

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(43) 国际公布日
2020 年 10 月 22 日 (22.10.2020)

(10) 国际公布号
WO 2020/211237 A1

(51) 国际专利分类号:
G06F 16/683 (2019.01)

安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(21) 国际申请号: PCT/CN2019/102189

(22) 国际申请日: 2019 年 8 月 23 日 (23.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910307611.3 2019 年 4 月 17 日 (17.04.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福

(72) 发明人: 曹靖康(CAO, Jingkang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。王义文(WANG, Yiwu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理事务所(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL

(54) Title: NEURAL NETWORK-BASED METHOD AND APPARATUS FOR GENERATING RHYTHM FROM LYRICS, AND STORAGE MEDIUM

(54) 发明名称: 基于神经网络的以词生成节奏的方法、装置及存储介质

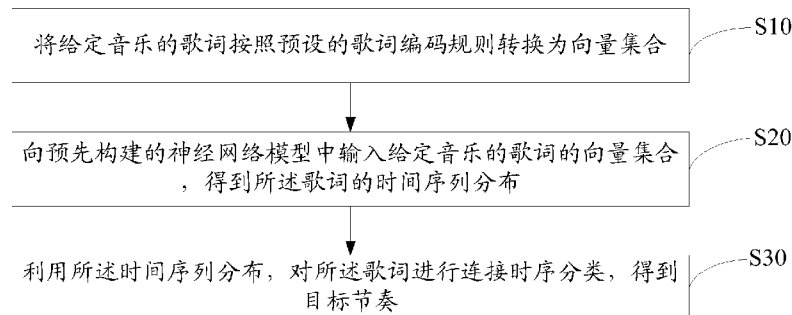


图 1

S10 Convert lyrics of given music into a vector set according to a preset lyrics encoding rule
S20 Input the vector set of the lyrics of the given music into a pre-constructed neural network model to obtain temporal sequence distribution of the lyrics
S30 Perform connectionist temporal classification on the lyrics by means of the temporal sequence distribution to obtain target rhythm

(57) Abstract: A neural network-based method for generating rhythm from lyrics, relating to the field of artificial intelligence. The method comprises: converting lyrics of given music into a vector set according to a preset lyrics encoding rule (S10); inputting the vector set of the lyrics of the given music into a pre-constructed neural network model to obtain temporal sequence distribution of the lyrics (S20); and performing connectionist temporal classification on the lyrics by means of the temporal sequence distribution to obtain target rhythm (S30). Further provided are a neural network-based apparatus for generating rhythm from lyrics, and a computer readable storage medium. The application of a deep learning network to the generation of music rhythm can obtain a reliable result, making the generated music more comply with the specification of the original music.

(57) 摘要: 一种基于神经网络的以词生成节奏方法, 涉及人工智能领域, 该方法包括: 将给定音乐的歌词按照预设的歌词编码规则转换为向量集合(S10); 向预先构建的神经网络模型中输入给定音乐的歌词的向量集合, 得到所述歌词的时间序列分布(S20); 利用所述时间序列分布, 对所述歌词进行连接时序分类, 得到目标节奏(S30)。还提出一种基于神经网络的以词生成节奏装置以及一种计算机可读存储介质。将深度学习网络应用到音乐节奏的生成并可以得到可靠结果, 使其生成的音乐更加符合原有音乐的规范。

PARTNERSHIP)); 中国广东省深圳市福田区
园岭街道八卦四路10号中浩大厦1528-1530
室于志光, Guangdong 518000 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,
JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,
LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

基于神经网络的以词生成节奏的方法、装置及存储介质

本申请基于巴黎公约申明享有 2019 年 4 月 17 日递交的申请号为 CN201910307611.3、名称为“基于神经网络的以词生成节奏方法、装置及存储介质”的中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请涉及人工智能技术领域，尤其涉及一种基于神经网络的以词生成节奏方法、装置及计算机可读存储介质。

背景技术

音乐节奏是音乐自动生成算法中重要的一环，音乐节奏能够规范歌词的分布，同时能够约束音高、旋律等，是连接歌词和音乐的桥梁。传统的语音识别和音乐模型的构建，都是采用状态建模，一个音素或者一个字被人为的分成多个没有物理意义的状态，然后采用离散或者连续高斯模型描述每个状态的输出分布。这种建模方式需要预先对连续序列中间的建模单元的边界进行切分，并且输入输出分布边缘对齐，计算速度慢。

人们对于将深度神经网络与各个领域相结合做出了很多努力，在音乐生成方面，概率生成算法、马尔科夫链能够较准确的生成原有的音乐节奏，但生成的音乐节奏的旋律过于简单；长短期记忆网络(Long Short-Term Memory , LSTM)的结构模型过于复杂，模型的训练时间长；而循环神经网络(Recurrent Neural Network , RNN)在处理距离较远的序列时，很有可能会出现梯度消失的问题。因此，如何将深度学习网络应用到音乐节奏的生成并得到可靠结果使其生成的音乐更加符合原有音乐的规范，且使系统具备稳定的鲁棒性，是一个亟需解决的问题。

发明内容

本申请提供一种基于神经网络的以词生成节奏方法、装置及计算机可读存储介质，其主要目的在于提供一种将深度学习网络应用到音乐节奏的生成

的技术方案。

为实现上述目的，本申请提供的一种基于神经网络的以词生成节奏方法，包括：

5 将给定音乐的歌词按照预设的歌词编码规则转换为向量集合，其中，所述预设的歌词编码规则包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充；

预先构建神经网络模型，其中，预先构建的所述神经网络模型包含三层时空卷积网络以及一层双向门控循环单元；

10 向所述三层时空卷积网络中输入给定音乐的歌词的向量集合，提取出特征向量；

利用所述双向门控循环单元对所述特征向量进行聚合操作，得到时间步长；及

对每一个时间步长进行线性变换，得到时间序列分布；

利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏。

15 此外，为实现上述目的，本申请还提供一种基于神经网络的以词生成节奏装置，所述装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的以词生成节奏程序，所述以词生成节奏程序被所述处理器执行时实现一种基于神经网络的以词生成节奏的方法，所述方法包括：

20 将给定音乐的歌词按照预设的歌词编码规则转换为向量集合，其中，所述预设的歌词编码规则包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充；

预先构建神经网络模型，其中，预先构建的所述神经网络模型包含三层时空卷积网络以及一层双向门控循环单元；

25 向所述三层时空卷积网络中输入给定音乐的歌词的向量集合，提取出特征向量；

利用所述双向门控循环单元对上所述特征向量进行聚合操作，得到时间步长；及

对每一个时间步长进行线性变换，得到时间序列分布；

利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏。

30 此外，为实现上述目的，本申请还提供一种计算机可读存储介质，所述

计算机可读存储介质上存储有以词生成节奏程序，所述的以词生成节奏程序可被一个或者多个处理器执行，以实现如上所述的基于神经网络的以词生成节奏方法的步骤。

5 本申请提出的基于神经网络的以词生成节奏方法、装置及计算机可读存储介质，将给定音乐的歌词按照预设的歌词编码规则转换为向量集合；向预先构建的神经网络模型中输入给定音乐的歌词的向量集合，得到所述歌词的时间序列分布；利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏。因此，本申请将深度学习网络应用到音乐节奏的生成并可以得到可靠结果使其生成的音乐更加符合原有音乐的规范。

10

附图说明

图1为本申请一实施例提供的基于神经网络的以词生成节奏方法的流程示意图；

15 图2为本申请一实施例提供的基于神经网络的以词生成节奏方法中所述的双向门控循环单元的示意图；

图3为本申请一实施例提供的基于神经网络的以词生成节奏方法中数据流的示意图；

20 图4为本申请一实施例提供的基于神经网络的以词生成节奏装置的内部结构示意图；

图5为本申请一实施例提供的基于神经网络的以词生成节奏装置中以词生成节奏程序的模块示意图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

25

具体实施方式

应当理解，此处所描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供一种基于神经网络的以词生成节奏方法。参照图 1 所示，为本申请一实施例提供的基于神经网络的以词生成节奏方法的流程示意图。该

30

方法可以由一个装置执行，该装置可以由软件和/或硬件实现。

在本实施例中，所述基于神经网络的以词生成节奏方法包括：

S10、将给定音乐的歌词按照预设的歌词编码规则转换为向量集合。

本申请较佳实施例中，所述预设的歌词编码规则包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充。

本申请较佳实施例中，生成的向量形式为 $x_i = [\text{time}, 1, 1, \text{channel}]$ 。其中，“time”为歌词出现在音乐中的时间的 BCD 码（Binary-Coded Decimal，亦称二进码十进数或二-十进制代码）；“1”引申指的是图像的高和宽，在音乐中一个字符相当于一个像素，所以设置宽高均为 1；“channel”为上述的歌词编码，如上所述单个字符的 channel 值为 1，单个标点符号的 channel 值为 0 等。于是，给定音乐中的歌词可以转换为向量集合为 $X = \{x_1, \dots, x_i, \dots, x_t\}$ 。

S20、向预先构建的神经网络模型中输入给定音乐的歌词的向量集合，得到所述歌词的时间序列分布。

本申请所述预先构建的神经网络模型包含三层时空卷积网络（SpatioTemporal convolutional neural networks，STCNNs）以及一层双向门控循环单元（Bi-GRU，Bidirectional Gated Recurrent Unit）。

卷积神经网络（Convolutional Neural Networks，CNNs）是一种前馈神经网络，其可以在图像空间上进行卷积堆叠操作，有助于提高计算机视觉任务的性能。而所述时空卷积网络 STCNNs 可以通过在时间和空间维度上进行卷积运算从而可以对音视频数据进行处理。

本申请所述时空卷积网络 STCNNs 的每一层从输入到输出的计算方式为：

$$y = \sigma \left(\sum_k \sum_j \sum_i \left(\overrightarrow{x_{ijk}} * \bar{w} + b \right) \right),$$

其中， y 表示某一层的输出， σ 表示激活函数， i, j, k 表示样本上对应位置的坐标， $\overrightarrow{x_{ijk}}$ 表示每一层输入到对应于上述 (i, j, k) 处与对应卷积核尺寸大小相等的局部区域， $\bar{w}$ 表示卷积核的权重矩阵， b 表示对应卷积核的偏置值。

本申请较佳实施例中，三层的 STCNNs 卷积核形状为 $\begin{bmatrix} 9 & 1 & 1 & 8 \\ 9 & 1 & 1 & 16 \\ 9 & 1 & 1 & 32 \end{bmatrix}$ ，四个维度分别是时间、高、宽和特征数。

在经过训练后，向所述三层时空卷积网络中输入给定音乐的歌词的向量

集合 X 后可以提取出特征向量 z 。

进一步地，本申请利用 Bi-GRU 进一步对 STCNNs 提取到的特征向量 z 进行进一步聚合，得到时间步长。

本申请较佳实施例在 STCNNs 后连接一层双向门控循环单元 (Bi-GRU)。

5 GRU 是一种循环神经网络 (RNN) 的一种变体，GRU 的重复单元模型如下，它有两个门，分别为更新门 u_t 和重置门 r_t 。更新门用于控制前一时刻的状态信息被带入到当前状态中的程度，更新门的值越大说明前一时刻的状态信息带入越多。重置门用于控制忽略前一时刻的状态信息的程度，重置门的值越小说明忽略得越多。而 Bi-GRU 主要特点是增加了对未来的学习能力，克服了仅
10 能处理历史信息的缺陷。Bi-GRU 将一个普通 GRU 拆成两个方向，一个按照序列顺序向前，一个逆时序反向，但是两个 GRU 连接同一个输入层和输出层，结构如图 2 所示。本申请较佳实施例中，所述 Bi-GRU 的神经元数为 256 个。

本申请所述 Bi-GRU 采用如下公式得到时间步长：

$$\begin{aligned} 15 \quad & r_t = \sigma(W_r \cdot [h_{t-1}, z]) ; \\ & u_t = \sigma(W_u \cdot [h_{t-1}, z]) ; \\ & \tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]) ; \\ & h_t = (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t , \end{aligned}$$

其中： u_t 和 r_t 分别为更新门和重置门， $[]$ 表示两个向量相连接， $*$ 表示矩阵
20 元素相乘， σ 为 sigmoid 函数， $z = \{z_1, \dots, z_t\}$ 为 Bi-GRU 的输入即 STCNNs 的输出特征， W_r 和 W_u 分别为重置门和更新门的权重， $\tilde{h}_t$ 表示 t 时刻的候选状态， $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值， h_t 为 t 时刻的输出状态。

Bi-GRU 两个方向的映射分别为：

$$\begin{aligned} 25 \quad & \{z_1, \dots, z_t\} \rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\} ; \\ & \{z_t, \dots, z_1\} \rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\} , \end{aligned}$$

由此得到 t 时刻的时间步长为：

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}]。$$

进一步地，本申请对每一个时间步长 h_t 进行线性变换，得到时间序列分布。

为了参数化序列分布，本申请对于对每一个时间步 t ，使得 $p(u_t|z) =$
30 $\text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$ ，其中 softmax 是一种归一化指数函数， mlp 是含有权重 W_{mlp} 的前馈神经网络，然后定义时间序列分布：

$$p(u_t, \dots, u_T|z) = \prod_{1 \leq t \leq T} p(u_t|z) ,$$

在该模型中 z 为 GRU 的输入，即 STCNNs 的输出。即在输入为 z 时，将

t 时刻的输出状态进行反向传输，得到对每个 t 时刻状态的分类。最终根据定义得到所有时间步长个数 T（即为 z 的向量长度）的时间序列分布 p。

S30、利用所述时间序列分布，对所述歌词进行连接时序分类（Connectionist temporal classification，CTC），得到目标节奏。

5 所述 CTC 是为 RNN 专门设计的顶层（top layer），用于序列学习所用，它可以消除输入与目标输出对齐的步骤。

本申请较佳实施例中，对所述歌词进行 CTC 的主要流程如下：

1) 在给定音乐的歌词 V 的基础上加上空白标签得到字符串 $\tilde{V}$ ，即 $\tilde{V} = V \cup \{blank\}$ ；

10 2) 定义函数 $B: \tilde{V} \rightarrow V^*$ ，其中， V^* 为 $\tilde{V}$ 执行下述操作获得 1) 合并连续的相同符号；2) 去掉空白字符；

对于一个字符串序列 $y \in V^*$ ，定义：

$$p(y|z) = \sum_{u \in B^{-1}(y).s.t.|u|=T} p(u_t, \dots, u_T|z)$$

15 其中， V^* 的所有元素被称为路径， V^* 是所有路径的集合， $p(y|z)$ 代表目标歌词集合 V 对应路径的概率之和，z 为所述三层时空卷积网络输出的特征向量，T 为所有时间步长个数， $p(u_t, \dots, u_T|z)$ 为所有时间步长个数 T 的时间序列分布，s.t. $|u| = T$ 是一个条件函数，表述需要满足 u 是在所有时间步长 T 之中这个条件， $B^{-1}(y).s.t.|u|=T$ 表示长度为 T 且示经过函数 B 变换结果为字符串 y 的集合；

20 3) 根据输入的特征向量 z，计算最大概率和，得到输入序列对应的目标歌词序列 h(x)，即为目标歌词在给定音乐下生成的节奏：

$$h(x) = \arg \max_{y_{1 \leq y \leq T}} p(y|z)。$$

25 综上所述，参阅图 3 所示，本申请的数据流如下：对于一段音乐，本申请较佳实施例将音乐中的歌词转换为向量，并将其传输到该构建的神经网络中得到时间序列，其中，本申请所述神经网络包括三层时空卷积网络以及一层双向门控循环单元；将得到的时间序列输入到连接时序分类当中，同时输入目标歌词，最终得到对应的目标歌词的序列则为这段音乐对应的目标歌词的节奏。

本申请还提供一种基于神经网络的以词生成节奏装置。参照图 4 所示，

为本申请一实施例提供的基于神经网络的以词生成节奏装置的内部结构示意图。

在本实施例中，基于神经网络的以词生成节奏装置 1 可以是 PC (Personal Computer，个人电脑)，也可以是智能手机、平板电脑、便携计算机等终端设备。该基于神经网络的以词生成节奏装置 1 至少包括存储器 11、处理器 12，通信总线 13，以及网络接口 14。

其中，存储器 11 至少包括一种类型的可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）磁性存储器、磁盘、光盘等。存储器 11 在一些实施例中可以是基于神经网络的以词生成节奏装置 1 的内部存储单元，例如该基于神经网络的以词生成节奏装置 1 的硬盘。存储器 11 在另一些实施例中也可以是基于神经网络的以词生成节奏装置 1 的外部存储设备，例如基于神经网络的以词生成节奏装置 1 上配备的插接式硬盘，智能存储卡 (Smart Media Card, SMC)，安全数字 (Secure Digital, SD) 卡，闪存卡 (Flash Card) 等。进一步地，存储器 11 还可以既包括以词生成节奏装置 1 的内部存储单元也包括外部存储设备。存储器 11 不仅可以用于存储安装于基于神经网络的以词生成节奏装置 1 的应用软件及各类数据，例如以词生成节奏程序 01 的代码等，还可以用于暂时地存储已经输出或者将要输出的数据。

处理器 12 在一些实施例中可以是一中央处理器 (Central Processing Unit, CPU) 控制器、微控制器、微处理器或其他数据处理芯片，用于运行存储器 11 中存储的程序代码或处理数据，例如执行所述以词生成节奏程序 01 等。

通信总线 13 用于实现这些组件之间的连接通信。

网络接口 14 可选的可以包括标准的有线接口、无线接口 (如 WI-FI 接口)，通常用于在该装置 1 与其他电子设备之间建立通信连接。

可选地，该装置 1 还可以包括用户接口，用户接口可以包括显示器 (Display) 输入单元比如键盘 (Keyboard)，可选的用户接口还可以包括标准的有线接口、无线接口。可选地，在一些实施例中，显示器可以是 LED 显示器、液晶显示器、触控式液晶显示器以及 OLED (Organic Light-Emitting Diode，有机发光二极管) 触摸器等。其中，显示器也可以适当的称为显示屏或显示单元，用于显示在基于神经网络的以词生成节奏装置 1 中处理的信息

以及用于显示可视化的用户界面。

图 4 仅示出了具有组件 11-14 以及以词生成节奏程序 01 的基于神经网络的以词生成节奏装置 1，本领域技术人员可以理解的是，图 4 示出的结构并不构成对基于神经网络的以词生成节奏装置 1 的限定，可以包括比图示更少或者更多的部件，或者组合某些部件，或者不同的部件布置。

在图 4 所示的装置 1 实施例中，存储器 11 中存储有以词生成节奏程序 01；处理器 12 执行存储器 11 中存储的以词生成节奏程序 01 时实现如下步骤：

步骤一、将给定音乐的歌词按照预设的歌词编码规则转换为向量集合。

本申请较佳实施例中，所述预设的歌词编码规则包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充。

本申请较佳实施例中，生成的向量形式为 $x_i = [\text{time}, 1, 1, \text{channel}]$ 。其中，“time”为歌词出现在音乐中的时间的 BCD 码（Binary-Coded Decimal，亦称二进制十进数或二-十进制代码）；“1”引申指的是图像的高和宽，在音乐中一个字符相当于一个像素，所以设置宽高均为 1；“channel”为上述的歌词编码，如上所述单个字符的 channel 值为 1，单个标点符号的 channel 值为 0 等。于是，给定音乐中的歌词可以转换为向量集合为 $X = \{x_1, \dots, x_i, \dots, x_t\}$ 。

步骤二、向预先构建的神经网络模型中输入给定音乐的歌词的向量集合，得到所述歌词的时间序列分布。

本申请所述预先构建的神经网络模型包含三层时空卷积网络（SpatioTemporal convolutional neural networks，STCNNs）以及一层双向门控循环单元（Bi-GRU，Bidirectional Gated Recurrent Unit）。

卷积神经网络（Convolutional Neural Networks，CNNs）是一种前馈神经网络，其可以在图像空间上进行卷积堆叠操作，有助于提高计算机视觉任务的性能。而所述时空卷积网络 STCNNs 可以通过在时间和空间维度上进行卷积运算从而可以对音视频数据进行处理。

本申请所述时空卷积网络 STCNNs 的每一层从输入到输出的计算方式为：

$$y = \sigma \left(\sum_k \sum_j \sum_i \left(\overrightarrow{x_{ijk}} * \bar{w} + b \right) \right),$$

其中， y 表示某一层的输出， σ 表示激活函数， i, j, k 表示样本上对应位置的坐标， $\overrightarrow{x_{ijk}}$ 表示每一层输入到对应于上述 (i, j, k) 处与对应卷积核尺寸大小相等的局部区域， $\bar{w}$ 表示卷积核的权重矩阵， b 表示对应卷积核的偏置值。

本申请较佳实施例中,三层的 STCNNs 卷积核形状为 $\begin{bmatrix} 9 & 1 & 1 & 8 \\ 9 & 1 & 1 & 16 \\ 9 & 1 & 1 & 32 \end{bmatrix}$, 四个维度分别是时间、高、宽和特征数。

在经过训练后,向所述三层时空卷积网络中输入给定音乐的歌词的向量集合 X 后可以提取出特征向量 z。

进一步地,本申请利用 Bi-GRU 进一步对 STCNNs 提取到的特征向量 z 进行进一步聚合,得到时间步长。

本申请较佳实施例在 STCNNs 后连接一层双向门控循环单元(Bi-GRU)。GRU 是一种循环神经网络(RNN)的一种变体,GRU 的重复单元模型如下,它有两个门,分别为更新门 u_t 和重置门 r_t 。更新门用于控制前一时刻的状态信息被带入到当前状态中的程度,更新门的值越大说明前一时刻的状态信息带入越多。重置门用于控制忽略前一时刻的状态信息的程度,重置门的值越小说明忽略得越多。而 Bi-GRU 主要特点是增加了对未来的学习能力,克服了仅能处理历史信息的缺陷。Bi-GRU 将一个普通 GRU 拆成两个方向,一个按照序列顺序向前,一个逆时序反向,但是两个 GRU 连接同一个输入层和输出层,结构如图 2 所示。本申请较佳实施例中,所述 Bi-GRU 的神经元数为 256 个。

本申请所述 Bi-GRU 采用如下公式得到时间步长:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, z]); \\ u_t &= \sigma(W_u \cdot [h_{t-1}, z]); \\ \tilde{h}_t &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]); \\ h_t &= (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t, \end{aligned}$$

其中: u_t 和 r_t 分别为更新门和重置门, $[\]$ 表示两个向量相连接,*表示矩阵元素相乘, σ 为 sigmoid 函数, $z = \{z_1, \dots, z_t\}$ 为 Bi-GRU 的输入即 STCNNs 的输出特征, W_r 和 W_u 分别为重置门和更新门的权重, $\tilde{h}_t$ 表示 t 时刻的候选状态, $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值, h_t 为 t 时刻的输出状态。

Bi-GRU 两个方向的映射分别为:

$$\begin{aligned} \{z_1, \dots, z_t\} &\rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\}; \\ \{z_t, \dots, z_1\} &\rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\}, \end{aligned}$$

由此得到 t 时刻的时间步长为:

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}].$$

进一步地,本申请对每一个时间步长 h_t 进行线性变换,得到时间序列分布。

为了参数化序列分布，本申请对于对每一个时间步 t ，使得 $p(u_t|z) = \text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$ ，其中 softmax 是一种归一化指数函数， mlp 是含有权重 W_{mlp} 的前馈神经网络，然后定义时间序列分布：

$$p(u_t, \dots, u_T|z) = \prod_{1 \leq t \leq T} p(u_t|z),$$

5 在该模型中 z 为 GRU 的输入，即 STCNNs 的输出。即在输入为 z 时，将 t 时刻的输出状态进行反向传输，得到对每个 t 时刻状态的分类。最终根据定义得到所有时间步长个数 T （即为 z 的向量长度）的时间序列分布 p 。

步骤三、利用所述时间序列分布，对所述歌词进行连接时序分类（Connectionist temporal classification，CTC），得到目标节奏。

10 所述 CTC 是为 RNN 专门设计的顶层（top layer），用于序列学习所用，它可以消除输入与目标输出对齐的步骤。

本申请较佳实施例中，对所述歌词进行 CTC 的主要流程如下：

1) 在给定音乐的歌词 V 的基础上加上空白标签得到字符串 $\tilde{V}$ ，即 $\tilde{V} = V \cup \{\text{blank}\}$ ；

15 2) 定义函数 $B: \tilde{V} \rightarrow V^*$ ，其中， V^* 为 $\tilde{V}$ 执行下述操作获得 1) 合并连续的相同符号；2) 去掉空白字符；

对于一个字符串序列 $y \in V^*$ ，定义：

$$p(y|z) = \sum_{u \in B^{-1}(y) \text{ s.t. } |u|=T} p(u_t, \dots, u_T|z)$$

其中， V^* 的所有元素被称为路径， V^* 是所有路径的集合， $p(y|z)$ 代表目标歌词集合 V 对应路径的概率之和， z 为所述三层时空卷积网络输出的特征向量， T 为所有时间步长个数， $p(u_t, \dots, u_T|z)$ 为所有时间步长个数 T 的时间序列分布， $\text{s.t. } |u|=T$ 是一个条件函数，表述需要满足 u 是在所有时间步长 T 之中这个条件， $B^{-1}(y) \text{ s.t. } |u|=T$ 表示长度为 T 且示经过函数 B 变换结果为字符串 y 的集合；

3) 根据输入的特征向量 z ，计算最大概率和，得到输入序列对应的目标歌词序列 $h(x)$ ，即为目标歌词在给定音乐下生成的节奏：

$$h(x) = \arg \max_{y_{1 \leq y \leq T}} p(y|z)。$$

可选地，在其他实施例中，所述的以词生成节奏程序还可以被分割为一个或者多个模块，一个或者多个模块被存储于存储器 11 中，并由一个或多个

处理器（本实施例为处理器 12）所执行以完成本申请，本申请所称的模块是指能够完成特定功能的一系列计算机程序指令段，用于描述以词生成节奏程序在基于神经网络的以词生成节奏装置中的执行过程。

例如，参照图 5 所示，为本申请基于神经网络的以词生成节奏装置一实施例中的以词生成节奏程序的程序模块示意图，该实施例中，以词生成节奏程序 01 可以被分割为歌词转换模块 10、模型计算模块 20、节奏生成模块 30，示例性地：

歌词转换模块 10 用于：将给定音乐的歌词按照预设的规则转换为向量集合。

可选地，所述预设的歌词编码规则，包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充。

模型计算模块 20 用于：向预先构建的神经网络模型中输入给定音乐的歌词的向量集合，得到所述歌词的时间序列分布。

可选地，所述预先构建的神经网络模型包含三层时空卷积网络（SpatioTemporal convolutional neural networks，STCNNs）以及一层双向门控循环单元（Bi-GRU，Bidirectional Gated Recurrent Unit）。

可选地，所述向预先构建的神经网络模型中输入给定音乐的歌词的向量集合 X，得到所述歌词的时间序列分布，包括：

向所述三层时空卷积网络中输入给定音乐的歌词的向量集合 X，提取出特征向量 z；

利用所述双向门控循环单元对上述提取到的特征向量 z 进行聚合操作，得到时间步长；

对每一个时间步长 h_t 进行线性变换，得到时间序列分布。

可选地，所述双向门控循环单元采用如下公式得到时间步长：

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, z]) ; \\ u_t &= \sigma(W_u \cdot [h_{t-1}, z]) ; \\ \tilde{h}_t &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]) ; \\ h_t &= (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t , \end{aligned}$$

其中： u_t 和 r_t 分别为更新门和重置门， $[]$ 表示两个向量相连接， $*$ 表示矩阵元素相乘， σ 为 sigmoid 函数， $z = \{z_1, \dots, z_t\}$ 为 Bi-GRU 的输入即 STCNNs 的输出特征， W_r 和 W_u 分别为重置门和更新门的权重， $\tilde{h}_t$ 表示 t 时刻的候选状

态, $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值, h_t 为t时刻的输出状态。

双向门控循环单元两个方向的映射分别为:

$$\begin{aligned} \{z_1, \dots, z_t\} &\rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\}; \\ \{z_t, \dots, z_1\} &\rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\}, \end{aligned}$$

由此得到t时刻的时间步长为:

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}].$$

可选地, 所述时间序列分布的计算公式为:

$$p(u_t, \dots, u_T | z) = \prod_{1 \leq t \leq T} p(u_t | z),$$

其中, t为时间步, $p(u_t | z) = \text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$, 所述softmax是一种归一化指数函数, mlp是含有权重 W_{mlp} 的前馈神经网络, z为三层时空卷积网络的输出, T为所有时间步长个数。

节奏生成模块30用于: 利用所述时间序列分布, 对所述歌词进行连接时序分类 (Connectionist temporal classification, CTC), 得到目标节奏。

可选地, 所述对所述歌词进行连接时序分类包括:

在给定音乐的歌词V的基础上加上空白标签得到字符串 $\tilde{V}$, 即 $\tilde{V} = V \cup \{\text{blank}\}$;

定义函数 $B: \tilde{V} \rightarrow V^*$, 其中, V^* 为 $\tilde{V}$ 执行下述操作获得1) 合并连续的相同符号; 2) 去掉空白字符;

对于一个字符串序列 $y \in V^*$, 定义:

$$p(y | z) = \sum_{u \in B^{-1}(y) \text{ s.t. } |u|=T} p(u_t, \dots, u_T | z)$$

其中, V^* 的所有元素被称为路径, V^* 是所有路径的集合, $p(y | z)$ 代表目标歌词集合V对应路径的概率之和, z为所述三层时空卷积网络输出的特征向量, T为所有时间步长个数, $p(u_t, \dots, u_T | z)$ 为所有时间步长个数T的时间序列分布, s.t. $|u|=T$ 是一个条件函数, 表述需要满足u是在所有时间步长T之中这个条件, $B^{-1}(y) \text{ s.t. } |u|=T$ 表示长度为T且示经过函数B变换结果为字符串y的集合;

根据输入的特征向量z, 计算最大概率和, 得到输入序列对应的目标歌词序列h(x), 即为目标歌词在给定音乐下生成的节奏:

$$h(x) = \arg \max_{y_{1 \leq y \leq T}} p(y | z).$$

上述歌词转换模块10、模型计算模块20、节奏生成模块30等程序模块

被执行时所实现的功能或操作步骤与上述实施例大体相同，在此不再赘述。

此外，本申请实施例还提出一种计算机可读存储介质，所述计算机可读存储介质上存储有以词生成节奏程序，所述的以词生成节奏程序可被一个或多个处理器执行，以实现如下操作：

将给定音乐的歌词按照预设的规则转换为向量集合；

向预先构建的神经网络模型中输入给定音乐的歌词的向量集合，得到所述歌词的时间序列分布；

利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏。

本申请计算机可读存储介质具体实施方式与上述基于神经网络的以词生成节奏装置和方法各实施例基本相同，在此不作累述。

需要说明的是，上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。并且本文中的术语“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、装置、物品或者方法不仅包括那些要素，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、装置、物品或者方法所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

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上述实施例方法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，但很多情况下前者是更佳的实施方式。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如上所述的一个存储介质(如 ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，或者网络设备等)执行本申请各个实施例所述的方法。

以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权 利 要 求 书

1、一种基于神经网络的以词生成节奏方法,其特征在于,所述方法包括:
将给定音乐的歌词按照预设的歌词编码规则转换为向量集合,其中,所
述预设的歌词编码规则包括:规定歌词中的单个字符为1,单个标点符号为0,
字符之间用0填充;

预先构建神经网络模型,其中,预先构建的所述神经网络模型包含三层
时空卷积网络以及一层双向门控循环单元;

向所述三层时空卷积网络中输入给定音乐的歌词的向量集合,提取出特
征向量;

利用所述双向门控循环单元对所述特征向量进行聚合操作,得到时间步
长;及

对每一个时间步长进行线性变换,得到时间序列分布;

利用所述时间序列分布,对所述歌词进行连接时序分类,得到目标节奏。

2、如权利要求1所述的基于神经网络的以词生成节奏方法,其特征在于,
所述时空卷积网络的每一层的计算方式为:

$$y = \sigma \left(\sum_k \sum_j \sum_i \left(\overrightarrow{x_{ijk}} * \bar{w} + b \right) \right)$$

其中,y表示某一层的输出, σ 表示激活函数,i,j,k表示样本上对应位
置的坐标, $\overrightarrow{x_{ijk}}$ 表示每一层输入到对应于上述(i,j,k)处与对应卷积核尺寸
大小相等的局部区域, $\bar{w}$ 表示卷积核的权重矩阵,b表示对应卷积核的偏置值。

3、如权利要求1所述的基于神经网络的以词生成节奏方法,其特征在于,
所述三层时空卷积网络的卷积核形状为 $\begin{bmatrix} 9 & 1 & 1 & 8 \\ 9 & 1 & 1 & 16 \\ 9 & 1 & 1 & 32 \end{bmatrix}$,四个维度分别是时间、
高、宽和特征数。

4、如权利要求1所述的基于神经网络的以词生成节奏方法,其特征在于,
所述利用所述时间序列分布,对所述歌词进行连接时序分类,得到目标节奏,
包括:

在给定音乐的歌词V的基础上加上空白标签得到字符串 $\tilde{V}$,即
 $\tilde{V} = V \cup \{blank\}$;

定义函数 $B: \tilde{V} \rightarrow V^*$,其中, V^* 为 $\tilde{V}$ 执行下述操作获得:1)合并连续的

相同符号；2) 去掉空白字符；

对于一个字符串序列 $y \in V^*$ ，定义：

$$p(y|z) = \sum_{u \in B^{-1}(y) \text{ s.t. } |u|=T} p(u_t, \dots, u_T|z)$$

其中， V^* 的所有元素被称为路径， V^* 是所有路径的集合， $p(y|z)$ 代表目标歌词集合 V 对应路径的概率之和， z 为所述三层时空卷积网络输出的特征向量， T 为所有时间步长个数， $p(u_t, \dots, u_T|z)$ 为所有时间步长个数 T 的时间序列分布， $s.t. |u| = T$ 是一个条件函数，表述需要满足 u 是在所有时间步长 T 之中这个条件， $B^{-1}(y) \text{ s.t. } |u|=T$ 表示长度为 T 且示经过函数 B 变换结果为字符串 y 的集合；

根据输入的特征向量 z ，计算最大概率和，得到输入序列对应的目标歌词序列 $h(x)$ ，即为目标歌词在给定音乐下生成的节奏：

$$h(x) = \arg \max_{y_{1 < y \leq T}} p(y|z)。$$

5、如权利要求 1 所述的基于神经网络的以词生成节奏方法，其特征在于，所述双向门控循环单元连接在所述三层时空卷积网络后，所述双向门控循环单元的神经元数为 256 个。

6、如权利要求 1 所述的基于神经网络的以词生成节奏方法，其特征在于，所述双向门控循环单元采用如下公式得到时间步长：

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, z]) ; \\ u_t &= \sigma(W_u \cdot [h_{t-1}, z]) ; \\ \tilde{h}_t &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]) ; \\ h_t &= (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t , \end{aligned}$$

其中： u_t 和 r_t 分别为更新门和重置门， $[]$ 表示两个向量相连接， $*$ 表示矩阵元素相乘， σ 为 sigmoid 函数， $z = \{z_1, \dots, z_t\}$ 为所述三层时空卷积网络输出的特征向量， W_r 和 W_u 分别为重置门和更新门的权重， $\tilde{h}_t$ 表示 t 时刻的候选状态， $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值， h_t 为 t 时刻的输出状态，

双向门控循环单元两个方向的映射分别为：

$$\begin{aligned} \{z_1, \dots, z_t\} &\rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\} ; \\ \{z_t, \dots, z_1\} &\rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\} , \end{aligned}$$

由此得到 t 时刻的时间步长为：

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}]。$$

7、如权利要求 1 至 6 中的任意一项所述的基于神经网络的以词生成节奏

方法，其特征在于，所述时间序列分布的计算公式为：

$$p(u_t, \dots, u_T | z) = \prod_{1 \leq t \leq T} p(u_t | z),$$

其中， t 为时间步长， $p(u_t | z) = \text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$ ，所述 softmax 是一种归一化指数函数， mlp 是含有权重 W_{mlp} 的前馈神经网络， z 为三层时空卷积网络输出的特征向量， T 为所有时间步长个数。

8、一种基于神经网络的以词生成节奏装置，其特征在于，所述装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的以词生成节奏程序，所述以词生成节奏程序被所述处理器执行时实现一种基于神经网络的以词生成节奏的方法，所述方法包括：

将给定音乐的歌词按照预设的歌词编码规则转换为向量集合，其中，所述预设的歌词编码规则包括：规定歌词中的单个字符为 1，单个标点符号为 0，字符之间用 0 填充；

预先构建神经网络模型，其中，预先构建的所述神经网络模型包含三层时空卷积网络以及一层双向门控循环单元；

向所述三层时空卷积网络中输入给定音乐的歌词的向量集合，提取出特征向量；

利用所述双向门控循环单元对所述特征向量进行聚合操作，得到时间步长；及

对每一个时间步长进行线性变换，得到时间序列分布；

利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏。

9、如权利要求 8 所述的基于神经网络的以词生成节奏装置，其特征在于，所述时空卷积网络的每一层的计算方式为：

$$y = \sigma \left(\sum_k \sum_j \sum_i \left(\overrightarrow{x_{ijk}} * \bar{w} + b \right) \right),$$

其中， y 表示某一层的输出， σ 表示激活函数， i, j, k 表示样本上对应位置的坐标， $\overrightarrow{x_{ijk}}$ 表示每一层输入到对应于上述 (i, j, k) 处与对应卷积核尺寸大小相等的局部区域， $\bar{w}$ 表示卷积核的权重矩阵， b 表示对应卷积核的偏置值。

10、如权利要求 8 所述的基于神经网络的以词生成节奏装置，其特征在于，所述三层时空卷积网络的卷积核形状为 $\begin{bmatrix} 9 & 1 & 1 & 8 \\ 9 & 1 & 1 & 16 \\ 9 & 1 & 1 & 32 \end{bmatrix}$ ，四个维度分别是时间、高、宽和特征数。

11、如权利要求 8 所述的基于神经网络的以词生成节奏装置，其特征在于，所述利用所述时间序列分布，对所述歌词进行连接时序分类，得到目标节奏，包括：

在给定音乐的歌词 V 的基础上加上空白标签得到字符串 $\tilde{V}$ ，即
5 $\tilde{V} = V \cup \{blank\}$ ；

定义函数 $B: \tilde{V} \rightarrow V^*$ ，其中， V^* 为 $\tilde{V}$ 执行下述操作获得 1) 合并连续的相同符号；2) 去掉空白字符；

对于一个字符串序列 $y \in V^*$ ，定义：

$$p(y|z) = \sum_{u \in B^{-1}(y) \text{ s.t. } |u|=T} p(u_t, \dots, u_T|z)$$

其中， V^* 的所有元素被称为路径， V^* 是所有路径的集合， $p(y|z)$ 代表目标
10 歌词集合 V 对应路径的概率之和， z 为所述三层时空卷积网络输出的特征向量， T 为所有时间步长个数， $p(u_t, \dots, u_T|z)$ 为所有时间步长个数 T 的时间序列分布， $\text{s.t. } |u| = T$ 是一个条件函数，表述需要满足 u 是在所有时间步长 T 之中这个条件， $B^{-1}(y) \text{ s.t. } |u|=T$ 表示长度为 T 且示经过函数 B 变换结果为字符串 y 的集合；

15 根据输入的特征向量 z ，计算最大概率和，得到输入序列对应的目标歌词序列 $h(x)$ ，即为目标歌词在给定音乐下生成的节奏：

$$h(x) = \arg \max_{y_{1 \leq y \leq T}} p(y|z)。$$

12、如权利要求 8 所述的基于神经网络的以词生成节奏装置，其特征在于，所述双向门控循环单元连接在所述三层时空卷积网络后，所述双向门控
20 循环单元的神经元数为 256 个。

13、如权利要求 8 所述的基于神经网络的以词生成节奏装置，其特征在于，所述双向门控循环单元采用如下公式得到时间步长：

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, z]) ; \\ u_t &= \sigma(W_u \cdot [h_{t-1}, z]) ; \\ \tilde{h}_t &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]) ; \\ h_t &= (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t , \end{aligned}$$

25

其中： u_t 和 r_t 分别为更新门和重置门， $[]$ 表示两个向量相连接， $*$ 表示矩阵元素相乘， σ 为 sigmoid 函数， $z = \{z_1, \dots, z_t\}$ 为所述三层时空卷积网络输出的特征向量， W_r 和 W_u 分别为重置门和更新门的权重， $\tilde{h}_t$ 表示 t 时刻的候选

状态, $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值, h_t 为 t 时刻的输出状态,

双向门控循环单元两个方向的映射分别为:

$$\{z_1, \dots, z_t\} \rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\};$$

$$\{z_t, \dots, z_1\} \rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\},$$

5 由此得到 t 时刻的时间步长为:

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}].$$

14、如权利要求 8 至 13 中的任意一项所述的基于神经网络的以词生成节奏装置, 其特征在于, 所述时间序列分布的计算公式为:

$$p(u_t, \dots, u_T | z) = \prod_{1 \leq t \leq T} p(u_t | z),$$

10 其中, t 为时间步长, $p(u_t | z) = \text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$, 所述 softmax 是一种归一化指数函数, mlp 是含有权重 W_{mlp} 的前馈神经网络, z 为三层时空卷积网络输出的特征向量, T 为所有时间步长个数。

15 15、一种计算机可读存储介质, 其特征在于, 所述计算机可读存储介质上存储有以词生成节奏程序, 所述的以词生成节奏程序可被一个或者多个处理器执行, 以实现如下步骤:

将给定音乐的歌词按照预设的歌词编码规则转换为向量集合, 其中, 所述预设的歌词编码规则包括: 规定歌词中的单个字符为 1, 单个标点符号为 0, 字符之间用 0 填充;

20 预先构建神经网络模型, 其中, 预先构建的所述神经网络模型包含三层时空卷积网络以及一层双向门控循环单元;

向所述三层时空卷积网络中输入给定音乐的歌词的向量集合, 提取出特征向量;

利用所述双向门控循环单元对所述特征向量进行聚合操作, 得到时间步长; 及

25 对每一个时间步长进行线性变换, 得到时间序列分布;

利用所述时间序列分布, 对所述歌词进行连接时序分类, 得到目标节奏。

16、如权利要求 15 所述的计算机可读存储介质, 其特征在于, 所述时空卷积网络的每一层的计算方式为:

$$y = \sigma \left(\sum_k \sum_j \sum_i (\overleftarrow{x_{ijk}} * \bar{w} + b) \right),$$

其中, y 表示某一层的输出, σ 表示激活函数, i, j, k 表示样本上对应位

置的坐标, $\overrightarrow{x_{ijk}}$ 表示每一层输入到对应于上述 (i, j, k) 处与对应卷积核尺寸大小相等的局部区域, w 表示卷积核的权重矩阵, b 表示对应卷积核的偏置值。

17、如权利要求 15 所述的计算机可读存储介质, 其特征在于, 所述利用所述时间序列分布, 对所述歌词进行连接时序分类, 得到目标节奏, 包括:

5 在给定音乐的歌词 V 的基础上加上空白标签得到字符串 $\tilde{V}$, 即 $\tilde{V} = V \cup \{blank\}$;

定义函数 $B: \tilde{V} \rightarrow V^*$, 其中, V^* 为 $\tilde{V}$ 执行下述操作获得 1) 合并连续的相同符号; 2) 去掉空白字符;

对于一个字符串序列 $y \in V^*$, 定义:

$$p(y|z) = \sum_{u \in B^{-1}(y) \text{ s.t. } |u|=T} p(u_t, \dots, u_T|z)$$

10 其中, V^* 的所有元素被称为路径, V^* 是所有路径的集合, $p(y|z)$ 代表目标歌词集合 V 对应路径的概率之和, z 为所述三层时空卷积网络输出的特征向量, T 为所有时间步长个数, $p(u_t, \dots, u_T|z)$ 为所有时间步长个数 T 的时间序列分布, $\text{s.t. } |u| = T$ 是一个条件函数, 表述需要满足 u 是在所有时间步长 T 之中这个条件, $B^{-1}(y) \text{ s.t. } |u|=T$ 表示长度为 T 且示经过函数 B 变换结果为字符串 y 的集合;

根据输入的特征向量 z , 计算最大概率和, 得到输入序列对应的目标歌词序列 $h(x)$, 即为目标歌词在给定音乐下生成的节奏:

$$h(x) = \arg \max_{y_{1 \leq y \leq T}} p(y|z)。$$

18、如权利要求 15 所述的计算机可读存储介质, 其特征在于, 所述双向门控循环单元连接在所述三层时空卷积网络后, 所述双向门控循环单元的神经元数为 256 个。

19、如权利要求 15 所述的计算机可读存储介质, 其特征在于, 所述双向门控循环单元采用如下公式得到时间步长:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, z]); \\ u_t &= \sigma(W_u \cdot [h_{t-1}, z]); \\ \tilde{h}_t &= \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, z_t]); \\ h_t &= (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t, \end{aligned}$$

其中: u_t 和 r_t 分别为更新门和重置门, $[]$ 表示两个向量相连接, $*$ 表示矩阵元素相乘, σ 为 sigmoid 函数, $z = \{z_1, \dots, z_t\}$ 为所述三层时空卷积网络输

出的特征向量， W_r 和 W_u 分别为重置门和更新门的权重， $\tilde{h}_t$ 表示 t 时刻的候选状态， $W_{\tilde{h}}$ 表示 $\tilde{h}_t$ 的权值， h_t 为 t 时刻的输出状态，

双向门控循环单元两个方向的映射分别为：

$$\begin{aligned} \{z_1, \dots, z_t\} &\rightarrow \{\overrightarrow{h_1}, \dots, \overrightarrow{h_T}\}; \\ \{z_t, \dots, z_1\} &\rightarrow \{\overleftarrow{h_1}, \dots, \overleftarrow{h_T}\}, \end{aligned}$$

由此得到 t 时刻的时间步长为：

$$h_t = [\overrightarrow{h_T}, \overleftarrow{h_T}]。$$

20、如权利要求 15 至 19 中的任意一项所述的计算机可读存储介质，其特征在于，所述时间序列分布的计算公式为：

$$p(u_t, \dots, u_T | z) = \prod_{1 \leq t \leq T} p(u_t | z),$$

其中 t 为时间步长 $p(u_t | z) = \text{softmax}(\text{mlp}(h_t; W_{\text{mlp}}))$ ，所述 softmax 是一种归一化指数函数， mlp 是含有权重 W_{mlp} 的前馈神经网络， z 为三层时空卷积网络输出的特征向量， T 为所有时间步长个数。

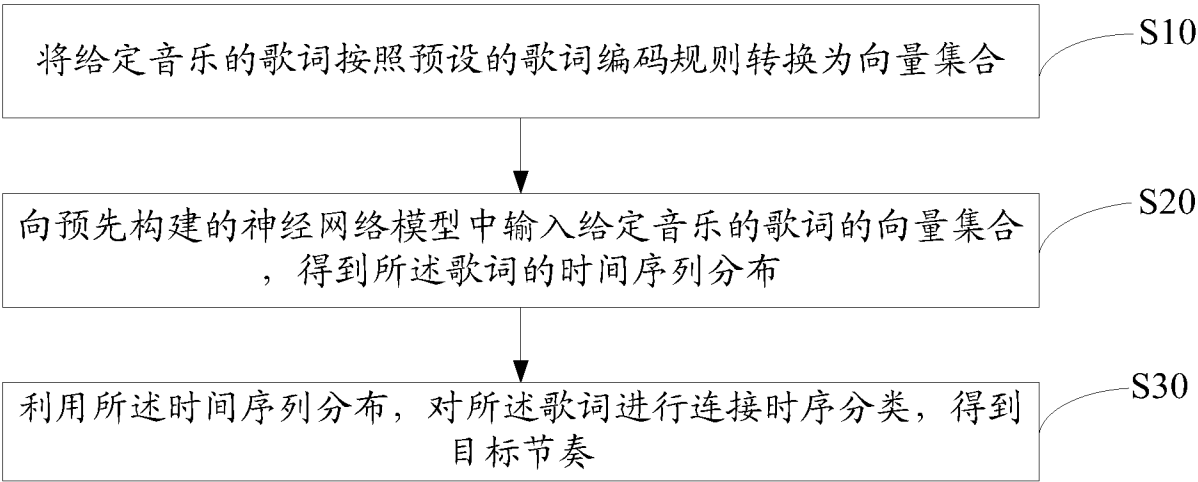


图 1

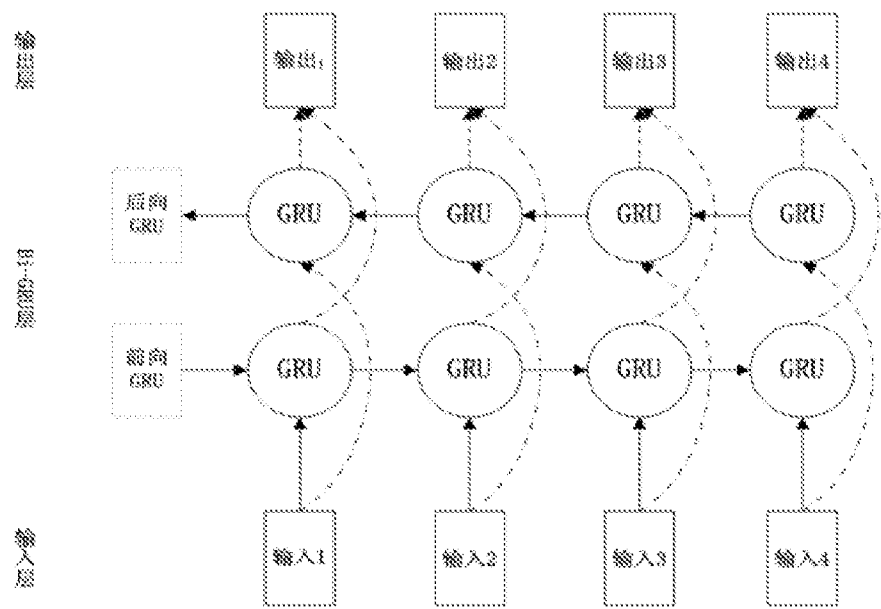


图 2

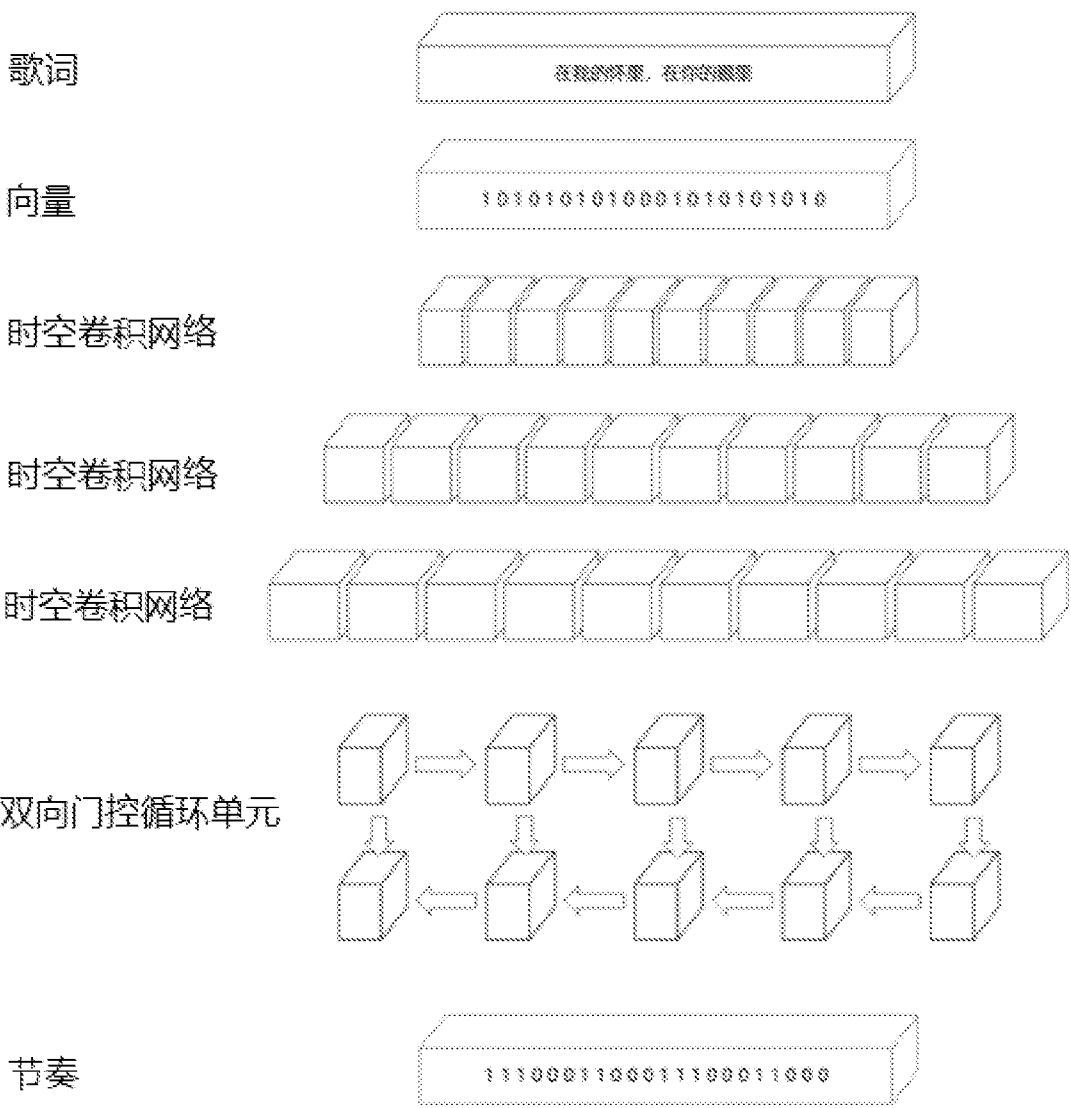


图 3

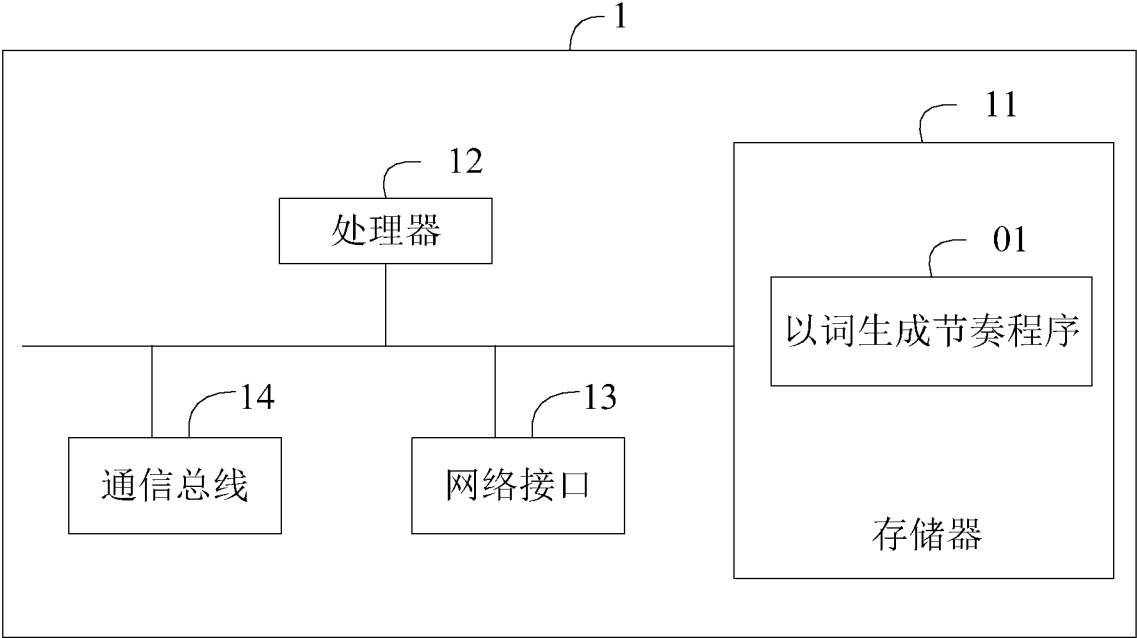


图 4

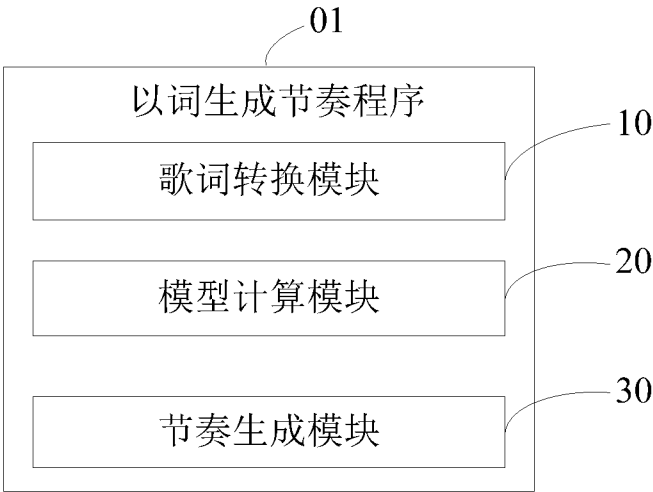


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102189

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/683(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI: 音乐, 歌词, 节奏, 生成, 神经网络, 卷积, 分词, 向量, 时间序列, music, network, vector, time, word, neural, step, sequence, generat+, convolution

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 109346045 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 15 February 2019 (2019-02-15) description, paragraphs [0026]-[0063]	1-20
A	CN 108509534 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 07 September 2018 (2018-09-07) entire document	1-20
A	CN 109637509 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 16 April 2019 (2019-04-16) entire document	1-20
A	US 8874249 B2 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 28 October 2014 (2014-10-28) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

10 January 2020

Date of mailing of the international search report

15 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/102189

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109346045	A	15 February 2019	None			
CN	108509534	A	07 September 2018	None			
CN	109637509	A	16 April 2019	None			
US	8874249	B2	28 October 2014	US	8626328	B2	07 January 2014

国际检索报告

国际申请号

PCT/CN2019/102189

A. 主题的分类

G06F 16/683 (2019. 01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI, EP0DOC, CNPAT, CNKI: 音乐, 歌词, 节奏, 生成, 神经网络, 卷积, 分词, 向量, 时间序列, music, network, vector, time, word, neural, step, sequence, generat+, convolution

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 109346045 A (平安科技深圳有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第[0026]-[0063]段	1-20
A	CN 108509534 A (华南理工大学) 2018年 9月 7日 (2018 - 09 - 07) 全文	1-20
A	CN 109637509 A (平安科技深圳有限公司) 2019年 4月 16日 (2019 - 04 - 16) 全文	1-20
A	US 8874249 B2 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 10月 28日 (2014 - 10 - 28) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 1月 10日

国际检索报告邮寄日期

2020年 1月 15日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

田晶

电话号码 86-(10)-53961334

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/102189

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109346045	A	2019年 2月 15日	无	
CN	108509534	A	2018年 9月 7日	无	
CN	109637509	A	2019年 4月 16日	无	
US	8874249	B2	2014年 10月 28日	US 8626328 B2	2014年 1月 7日