



(12)发明专利申请

(10)申请公布号 CN 111414513 A

(43)申请公布日 2020.07.14

(21)申请号 202010183719.9

(22)申请日 2020.03.16

(71)申请人 腾讯音乐娱乐科技(深圳)有限公司

地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室

(72)发明人 林梅露 吴康健 吴斌 王征韬

夏志强 雷兆恒

(74)专利代理机构 深圳翼盛智成知识产权事务

所(普通合伙) 44300

代理人 彭绪坤

(51)Int.Cl.

G06F 16/65(2019.01)

G06F 16/683(2019.01)

G06F 40/30(2020.01)

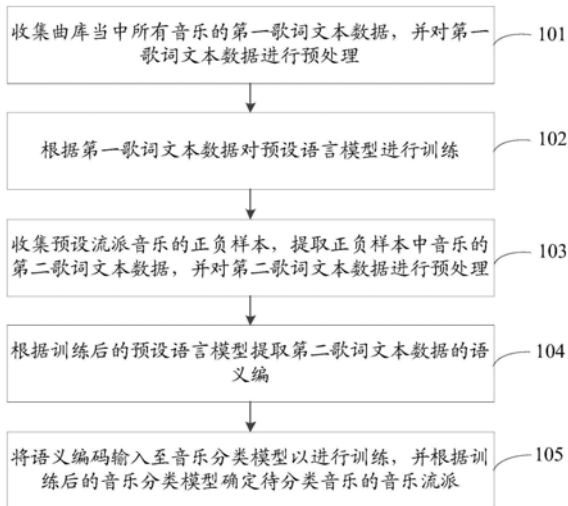
权利要求书2页 说明书13页 附图4页

(54)发明名称

音乐流派的分类方法、装置及存储介质

(57)摘要

本发明实施例公开了一种音乐流派的分类方法、装置及存储介质。该方案通过收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,提取正负样本中音乐的音频数据,并对音频数据进行预处理,对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征,根据训练后的预设语言模型提取第一歌词文本数据的语义编码,将语义编码与第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案通过训练完成后的分类模型确定待分类音乐的流派,提升了音乐流派分类的准确性。



1. 一种音乐流派的分类方法,其特征在于,包括:
  - 收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理;
  - 提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理;
  - 对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;
  - 根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;
  - 将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。
2. 如权利要求1所述的音乐流派的分类方法,其特征在于,在收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据之前,所述方法还包括:
  - 收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理;
  - 根据所述第二歌词文本数据对预设语言模型进行训练。
3. 如权利要求2所述的音乐流派的分类方法,其特征在于,在收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理之前,所述方法还包括:
  - 从文本语料库当中提取语句序列;
  - 将所述语句序列输入至预设语言模型以进行预训练。
4. 如权利要求1-2任一项所述的音乐流派的分类方法,其特征在于,对所述第二歌词文本数据或第一歌词文本数据进行预处理,包括:
  - 删除所述第二歌词文本或第一歌词文本中的非标准字符;和/或
  - 对所述第二歌词文本或第一歌词文本进行截取,以保留预设数量的字符。
5. 如权利要求2所述的音乐流派的分类方法,其特征在于,根据所述第二歌词文本数据对预设语言模型进行训练,包括:
  - 根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值;
  - 根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。
6. 如权利要求1所述的音乐流派的分类方法,其特征在于,对所述音频数据进行预处理,包括:
  - 在所述正负样本中选取所述音频数据满足预设时长的目标音频数据;和/或
  - 提取所述音频数据的第二音频特征,并对所述第二音频特征进行放大处理。
7. 一种音乐流派的分类装置,其特征在于,包括:
  - 第一收集单元,用于收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理;
  - 第一训练单元,用于根据所述第二歌词文本数据对预设语言模型进行训练;
  - 第二收集单元,用于收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理,提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理,对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;
  - 编码提取单元,用于根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;

第二训练单元,用于将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

8. 如权利要求7所述的音乐流派的分类装置,其特征在于,所述装置还包括:

序列提取单元,用于在所述第一收集单元收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理之前,从文本语料库当中提取语句序列;

预训练单元,用于将所述语句序列输入至预设语言模型以进行预训练。

9. 如权利要求7所述的音乐流派的分类装置,其特征在于,所述第一训练单元包括:

计算子单元,用于根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值;

估计子单元,用于根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。

10. 一种存储介质,其特征在于,所述存储介质存储有多条指令,所述指令适于处理器进行加载,以执行权利要求1至6任一项所述的音乐流派的分类方法。

## 音乐流派的分类方法、装置及存储介质

### 技术领域

[0001] 本发明涉及数据处理技术领域,具体涉及一种音乐流派的分类方法、装置及存储介质。

### 背景技术

[0002] 目前,随着互联网的发展,人们可以越来越多的接触到互联网提供的音乐内容,同时对音乐服务的要求也越来越高。高效的音乐信息检索可以帮助人们从海量数据中,找出满足个性化需求的音乐资源。一种普遍的检索方式,可以通过曲风和流派的标签进行歌曲的搜索。但由于曲库歌曲数量庞大及人工分配标签成本过高的原因,仍然存在大量歌曲没有曲风或流派的标签,因此各大音乐平台通常会采用构建曲风/流派分类器的方式,为歌曲自动分配相应标签。

[0003] 目前,对歌曲进行流派分类主要是基于音频内容的方式。但这种方式对于一些特定流派的分类效果并不理想,比如古风流派。古风歌曲最重要的特征,是它的歌词有别于其他流派歌曲,具有独特的艺术特点。因此,如果采用基于纯音频内容的方式进行古风流派的分类,会由于忽略了古风歌曲最本质的特点,而造成分类准确率较低的问题。

### 发明内容

[0004] 本发明实施例提供一种音乐流派的分类方法、装置及存储介质,旨在提升音乐流派分类的准确性。

[0005] 本发明实施例提供一种音乐流派的分类方法,包括:

[0006] 收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理;

[0007] 提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理;

[0008] 对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;

[0009] 根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;

[0010] 将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0011] 本发明实施例还提供一种音乐流派的分类装置,包括:

[0012] 第一收集单元,用于收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理;

[0013] 第一训练单元,用于根据所述第二歌词文本数据对预设语言模型进行训练;

[0014] 第二收集单元,用于收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理,提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理,对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;

[0015] 编码提取单元,用于根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;

[0016] 第二训练单元,用于将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0017] 本发明实施例还提供一种存储介质,所述存储介质存储有多条指令,所述指令适于处理器进行加载,以执行本发明实施例所提供的任一音乐流派的分类方法。

[0018] 本发明实施例提供的音乐流派的分类方案,通过收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,提取正负样本中音乐的音频数据,并对音频数据进行预处理,对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征,根据训练后的预设语言模型提取第一歌词文本数据的语义编码,将语义编码与第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案从古风流派歌曲最为重要的歌词特征入手,对音乐分类模型进行训练,通过训练完成后的分类模型确定待分类音乐的风格流派,可以有效提升音乐流派分类的准确性。

## 附图说明

[0019] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域技术人员来讲,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。

[0020] 图1a是本发明实施例提供的音乐流派的分类方法的第一流程示意图;

[0021] 图1b是本发明实施例提供的音乐流派的分类方法的第二流程示意图;

[0022] 图2是本发明实施例提供的音乐分类模型的结构示意图;

[0023] 图3a是本发明实施例提供的音乐流派的分类装置的第一种结构示意图;

[0024] 图3b是本发明实施例提供的音乐流派的分类装置的第二种结构示意图;

[0025] 图4是本发明实施例提供的服务器的结构示意图。

## 具体实施方式

[0026] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域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,都属于本发明保护的范围。

[0027] 在本文中提及“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发明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。

[0028] 本发明实施例提供一种音乐流派的分类方法,该音乐流派的分类方法的执行主体

可以是本发明实施例提供的音乐流派的分类装置,或者集成了该音乐流派的分类装置的服务器,其中该音乐流派的分类装置可以采用硬件或者软件的方式实现。

[0029] 目前,现有技术当中针对音乐流派的分类方法主要有两种,分别是基于音频内容的流派分类方法和基于歌词文本的情感分类方法。具体的,基于音频内容的流派分类方法通常包含以下模块:

[0030] 1. 音频特征提取模块

[0031] 音频特征提取模块可以提供一种,或多种由音频数据直接产生的特征,例如mel频率倒谱系数(MFCC),频谱平坦度,能量等;音频特征提取模块也可以不进行特征提取,而直接使用原始音频数据作为后续模块的输入。

[0032] 2. 深度学习模型

[0033] 对于流派分类问题,可以利用输入的音频特征和其对应的流派标签来进行监督学习的训练。常用于处理音频输入数据的深度学习模型,包括卷积神经网络(CNN),长短时记忆(LSTM)等。通过梯度下降的方式优化分类损失,使模型学会对目标流派的区分。但是上述这种基于音频内容的流派分类方法对于一些区分度不完全在音频的流派(如古风流派),这种方式无法提供令人满意的分类准确率。

[0034] 基于歌词文本的情感分类方法,一种做法是,通过标注了感情类型的歌词样本获取各类型的感情词库,再利用感情词库建立歌词样本的特征向量,最后利用各歌词的特征向量及其对应的情感标签进行分类模型的训练。除此之外,利用歌词进行情感分类,同样也包括使用CNN,LSTM等模型进行目标任务训练的方法。

[0035] 对于流派分类,无法采用类似构建感情词库的方法,因为大多数流派并没有特别的词汇表征。另外,对于一些歌曲,部分词组具有古意也不能代表其为古风歌曲。

[0036] 对于使用CNN进行歌词分析的方法,考虑到CNN的卷积核机制,会使得网络受到部分词组的影响比较大,无法关注到整体的内容。类似的,使用LSTM的模型,由于其模型结构存在时序上的依赖,使得训练时间会很长,很难训练深一些的模型,因此在长时间序列的任务上也不是很理想。

[0037] 对于古风流派的歌曲,通常情况下需要一首歌的歌词在整体语义上具备古风特点,因此选取可以处理长时间序列,且能够关注到整体语义语境的模型,能保证模型获得更全面的信息,进而提高古风流派分类的准确率。

[0038] 基于上述的问题,本申请实施例提供了一种音乐流派的分类方法。如图1a所示,图1a是本发明实施例提供的音乐流派的分类方法的第一流程示意图,该音乐流派的分类方法的具体流程可以如下:

[0039] 101、收集曲库当中音乐的第二歌词文本数据,并对第二歌词文本数据进行预处理。

[0040] 在实际应用当中,有许多纯音乐比如钢琴曲、吉他曲等等是不包含歌词的,因此对于这些音乐就无法获取其歌词文本数据。在一实施例中,可以先确定曲库中的音乐是否包括歌词,具体可以通过判断音乐当中是否包含人声的方式来进行确定,若包括则继续获取该音乐的歌词文本数据。在其他实施例中,还可以通过查询是否有歌词文件的方式来进行判断,在此不再一一列举。

[0041] 在一实施例中,可以通过爬虫技术在互联网中爬取音乐的第二歌词文本数据。例

如,电子设备通过运行预先设定的抓虫脚本,抓取各个音乐平台的歌曲的歌词文本数据。其中,预先设定的抓虫脚本可以由用户根据实际需求来进行编写。例如,用户编写抓虫脚本的过程为:梳理需要爬取的音乐平台列表和歌曲列表,整理在各音乐平台上需要爬取的指标,例如,抓取的指标为歌词文本数据;可以采用计算机程序设计语言Python编写爬虫脚本,爬取各平台曲库当中音乐的第二歌词文本数据。

[0042] 其中,需要说明的是,音乐的第二歌词文本数据也可以由用户直接导入,本领域技术人员可以理解,实际应用中,可以采用多种方式获取音乐的第二歌词文本数据,本实施方式不限制获取音乐的第二歌词文本数据的具体方式。

[0043] 进一步的,在获取到音乐的第二歌词文本数据之后,还可以对上述第二歌词文本数据进行预处理,该预处理具体可以为对第二歌词文本数据当中非标准字符的删除,其中,该非标准字符可以包括标点符号,特殊符号,时间戳等。在其他实施例中,上述预处理还可以为对歌词文本的长度进行截取,比如保留1-512个字符的歌词文本数据,进一步的,还可以在该第二歌词文本数据当中添加一个CLS字符。也即对所述第二歌词文本数据进行预处理的步骤,可以包括:

[0044] 删除所述第二歌词文本中的非标准字符;和/或

[0045] 对所述第二歌词文本进行截取,以保留预设数量的字符。

[0046] 102、根据第二歌词文本数据对预设语言模型进行训练。

[0047] 在一实施例中,上述预设语言模型可以为BERT(Bidirectional Encoder Representations from Transformers)模型,BERT是一个语言表征模型(language representation model),通过超大数据、巨大模型、和极大的计算开销训练而成。具体可以利用步骤102中获取到的第二歌词文本数据对该BERT模型进行训练。

[0048] 在一实施例中,在根据第二歌词文本数据对预设语言模型进行训练之前,还可以对BERT模型进行预训练,也即,利用第二歌词文本进行训练,可以使用预训练的BERT模型参数。其中,上述BERT模型代表Transformer的双向编码器表示。与最近的其他语言表示模型不同,BERT模型旨在通过联合调节所有层中的上下文来预先训练深度双向表示。因此,预训练的BERT模型表示可以通过一个额外的输出层进行微调(fine-tuning),适用于广泛任务的最先进模型的构建。

[0049] 在一申请实施例中,根据所述第二歌词文本数据对预设语言模型进行训练的的步骤可以包括:

[0050] 根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值;

[0051] 根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。

[0052] 其中,损失函数(loss function)是用来估量模型的预测值 $f(x)$ 与真实值 $Y$ 的不一致程度,它是一个非负实值函数,通常使用 $L(Y, f(x))$ ,或者 $L(w)$ 来表示,损失函数越小,模型的鲁棒性就越好。损失函数是经验风险函数的核心部分,也是结构风险函数重要组成部分。其中, $w$ 为模型参数。在一实施例中,可以采用梯度下降法或其他优化方法训练模型参数。

[0053] 在一申请实施例中,在对BERT模型进行训练的过程中,其损失函数包含两部分,一部分是对输入数据进行mask操作,让模型去预测mask掉的字符;另一部分是让模型去预测前后两个句子是不是属于同一上下文。第一部分的损失可以帮助模型学会根据前后文关系

更好的表示语义,第二部分的损失可以让模型更好的把控整体的语义。对于第二部分的损失,我们可以认为一首歌的歌词在整体上应该有连贯的语义。因此,对这部分损失,我们将前后两个句子的限定扩展到同一首歌中任意两个句子即可。

[0054] 103、收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理。

[0055] 在一实施例中,上述预设音乐流派可以为古风流派,其中,针对与分类问题,正样本则是我们想要正确分类出的类别所对应的样本,在本申请当中可以将古风流派的音乐作为正样本,相对的,可以将非古风流派的音乐作为负样本。上述正负样本可以包括多个,比如包括多个正样本以及多个负样本。

[0056] 进一步的,在得到上述正负样本后,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理。其中,本申请当中提取音乐的第一歌词文本数据,以及对所述第一歌词文本数据进行预处理的步骤可以参考步骤101当中对第二歌词文本的处理过程,比如也可以对第一歌词文本数据删除非标准字符,和/或对第一歌词文本进行截取,在此不作进一步赘述。

[0057] 104、根据训练后的预设语言模型提取第一歌词文本数据的语义编码。

[0058] 在一实施例中,上述BERT语言模型可以包括Embedding层、双向Transformer编码器、输出层,其中Embedding层是词嵌入,位置嵌入,类型嵌入之和,分别表示词信息,位置信息,句子对信息。双向Transformer编码器是多层编码单元的堆叠,每个编码单元包括自注意力模块,残差网络,层归一化结构,DropOut层,用于提取语义信息。通过上述训练后的BERT模型可以对句子进行编码,也即对第一歌词文本数据进行编码,得到第一歌词文本语义的编码。

[0059] 105、将语义编码输入至音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0060] 在一实施例中,上述音乐分类模型为基于BERT模型的,比如该音乐分类模型的结构可以包括输入层、BERT层、交互层、特征选择层、编码层、匹配层以及输出层组成。其中,本实施例提供的方法并不限于该结构的深度匹配模型,还可以为其它结构,但是以输入层、BERT层、交互层、匹配层以及输出层为基础结构。

[0061] 根据上述第一歌词文本数据的语义编码对音乐分类模型进行训练,需要说明的是,在训练的过程中,上述BERT模型也会进行微调学习。进一步的,在对上述音乐模型进行训练后,可以利用训练好的模型对训练数据进行预测,排除预测错误的数据。在其他实施例中,还可以利用训练好的模型对训练数据进行预测,在训练集中添加与预测错误数据流派相同/相近的音乐。

[0062] 进一步的,对上述训练过程进行多次迭代,可以以最小化损失值为目标,对模型不断训练以得到最终的音乐分类模型,从而使用该最终的音乐分类模型确定待分类音乐的音乐流派。

[0063] 比如,确定待分类音乐后,可以先收集该音乐的歌词文本数据,若不存在歌词文本数据(例如纯音乐等),则直接退出,若存在歌词文本数据,则进行与上述训练数据相同的预处理操作,并将数据输入到训练好的基于BERT模型的音乐分类模型当中,得到分类结果。具体的,将数据输入到上述音乐分类模型后,输出第一分类结果的概率以及第二分类结果的



概率,比如该第一分类结果可以为古风流派,第二分类结果可以为非古风流派,然后将上述第一分类结果的概率以及第二分类结果的概率进行对比,若第一分类结果的概率大于第二分类结果的概率,则可以预测待分类音乐为古风流派,若第一分类结果的概率不大于第二分类结果的概率,则可以预测待分类音乐为非古风流派。

[0064] 由上所述,本发明实施例提出的音乐流派的分类方法可以收集曲库当中音乐的第二歌词文本数据,并对第二歌词文本数据进行预处理,根据第二歌词文本数据对预设语言模型进行训练,收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,根据训练后的预设语言模型提取第一歌词文本数据的语义编码,将语义编码输入至音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案从古风流派歌曲最为重要的歌词特征入手,对音乐分类模型进行训练,通过训练完成后的分类模型确定待分类音乐的风格流派,可以有效提升音乐流派分类的准确性。

[0065] 根据前面实施例所描述的方法,以下将作进一步详细说明。

[0066] 请参阅图1b,图1b是本发明实施例提供的音乐流派的分类方法的第二流程示意图。所述方法包括:

[0067] 201、从文本语料库当中提取语句序列,将语句序列输入至预设语言模型以进行预训练。

[0068] 目前,缺少训练数据是自然语言处理(Natural Language Processing,NLP)面临的巨大挑战之一。由于NLP是一个具备不同任务的多样化领域,因此大多数任务特定数据集仅包含几千或几十万个人类标签的训练样例。然而,基于深度学习的NLP模型可以从更大量的数据中获益,在数百万或数十亿的带标签的训练样例中学习得到改善。为了帮助缩小在数据方面的差距,可以使用网络上无标签的文本训练一个通用的语言表示模型(称为预训练)。用预训练模型在小数据的NLP任务(如问答和情感分析)上进行微调,与从头开始训练相比,可以显著提高准确度。

[0069] 在本申请实施例中,BERT建立在最新的预训练与上下文相关的语言表示的工作之上—包括Semi-supervised Sequence Learning、Generative Pre-Training、ELMo和ULMFit。然而,与以前的模型不同,BERT是第一个深度、双向、无监督的语言表示模型,仅使用无标签的文本语料库(在本实施例中可以为维基百科)进行预训练。

[0070] 202、收集曲库当中音乐的第二歌词文本数据,并对第二歌词文本数据进行预处理。

[0071] 在一实施例中,可以通过爬虫技术在互联网中爬取音乐的第二歌词文本数据。需要说明的是,音乐的第二歌词文本数据也可以由用户直接导入。在获取到音乐的第二歌词文本数据之后,还可以对上述第二歌词文本数据进行预处理,该预处理具体可以为对第二歌词文本数据当中非标准字符的删除,其中,该非标准字符可以包括标点符号,特殊符号,时间戳等。在其他实施例中,上述预处理还可以为对歌词文本的长度进行截取,比如保留1-512个字符的歌词文本数据,进一步的,还可以在该第二歌词文本数据当中添加一个CLS字符。

[0072] 203、根据第二歌词文本数据对预设语言模型进行训练。

[0073] 在一实施例中,上述预设语言模型可以为BERT模型,根据所述第二歌词文本数据

对预设语言模型进行训练的步骤可以包括:根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值,根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。

[0074] 其中,在对BERT模型进行训练的过程中,其损失函数包含两部分,一部分是对输入数据进行mask操作,让模型去预测mask掉的字符;另一部分是让模型去预测前后两个句子是不是属于同一上下文。第一部分的损失可以帮助模型学会根据前后文关系更好的表示语义,第二部分的损失可以让模型更好的把控整体的语义。对于第二部分的损失,我们可以认为一首歌的歌词在整体上应该有连贯的语义。因此,对这部分损失,我们将前后两个句子的限定扩展到同一首歌中任意两个句子即可。

[0075] 204、收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理。

[0076] 在一实施例中,上述预设音乐流派可以为古风流派,其中,针对与分类问题,正样本则是我们想要正确分类出的类别所对应的样本,在本申请当中可以将古风流派的音乐作为正样本,相对的,可以将非古风流派的音乐作为负样本。

[0077] 进一步的,在得到上述正负样本后,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理。其中,本申请当中提取音乐的第一歌词文本数据,以及对所述第一歌词文本数据进行预处理的步骤可以参考步骤202当中对第二歌词文本的处理过程。

[0078] 205、提取正负样本中音乐的音频数据,并对音频数据进行预处理。

[0079] 在一实施例中,上述对所述音频数据进行预处理的步骤可以包括:

[0080] 在所述正负样本中选取所述音频数据满足预设时长的目标音频数据;和/或

[0081] 提取所述音频数据的第二音频特征,并对所述第二音频特征进行放大处理。

[0082] 具体的,上述对音频数据进行的预处理可以包括多种处理方式,比如对歌曲长度进行限制,保留时长在m分钟至n分钟之间的音乐,或者剔除>n分钟或/和<m分钟的音乐。还可以对音频数据提取特征,例如Mel Spectrogram,MFCC,ZCR,功率/能量等,可以选择其中的一种特征,也可以对多种特征进行拼接。另外,还可以上述提取到的音频特征进行放大/缩小/平移/取对数等操作。

[0083] 206、根据训练后的预设语言模型提取第一歌词文本数据的语义编码。

[0084] 具体的,通过上述训练后的BERT模型可以对句子进行编码,也即对第一歌词文本数据进行编码,得到第一歌词文本语义的编码。

[0085] 207、对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征。

[0086] 具体的,本申请实施例利用卷积滑窗操作对古风流派歌曲正负样本的音频输入进行压缩,并通过多层Transformer完成对音频特征的提取。

[0087] 208、将语义编码与第一音频特征进行结合,并将结合后的数据输入至音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0088] 在一实施例中,上述将语义编码与第一音频特征进行结合的方法可以有多种,比如可以将语义编码与第一音频特征进行拼接,举例来说,若音频特征有n维,歌词特征有m维,则可以拼接为n+m维。在其他实施例中,还可以将语义编码与第一音频特征进行点成或者加权相加等以进行结合,并将结合后的数据输入至音乐分类模型以进行训练。

[0089] 在本申请实施例中,请参阅图2,图2为本发明实施例提供的音乐分类模型的结构示意图。本实施例给出了一种歌词辅助音频的古风流派分类方法。目前对于音频内容的处理大多是基于CNN的模型结构,这使得模型不能很好满足音频率序列间的依赖,因此我们考虑使用多层Transformer的结构进行序列建模。由于Transformer是一个公开的最为先进的特征变换器,因此在此不对其进行详细描述。

[0090] 考虑到音频数据序列一般较长,直接作为多层Transformer结构的输入的话,会造成训练负担加重,并且音频数据本身的冗余也较大。因此,本申请实施例设计了下图左边的模型,采用卷积滑窗结合多层Transformer的结构,暂时称为Conv-Transformer。卷积滑窗的主要作用是对音频信号进行时间维度上的压缩,该部分可以是单层卷积滑窗,也可以是多层卷积滑窗,卷积核的大小及相关操作的参数不做限定。图2中的模型与上一实施例相比,多了进行音频特征和歌词特征的整合的部分,可以用于进行最终的分类。因此本申请实施例中的音乐分类模型是基于Conv-Transformer模型和BERT模型。

[0091] 需要说明的是,在训练的过程中,Conv-Transformer模型和BERT模型都会进行参数学习。进一步的,在对上述音乐模型进行训练后,可以利用训练好的模型对训练数据进行预测,排除预测错误的的数据。在其他实施例中,还可以利用训练好的模型对训练数据进行预测,在训练集中添加与预测错误数据流派相同/相近的音乐。对上述训练过程进行多次迭代以得到最终的音乐分类模型,从而使用该最终的音乐分类模型确定待分类音乐的音乐流派。

[0092] 在一实施例中,训练过程中可以设置学习率的。具体的,由于BERT模型是经过二次训练的模型,因此在分类任务的学习中,可以设置一个很小的学习率对其进行微调学习;而对于Conv-Transformer部分,由于不是预训练的模型,因此会设置较大一些的学习率,来加快其收敛速度。

[0093] 在一实施例中,确定待分类音乐后,可以先收集该音乐的歌词文本数据,并进行与训练数据相同的预处理操作,收集待预测歌曲的音频内容,并进行与训练数据相同的预处理操作,若歌词或音频两者均不存在,则退出。若均存在,则将歌词数据和音频数据输入训练好的基于Conv-Transformer的音频+基于BERT的歌词古风流派分类模型中,从而得到分类结果。

[0094] 由上所述,本发明实施例提出的音乐流派的分类方法可以从文本语料库当中提取语句序列,将语句序列输入至预设语言模型以进行预训练,收集曲库当中音乐的第二歌词文本数据,并对第二歌词文本数据进行预处理,根据第二歌词文本数据对预设语言模型进行训练,收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,提取正负样本中音乐的音频数据,并对音频数据进行预处理,根据训练后的预设语言模型提取第一歌词文本数据的语义编码,对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征,将语义编码与第一音频特征进行结合,并将结合后的数据输入至音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案从古风流派歌曲最为重要的歌词特征入手,对音乐分类模型进行训练,通过训练完成后的分类模型确定待分类音乐的风格流派,可以有效提升音乐流派分类的准确性。

[0095] 为了实施以上方法,本发明实施例还提供一种音乐流派的分类装置,该音乐流派

的分类装置具体可以集成在终端设备如手机、平板电脑等设备中。

[0096] 例如,如图3a所示,是本发明实施例提供的音乐流派的分类装置的第一种结构示意图。该音乐流派的分类装置可以包括:

[0097] 第一收集单元301,用于收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理。

[0098] 在一实施例中,可以先确定曲库中的音乐是否包括歌词,具体可以通过判断音乐当中是否包含人声的方式来进行确定,若包括则由第一收集单元301继续获取该音乐的歌词文本数据。

[0099] 在获取到上述歌词文本数据后,还可以对上述第二歌词文本数据进行预处理,该预处理具体可以为对第二歌词文本数据当中非标准字符的删除,其中,该非标准字符可以包括标点符号,特殊符号,时间戳等。在其他实施例中,上述预处理还可以为对歌词文本的长度进行截取,比如保留1-512个字符的歌词文本数据,进一步的,还可以在该第二歌词文本数据当中添加一个CLS字符。

[0100] 第一训练单元302,用于根据所述第二歌词文本数据对预设语言模型进行训练。

[0101] 在一实施例中,上述预设语言模型可以为BERT模型。该第一训练单元302根据所述第二歌词文本数据对BERT模型进行训练的步骤可以包括:根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值,根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。

[0102] 其中,损失函数包含两部分,一部分是对输入数据进行mask操作,让模型去预测mask掉的字符;另一部分是让模型去预测前后两个句子是不是属于同一上下文。第一部分的损失可以帮助模型学会根据前后文关系更好的表示语义,第二部分的损失可以让模型更好的把控整体的语义。对于第二部分的损失,我们可以认为一首歌的歌词在整体上应该有连贯的语义。因此,对这部分损失,我们将前后两个句子的限定扩展到同一首歌中任意两个句子即可。

[0103] 第二收集单元303,用于收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,提取正负样本中音乐的音频数据,并对音频数据进行预处理,对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征。

[0104] 在一实施例中,上述预设音乐流派可以为古风流派,其中,针对与分类问题,正样本则是我们想要正确分类出的类别所对应的样本,在本申请当中第二收集单元303可以将古风流派的音乐作为正样本,相对的,可以将非古风流派的音乐作为负样本。上述正负样本可以包括多个,比如包括多个正样本以及多个负样本。

[0105] 编码提取单元304,用于根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码。

[0106] 具体的,编码提取单元304通过上述训练后的BERT模型可以对句子进行编码,也即对第一歌词文本数据进行编码,得到第一歌词文本语义的编码。

[0107] 第二训练单元305,用于将所述语义编码输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0108] 在一实施例中,第二训练单元305根据上述第一歌词文本数据的语义编码对音乐

分类模型进行训练,需要说明的是,在训练的过程中,上述BERT模型也会进行微调学习。

[0109] 进一步的,对上述训练过程进行多次迭代,可以以最小化损失值为目标,对模型不断训练以得到最终的音乐分类模型,从而使用该最终的音乐分类模型确定待分类音乐的音乐流派。

[0110] 在一实施例中,请参阅图3b,其中,所述音乐流派的分类装置还可以包括:

[0111] 序列提取单元306,用于在所述第一收集单元301收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理之前,从文本语料库当中提取语句序列;

[0112] 预训练单元307,用于将所述语句序列输入至预设语言模型以进行预训练。

[0113] 在一实施例中,上述第一训练单元302可以包括:

[0114] 计算子单元3021,用于根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值;

[0115] 估计子单元3022,用于根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。

[0116] 具体实施时,以上各个单元可以作为独立的实体来实现,也可以进行任意组合,作为同一或若干个实体来实现,以上各个单元的具体实施可参见前面的方法实施例,在此不再赘述。

[0117] 应当说明的是,本发明实施例提供的音乐流派的分类装置与上文实施例中的音乐流派的分类方法属于同一构思,在音乐流派的分类装置上可以运行音乐流派的分类方法实施例中提供的任一方法,其具体实现过程详见音乐流派的分类方法实施例,此处不再赘述。

[0118] 本发明实施例提出的音乐流派的分类装置,通过第一收集单元301收集曲库当中音乐的第二歌词文本数据,并对第二歌词文本数据进行预处理,第一训练单元302根据第二歌词文本数据对预设语言模型进行训练,第二收集单元303收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,编码提取单元304根据训练后的预设语言模型提取第一歌词文本数据的语义编码,第二训练单元305将语义编码输入至音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案从古风流派歌曲最为重要的歌词特征入手,对音乐分类模型进行训练,通过训练完成后的分类模型确定待分类音乐的风格流派,可以有效提升音乐流派分类的准确性。

[0119] 本发明实施例还提供一种服务器,如图4所示,其示出了本发明实施例所涉及的服务器的结构示意图,具体来讲:

[0120] 该服务器可以包括一个或者一个以上处理核心的处理器401、一个或一个以上计算机可读存储介质的存储器402、电源403和输入单元404等部件。本领域技术人员可以理解,图4中示出的服务器结构并不构成对服务器的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0121] 其中:

[0122] 处理器401是该电子设备的控制中心,利用各种接口和线路连接整个电子设备的各个部分,通过运行或执行存储在存储器402内的软件程序和/或单元,以及调用存储在存储器402内的数据,执行电子设备的各种功能和处理数据,从而对电子设备进行整体监控。可选的,处理器401可包括一个或多个处理核心;优选的,处理器401可集成应用处理器和调

制解调处理器,其中,应用处理器主要处理操作系统、用户界面和应用程序等,调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器401中。

[0123] 存储器402可用于存储软件程序以及单元,处理器401通过运行存储在存储器402的软件程序以及单元,从而执行各种功能应用以及数据处理。存储器402可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据电子设备的使用所创建的数据等。此外,存储器402可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地,存储器402还可以包括存储器控制器,以提供处理器401对存储器402的访问。

[0124] 服务器还包括给各个部件供电的电源403,优选的,电源403可以通过电源管理系统与处理器401逻辑相连,从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源403还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

[0125] 该服务器还可包括输入单元404,该输入单元404可用于接收输入的数字或字符信息,以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。

[0126] 尽管未示出,服务器还可以包括显示单元等,在此不再赘述。具体在本实施例中,服务器中的处理器401会按照如下的指令,将一个或一个以上的应用程序的进程对应的可执行文件加载到存储器402中,并由处理器401来运行存储在存储器402中的应用程序,从而实现各种功能,如下:

[0127] 收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理;

[0128] 提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理;

[0129] 对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;

[0130] 根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;

[0131] 将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。

[0132] 在一些实施例中,在收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据之前,处理器401运行存储在存储器402中的应用程序,还可以实现如下功能:

[0133] 收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理;

[0134] 根据所述第二歌词文本数据对预设语言模型进行训练。

[0135] 在一些实施例中,在收集曲库当中音乐的第二歌词文本数据,并对所述第二歌词文本数据进行预处理之前,处理器401运行存储在存储器402中的应用程序,还可以实现如下功能:

[0136] 从文本语料库当中提取语句序列;

- [0137] 将所述语句序列输入至预设语言模型以进行预训练。
- [0138] 在一些实施例中,在对所述第二歌词文本数据或第一歌词文本数据进行预处理时,处理器401运行存储在存储器402中的应用程序,还可以实现如下功能:
- [0139] 删除所述第二歌词文本或第一歌词文本中的非标准字符;和/或
- [0140] 对所述第二歌词文本或第一歌词文本进行截取,以保留预设数量的字符。
- [0141] 在一些实施例中,在根据所述第二歌词文本数据对预设语言模型进行训练时,处理器401运行存储在存储器402中的应用程序,还可以实现如下功能:
- [0142] 根据所述第二歌词文本数据计算所述预设语言模型损失函数的损失值;
- [0143] 根据所述损失函数的损失值,调整所述预设语言模型中的模型参数。
- [0144] 在一些实施例中,对所述音频数据进行预处理时,处理器401运行存储在存储器402中的应用程序,还可以实现如下功能:
- [0145] 在所述正负样本中选取所述音频数据满足预设时长的目标音频数据;和/或
- [0146] 提取所述音频数据的第二音频特征,并对所述第二音频特征进行放大处理。
- [0147] 以上各个操作的具体实施可参见前面的实施例,在此不再赘述。
- [0148] 本领域普通技术人员可以理解,上述实施例的各种方法中的全部或部分步骤可以通过指令来完成,或通过指令控制相关的硬件来完成,该指令可以存储于一计算机可读存储介质中,并由处理器进行加载和执行。
- [0149] 由上所述,本发明实施例提出的服务器,通过收集预设流派音乐的正负样本,提取正负样本中音乐的第一歌词文本数据,并对第一歌词文本数据进行预处理,提取正负样本中音乐的音频数据,并对音频数据进行预处理,对音频数据进行压缩,并提取压缩后的音频数据中的第一音频特征,根据训练后的预设语言模型提取第一歌词文本数据的语义编码,将语义编码与第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。本申请实施例所提供的方案从古风流派歌曲最为重要的歌词特征入手,对音乐分类模型进行训练,通过训练完成后的分类模型确定待分类音乐的流派,提升了音乐流派分类的准确性。
- [0150] 另外,本发明实施例还提供一种存储介质,其中存储有多条指令,该指令能够被处理器进行加载,以执行本发明实施例所提供的任一种音乐流派的分类方法中。例如,该指令可以执行:
- [0151] 收集预设流派音乐的正负样本,提取所述正负样本中音乐的第一歌词文本数据,并对所述第一歌词文本数据进行预处理;
- [0152] 提取所述正负样本中音乐的音频数据,并对所述音频数据进行预处理;
- [0153] 对所述音频数据进行压缩,并提取所述压缩后的音频数据中的第一音频特征;
- [0154] 根据所述训练后的预设语言模型提取所述第一歌词文本数据的语义编码;
- [0155] 将所述语义编码与所述第一音频特征进行结合,并将结合后的数据输入至基于语言模型生成的音乐分类模型以进行训练,并根据训练后的音乐分类模型确定待分类音乐的音乐流派。
- [0156] 以上操作的具体实施可参见前面的实施例,在此不再赘述。
- [0157] 其中,该存储介质可以包括:只读存储器(ROM,Read Only Memory)、随机存取记忆体(RAM,Random Access Memory)、磁盘或光盘等。

[0158] 由于该存储介质中所存储的指令,可以执行本发明实施例所提供的任一种音乐流派的分类方法,因此,可以实现本发明实施例所提供的任一种音乐流派的分类方法所能实现的有益效果,详见前面的实施例,在此不再赘述。以上对本发明实施例所提供的一种音乐流派的分类方法、装置及存储介质进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上,本说明书内容不应理解为对本发明的限制。



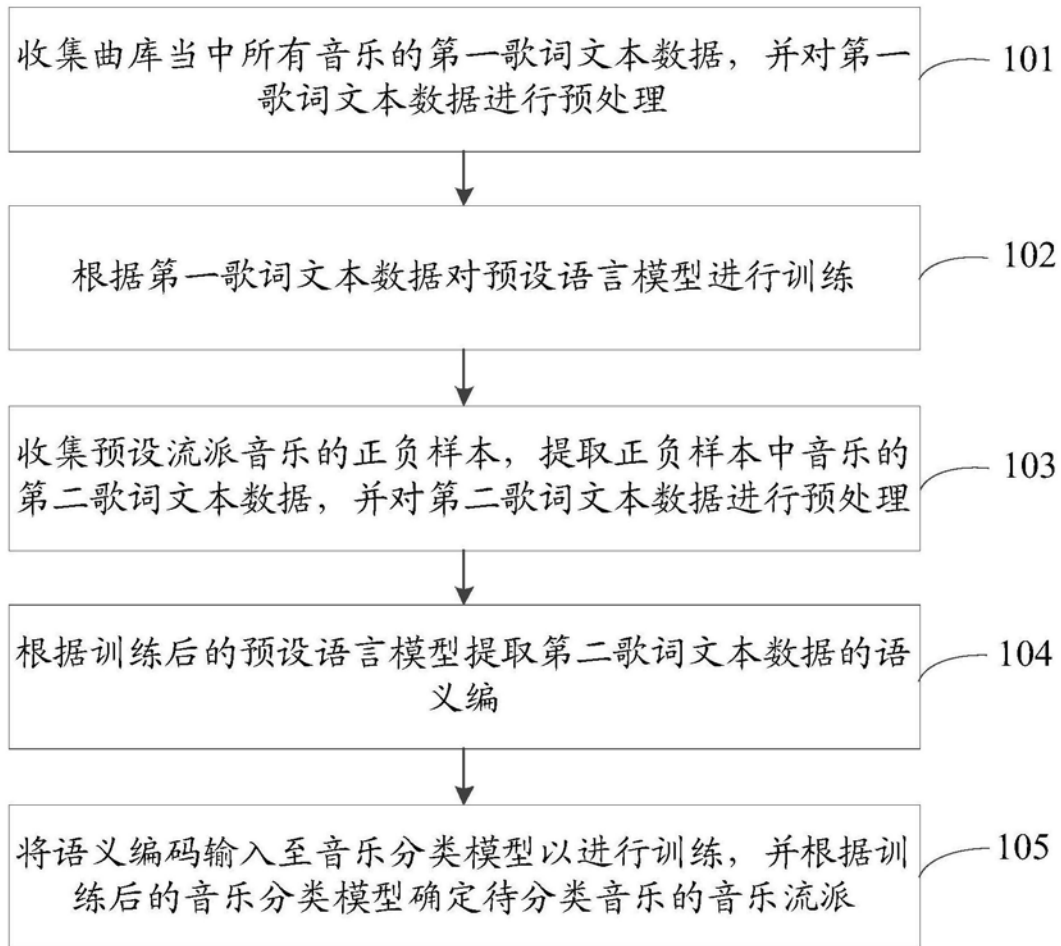


图1a

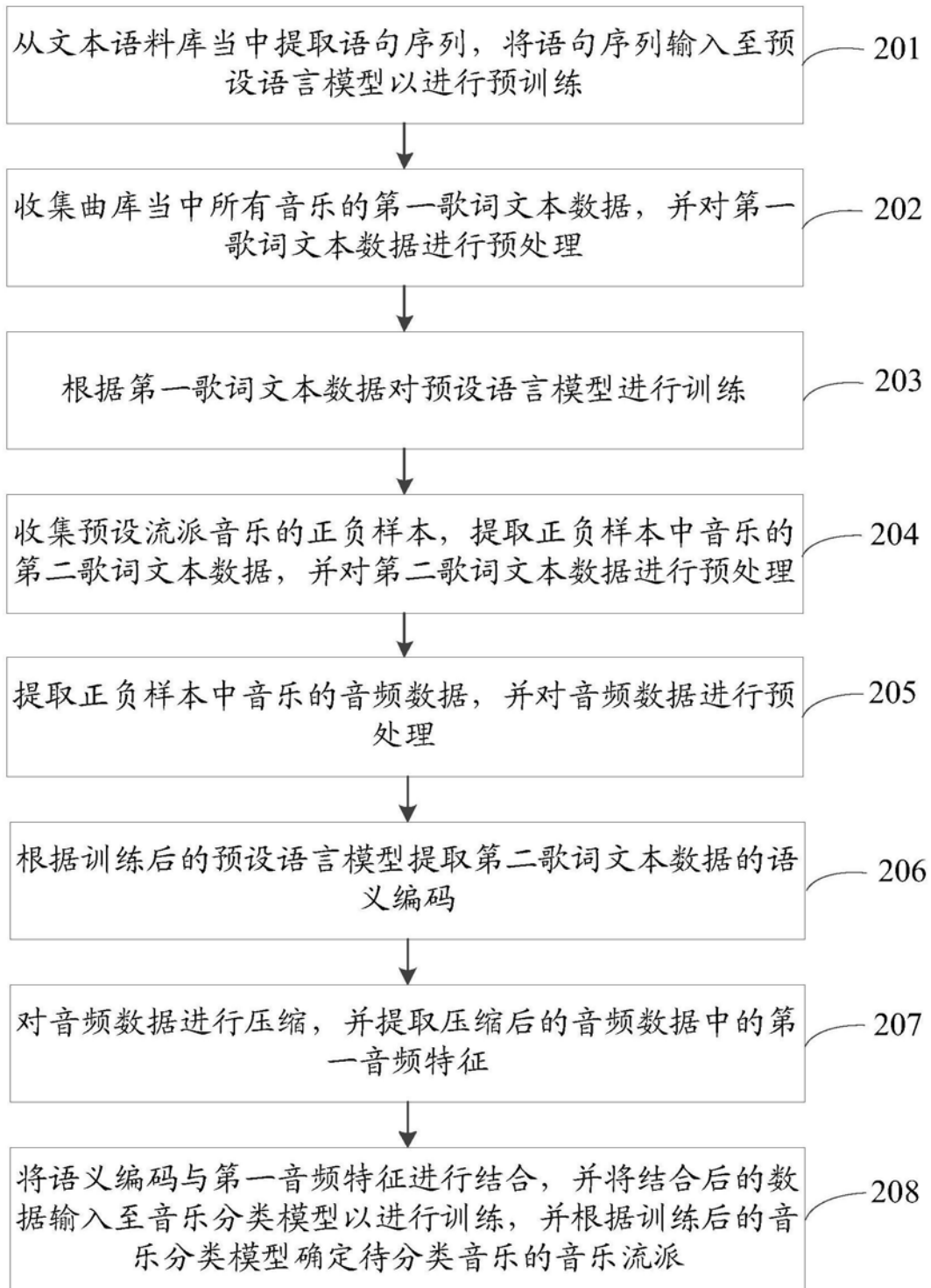


图1b

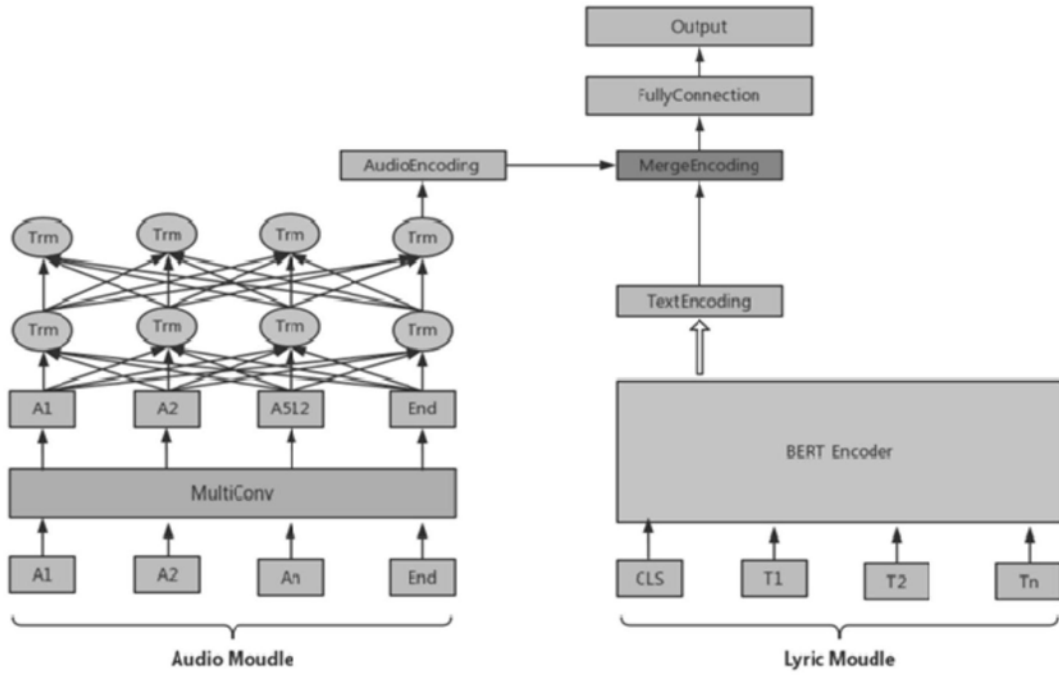


图2



图3a

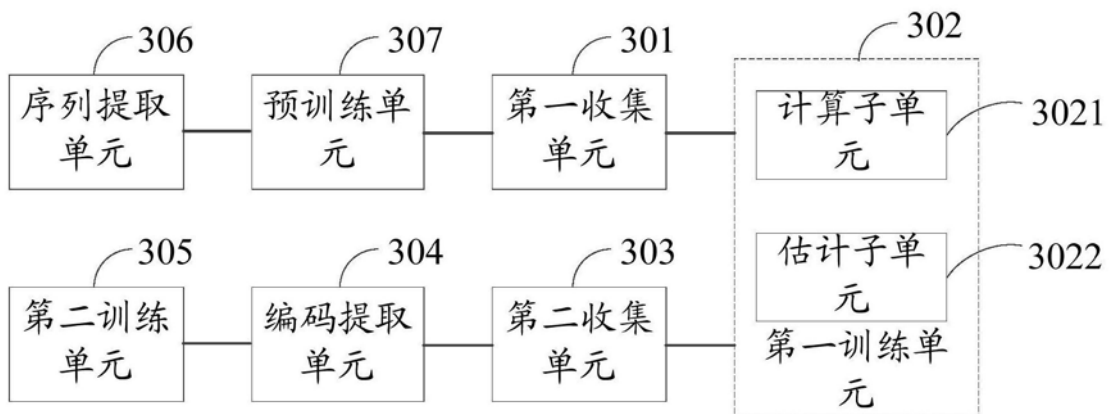


图3b

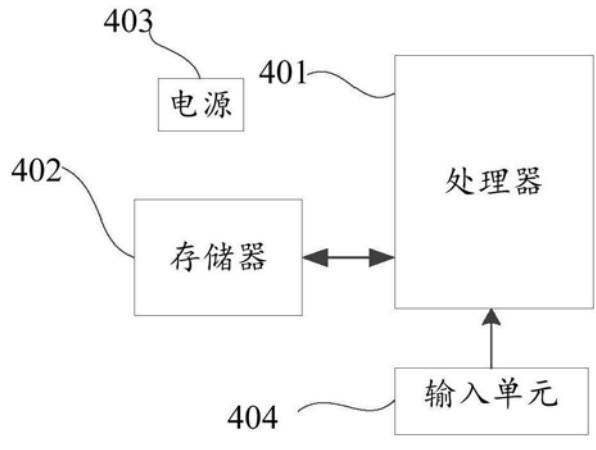


图4