记数据，获得预测损失函数。

其中，预测损失函数是配置成衡量真实语音标记数据和预测语音标记数据之间的差异。

在一些实施方式中，电子设备可以计算真实语音标记数据与预测语音标记数据之间的分布距离，获得预测损失函数。

在一些实施方式中，电子设备可以通过以下公式计算真实语音标记数据与预测语音标记数据之间的交叉熵，来计算真实语音标记数据与预测语音标记数据之间的分布距离，获得预测损失函数。

$$Loss_{ce}=Cross_Entropy(speech_token_{truth}，speech_token_{predict})$$

其中，$Loss_{ce}$ 表示预测损失函数，Cross_Entropy ($speech_token_{truth}$，$speech_token_{predict}$) 表示真实语音标记数据与预测语音标记数据之间的交叉熵损失，$speech_token_{truth}$ 表示真实语音标记数据，$speech_token_{predict}$ 表示预测语音标记数据。

步骤 142：根据真实风格特征以及预测风格特征，获得风格损失函数。

在一些实施方式中，计算真实风格特征与预测风格特征之间的相似度，获得风格损失函数。

其中，电子设备可以通过余弦相似度、KL 散度以及动态时间规整（Dynamic Time Warping， DTW）等技术，计算真实风格特征与预测风格特征之间的相似度。

在一些实施方式中，电子设备可以通过以下公式计算真实风格特征与预测风格特征之间的余弦相似度，获得预测损失函数。

$$Loss_{style}=Cosine_Embedding(style_hidden_{truth}，style_hidden_{predict})$$

其中，Lossstyle 表示风格损失函数，$Cosine_Embedding(style_hidden_{truth}，style_hidden_{predict})$ 表示真实风格特征与预测风格特征之间的余弦相似度，$style_hidden_{truth}$ 表示真实风格特征，$style_hidden_{predict}$ 表示预测风格特征。

步骤 143：根据预测损失函数以及风格损失函数，获得总损失函数。

在一些实施方式中，对预测损失函数以及风格损失函数进行加权计算，获得总损失函数。

在一些实施方式中，电子设备中预先设置并存储有预设加权值，电子设备可以根据以下公式对预测损失函数以及风格损失函数进行加权计算，获得总损失函数。

$$Loss_{llm=}\alpha \times Loss_{ce} + (1-\alpha) \times Loss_{style}$$

其中，$Loss_{llm}$ 表示总损失函数，α 表示预设加权值，$Loss_{ce}$ 表示预测损失函数，$Loss_{style}$ 表示风格损失函数。

需要说明的是，预测损失函数优化的是真实语音标记数据和预测语音标记数据之间的硬匹配，当真实语音标记数据和预测语音标记数据之间的差异较小时，语音标记数据所包含的风格信息可能并未被准确的预测，因此通过加权计算在语义信息和风格信息的差异方面做一个权衡。

本公开实施例提供的语言处理模型训练方法，通过引入额外的风格编码模块以及风格损失函数，确保语音合成大模型生成的语音在语义和风格层面都能与训练文本数据和真实语音标记数据（参考语音对应的语音标记数据）的风格保持一致，从而实现更精确的语音风格控制，进而提升用户的体验感。

为了执行上述实施例及各个可能的方式中的相应步骤，下面给出一种语言处理模型训练装置的实现方式。可选地，请参阅图 5，图为本公开实施例提供的一种语言处理模型训练装置的功能模块图。需要说明的是，本实施例所提供的语言处理模型训练装置，其基本原理及产生的技术效果和上述实施例相同，为简要描述，本实施例部分未提及之处，可参考上述的实施例中相应内容。该语言处理模型训练装置 200 包括：预测语音标记数据获得模块 210、风格特征获得模块 220、损失函数获得模块 230 以及优化训练模块 240，其中：

预测语音标记数据获得模块 210，配置成获得训练文本数据以及真实语音标记数据；将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得预测模块输出的预测语音标记数据。

风格特征获得模块 220，配置成根据真实语音标记数据获得真实风格特征，并根据预测语音标记数据获得预测风格特征。

损失函数获得模块 230，配置成根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征，获得总损失函数。

优化训练模块 240，配置成根据总损失函数对语言处理模型进行优化训练。

可选地，语言处理模型包括风格编码模块，风格特征获得模块还配置成通过风格编码模块对真实语音标记数据进行风格提取，获得真实风格特征；通过风格编码模块对预测语音标记数据进行风格提取，获得预测风格特征。

可选地，损失函数获得模块 230 具体还配置成根据真实语音标记数据以及预测语音标记数据，获得预测损失函数；根据真实风格特征以及预测风格特征，获得风格损失函数；根据预测损失函数以及风格损失函数，获得总损失函数。

可选地，损失函数获得模块 230 具体还配置成计算真实语音标记数据与预测语音标记数据之间的分布距离，获得预测损失函数。

可选地，损失函数获得模块 230 具体还配置成计算真实风格特征与预测风格特征之间的相似度，获得风格损失函数。

可选地，损失函数获得模块 230 具体还配置成对预测损失函数以及风格损失函数进行加权计算，获得总损失函数。

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，上述描述装置和模块的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本公开所提供的几个实施例中，模块相互之间的耦合可以是电性，机械或其它形式的耦合。

另外，在本公开各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。

请参照图 6，是本公开实施例提供的电子设备 100 的方框示意图。电子设备 100 包括存储器 110、处理器 120 及通信模块 130。存储器 110、处理器 120 以及通信模块 130 各元件相互之间直接或间接地电性连接，以实现数据的传输或交互。例如，这些元件相互之间可通过一条或多条通讯总线或信号线实现电性连接。

其中，存储器 110 配置成存储程序或者数据。存储器 110 可以是，但不限于，随机存取存储器（Random Access Memory，RAM），只读存储器（Read Only Memory，ROM），可编程只读存储器（Programmable Read-Only Memory，PROM），可擦除只读存储器（Erasable Programmable Read-Only Memory，EPROM），电可擦除只读存储器（Electric Erasable Programmable Read-Only Memory，EEPROM）等。

处理器 120 配置成读/写存储器中存储的数据或程序，并执行相应地功能。例如，当存储器 110 中存储的计算机程序被处理器 120 执行时，可以实现上述各实施例所揭示的语言处理模型训练方法。

通信模块 130 配置成通过网络建立电子设备 100 与云服务器之间的通信连接，并配置成通过网络收发数据。

应当理解的是，图 6 所示的结构仅为电子设备的结构示意图，电子设备还可包括比图 6 中所示更多或者更少的组件。图 6 中所示的各组件可以采用硬件、软件或其组合实现。

本公开实施例还提供了一种计算机可读存储介质，该计算机可读介质中存储有计算机程序，计算机程序可被处理器执行时实现上述方法实施例中所描述的语言处理模型训练方法。

计算机可读存储介质可以是诸如闪存、EEPROM（电可擦除可编程只读存储器）、EPROM、硬盘或者 ROM 之类的电子存储器。可选地，计算机可读存储介质包括非易失性计算机可读介质（non-transitory computer-readable storage medium）。计算机可读存储介质具有执行上述方法中的任何方法步骤的程序代码的存储空间。这些程序代码可以从一个或者多个计算机程序产品中读出或者写入到这一个或者多个计算机程序产品中。程序代码可以例如以适当形式进行压缩。

在本公开所提供的几个实施例中，应该理解到，所揭露的装置和方法，也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的，例如，附图中的流程图和框图显示了根据本公开的多个实施例的装置、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，模块、程序段或代码的一部分包含一个或多个配置成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方式中，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，两个连续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。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动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。

功能如果以软件功能模块的形式实现并作为独立的产品销售或使用时，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，本公开的技术方案本质上或者说对相关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可以存储程序代码的介质。

以上仅为本公开的可选实施例而已，并不配置成限制本公开，对于本领域的技术人员来说，本公开可以有各种更改和变化。凡在本公开的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本公开的保护范围之内。

工业实用性

采用上述方案，获取训练文本数据和真实语音标记数据并输入至语言处理模型中的预测模块，获得预测语音标记数据，分别提取真实和预测语音标记数据中的风格特征，再根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征构建总损失函数，再根据总损失函数对预测模块进行优化训练，将风格特征损失引入总损失函数，确保语言处理模型能够学习风格信息，使得语言处理模型更准确地复现语音风格，从而提高语言处理模型在处理语音风格信息的表现，并且语音处理模型是语音合成模型中的关键部分，进而能够提高语音合成大模型在处理语音风格信息的表现。

权利要求书

1.一种语言处理模型训练方法，其特征在于，所述方法包括：

获得训练文本数据以及真实语音标记数据；

将所述训练文本数据以及所述真实语音标记数据输入至语言处理模型中的预测模块，获得所述预测模块输出的预测语音标记数据；

根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征；

根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数；

根据所述总损失函数对所述预测模块进行优化训练。

2.根据权利要求 1 所述的方法，其特征在于，所述语言处理模型包括风格编码模块，所述根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征，包括：

通过所述风格编码模块对所述真实语音标记数据进行风格提取，获得所述真实风格特征；

通过所述风格编码模块对所述预测语音标记数据进行风格提取，获得所述预测风格特征。

3.根据权利要求 1 或 2 所述的方法，其特征在于，所述根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数，包括：

根据所述真实语音标记数据以及所述预测语音标记数据，获得预测损失函数；

根据所述真实风格特征以及所述预测风格特征，获得风格损失函数；

根据所述预测损失函数以及所述风格损失函数，获得所述总损失函数。

4.根据权利要求 3 所述的方法，其特征在于，所述根据所述真实语音标记数据以及所述预测语音标记数据，获得预测损失函数，包括：

计算所述真实语音标记数据与所述预测语音标记数据之间的分布距离，获得所述预测损失函数。

5.根据权利要求 3 或 4 所述的方法，其特征在于，所述根据所述真实风格特征以及所述预测风格特征，获得风格损失函数，包括：

计算所述真实风格特征与所述预测风格特征之间的相似度，获得所述风格损失函数。

6.根据权利要求 3-5 任一项所述的方法，其特征在于，所述根据所述预测损失函数以及所述风格损失函数，获得所述总损失函数，包括：

对所述预测损失函数以及所述风格损失函数进行加权计算，获得所述总损失函数。

7.根据权利要求 2-6 任一项所述的方法，其特征在于，所述通过所述风格编码模块对所述真实语音标记数据进行风格提取，获得所述真实风格特征，包括：

通过所述风格编码模块对所述真实语音标记数据进行风格提取，得到所述真实语音标记数据的线性预测系数，所述真实语音标记数据的线性预测系数表征所述真实语音标记数据的频谱特性；

根据所述真实语音标记数据的频谱特性，获取所述真实风格特征。

8.根据权利要求 2-7 任一项所述的方法，其特征在于，所述通过所述风格编码模块对所述预测语音标记数据进行风格提取，获得所述预测风格特征，包括：

通过所述风格编码模块对所述预测语音标记数据进行风格提取，得到所述预测语音标记数据的线性预测系数，所述预测语音标记数据的线性预测系数表征所述预测语音标记数据的频谱特性；

根据所述预测语音标记数据的频谱特性，获取所述预测风格特征。

9.根据权利要求 4-8 任一项所述的方法，其特征在于，所述计算所述真实语音标记数据与所述预测语音标记数据之间的分布距离，获得所述预测损失函数，包括：

计算所述真实语音标记数据与所述预测语音标记数据之间的交叉熵；

根据所述交叉熵，获取所述预测损失函数。

10.根据权利要求 2-8 任一项所述的方法，其特征在于，所述方法还包括：

根据所述总损失函数对所述风格编码模块进行优化训练。

11.一种语言处理模型训练装置，其特征在于，所述装置包括：

预测语音标记数据获得模块，配置成获得训练文本数据以及真实语音标记数据；将所述训练文本数据以及所述真实语音标记数据输入至语言处理模型中的预测模块，获得所述预测模块输出的预测语音标记数据；

风格特征获得模块，配置成根据所述真实语音标记数据获得真实风格特征，并根据所述预测语音标记数据获得预测风格特征；

损失函数获得模块，配置成根据所述真实语音标记数据、所述预测语音标记数据、所述真实风格特征以及所述预测风格特征，获得总损失函数；

优化训练模块，配置成根据所述总损失函数对所述预测模块进行优化训练。

12.根据权利要求 11 所述的装置，其特征在于，所述语言处理模型包括风格编码模块，所述风格特征获得模块还配置成通过所述风格编码模块对所述真实语音标记数据进行风格提取，获得所述真实风格特征；通过所述风格编码模块对所述预测语音标记数据进行风格提取，获得所述预测风格特征。

13.一种电子设备，其特征在于，包括处理器和存储器，所述存储器存储有能够被所述处理器执行的计算机程序，所述处理器可执行的计算机程序以实现权利要求 1-10 任一所述的语言处理模型训练方法。

14.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求 1-10 中任一项所述的语言处理模型训练方法。