

浅谈人工智能技术在音乐创作中的应用

■ 陈世哲

A PRELIMINARY EXPLORATION OF THE APPLICATION OF AI TECHNOLOGY
IN MUSIC COMPOSITION

摘要:随着人工智上升为国家战略,如何在音乐领域应用人工智能技术变成一个新研究的研究领域。在这个领域中,利用计算机来创作音乐是一个重点研究方向。近年来有很多相关技术涌现并得到了应用,在这个过程中,传统算法作曲的相关理论为人工智能技术在音乐创作中的实现提供了参考和基础,而人工智能技术也在音乐的创作思维、制作流程等方面对传统理论产生了影响。结合已有的人工智能作曲技术,总结相关算法作曲理论与其的关系,通过具体的应用实例,分析这种新方式的趋势和特点,希望通过此类研究使科技更好地为音乐服务。

关键词:人工智能;电子音乐;算法作曲

中图分类号: J60-05

文献标识码: A

文章编号: 1004-2172(2020)01-0125-08

DOI:10.15929/j.cnki.1004-2172.2020.01.016

引言

人工智能(Artificial Intelligence)并不是一个新名词。早在20世纪50年代,计算机领域就引起了第一次科技热潮,中间也历经了多次低谷,但随着谷歌DeepMind团队的AlphaGo击败人类围棋冠军,一个计算机学科中的人工智能技术迅速引发人类下一个科技浪潮,并在各行业产生新的机遇。

在传统意义上,机器的优势在于能够帮助人类完成机械而重复性的劳动,对于创造性的工作则参与较少。但随着人工智能技术的发展,它正逐渐应用于音乐的创作、制作、分析和教育

等领域。其中,利用AI作曲或者辅助作曲变成关注度最高的应用方向之一,很多疑问也就此产生。如AI作曲技术能否在一定程度上代替人类作曲家?AI作曲技术会如何影响传统作曲的思维?这些将变成一个个有趣而又严肃的课题。

在人工智能于各行业应用的实际过程中,机器学习^①可以说是AI最重要的子集之一,它深刻影响着AI中的其他领域。用AI技术自动作曲并不是一个新的课题,相关的研究很多年前就已经开始,但是一直受技术所限。人工智能作曲的主要原理同下围棋的原理类似,主要运用遗传算法^②、神经网络、马尔可夫链^③和混合型算法等,利用音乐规律给计算机制定规则、建

① 机器学习是一项专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身性能的人工智能技术。

② 遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型。

③ 马尔可夫链是概率论和数理统计中的一种结构。

立海量数据库,继而进行深度学习^①,分析作曲规则、结构等各项信息,然后重新生成音乐。目前,有多家国内外机构和公司开始了该领域的研究。传统的方案是完全建立在用规则构建智能系统的基础上,而新的方案是更多地使用神经网络的方式,即使用“学习”的方式来实现。

本文主要讨论将 AI 技术应用在音乐创作领域所主要使用的技术方案和策略算法,并分析目前所碰到的问题以及与算法作曲的关系,讨论其应用方式,以促进此技术能更好地为音乐服务。

二、相关技术手段分析

(一)神经网络(NN)

人工智能的传统方式是利用规则,即以一种自上而下的思路来解决问题;而神经网络(Neural Network,简称 NN)则是以一种自下而上的思路来解决问题,它的基本特点是模仿人类大脑的神经元^②之间的信息传递和模式。

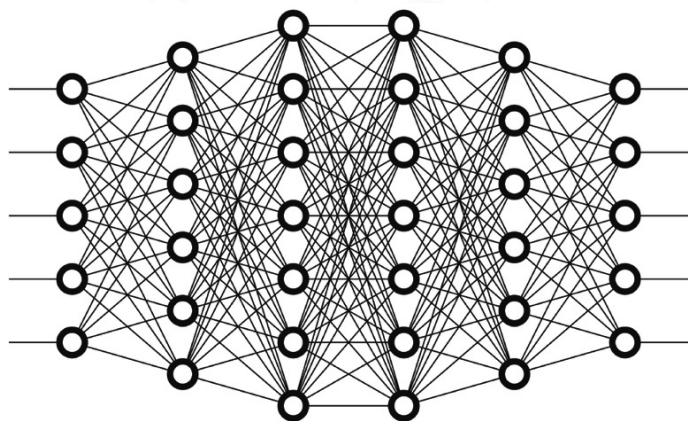
神经网络有两个特性,一是每个神经元通过对应的输出函数,计算和处理来自相邻神经元的加权输入值;二是通过加权值来定义神经元之间

信息传递的关系,算法在处理过程中不断地自我学习、不断地优化和调整这个加权值。此外,神经网络的处理过程需要依靠大量的数据来训练。所以,神经网络的处理过程具有非线性、分布式、并行计算、自适应和自组织的特点。

以人类学习创作音乐的过程为例,一般要经历音乐感知(欣赏),音乐模仿写作,最后达到独立创作。在创作过程中,也包含对作曲技法、和声理论等的学习,学习者在不断地练习中,通过教师地批改引导,不断完善自己的创作思路等。这些学习过程基本可以通过神经网络的架构模拟出来,这也是这项技术得以应用的一个基础。

神经网络的运作过程需要有输入与输出,权重和阈值以及多层的感知器^③的结构(图1)。神经网络可以看成是一个“黑盒子”,给定足够的训练集,即可以在输入端给定 X 后,得到预期的 Y。具体来说,神经网络的运作过程需要确定输入和输出,然后找到一种或多种算法,可以从输入得到输出,再找到一组已知答案的数据集,用来训练模型,此后,重复这个过程,输入模型,就可以得到结果,不断修正这个模型^④。

图1 神经网络示意图



① 深度学习是机器学习中一种基于对数据进行表征学习的方法。

② 神经元即神经细胞,是神经系统最基本的结构和功能单位。

③ 感知器是人工神经网络中的一种典型结构。

④ Allen Huang, Raymond Wu, *Deep learning for music*, arXiv:1606.04930 (2016).

音乐是时间的艺术,许多信息是基于时间轴建立的。而神经网络实现的机制有很多种,其中能够较好地处理时间轴信息的一种技术就是递归神经网络(Recursive Neural Network,简称RNN)。RNN是一种(前馈)神经网络,通过新增表示时间维度信息的参数以及相关机制,使神经网络不仅可以基于当前数据而且可以基于先前数据来学习。和之前的模式不同,RNN系统中,前一个输入和后一个输入是有关联的,RNN是一个在时间上传递的神经网络,时间作为其深度的度量。循环网络通常具有相同的输入层和输出层,因为循环网络预测下一个项目是以迭代的方式用作下一个输入,以便产生序列,因此RNN是音乐创作中一项重要的实现方式。

(二)LSTM长短期记忆单元

LSTM(全称 Long Short-Term Memory)是一种特殊的RNN结构,为RNN的变种结构,属于反馈神经网络的范畴。LSTM是为了克服RNN循环神经网络的梯度消失^①或爆炸而产生的神经网络,它除了继承RNN模型的特性外还具备了自身的优点。RNN虽然可以兼顾时间维度信息的处理,但是若时间间隔拉长后,长期保存信息并在其中开展学习得到的结果并不理想,这对于音乐信息处理是一个致命的问题,而解决

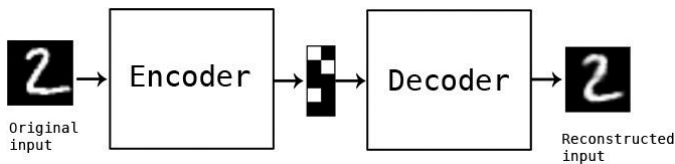
该问题的一个重要方向便是增大网络存储。因此,采用特殊隐式单元的LSTM被首先提出,其自然行为是长期的保存输入。LSTM主要的改变是增加了3个门,分别是输入门、输出门和忘记门。在实践中,LSTM被证明比传统的RNN更加有效,它最先应用于机器翻译领域、对话生成以及编解码领域^②。LSTM能够表征更复杂的人类的逻辑发展和认知过程,所以也是音乐生成目前最值得研究的方向。

(三)VAE

Autoencoder是通过对特征数据进行压缩和解压来实现非监督学习的过程^③,它本身既是一个多层神经网络,也是一个对数据的非监督学习^④的模型。^⑤音乐数据没有明确的优与劣之分,所以更适合于非监督的模式。而其中的变分自动编码器(Variational Autoencoder,简称VAE)是自动编码器的升级版,结构与自动编码器相似,其原理是在编码过程中增加一些限制(图2)。这种处理原理同作曲的思维过程有相似之处,作曲在某种意义上是一个创作和规则并存的过程,而VAE的机制可以很好地与之相符合。

在实践过程中,VAE已经可以在对多声部音乐的音高动态和乐器等信息分析与生成中得到很好的应用,尤其在古典音乐和爵士乐方面。

图2 VAE原理示意图



① 在神经网络中,随着隐藏层数目的增加,准确率反而下降了,这种现象简称梯度消失。

② Jean-Pierre Briot, Ga tan Hadjeres, Fran ois-David Pachet, *Deep learning techniques for music generation--a survey*, arXiv:1709.01620 (2017).

③ Gino Brunner, Andres Konrad, Yuyi Wang, Roger Wattenhofer, *MIDI-VAE: Modeling Dynamics and Instrumentation of Music with Applications to Style Transfer*, arXiv:1809.07600 (2018).

④ 非监督学习是利用未加标签的数据来完成相关的机器学习过程的方式。

⑤ Jay A. Hennig, Akash Umakantha, Ryan C. Williamson, *A Classifying Variational Autoencoder with Application to Polyphonic Music Generation*, arXiv:1711.07050 (2017).

它甚至可以用爵士风格演绎莫扎特的作品,产生新的混搭方式。

变分自动编码器是目前生成内容最优的方法之一,它可以成功生成各种形式的数据,特别是在多声部音乐的生成工作过程方面。但是,当数据是多模式时,VAE并不能提供明确的机制,它不能对离散值潜在变量进行推理,所以这是制约其发展的一个重要因素。比如在C大调与c小调的处理过程中,对音阶中各种音的使用倾向是不同的,特别是扩展到24音及以上时,VAE处理起来就比较困难。所以现在还有一些研究思路是设计一种VAE+LSTM方式,这种方式可以避免部分问题。此外,在和弦问题上还可以用受限玻尔兹曼机(Restricted Boltzmann Machine,简称RBM)^①建模等等。

三、深度学习与算法作曲的关系

算法作曲(Algorithmic Composition)是电脑作曲的一种重要方式,也可称为自动作曲(Automatic Composition),主要利用算法减少音乐创作时的人类干预。在传统意义上,作曲可以从多个维度来理解,比如旋律、节奏、和声、编曲或者配器等等,虽然这些维度不能完全表征音乐,但可以作为数据化的起点,计算机可通过相应的语言和技术来表征这些,即计算机辅助作曲。但本文讨论的AI作曲比之已有的方式更缩减了人类干预,并不属于传统意义上的计算机辅助作曲。

传统算法作曲领域的技术实现方案,除了前文提到的神经网络方式,还包括以下几种模式^②。

1. 翻译模式(Translational Models)

这一类模式主要是将一些其他媒介的信息转换成音乐的相关信息,可以是因一定规则生成的,也可以是随机产生的。例如将图片的色彩和明暗对应到音色的色彩和明暗,但这种直接、简单的对应,其成果和表现形式往往过于牵强。

2. 数学模型方式(Mathematical Models)

这种方式主要通过数学模型来产生音乐,在此模型下,很多音乐的元素是通过非确定性的方法构成的。这种随机的方式在很多传统音乐家的创作过程中都有所应用,而这里的随机是完全基于数学方式的。在创作过程中,作曲家可以参与算法相关的参数配置。数学模型方式主要运用马尔可夫链、高斯分布^③及分形理论^④等实现。这种方式存在的问题是:其算法完全基于规则的设定,对于创作音乐这项复杂的工作,公式化实现只能完成其中一个过程,完全由数学模型方式实现的创作结果很难完美。

3. 语法模式(Grammars)

这种研究方式认为音乐中存在着语法,如同语言中的语法。这种方式借鉴语言学的概念,将音乐当作一种特殊的语音对待,其研究方向在于如何以这种思路从音乐中提炼出算法,目前有一定的研究成果,也可以借鉴其他学科的相关成果。这种方式的问题是:第一,语法是存在层次结构的,或者说语言的规则是相对固定的,但音乐的表达是存在一定的即兴性或是模糊性的;第二,在语法分析和规则较多时,尤其存在模糊性的情况下,计算量要求比较高。

① 受限玻尔兹曼机是一种可以用于降维、分类、回归、协同过滤、特征学习以及主题建模的算法。

② Ramon Lopez de Mantaras, Josep Lluís Arcos, "AI and Music: From Composition to Expressive Performance," *AI magazine* 23.3 (2002): 43-43.

③ 高斯分布是一个重要的概率分布,也称为正态分布。

④ 分形理论的最基本特点是用分数维度的视角和数学方法描述和研究客观事物,也就是用分形分维的数学工具来描述研究客观事物。

4. 演化模式(Evolutionary Methods)

这种方式可以通过一定的方法对信息做筛选,利用算法把好的方案筛选出来并不断优化,得到相应的结果。该方式主要利用遗传算法来模拟生物进化的过程,因为遗传算法已被证明是解决大规模搜索空间的有效解决方式,可以找到多个解法,这些逻辑很符合音乐创作的某些逻辑^①。这种方式的问题是:遗传算法并不能完全模拟作曲的思路,特别是在实际操作过程中,为了节约资源通常会采用简化的方式,而一旦简化就会大大影响最终的结果。

5. 混合模式(Hybrid Systems)

混合模式就是混合以上方式甚至更多方式,一起来生成音乐。从理论上来说,这种方式有其合理性,研究十分复杂。但在实际应用中,传统基于规则和其他相关信息生成音乐的实现方法相对简单,对算法本身的要求太高,整个的运作模式并不存在“智能”和“学习”的成分^②。神经网络的方式从运作模式上来说是最接近人类思维的,传统算法作曲的思路则各有各的特点,目前的各种主要技术方案还有一定的问题,所以结合规则和神经网络的方案在相当长一段时间里都是最优的一种解决方案。

四、应用实例与分析比较

人工智能应用于音乐创作时,要确定处理的音乐信号种类的问题,通常使用信号类信息或符号类信息。但在深度学习的过程中,使用符号信息的方式相对更加普遍。

信号类信息中的第一种是音频信号,它可以是波形文件,也可以是通过傅里叶变换处理后的音频频谱信息。符号类信息主要是用 MIDI

信息。MIDI 已经是一种成熟的并被广泛应用的格式,主要使用音符信息中的 Note On 信息和 Note Off 信息,利用二者在 0~127 之间的取值来表示主要的音乐信息。符号类信息另一个使用的信息就是 MIDI 里面衡量时间点的 Tick 值。

在深度学习过程中,还有一种和计算机交换信息的方式,是直接用文本信息来表示音乐,当然它的规则也有很多。此外,还有以和弦、节奏或者总谱的方式用于深度学习。

就目前的研究水平来看,首先,绝大多数的研究不考虑自动作曲的音乐表情问题,也就是说,生产出来的音乐基本是比较机械的,或者说是多种音频采样的组合。其次,由于音色采样和声音合成目前在商业应用领域发展的比较完整,并没有相关的 AI 研究去关注这些方面。再次,游戏音乐因其结构完全取决于游戏场景,所以大部分研究也没有涉及到此类型。

实际上,以“AI 为音乐服务”为宗旨的相关产品已经开始为音乐服务。

例如,人工智能作曲系统 DeepBach 用于复调音乐特别是圣咏类作品的创作。为保证最终效果,该系统主要围绕四声部合唱来工作,并专注于巴赫的四声部合唱类作品的创作。系统采用灵活高效的采样抽样方法,除了通过机器学习的方式生成,使用者还可以在过程中添加音符生成,称为增加一元约束,通过节奏或相关信息对模型进行控制。这种让用户干预过程的方式,是在音乐概率模型中经常被忽略的一种方式。与基于 RNN 的模型相反,DeepBach 不从左到右进行采样。在考虑单个时间方向的情况下,DeepBach 架构会考虑时间上向前和向后的两个方向,使用两个循环网络:一个用于总结过去

^① Fernández, Jose D., Francisco Vico, "AI Methods in Algorithmic Composition: A Comprehensive Survey," *Journal of Artificial Intelligence Research* 48 (2013): 513-582.

^② Ivan P. Yamshchikov, Alexey Tikhonov, *Music generation with variational recurrent autoencoder supported by history*, arXiv:1705.05458 (2017).

的信息,另一个用于汇总来自未来的信息,以及用于同时发生的音符的非递归神经网络。DeepBach 能够生成连贯的音乐短语,并提供各种旋律的重新调和,而不会出现重复。这个系统的工作困难来自于和声与旋律之间错综复杂的相互作用。此外,每个声音都有自己的“风格”和自己的连贯性。找到一种像巴赫式的和声进行,并与音乐上有趣的旋律运动相结合,最终生成类似合唱的音乐是这个系统的目标。对此,我们通过网上调查问卷的方式,对 1200 多人(包括音乐专业人士和业余人士)进行作品听辨测试,结果显示几乎所有人都难以区分这些作品是巴赫创作的还是 DeepBach 创作的。谱例 1 为 AI 生成的作品中的一个片段。

AIVA(Artificial Intelligence Virtual Artist)系统^①则瞄准古典音乐方向,通过大量的巴赫、贝

多芬和莫扎特等经典作曲家的作品教会深度神经网络,从其中提炼出一定的规则并生成音乐。

AIVA 的第一张专辑 *Genesis* 便登陆了 Sound Cloud^②,同时也是作为史上第一位 AI 作曲家获得发行权和版权。目前,AIVA 是一个纯粹的作曲者而不是一个演奏者,它只能生成纸面的音乐,需要人类乐团来演奏,这些音乐可以用于电影音乐、电子游戏和其他相关的商业领域。AIVA 基于深度学习算法并使用强化学习^③的方法,这种方式不需要为输入输出的数据进行标注,所以无需详细说明,机器就可以自主提高性能。

2017 年 Spotify^④ 全球音乐榜排名第二的歌曲是由 IBM 的 Waston Beat^⑤ 人工智能系统参与创作的 *Not Easy*,整个制作过程从多个方面应用了相关的技术来辅助完成。该系统能够

谱例 1 DeepBach 创作的四声部作品



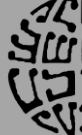
① AIVA系统相关信息参见网站<https://www.aiva.ai/>。

② SoundCloud是全球知名的提供音乐分享社区服务的网站。

③ 强化学习是重要的机器学习方法之一。

④ Spotify是全球知名的正版流媒体音乐服务平台。

⑤ Waston Beat系统相关信息参见网站<https://www.ibm.com/case-studies/ibm-watson-beat>。



在听到大于等于 20 秒的音乐后,基于原始的采样生成一段新的旋律或相关的声音和节奏。整个制作流程中,Watson 首先对近 5 年热门的新闻、文化等音乐相关的数据进行分析、提炼,将音乐创作的核心词定为“心碎”。其次,在具体歌词的创作阶段,Waston 分析了 5 年来近 26000 首歌曲的歌词,提取了博客、推特等社交媒体中的大量信息,分析大众对“心碎”这个词的反应,辅助人类创作歌词。再次,Watson 分析了这些歌曲的风格、乐器、音高、节奏等相关特性,然后建立模型给人类作曲家提供参考。最后,Waston 进行专辑封面的设计。实际上,AI 技术用于音乐的识别效果更好,因为 Waton 每秒钟阅读网页的速度是 700000 页。这个实例证明了目前 AI 技术能够在各个方面影响音乐制作的流程。

由 3 位好莱坞作曲家创立的 AmperMusic^①则瞄准了更广泛的商用音乐领域,他们的目标是生产低成本的背景音乐。AmperMusic 不是作曲家,它更应该被称作演奏者或是制作人。即使对音乐一无所知的人,也可以在 AmperMusic 的帮助下创作自有版权的音乐,只需要点击几个键。和其他系统不一样的是,AmperMusic 没有使用神经网络,也没有使用乐谱的训练,它只是被教会使用音乐的规则和理论,

并通过情绪识别的方式来触发,但也获得了不错的效果。

Magenta^②是 Google 公司的音乐深度学习开发项目,这个项目从 2016 年 6 月份开源以来,为大众提供了很多模型和数据集,为音乐创作、演奏、声音合成等提供方案并持续高速更新。该项目大大促进了世界范围内该领域的研究、应用和普及。

通过比较上述项目和系统后,可以发现深度学习的应用从输入信息上可分为两个种类,即音符输入和波形输入。它们各自在应用中有着不同的特性,参见表 1。

五、结论与思考

从上述应用的实例我们可以看出,利用规则和相关的 AI 算法模型结合的方式来实现音乐创作是相对比较合理的方案之一。在 AI 参与音乐创作的过程中,不仅仅是旋律、歌词,听众的用户数据也会变成它的目标,AI 在全面参与和改变音乐的制作流程,从和声、编曲到最终声音的合成,辅助人类工作并提高效率。

尽管 AI 音乐技术发展迅猛,但距离能真正影响到人类作曲家还有很长一段路要走。AI 依然需要人类来输入音乐规则,更多的情况还是在帮助人们在重复性的劳动里节省时间,特

表 1 不同处理信息的特点比较表

特性 \ 种类	音符数据(包括 MIDI、文字、数字乐谱等)	录音档或波形
计算资源消耗	相对低	相对高
结果可编辑性	可编辑	不可编辑
音乐处理复杂度	等同于单声部音乐	等同于多声部音乐
可获取资源量	少	多
最终音响呈现效果	一般	较好

① AmperMusic 系统相关信息参见网站 <https://www.ampermusic.com/>。

② Magenta 系统相关信息参见网站 <https://magenta.tensorflow.org/>。

别是在商业音乐领域。在快节奏的制作环境中,不断变化的音乐需求需要更高效的手段,这种需求给 AI 提供了应用场景。对于 AI 技术本身来说,最大的挑战是如何理解创造性的艺术思维。目前,计算机很难通过程序来获得理解力,AI 作曲系统也无法取代人类,但是它确实从实质上影响了音乐的生产和创作流程。或许未来在人类音乐家的帮助下,AI 技术会带来更多的冲击。

关于 AI 音乐的远景尚未有定论,但我们更愿意相信 AI 能够解决传统音乐行业的问题,比如降低音乐教育的门槛,提高教学的效率,解决音乐版权的问题等等。相信随着国内外相关研究的投入和关注的迅速增加,会涌现出更多的理论和应用。

◎ 本篇责任编辑 张放

参考文献:

- [1] Fernández, Jose D., Francisco Vico. AI Methods in Algorithmic Composition: A Comprehensive Survey [J]. Journal of Artificial Intelligence Research, 48 (2013): 513-582.
- [2] Ramon Lopez de Mantaras, Josep Lluís Arcos. AI and Music: From Composition to Expressive Performance [J]. AI magazine, 23.3 (2002): 43-43.

收稿日期:2019-03-08

基金项目:2018 年上海音乐艺术发展协同创新中心资助项目;2019 年上海市音乐声学艺术重点实验室资助项目(SKLMA-2019-06)。

作者简介:陈世哲(1983—),男,硕士,上海音乐学院音乐工程系讲师(上海 200031)。

四川音乐学院学报《音乐探索》喜获 第六届“全国高校社科优秀期刊”奖 和“第四届四川省社会科学学术期刊·特色期刊”称号

2019 年 11 月 25 日,经教育部社科司批准认定,由全国高等学校文科学报研究会组织举办的“2019 年第六届高校社科期刊评优活动”正式公布获奖名单。四川音乐学院学报《音乐探索》荣获“全国高校社科优秀期刊”奖,“民族民间音乐”栏目被评为“全国高校社科期刊特色栏目”,学报社副主编李姝荣获“全国高校社科期刊优秀编辑”称号。

2019 年 12 月 20 日,“四川省社科学术期刊协会 2019 年年会”暨“第四届四川省社会科学学术期刊”评奖活动在成都顺利召开。四川音乐学院学报《音乐探索》荣获“特色期刊”称号,学报社副主编李姝荣获“优秀编辑”称号。该评奖活动由四川省社会科学学术期刊协会组织,自 2013 年开始,每两年一次,评选标准严格。

学报社此次获奖,是对川音学报近年来工作成绩的极大肯定。今后,学报社将继续秉持求真务实的工作作风,坚持严谨专业的态度,发扬艺术专业的学科特点,努力打造更高质量的艺术类高校学报。

(学报社)