基于主题条件 CNN-BiLSTM 的旋律自动生成方法

曹西征,张航,李伟

(河南师范大学 计算机与信息工程学院;智慧商务与物联网技术河南省工程实验室;
河南省教育人工智能与个性化学习重点实验室,河南 新乡 453007)

摘 要:为了有效地生成结构化的旋律,提出了一种基于主题条件 CNN-BiLSTM 的旋律自动生成方法.将旋律表示为钢琴卷帘窗的形式,使用定长、变长相结合的方法分割钢琴卷帘窗;通过 Ward 聚类算法对钢琴卷帘窗片段进行聚类分析,将获取的最大簇作为歌曲的旋律主题;以旋律主题作为条件使用基于 CNN-BiLSTM 结构的模型进行旋律生成,其上半部分 CNN 可以有效地提取钢琴卷帘窗中所包含时间和音高之间的信息,下半部分利用 LSTM 和 BiLSTM 更好地捕捉到序列中的时序信息.结果表明,相较于现有的 MidiNet 模型,使用的旋律主题条件 CNN-BiLSTM 模型在准确率、归一化 KL 散度方面分别高出 23%和 0.17,生成的乐曲在连贯性和情感表达方面也优于传统的模型.

关键词:音乐生成;自动作曲;CNN-BiLSTM;旋律主题提取;聚类

中图分类号:TP391.9 **文献标志码**:A **文章编号**:1000-2367(2025)03-0135-08

随着人工智能技术的不断发展,AI 音乐生成技术逐渐成为人们关注的热点之一.AI 音乐生成技术是通过计算机算法和模型生成优美的音乐,减少人工干预对音乐生成带来的主观性.其中,基于深度学习的音乐生成模型已经取得了一定的进展,但仍然存在许多问题和挑战.

在现有的研究中,SHI 等[1]使用基于隐马尔可夫(hidden markov model,HMM)的方法来对钢琴音乐进行分析和分类;YANG 等[2]提出了 MidiNet 一种基于卷积神经网络(convolutional neural network,CNN)的方法来生成音乐,特别是以钢琴曲为例,设计了一个卷积神经网络来对音乐的节奏、和弦、旋律等特征进行建模,并在这个基础上进行乐曲生成;MINU 等[3]提出了一种基于长短时记忆网络(long short-term memory,LSTM)和循环神经网络(recurrent neural network,RNN)的方法来生成多轨钢琴曲;WANG 等[4]提出一种使用 RNN 模型来生成中国风格的音乐;JAYATHARAN 等[5]提出了一种使用 CNN 的对抗神经网络(generative adversarial network,GAN)来生成符号领域音乐的方法.以往的研究中,存在着音符序列长时依赖,局部和全局特征难以捕捉[6],难以生成表达特定情感的音乐这 3 个问题.

针对以往研究中的不足,提出了一种使用旋律主题作为条件、混合 CNN 和双向长短时记忆网络(bi-directional long short-term memory,BiLSTM)的 CNN-BiLSTM 模型来生成音乐.首先,将转换成钢琴卷帘窗的乐器数字化接口(musical instrument digital interface,MIDI)文件,按照章节分割为大段落之后,对每个大段落使用定长分割的方法分割为定长片段.其次,使用 Ward 聚类对旋律片段进行聚类,将其中最大中心对应乐谱片段作为旋律主题;将旋律主题和旋律片段拼接起来,并将其输入到 CNN-BiLSTM 模型中.最后,模型将会输出接下来时间片的对应音高.与 LSTM-RNN 模型[3]和 CNN-GAN 模型[5]不同,在本文提出的模型中

收稿日期:2023-09-04;**修回日期**:2024-06-08.

基金项目:国家自然科学基金(U1604154);河南省重点科技攻关项目(252102211035).

作者简介(通信作者):曹西征(1977—),男,山东莒县人,河南师范大学副教授,博士,主要研究领域为智能作曲、智能编曲、音乐信息处理,E-mail:287282578@qq.com.

引用本文:曹西征,张航,李伟.基于主题条件 CNN-BiLSTM 的旋律自动生成方法[J].河南师范大学学报(自然科学版),2025,53(3):135-142.(Cao Xizheng,Zhang Hang,Li Wei.Automatic melody generation method based on conditional CNN-BiLSTM[J].Journal of Henan Normal University(Natural Science Edition),2025,53(3):135-142.DOI:10.16366/j.cnki.1000-2367.2023.09.04.0002.)

采用了 Bi-LSTM,使其能够同时考虑到音符序列的过去和未来信息,从而更全面地理解音乐上下文,这种双向分析有助于提高音乐生成的连贯性和表现力.与 LI 等[6]提出的情感条件模型中的情感标签相比,使用主题旋律作为生成音乐的条件,解决了手动标注标签所带来的主观性和可扩展性方面的问题.

1 数据处理

1.1 钢琴卷帘窗分割

钢琴卷帘窗如图 1 所示,是一种将音符序列[7]映射到二维空间音乐表示方法,行表示时间,列表示音高,每个元素表示该时刻该音高是否被触发,这样的表示方式直观地反映了音乐信息的时间和音高维度.

使用钢琴卷帘窗的原因有以下 2 点:分片后看作二维图像,使得模型可以利用卷积神经网络来提取和学习音乐信息的时空特征;方便进行翻转、随机裁剪等数据增强操作,增加训练数据的多样性和模型的泛化能力.

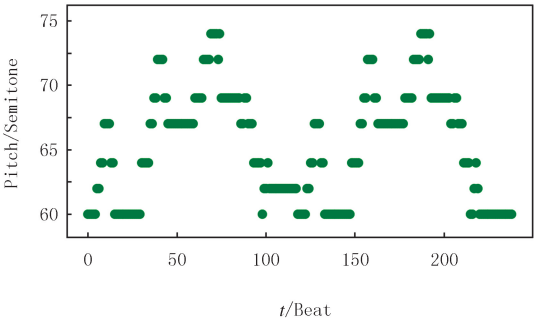

图1 钢琴卷帘窗

Fig.1 Piano roll

在钢琴卷帘窗分割中,SHI 等[1]使用了定长分割方法;YANG 等[2]使用了可变长分割方法保留时序信息.另外,ZHANG[8]阐述了一些音乐分割和编码方法,例如小节、乐句等,以解决符号音乐的生成问题.

鉴于现有的分割方法中存在的不足之处,本文使用了定、变长相结合的分割方法.具体地,将钢琴卷帘窗按乐谱章节分割大段落,然后在大段落中使用定长分割的方法分割成定长片段.使用此种分割方法,避免定长片段分割中乐谱缺失章节结构完整性,同时将章节内的旋律分割成易于训练的定长音符.

1.2 使用 Ward 聚类方法获取主题片段

在已有的条件音乐生成模型[9]中,往往使用乐谱和人工情感信息作为模型的输入,然而人工情感信息的标记缺乏客观性,这使得生成效果受到一定的限制.

为了解决这个问题,对钢琴卷帘窗 π(其中片段记作 $\pi_1,\cdots,\pi_2$)使用 Ward 聚类算法[10].选取其中的最大聚类中心作为歌曲的旋律主题,将其作为一个新的输入特征一起输入到 CNN-BiLSTM 模型中,减少人工情感标记的主观性的影响.

选择聚类获取主题片段的优势,无监督学习[10],不需要预先标注的数据进行训练;聚类分析可以帮助发现音乐数据中的隐含结构和模式;聚类分析可以生成多个不同的旋律主题.相比之下,使用卷积模块则需要数据量和标注,而带有主题标注的音乐数据相对缺乏.具体来说,使用 Ward 聚类[11]对乐谱的钢琴卷帘窗片段进行聚类,将其中最大的聚类中心作为整首歌曲的旋律主题.使用聚类算法能够有效地提取出乐谱中多次出现的旋律,而这个旋律往往代表乐谱风格和情感.

2 旋律主题条件 CNN-LSTM 模型

旋律主题是指在一首乐谱中反复出现的片段[12],带有充分的情感信息的片段,因此利用旋律主题作为 CNN-BiLSTM 模型[11]的条件[6],使得模型能够更好控制生成乐谱的情感特征.

模型如图 2 所示,由乐谱特征提取网络、旋律主题特征提取网络和乐谱生成网络 3 个子网络组成,每个子网络完成不同的工作.通过这 3 个子网络,旋律主题条件 CNN-BiLSTM 模型可以学习到音乐片段的模式,并生成新的音乐片段,从而实现音乐创作的自动化.

使用 CNN 作为模型的前半部分[6],有效地提取出钢琴卷帘窗中关于旋律之间的时间和音高对应特征[13],从而使得 BiLSTM 在后半部分[14]更好地捕捉到序列中的时序信息,提高模型的性能和表现.

具体而言,乐谱特征提取网络将输入的乐谱映射为特征向量 f_x[15],即 $f_x=\mathrm{CNN}_x(x)$;旋律主题特征提取网络将输入的旋律主题 m 映射[16]为特征向量 f_m,即 $f_m=\mathrm{CNN}_m(m)$;乐谱生成网络根据当前时刻 t 的输入 f_t,前一时刻生成的旋律 y_{i-1} 和整合后的特征向量(f_x,f_m)来生成下一个旋律 y_t 的概率分布,即 $p(y_t \mid x,y_{<t},m)=\mathrm{BiLSTM}(f_t,y_{t-1},(f_x,f_m))$.

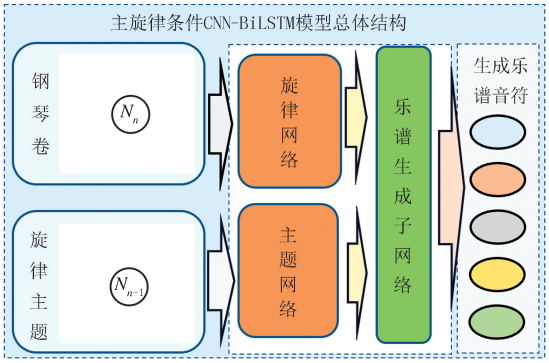

图2　旋律主题条件CNN-BiLSTM模型结构

Fig. 2　Melody theme conditional CNN-BiLSTM structure

模型的总体概率

$$p(y \mid x,m)=\prod_{t=1}^{T} p(y_i \mid x,y_{<t},m),$$

其中,y 是模型输出的旋律序列,x 是输入的乐谱,m 是输入的旋律主题,$y_{<t}$ 表示生成的序列中第 t 个时刻之前的旋律片段,T 是总序列时刻数. 整个模型的学习目标是最大化训练集上的对数似然:

$$L=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\log_2 p(y_{i,t} \mid x_i,y_{i<t},m_i),$$

其中,N 是训练集中旋律序列的数量,T_i 是第 i 个序列的长度.

将旋律主题 m 作为条件[16]使得模型可以控制生成特定情感和乐谱结构的音乐,同时也避免了需要通过人工标记来获取感情标签的主观性缺陷.相比于传统的生成模型,旋律主题条件 CNN-BiLSTM 模型更具针对性,更加准确地生成符合旋律主题和情感的音乐作品.

2.1　CNN 特征提取网络提取乐谱与旋律主题特征

由于乐谱钢琴卷帘窗片段特征提取网络和旋律主题钢琴卷帘窗特征提取网络所处理的数据具有相似性,因此对这 2 个子网络使用相同的构架方式.

特征提取网络结构如图 3 所示,由 4 层卷积层和全连接特征输出层组成,使用混合构架能更好地利用不同结构网络的优势,提高乐谱生成的效果.其中,带残差的网络结构[17],在保持网络深度和复杂度的同时,增加网络的非线性表示能力;最大池和平均池交替使用,在同时提取局部和全局特征的基础上,提高特征的多样性和丰富度;串并联相结合的 CNN 构架兼顾网络的深度和宽度,并联结构增加了网络的非线性表示能力和泛化性能,串联结构实现了特征的分层提取和加工.

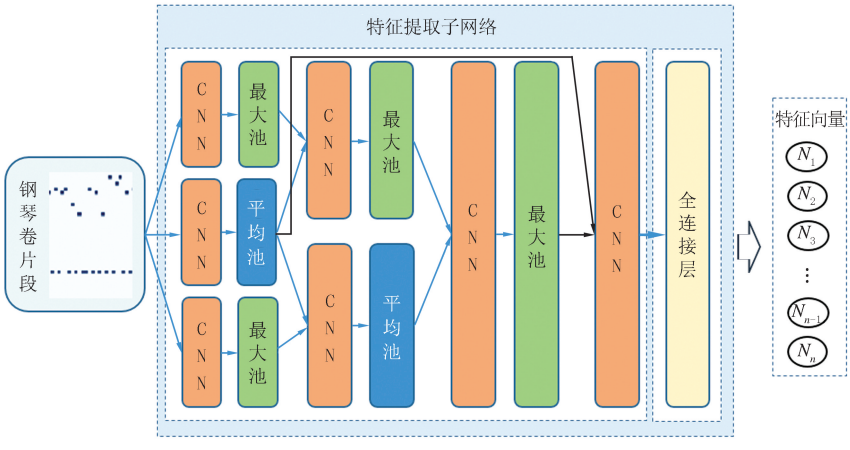

图3　特征提取子网络结构

Fig. 3　Feature extraction sub-network structure

2.2　整合旋律主题特征的生成乐谱子网络

乐谱生成子网络结构如图 4 所示,通过 LSTM 和 BiLSTM[18]混合结构,2 组 LSTM 的设计分别对乐谱

片段和旋律主题进行建模.

子网络结构采用了 LSTM 和 BiLSTM 相结合的结构,更好地学习和捕捉乐谱的整体和局部特征,乐谱片段和旋律主题之间的关系[19],从而生成符合旋律主题的新乐谱片段.

3 结果与评价分析

对于音乐生成模型的评价一般从客观和主观[20]2 个角度进行.客观评价通常基于一些音乐学上的规则和统计指标,如音高、节奏、和弦等.主观评价[21]则基于人的感受和情感[22],主观评价往往会涉及更多的审美和情感因素,因此比较主观和难以量化[23].本文提出的基于旋律主题条件 CNN-BiLSTM 模型分别从主观和客观 2 个方面进行评价分析.

3.1 数据集与客观评价指标

使用的 LMD(Lakh MIDI Dataset)[19]和 GiantMIDI-Piano[18]这 2 个数据集,LMD 是一个用于音乐信息检索和音乐信息提取的公共数据集,它包含了超过 10 万个来自不同的音乐时期、风格和文化背景的音乐作品.针对同一首歌曲,往往 LMD 数据集中会存在有多个不同版本.鉴于 LMD 存在的问题,使用的数据集在 LMD 数据集挑选出 Pop,Classical,Jazz3 种风格的 MIDI 格式的歌曲作为本文模型的训练数据.GiantMIDI-Piano 数据集的标注信息非常详细,包括每个音符的音高,起始时间,持续时间,音量,可以准确地分析和处理数据,并用于训练和评估音乐生成模型.

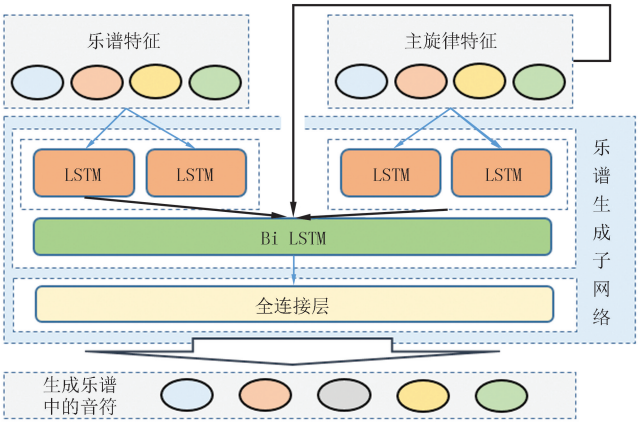

图4 乐谱生成子网络结构

Fig.4 Music score generation sub-network structure

在音乐生成任务中,常用的模型评价指标有准确率、KL 散度、Empty Bar(EB)、Pitch Range(PR)和 Number of Unique Note Pitch(NUP).它们反映模型生成的音乐的多样性与真实音乐的相似度[18].因此,采用归一化后的 KL 散度、准确率、Empty Bar(EB)、Pitch Range(PR)和 Number of Unique Note Pitch(NUP)作为客观评价指标,评估模型生成的音乐与输入旋律主题的匹配程度.

归一化 KL 散度

$$K_{\text{归一化}}=\frac{L_i}{\max\limits_i\{K_i\}},$$

其中,K_i 为模型 i 的 KL 散度.

模型准确率

$$A=\frac{\sum\limits_{i\in\theta}^{\theta}w_iN_{\text{TP},i}}{\sum\limits_{i\in\theta}^{\theta}(w_{\text{pop}}+w_{\text{classical}}+w_{\text{jazz}})(N_{\text{TP},i}+N_{\text{FN},i})},(\theta\in(\text{Pop},\text{Classical},\text{Jazz})),$$

其中,w_{pop}、$w_{\text{classical}}$、w_{jazz}分别表示 3 类数据在数据集中的比例,$N_{\text{TP},i}$表示生成与真实一致的音符数,$N_{\text{FN},i}$表示生成的音符与实际音符不一致的数目.

Empty Bar(EB)表示生成音乐中的空拍率;Pitch Range(PR)表示音高的范围,用最高音高和最低音高之间的音程来衡量;Number of Unique Note Pitch(NUP)表示音高的种类数量,用于评估音乐的多样性.

3.2 消融实验与客观评价分析

使用模型生成 30 首旋律,其中一首生成的旋律如图 5 所示,蓝色部分为输入的旋律片段,之后是模型对

给定片段生成的旋律.从图 5 中,可以看出输入旋律与生成旋律部分的结构具有明显的关系性.

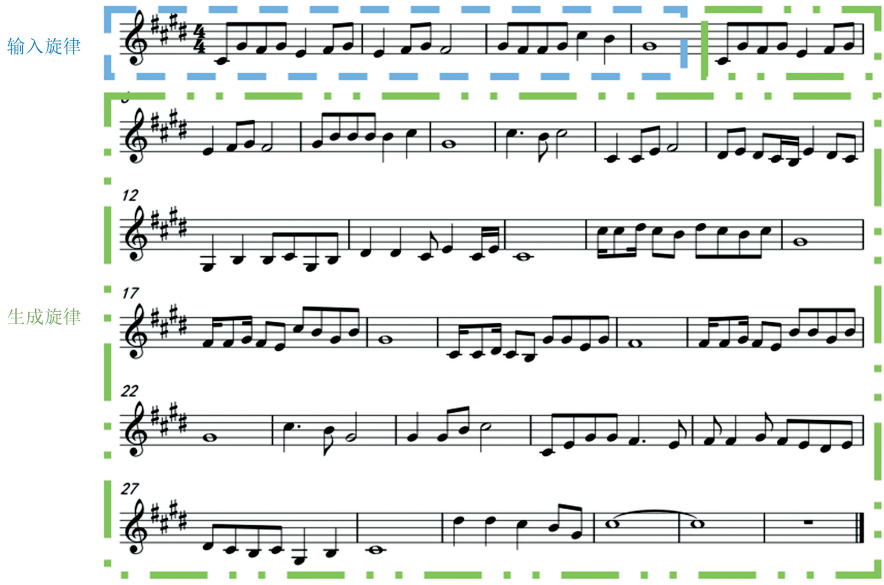

图5　本文模型生成的一首旋律

Fig.5　A melody generated by our model

为了验证子网络的有效性,设计一个消融实验[24]比较完整模型,去除乐谱特征提取网络和去除旋律主题特征提取网络的效果,使用模型在 LMD 和 GiantMIDI-Piano[18]这 2 个数据集上的准确率和归一化 KL 散度作为衡量指标,实验结果如表 1 所示.

通过表 1 和表 2 分析得出,基准模型与去除子网络的消融模型在 LMD(L)和 GiantMIDI-Piano(G)[18]这 2 个数据集上的准确率相比于去除乐谱特征提取网络模型分别高出约 9%和 7%,相比于去除旋律主题特征提取网络模型分别高出约 7%和 9%.同时消融实验表明,基准模型的归一化 KL 散度值相比于消融模型高出约 0.13,EB、PR 和 NUP 相比高出约 0.03、3.15 和 9.15.上述实验结果说明子网络可以有效地提升模型的效果.

表 1　准确率、归一化 KL 散度的消融实验结果对比

Tab. 1　Comparison of ablation experimental results for accuracy and normalized KL divergence

模型	准确率(L)	归一化 KL 散度(L)	准确率(G)	归一化 KL 散度(G)
基准模型	0.948 1	1.000 0	0.951 6	1.000 0
去除乐谱特征提取网络模型	0.862 0	0.870 4	0.879 1	0.928 6
去除旋律主题特征提取网络模型	0.879 0	0.949 2	0.865 8	0.952 4

表 2　EB、PR 和 NUP 的消融实验结果对比

Tab. 2　Comparison of ablation experimental results for EB, PR and NUP

模型	归一化 KL 散度					
	EB(L)	PR(L)	NUP(L)	EB(G)	PR(G)	NUP(G)
基准模型	0.036 1	16.62	26.62	0.039 8	16.15	26.55
去除乐谱特征提取网络模型	0.072 7	13.47	17.47	0.074 5	13.01	17.49
去除旋律主题特征提取网络模型	0.062 5	13.95	20.50	0.072 6	13.08	20.52

本文提出的旋律主题 CNN-BiLSTM 模型,在 LMD 数据集中的准确率和归一化 KL 散度如下表 3 所示,其中 Pop,Classical,Jazz 准确率是模型对数据集中 3 种不同类型音乐的准确率.

通过表 3 和表 4 分析得出,相较于 MidiNet 模型[2]、LSTM-RNN 模型[3]和 CNN-GAN 模型[5],旋律主题条件 CNN-BiLSTM 模型的准确率分别高出约 23%、13%、7%.在归一化后的 KL 散度方面,相比于已有模型,本文提出的模型高出约 0.17,在 EB、PR、NUP 方面,相比提高约 8%、8%、14%.说明旋律主题条件

CNN-BiLSTM 模型可以更好地捕捉旋律主题和音符之间的关系,生成更加符合音乐规律的音乐片段.

表 3 模型在数据集中不同音乐风格准确率、总准确率和归一化 KL 散度的结果对比

Tab. 3 The results of different music style accuracy, total accuracy and normalized KL

模型	Pop 准确率	Classical 准确率	Jazz 准确率	模型准确率	归一化 KL 散度
MidiNet	0.717 7	0.722 8	0.704 3	0.714 9	0.832 4
LSTM-RNN	0.817 8	0.811 0	0.812 3	0.813 7	0.907 5
CNN-GAN	0.874 0	0.870 4	0.875 5	0.873 3	0.962 5
旋律主题条件 CNN-LSTM	0.946 2	0.949 2	0.948 8	0.948 1	1.000 0

表 4 模型在数据集中 EB、PR 和 NUP 的结果对比

Tab. 4 Comparison of the results of EB, PR and NUP in the data set of the model

模型	归一化 KL 散度		
	Empty Bar(EB)	Pitch Range(PR)	Number of Unique Note Pitch(NUP)
MidiNet	0.112 0	9.82	14.74
LSTM-RNN	0.107 8	7.91	13.10
CNN-GAN	0.092 4	12.75	19.45
旋律主题条件 CNN-LSTM	0.035 3	15.62	26.82

3.3 模型主观评价及分析

在计算机生成音乐领域,客观指标仅能反映生成音乐与目标音乐的相似度,而不能充分评价生成音乐的音乐性和美感[14],因此需要使用人类听感评价[25]对模型进行主观评价.

本文采用音乐研究中主要使用的用户调查[2]作为模型的主观评价[22]方法.为了使得实验结果的评价更具有公平性,因此采用多样化的评价者群体,明确和统一评价指标,结合客观指标这几种手段提高评价结果的公平性.具体地,从有不同经验的人群选取的 20 人作为音乐的评价者,其中 10 名男性,10 名女性.对于音乐歌曲评价指标分为情感表达和连贯性 2 个方面.每个方面均分为 3 个档次,分别是优秀、中等、较差.

首先,使用旋律主题条件 CNN-BiLSTM 模型、MidiNet 模型[2]、LSTM-RNN 模型[3]和 CNN-GAN 模型[5]4 种不同的算法生成 20 首音乐作品(每个模型 5 首).将这些音乐作品随机编号,形成 20 个序号,用于测试和评价.其次,对于每个模型生成的作品,20 名评价者随机抽取一首音乐进行评价,并在包括歌曲编号、评价者编号、评价的情感表达和连贯性的评价表上进行标记.

分析图 6 中的评价者的主观评价,旋律主题条件 CNN-BiLSTM 生成的音乐,在连贯性上,音符之间更为自然流畅,听起来接近于一首完整的曲子;在情绪表达方面,更加准确地表现出音乐中包含的情感、气氛和情绪.由上述分析可知,旋律主题条件 CNN-BiLSTM 在连贯性和情绪表达方面相较于其他 3 个模型有明显的优势.

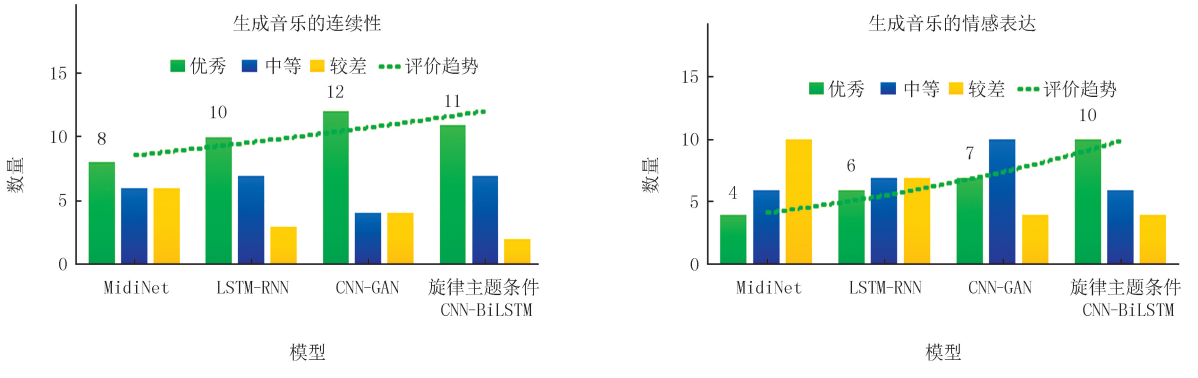

图6 生成音乐在连贯性和情感表达方面的主观评价

Fig.6 Generated melody evaluations in terms of coherence and emotional expression

4 总　结

本文提出旋律主题条件 CNN-BiLSTM 模型，在主观评价和客观评价 2 个方面，均优于对比的 3 种乐谱生成模型，且生成的音乐在连贯和情绪 2 个方面都具有较为不错的表现.CNN-BiLSTM 结构，通过结合 CNN 对钢琴卷帘窗的特征提取和 LSTM、BiLSTM 对序列具有长短时记忆 2 个优势，使得模型可以更好地提取音乐时序和音高的关系.引入旋律主题条件机制，即可以控制生成音乐的情绪表达，又避免了人工标记情感的主观性，使得生成的音乐具有可控性和准确性.数据集中使用的多是钢琴，对于鼓乐器的数据相对不足，下一步可以增加鼓乐器的数据进行训练.模型仅使用旋律主题作为条件，未来可以增加乐器、流派、演奏标记等作为模型条件，以提升模型的性能.本文相关的成果已经开发为软件，以后将逐步完善、更新，读者可以到本创业团队官网 https://sdzqj.com 查看、下载.

参　考　文　献

[1] SHI W J，LI Y H，GUAN Y S，et al.et al.Optimized fingering planning for automatic piano playing using dual-arm robot system[C]// 2022 IEEE International Conference on Robotics and Biomimetics(ROBIO).Piscataway：IEEE Press，2022：933-938.

[2] YANG L C，CHOU S Y，YANG Y H.MidiNet：A convolutional generative adversarial network for symbolic-domain music generation [EB/OL].[2023-05-26].https://arxiv.org/pdf/1703.10847v2.

[3] MINU R I，Nagarajan G，Borah S，et al.LSTM-RNN-Based Automatic Music Generation Algorithm[C]// Intelligent and Cloud Computing：Proceedings of ICICC 2021.Singapore：Springer，2022：327-339.

[4] WANG J X，LI C Q.Chinese style pop music generation based on recurrent neural network[C]//2022 IEEE 5th Advanced Information Management，Communicates，Electronic and Automation Control Conference.Piscataway：IEEE Press，2022：513-516.

[5] JAYATHARAN V，ALWIS D.Alapana Generation using Finite State Machines and Generative Adversarial Networks[C]//2023 International Research Conference on Smart Computing and Systems Engineering.Piscataway：IEEE Press，2023，6：1-6.

[6] LI S Y，SUNG Y.INCO-GAN：variable-length music generation method based on inception model-based conditional GAN[J].Mathematics，2021，9(4)：387.

[7] MIYAMOTO K，TANAKA H，NAKAMURA S.Online EEG-based emotion prediction and music generation for inducing affective states [J].IEICE Transactions on Information and Systems，2022，E105.D(5)：1050-1063.

[8] ZHANG N.Learning adversarial transformer for symbolic music generation[J].IEEE Transactions on Neural Networks and Learning Systems，2023，34(4)：1754-1763.

[9] DE BERARDINIS J，VAMVAKARIS M，CANGELOSI A，et al.Unveiling the hierarchical structure of music by multi-resolution community detection[J].Transactions of the International Society for Music Information Retrieval，2020，3(1)：82-97.

[10] GOVENDER P，SIVAKUMAR V.Application of k-means and hierarchical clustering techniques for analysis of air pollution：a review (1980-2019)[J].Atmospheric Pollution Research，2020，11(1)：40-56.

[11] YU Y，HARSCOËT F，CANALES S，et al.Lyrics-conditioned neural melody generation[C]//MultiMedia Modeling：26th International Conference.[S.l.]：Springer，2020：709-714.

[12] DUNGAN B M，FERNANDEZ P L.Next bar predictor：an architecture in automated music generation[C]//2020 International Conference on Communication and Signal Processing.Piscataway：IEEE Press，2020：109-113.

[13] CHEN Y H，LERCH A.Melody-conditioned lyrics generation with SeqGANs[C]//2020 IEEE International Symposium on Multimedia. Piscataway：IEEE Press，2020：189-196.

[14] DUA M，YADAV R，MAMGAI D，et al.An Improved RNN-LSTM based Novel Approach for Sheet Music Generation[J].Procedia Computer Science，2020，171：465-474.

[15] LIM Y Q，CHAN C S，LOO F Y.Style-conditioned music generation[C]//2020 IEEE International Conference on Multimedia and Expo. Piscataway：IEEE Press，2020：1-6.

[16] BANERJEE S，RATH M，SWAIN T，et al.Music Generation using Time Distributed Dense Stateful Char-RNNs[C]//2022 IEEE 7th International conference for Convergence in Technology.Piscataway：IEEE Press，2022：1-5.

[17] BANAR B，COLTON S.A systematic evaluation of GPT-2-based music generation[C]//Artificial Intelligence in Music，Sound，Art and Design：11th International Conference.Cham：Springer，2022：19-35.

[18] WU J，LIU X G，HU X L，et al.PopMNet：generating structured pop music melodies using neural networks[J].Artificial Intelligence，2020，286：103303.

[19] 陈吉尚，哈里旦木·阿布都克里木，梁蕴泽，等.深度学习在符号音乐生成中的应用研究综述[J].计算机工程与应用，2023，59(9)：27-45.

CHEN J S,ABUDUKELIMU H,LIANG Y Z,et al.Review of the application of deep learning in symbolic music generation[J].Computer Engineering and Applications:2023,59(9):27-45.

[20] 汪涛,靳聪,李小兵,等.基于 Transformer 的多轨音乐生成对抗网络[J].计算机应用,2021,41(12):3585-3589.
WANG T,JIN C,LI X B,et al.Multi-track music generative adversarial network based on Transformer[J].Journal of Computer Applications,2021,41(12):3585-3589.

[21] 严丹,何军,刘红岩,等.考虑评级信息的音乐评论文本自动生成[J].计算机科学与探索,2020,14(8):1389-1396.
YAN D,HE J,LIU H Y,et al.Considering grade information for music comment text automatic generation[J].Journal of Frontiers of Computer Science and Technology,2020,14(8):1389-1396.

[22] 贾宁,郑纯军.基于注意力 LSTM 的音乐主题推荐模型[J].计算机科学,2019,46(S2):230-235.
JIA N,ZHENG C J.Model of Music Theme Recommendation Based on Attention LSTM[J].Computer Science.,2019,46(S2):230-235.

[23] JIANG F Z,ZHANG L M,WANG K X,et al.BoYaTCN:research on music generation of traditional Chinese pentatonic scale based on bi-directional octave your attention temporal convolutional network[J].Applied Sciences,2022,12(18):9309.

[24] WU G W,LIU S P,FAN X Y.The power of fragmentation:a hierarchical transformer model for structural segmentation in symbolic music generation[J].ACM Transactions on Audio,Speech,and Language Processing,2023,31:1409-1420.

[25] 曹西征,牛靖雯,秦杰,等.面向抒情歌曲旋律的钢琴自动伴奏算法[J].河南师范大学学报(自然科学版),2016,44(4):137-142.
CAO X Z,NIU J W,QIN J,et al.Automatic piano accompaniment algorithm for the melodies of lyric songs[J].Journal of Henan Normal University(Natural Science Edition),2016,44(4):137-142.

Automatic melody generation method based on conditional CNN-BiLSTM

Cao Xizheng, Zhang Hang, Li Wei

(College of Computer and Information Engineering; Engineering Lab of Intelligence Business & Internet of Things; Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Henan Normal University, Xinxiang 453007, China)

Abstract: To effectively generate structured melodies, a melody auto-generation method based on theme-conditioned CNN-BiLSTM is proposed. Melodies are represented in the form of piano roll windows, and the piano roll windows are segmented using a combination of fixed-length and variable-length methods. The Ward clustering algorithm is used to perform cluster analysis on the piano roll window segments, and the largest cluster obtained is taken as the melody theme of the song. The melody theme is used as a condition to generate melodies using a model based on the CNN-BiLSTM structure. The upper part of the CNN can effectively extract the information between time and pitch contained in the piano roll window, and the lower part uses LSTM and BiLSTM to capture the temporal information better in the sequence. The results show that, compared to the existing MidiNet model, the melody theme-conditioned CNN-BiLSTM model achieves improvements of 23% in accuracy and 0.17 in normalized KL divergence. The generated music is also superior to traditional models in terms of coherence and emotional expression.

Keywords: music generation; automatic composition; CNN-BiLSTM; main melody extraction; clustering

[责任编校 杨浦 刘洋]