arXiv:2301.05908v1 [cs.SD] 14 Jan 2023

用于符号化主调音乐乐谱审美质量评估的序复杂度模型

金鑫、周武、王金宇、徐铎、荣怡清、崔帅

摘要

计算美学评估在视觉艺术领域取得了显著成果，但音乐领域的相关研究仍有待深入。尽管现有的音乐生成研究成果颇丰，但与人类作曲家创作的乐谱相比，人工智能生成的乐谱质量仍然较差。人工智能创作的乐谱通常单调乏味、缺乏情感。基于 Birkhoff 的美学度量，本文提出一种用于主调音乐乐谱审美质量评估的客观定量方法。我们的工作主要贡献如下：首先，提出一种主调音乐乐谱审美模型，作为客观评估乐谱质量的基准模型；其次，提出八种基本音乐特征和四种音乐审美特征。

索引词— 计算美学、乐谱评估、伯克霍夫度量、音乐审美特征

1. 引言

计算美学评估 [1]使计算机能够对艺术作品作出定性或定量的审美判断。这些艺术作品通常包括绘画、音乐和设计。让计算机理解美具有重要意义，因为这可以指导人工智能生成任务。

尽管现有的音乐生成技术已相当成熟，但与人类作曲家创作的乐谱相比，人工智能生成的乐谱质量仍然较差。这可能是因为人工智能生成任务的本质是预测下一个将被演奏的音乐单元出现的概率，而缺乏先验音乐理论知识会导致人工智能生成的音乐听起来不悦耳。

流行音乐的制作主要分为三个步骤：乐谱创作、编曲，最后由演奏者演奏。我们希望能够在乐谱创作阶段评估乐谱质量，从而消除不同演奏者的演奏水平对乐谱质量评估造成的干扰。

由于缺乏像图像审美领域中的 AVA [2] 那样针对乐谱的审美评分标注，我们采用传统审美度量方法来研究审美模型。传统审美度量可以从对象所包含的信息的角度分析其美感。

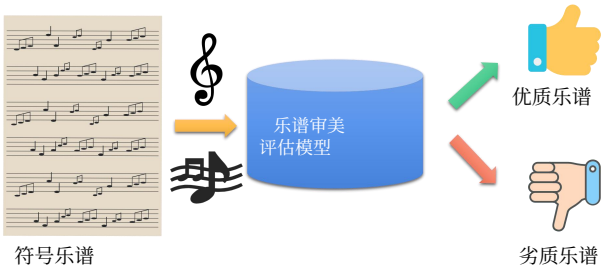

图1。通过乐谱审美评估模型（SAAM），可以轻松评估符号乐谱的质量。

我们的目标是创建一个乐谱审美评估模型，能够客观地区分优劣。

本文选用伯克霍夫方法 [3] 从信息论的角度研究乐谱的审美质量评估。Birkhoff 将对象的审美度量形式化为有序性与复杂度的比值：

$$M = \frac{O}{C} \tag{1}$$

图 1 简要介绍了我们的工作内容。我们的主要贡献如下：

- 我们提出了一种乐谱审美评估模型，作为客观评估主调音乐乐谱质量的基准模型。
- 我们结合信息论和音乐理论，提出并更新了八种基本音乐特征和四种音乐审美特征。

2. 相关工作

目前仅有一项运用信息论的音乐审美度量研究，即音频的审美度量。Audio Oracle [4] (AO) 将信息率（IR）用作审美度量。然而，由于重复或冗余（由 IR 提出）在艺术中具有本质不同的含义、诠释和价值，因此该系统无法明确其具体的审美意图；将信息率用作审美度量也值得质疑。

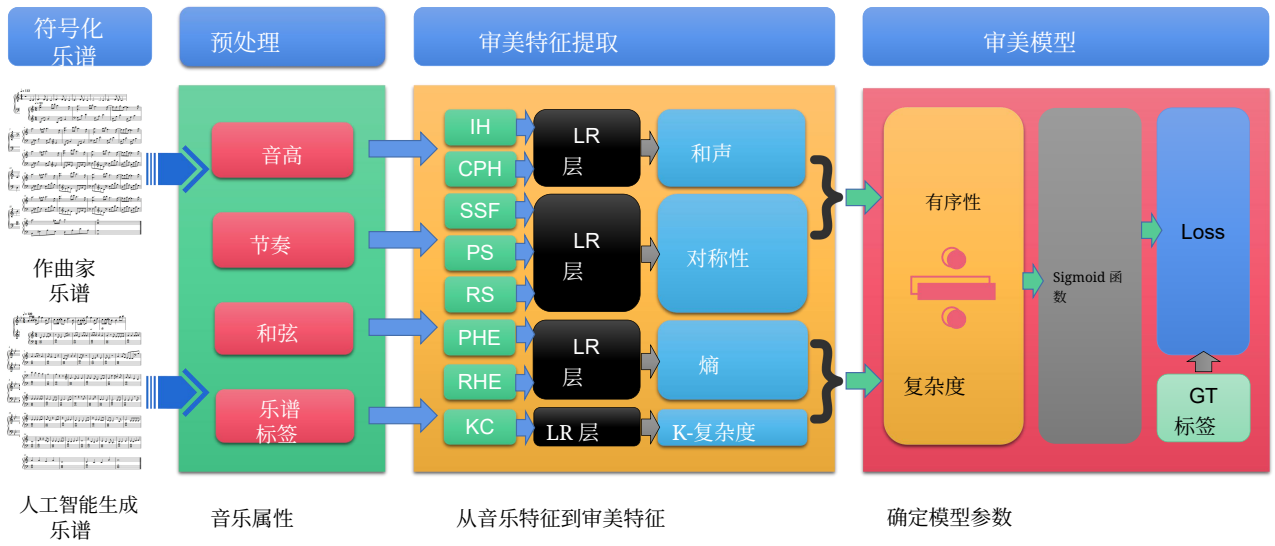

图 2。首先，我们不对符号乐谱进行分词，而是提取符号乐谱中的属性，这需要进行预处理。预处理后，我们得到乐谱的音高、节奏和和弦属性，以及作为分类真实标签的乐谱标签。接着，我们处理音乐属性并提取 8 个音乐特征（绿色小方框）。然后，我们通过四个逻辑回归模型（LR 表示逻辑回归）训练样本，并将其组合以提取四个审美特征（浅蓝色方框）。最后，我们将这四个审美属性输入模型，使用 Sigmoid 函数根据真实标签计算损失，并计算审美模型的参数。

音乐生成分为三个层次：乐谱生成、演奏生成和音频生成。本节讨论的工作与音乐乐谱生成相关。为了充分利用音乐理论并研究复杂度，我们不讨论单声部音乐，而是讨论主调音乐。主调音乐是一种多声部音乐，包含旋律声部和伴奏声部。在主调音乐生成领域，存在诸如直接生成旋律和和弦 [5], 通过和弦生成旋律 [6], 以及根据旋律生成和弦 [7]等任务。尽管现有生成技术已相当成熟，人工智能创作的音乐乐谱质量仍然较低，听起来单调且缺乏情感。

为了评估音乐质量，在音乐生成领域，评估通常分为客观评价和主观评价。客观评价通常是为人工智能生成的音乐设定一些统计指标，其结果完全由计算机计算得出。许多音乐工具包都封装了可直接使用的客观评价指标，例如 MV2H [8] 和 MusPy [9], 等。MV2H 会评估生成音乐与真实标签之间存在多少错误。MusPy 提供了符号乐谱的基本统计指标。主观评价通常包括聆听测试和视觉分析；几乎所有人工智能生成任务都涉及主观评价实验，例如评分和图灵测试，这些都被认为是必要且不可或缺的。

3. 乐谱审美评估模型

3.1. 模型的形式化

基于伯克霍夫理论、信息论和音乐理论，我们提出了四种审美特征：和声、对称性、熵和 K-复杂度。我们将分子的有序度量与分母的复杂度度量进行线性组合。详细的度量说明将在第 3.2 节和第 3.3 节中介绍。图 2 展示了我们的工作流程。音乐审美度量公式如下：

$$Aesthetic\ Measure = \frac{\omega_1 H + \omega_2 S + \theta_1}{\omega_3 E + \omega_4 K + \theta_2} \quad (2)$$

其中， H 是和声， S 是对称性， E 是熵， k是 K 复杂度。 ω 是权重， θ 是常数。

3.2. 有序度量

当物体具有某些和声、对称性或有序性的特征时，往往会带给人一定的美感。我们从和声与对称性两个维度量化音乐的有序性。和声主要依据乐理知识进行计算，而对称性则主要依赖音乐中的一些统计信息。接下来，我们将和声与对称性的线性组合用作有序性的度量。

3.2.1. 音程和声

在音乐中，两个音符之间的距离称为音程。特别是，当音程为 12 个半音时，我们称之为一个八度。音程分类详见补充材料，音程分为五类。

数学和物理研究表明，当两个声音频率之比为简单整数比时，听起来更悦耳。因此，我们提出了一种音程和声的计算方法，公式如下。

$$Interval\ Harmony = \sum_{i=1}^{12} \alpha_i * pir_i + \theta_{ih} \tag{3}$$

其中，α_i 是音程权重，pir_i 是音程占总音程的比例，θ_{ih} 是常数。

3.2.2. 和弦进行和声

根据勋伯格的《和声学理论》 [10], ，内部和弦分为三种功能和声：主和弦（T）、下属和弦（S）和属和弦（D）。和弦功能及序列的示例见补充材料。

完整的和声进行从主和弦开始，接着进行到下属和弦，再进行到属和弦，最后回到主和弦，完成一个完整循环，这称为完整进行。通常来说，和声进行是调性音乐中一定和声范围内和弦的连接。

量化和声进行的方法有很多。本文参考了 Mar´ıa [11] 的方法。在我们的研究中，我们取进行张力的平均值，以获得量化的和弦进行和声。计算方法参见以下公式：

$$\begin{aligned} Chord\ Progression\ Harmony = \lambda_1 d_1(T_i, T_{i-1}) + \\ \lambda_2 d_2(T_i, T_{key}) + \lambda_3 d_3(T_i - T_{key}, T_f) + \lambda_4 c(T_i) + \\ \lambda_5 m(T_i, P) + \lambda_6 h(T_i, P) \end{aligned} \tag{4}$$

其中，T_i 是进行中的第 i 个和弦，P，λ 是权重。有关参数 c、m、h 的更多信息，请参见 [11]。

3.2.3. 自相似度适应度

在音乐生成领域，结构常被视为一项重要特征。几乎所有音乐都包含重复片段。我们将探讨重复结构对音乐美学的影响，并以自相似度适应度进行衡量。

受图像与艺术美学的启发， [12], 音乐的美学之美源于音乐作品中的对称性。因此，在本研究中，我们参考 M¨uller 的适应度方法 [13] 来衡量一部音乐作品的重复程度。适应度公式如下：

$$Self\ Similarity\ Fitness = 2 \cdot \frac{\bar{\sigma}(\alpha) \cdot \bar{\gamma}(\alpha)}{\bar{\sigma}(\alpha) + \bar{\gamma}(\alpha)} \tag{5}$$

$\bar{\sigma}(\alpha)$ 和 $\bar{\gamma}(\alpha)$ 均与 M¨uller 的方法 [13]所定义的一个概念有关。

3.2.4. 偏度

偏度是 jSymbolic 提出的一个概念 [14]。它提出，音乐的音高和节奏都有偏度这一概念。音乐中的音符离不开音高和节奏。偏度描述音高／节奏相对于平均音高／节奏值向左或向右偏斜的程度。

这些特征基于 jSymbolic 提取，本文不具体介绍音高偏度和节奏偏度的计算公式。我们将音高偏度和节奏偏度线性结合，得到偏度公式：

$$Skewness = \beta_1 * PS + \beta_2 * RS + \theta_{sk} \tag{6}$$

其中，PS 是音高偏度，RS 是节奏偏度，β表示它们的权重，θ_{sk} 是一个常数。

3.3. 复杂度度量

Bense [15] 首先使用伯克霍夫美学度量公式计算美学度量。他们将信息的统计度量应用于美学对象，并认为美学对象的客观度量与对象的复杂度有关。他们的理念需要用到信息论，而熵是其中的核心。我们的方法考虑了两个特征：香农熵和柯尔莫哥洛夫复杂度。

3.3.1. 香农熵

设 Ω 为一个有限集，X 为一个随机变量。Ω 中的值 x 服从分布 $p(x) = Pr[X = x]$。随机变量 X 的香农熵 $H(X)$ 定义如下：

$$H(X) = -\sum_{x \in \Omega} p(x) \log(x) \tag{7}$$

香农熵 $H(X)$用于衡量随机变量 X 的平均不确定性，常用于评估系统内部状态的混沌程度。为了计算音乐的熵，需要获得音乐属性直方图。

音高和节奏是音乐的两个基本要素。在我们的方法中，我们考虑音高熵和节奏直方图熵，并将二者线性组合，作为熵的度量，用于描述音乐中的不确定性。我们的熵公式如下：

$$Entropy = \eta_1 * PHE + \eta_2 * RHE + \theta_e \tag{8}$$

其中，PHE 和 RHS 分别为音高和节奏直方图熵，η 表示它们的权重，θ_e 为常数。

3.3.2. 柯尔莫哥洛夫复杂度

对于字符串 s，其柯尔莫哥洛夫复杂度 $K(s)s$ 指的是在计算机上计算该字符串 s 所需的最短程序。本质上，字符串的柯尔莫哥洛夫复杂度就是该字符串最终压缩版本的长度。然后，我们使用熵与柯尔莫哥洛夫复杂度的线性组合来衡量复杂度。

从美学角度来看，冗余会让人感到乏味，从而产生负面情绪。根据柯尔莫哥洛夫复杂度的定义，我们也借鉴了在图像美学中使用柯尔莫哥洛夫复杂度的方法 [16]。

我们认为，音乐中的柯尔莫哥洛夫复杂度也是可计算的。它实际上是音乐的无损压缩率，可形式化为以下公式：

$$Kolmogorov\ Complexity = \frac{NH_m - K}{NH_m} \tag{9}$$

其中，NH_m 是音乐的信息量，K 是压缩后最简化的音乐信息。

4. 实现

4.1. 数据集

流行音乐有许多数据集，例如 POP909 [17]和 Wikifonia[1]。但由于一些原因，我们没有使用它们。

POP909 数据集中的所有音乐数据格式都是演奏 MIDI，这会导致不同演奏者对同一乐谱产生不同的审美影响。尽管 Wikifonia 是一个乐谱数据集（单轨），但它无法用于有序性测量。

为了消除不同演奏者之间的演奏差异，并客观地测量一首音乐作品的审美价值，我们最终从 Musescore[2] 网站下载了 100 首流行歌曲的乐谱。然后，我们提取 Musescore 中乐谱的和弦，用于生成 100 份乐谱。具体而言，我们从 Musescore 的乐谱中提取和弦进行，并确定乐谱的调号（主音）作为输入参数，再将其输入 Magenta 的 in- prov rnn [6] ，使其能够根据和弦进行和调号生成音乐。这种方法的优点是能够将两组数据集的时长控制在大致相同的范围内。Musescore 中的乐谱由作曲家创作，作为正样本；而 Magenta 创作的乐谱由人工智能生成，作为负样本。

由于我们使用交叉验证，因此不设置验证集。我们按 7:3 的比例划分数据集，用于训练和测试。

4.2. 预处理与美学特征计算

我们使用 music21 [18]工具包加载乐谱。music21 具有音符和和弦属性，可轻松获取乐谱信息并进行计算。

首先，我们获取乐谱中的音高直方图和节奏直方图，以计算熵。然后，通过计算同时发生的音符事件，获取所有音程并生成直方图。由此可以计算音高、节奏和音程和声的直方图熵。

其次，我们尝试获取乐谱中的所有和弦和调号，以计算和弦进行和声。和弦进行和调号将以 JSON 格式保存，随后输入预训练模型。

第三，我们使用 Musescore3 批量处理作曲家乐谱和人工智能乐谱，将其转换为乐谱 MIDI，以便使用 jSymbolic [14]提取特征。从 XML 转换为乐谱 MIDI 不会丢失乐谱信息。这样，我们便能得到音高偏度和节奏偏度。

第四，我们使用 Musescore3 将乐谱渲染为音频，从而确保所有乐谱均由 Musescore3 播放，保证乐谱信息无损。然后，我们使用 WAV 文件格式计算自相似度适应度。接着，参考 Monkey's Audio 的[3] 无损压缩方法，将音乐压缩为 APE 格式，以计算柯尔莫哥洛夫复杂度。

第五，计算 8 个基本音乐特征的值（参见表1）后，我们对其进行归一化。接下来，我们使用逻辑回归方法确定和声、对称性、熵和柯尔莫哥洛夫复杂度的值。

最后，我们将四个归一化的审美特征作为输入，使用 Sigmoid 函数将审美度量映射到 0 和 1 之间以进行分类，并使用交叉熵损失构建损失函数，通过梯度下降的方式确定审美模型的参数。我们将学习率设为 0.01。经过 1000 次迭代后，损失函数收敛。

5. 实验

5.1. 图灵测试

我们需要验证人们是否真的认为作曲家创作的音乐比人工智能生成的音乐更优美。因此，我们进行了图灵测试，以了解人们是否真的能够区分作曲家的音乐和人工智能生成的音乐。

我们分别从作曲家创作的音乐和人工智能生成的音乐中随机抽取了10首曲子，每首时长约15秒。参与图灵测试的志愿者需要在这10对曲子中辨认出哪些是作曲家创作的，哪些是人工智能生成的。志愿者还需要从每对曲子中选出自己认为更优美的一首。

共有15名志愿者参与了图灵测试。在300个样本中，他们对音乐的分类准确率为91.3%（274个）。假设作曲家创作的音乐更具审美性，则87%（261个）样本被正确分类。这证明了我们的假设是正确的。

[1] http://www.wikifonia.org/ [2] https://musescore.org/

[3] https://www.monkeysaudio.com/

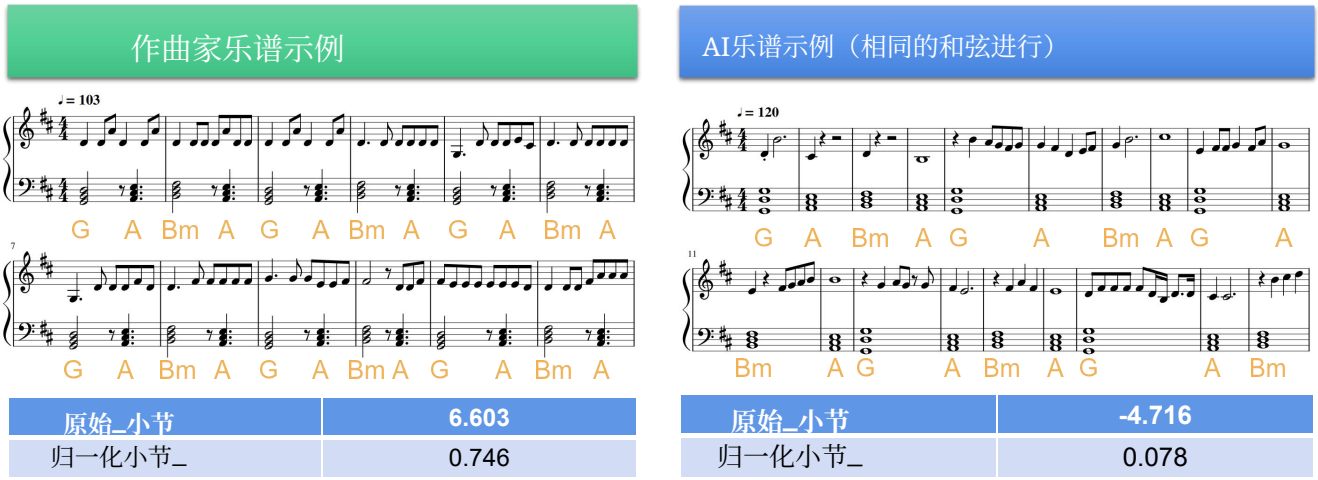

图3。如图所示，作曲家乐谱的旋律井然有序，并且与和弦进行十分契合。人工智能乐谱的旋律则显得十分随机、质量低下，具体表现为休止符出现不规则，且旋律不遵循和弦进行的张力。**注：归一化小节的总分为 1**。

5.2. 结果与讨论

由于作曲家创作的音乐比人工智能生成的音乐更具审美感，我们让机器根据标签值学习审美评分。这本质上是一个二元分类问题。

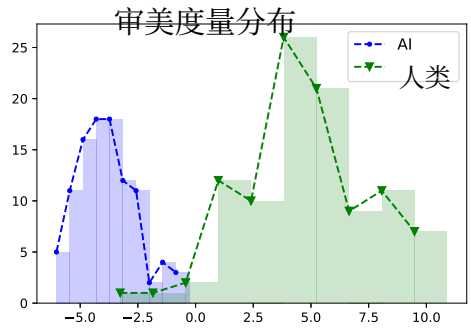

图4。各分布之间的交集很小，表明我们的模型能够很好地区分人工智能乐谱和作曲家乐谱。

通过梯度下降获得权重后，我们将其代入审美模型，以观察审美模型的分布。该分布如图4所示。

我们使用精确率和F1度量作为评估模型的指标。模型在测试集上的精确率为93.3%，F1度量为90.9%。这证明了模型的有效性。图3展示了一个乐谱比较示例。

表1展示了未经归一化的8项特征的计算结果。我们可以据此进行以下分析。无论是 IH 还是 CPH，作曲家的乐谱显然都优于人工智能生成的乐谱。这也符合音乐理论。就对称性而言，作曲家作品的 SSF 显然高于人工智能作品。这是因为人工智能创作的音乐过于随机，往往没有重复片段，这证明美与重复有关。虽然人工智能作品的 PS 小于作曲家的乐谱，但差异不大。作曲家作品的 RS 显然小于人工智能作品。这是因为与人工智能相比，作曲家倾向于创作节奏和音高更为规律的音乐。至于熵，作曲家乐谱中的 PE 和 HE 值显然高于人工智能乐谱。尽管这与 Birkhoff 的审美度量相悖，但也合乎情理。这是因为作曲家倾向于在音高和节奏上加入一些变化，而人工智能生成的音乐围绕主音的变化不大。至于 K-复杂度，作曲家乐谱与人工智能乐谱之间的差异并不显著，但从表中可以看出，人工智能创作的音乐在压缩率和复杂度方面相对较低。总之，适度的有序性和复杂度商可以在一定程度上量化审美感受。

数据集	IH	CPH	SSF	PS	RS	PHE	RHE	KC
AI	0.97	1.98	0.06	**0.49**	2.16	**1.34**	**1.60**	**0.69**
作曲家	**1.40**	**2.04**	**0.19**	0.53	**0.99**	2.16	1.79	0.73

表1。相似的CPH值表明，基于和弦生成的人工智能数据集具有合理性。**更详细的分布比较见补充材料**。加粗的结果具有更高的审美分数。

5.3. 消融研究

我们分别移除和声、对称性、熵和 K-复杂度，进行消融实验，训练出四个不同的模型，并将它们与原始模型进行比较。

如图 5 所示，我们观察了原始模型和分别移除一种审美特征后的四个模型的 ROC 曲线，并计算了它们的曲线下面积。我们的模型的曲线下面积为 0.93，明显高于这四个移除了审美特征的模型。

如果去除和声，曲线下面积值仅为 0.77，这表明和声的重要性，也进一步证明音乐理论在音乐美学中发挥着非常重要的作用。如果去除对称性，模型的曲线下面积值为 0.85，这表明对称性对美学的贡献可能略低于和声。至于熵，它显然很重要，但 K-复杂度似乎并非如此。

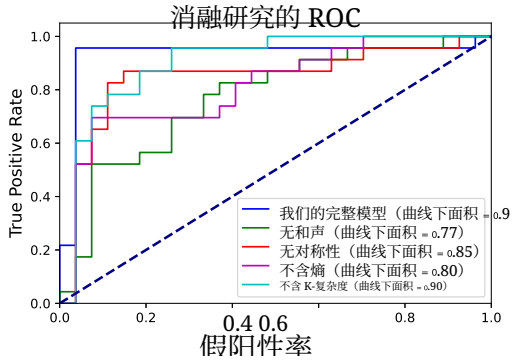

图 5。由于样本不足，ROC 曲线呈锯齿状。

6. 结论

总之，我们提出了一种乐谱审美评估模型，利用伯克霍夫审美度量量化乐谱的审美价值。我们还发现了四类音乐审美特征，共计八种基本审美特征。我们为提升乐谱质量做出了一些贡献，这可能有助于乐谱质量评估。然而，我们的方法仍有不足之处，例如尚未考虑创造力与音乐美学之间的关系。未来值得探索音乐音频质量评估的美学研究。

7. 参考文献

[1] Philip Galanter，《计算美学评估：迈向机器创造力》，载于*ACMSIGGRAPH2012 Courses*，第1–162页。2012年。[2] Naila Murray、Luca Marchesotti 和 Florent Perronnin，《AVA：一个大规模审美视觉数据库》，载于*2012*年*IEEE*计算机视觉与模式识别会议。IEEE，2012年，第2408–2415页。[3] George David Birkhoff，《审美度量》，载于审美度量。哈佛大学出版社，2013年。[4] Shlomo Dubnov、G´erard Assayag 和 Arshia Cont，《音乐信息率的音频预言机分析》，载于*2011*年*IEEE*第五届语义计算国际会议。IEEE，2011年，第567–571页。[5] Yu-Quan Lim、Chee Seng Chan 和 Fung Ying Loo，《风格条件音乐生成》，载于*2020*年*IEEE*多媒体与博览会国际会议（*ICME*）。IEEE，2020年，第1–6页。[6] 谷歌，《inprov rnn》，https://github.com/ _ magenta/magenta/tree/main/magenta/models/improv_rnn，2018年。[7] Hongyuan Zhu、Qi Liu、Nicholas Jing Yuan、Kun Zhang、Guang Zhou 和 Enhong Chen，《流行音乐生成：从旋律到多风格编曲》，*ACM*知识发现与数据挖掘汇刊（*TKDD*），第14卷，第5期，第1–31页，2020年。[8] Andrew McLeod 和 Mark Steedman，《自动复调音乐转录评估》，载于*ISMIR*，2018年，第42–49页。[9] Hao-Wen Dong、Ke Chen、Julian McAuley 和 Taylor Berg-Kirkpatrick，《MusPy：用于符号音乐生成的工具包》，*arXiv*预印本 *arXiv:2008.01951*，2020年。[10] Arnold Schoenberg，《和声学理论》，加州大学出版社，1983年。[11]Mar´ıa Navarro-C´aceres、Marcelo Caetano、Gilberto Bernardes、Mercedes S´anchez-Barba 和 Javier Merch´an S´anchez-Jara，《调性音程空间中和弦进行调性张力轮廓的计算模型》，熵，第22卷，第11期，第1291页，2020年。[12] Mohammad Majid al Rifaie、Anna Ursyn、Robert Zimmer 和 Mohammad Ali Javaheri Javid，《论对称性、美学与对称复杂度的量化》，载于国际进化与生物启发音乐及艺术会议。施普林格，2017年，第17–32页。[13] Meinard M¨uller、Nanzhu Jiang 和 Peter Grosche，《一种用于捕捉音乐录音中重复段的稳健适应度度量及其在音频缩略图生成中的应用》，*IEEE*《音频、语音与语言处理汇刊》，第21卷，第3期，第531–543页，2013年。[14]Cory McKay、Julie Cumming 和 Ichiro Fujinaga，《JSymbolic 2.2：提取符号音乐特征以用于音乐学与MIR研究》，载于*ISMIR*，2018年，第348–354页。[15] Max Bense，《美的一般文本理论与文本美学的程序化》，1960年。[16] Jaume Rigau、Miquel Feixas 和 Mateu Sbert，《信息美学度量》，*IEEE*《计算机图形学与应用》，第28卷，第2期，第24–34页，2008年。[17] Ziyu Wang、Ke Chen、Junyan Jiang、Yiyi Zhang、Maoran Xu、Shuqi Dai、Xianbin Gu 和 Gus Xia，《POP909：用于音乐编曲生成的流行歌曲数据集》，*arXiv*预印本 *arXiv:2008.07142*，2020年。[18] Michael Scott Cuthbert 和 Christopher Ariza，《music21：用于计算机辅助音乐学与符号音乐数据的工具包》，2010年。