

Melody 作曲助手：实现大模型生成修改 MIDI 文件

让人人都是音乐艺术家

Melody:WEB3 音乐革命领导者

摘要

随着深度学习和大模型在艺术创作领域的迅速发展，音乐生成与自动作曲成为近年来备受关注的方向。然而，大多数生成系统仅提供简单的旋律或和弦，难以兼顾演唱风格和伴奏等多维音乐要素。为解决这一痛点，本文提出了一种基于 **DeepseekR1** 大模型的“Melody 作曲助手”，这是世界上首个能够对 Midi 文件进行“千锤百炼”的、真正对艺术家友好且高效的作曲助手系统。该系统通过提示词工程（prompt engineering）与对用户声音特征进行个性化建模，不仅能生成旋律、和弦、伴奏，还能根据用户输入生成特定风格的人声演唱，从而满足艺术家对音乐创作多样化与高度定制化的需求。

1. 引言

1.1 研究背景

人工智能在音乐创作领域的应用已经有一定历史。从早期的基于规则的作曲系统，到近几年大火的深度学习生成音乐模型，人们不断探索如何让机器更好地“理解”音乐和协助音乐创作。深度神经网络的崛起极大促进了音乐 AI 的进步，其中对乐理的理解、旋律生成以及和声编配等功能日益完善。然而，传统系统往往只关注旋律或伴奏层面，无法与音乐创作者的个性化需求进行深度融合。

1.2 研究动机

在实际音乐制作场景中，歌曲生成不仅包括主旋律的构建，还包括对和声 (Chord Progression)、配器编配、速度节拍、演唱风格以及对歌手声音特征的还原等多方面要求。艺术家尤其注重最终成品的情感表达和可听度。常规的“自动作曲”或“AI 作曲”系统功能大多单一，若将它们集成到真正的音乐制作流程中，就需要人工额外二次或多次修改，从而耗费大量时间与精力。

本研究的目标在于构建一款面向艺术家的 AI 作曲助手系统——**Melody 作曲助手**：

1. 结合深度学习大模型 **DeepseekR1** 对歌词及创作指令进行理解，确定曲风、和声、编配形式和关键乐理信息；
2. 实现主旋律及伴奏自动生成；

3. 深度定制化的声音模型与演唱模型，将用户特定声线及风格体现在成品音乐中；
4. 通过交互式“千锤百炼”（对 Midi 进行多轮优化调整）的方式，显著减少人工修改成本，提升音乐成品质量。

1.3 论文结构

下文首先介绍 DeepseekR1 大模型相关的技术背景；接着阐述 Melody 作曲助手系统框架与工作流程；随后给出系统实现细节及关键算法描述，并在实验与案例分析部分展示了具体生成效果；最后对本系统的优势与不足进行讨论，并展望未来研究方向。

2. 相关研究

2.1 深度学习模型在音乐生成中的应用

自 GAN（生成对抗网络）、RNN（循环神经网络）以及变换器(Transformer)等深度学习框架兴起以来，音乐生成的研究变得更加多元化。Music Transformer、MuseNet 等模型在旋律与和声上的生成已经取得了较好的效果。但由于音乐本身是多层次、多维度的组合，不同乐器、和声风格和声学特性对最终音乐作品影响深远，传统模型往往难以一次性生成高度可用的 MIDI 文件或音频。

2.2 提示词工程 (Prompt Engineering)

随着大模型（Large Language Model, LLM）的强大语义理解和生成能力显著提升，提示词工程（Prompt Engineering）成为新一代 AI 应用研发的核心。通过精心设计的 Prompt，模型可以理解更复杂的需求与上下文，并生成贴合用户目标的内容。在音乐创作场景中，不仅要描述风格、速度、调性，更需要说明和弦走向、伴奏方式以及对演唱风格的期望。这正是提示词工程需要解决的多维度指令理解问题。

2.3 声音特征与个性化演唱生成

深度学习在语音领域的进步，使得基于歌手声音特征的演唱合成成为可能。通过收集歌手或用户样本，并进行特征提取与声纹建模，可以在训练好的声学模型中生成符合用户嗓音特质的演唱音色，这对于个性化作曲与演唱尤为关键。

3. Melody 作曲助手系统框架

本系统的核心在于 **DeepseekR1 大模型** 对作曲流程的管理与“千锤百炼”的多轮交互式 Midi 修改功能。整体框架如图所示（此处可附示意框图）：

1. 输入层

- **歌词与创作指令**：用户输入歌词文本、曲风（流行、摇滚、民谣等）、速度（BPM）偏好等，甚至可细化到和弦走向、编配形式（柱式、半柱式、分解和弦等）等具体需求。
- **用户声音特征**：从用户已录制的音频样本中提取声音特征，用于后续生成个性化演唱音色。

2. DeepseekR1 核心生成层

- **Prompt Engineering**: 将用户需求转化为可被 DeepseekR1 理解的提示词, 指示模型如何在旋律、和声、编配与演唱方面进行生成。
- **曲风识别与音乐特征规划**: DeepseekR1 根据提示自动输出速度、时长、key (调性)、风格、和声走向 (如 4536251) 等信息。
- **多轮迭代与 Midi 生成**: 模型初步生成 Midi 文件后, 根据用户或系统内部评估结果, 进行 “千锤百炼” 的反复迭代, 优化旋律线、和声衔接和编配方式。

3. 伴奏与人声合成模块

- **伴奏生成**: 利用从 DeepseekR1 输出的编配策略 (如分解、柱式、半柱式和弦走向等), 自动生成吉他、钢琴或其他指定乐器的伴奏轨道。
- **主旋律与和声合成**: 基于 Midi 音符信息, 引入音乐音色库进行合成与对齐。
- **个性化演唱生成**: 在声学模型中载入用户声音特征, 匹配 Midi 旋律, 把用户嗓音特征嵌入到演唱声音生成模型中, 从而实现 “AI 演唱”。

4. 输出层

- 生成**高质量 Midi 文件**, 并可导出 WAV 或 MP3 格式的**完整音频**。

- 。提供可视化界面，用于“千锤百炼”式的迭代修改，包括更换乐器、微调整奏或音高、变换和弦走向等。

4. 关键技术与实现细节

4.1 DeepseekR1 大模型的 Prompt 设计

为了让 DeepseekR1 准确理解音乐创作需求，需要精细化设计 Prompt。示例：

- 歌曲歌词：
[用户输入]
- 目标曲风：流行（Pop）
- 目标速度：120 BPM
- 伴奏形式：钢琴分解和弦
- 期望的和声走向：4536251
- 主旋律风格：明亮、朗朗上口
- 演唱对象：用户 A 的声音特征
- 需考虑的特殊指令：可插入 16 小节的间奏

通过上述 Prompt 示例，DeepseekR1 能够生成初步的 Midi 旋律和编配信息，包括**音高**、**节拍**、**时值**、**和弦**等。若对结果不满意，Melody 作曲助手可以自动或在用户指令下再次修改 Prompt，如调整和弦进行或节奏方式，并重新生成。

4.2 Midi 的“千锤百炼”迭代方法

本系统的创新之处之一，在于对 Midi 进行多轮迭代的优化机制。其核心流程如下：

1. **初稿生成**：DeepseekR1 根据初始 Prompt 给出 Midi 草稿；
2. **质量评估**：系统通过一系列规则和训练过的检测模型，对旋律的流畅度、和声进行的完整度、与歌词匹配度进行评分；
3. **误差反馈**：若出现过多不和谐音程、节拍错乱或歌词与旋律的韵律冲突，系统会将误差反馈写入下一个 Prompt；
4. **二次生成**：DeepseekR1 读取改进后的 Prompt，生成更优版本的 Midi；
5. **手动调节**（可选）：用户手动介入，如特定小节微调或更换和弦或乐器，系统更新 Prompt 并再次生成。

通过该多轮循环，不断打磨 Midi 文件，直到满足用户需求或达到可用成品的水平。

4.3 个性化演唱合成

当旋律和伴奏都稳定后，Melody 作曲助手会将最终 Midi 文件喂入**歌声合成 (Singing Voice Synthesis, SVS) **模块。在 SVS 模块中：

1. **用户声音特征提取**：利用 DNN 或 Transformer 声学模型，提取用户声纹特征和语音音色空间表示；

2. **旋律与歌词对齐**：根据 Midi 旋律音高及时值，与歌词在音素层面进行同步；
 3. **合成演唱**：将 Midi 音符作为条件输入，将用户音色向量注入声学解码器，实现个性化的演唱声音输出；
 4. **风格化处理**：若需要，可以在模型中添加转音、气声或颤音等风格细节，以更贴近真实歌者的演唱习惯。
-

5. 实验与案例分析

5.1 用例 1：流行情歌《你的眼神》

1. **输入**：用户提供浪漫爱情题材歌词，期望弹唱风格、速度 120 BPM，和弦走向设置为 6451 循环，希望伴奏以钢琴为主、吉他点缀。
2. **生成过程**：
 - DeepseekR1 解析歌词主题与情感基调，自动生成一条悠扬流畅的旋律；
 - 伴奏模块给出钢琴分解和弦，并在间奏部分添加吉他琶音；
 - 用户声音特征接入 SVS 模型，生成温柔且具有气息感的演唱音色。
3. **结果评价**：
 - 旋律与歌词情绪较匹配，听感浪漫；
 - 通过多轮迭代，对和弦细节进行微调，最后成品在主歌与副歌衔接、高潮处理等方面达到了较高水平。

5.2 用例 2：摇滚风格《前行的力量》

1. **输入：**励志题材歌词，期望速度 160 BPM，风格摇滚（Rock），希望鼓点节奏鲜明，吉他失真效果突出，和弦走向为 4536251。
 2. **生成过程：**
 - 在提示词中明确鼓的节奏类型和电吉他的失真度；
 - DeepseekR1 输出带有强烈鼓点和雄浑吉他 riff 的 Midi，并安排副歌部分带有和声的“呐喊”效果；
 - 演唱模型替换为用户 B 的声音特征，并增加演唱中的“破音”元素以彰显摇滚氛围。
 3. **结果评价：**
 - 最终生成的旋律富有张力，电吉他与鼓节奏融合度较高；
 - 多轮修改后，整体动态感更强，副歌处能突出励志气势。
-

6. 讨论与展望

6.1 系统优势

1. **多维度协同生成：**可同时生成旋律、伴奏和人声演唱，有效提升音乐创作效率。
2. **个性化与灵活性：**利用用户声音特征，深度定制化音色；通过提示词工程，轻松切换不同风格、速度、和声走向。

3. **迭代式优化**：通过“千锤百炼”的 Midi 多轮交互，大大降低人工修改成本。
4. **可扩展性强**：可以进一步扩充不同乐器库、风格模板以及歌词语言模型，形成更加多元化的曲风。

6.2 系统局限性

1. **风格极端化问题**：对于极端风格（如先锋实验音乐、无调性作品）可能需要更精细的 Prompt 和乐理指导。
2. **音色合成技术限制**：合成的人声仍可能与真人演唱存在细微差异，难以完全模拟某些高级的气息和情感变化。
3. **数据依赖度**：DeepseekR1 和 SVS 模型的效果与训练数据质量密切相关，需要持续维护与更新。

6.3 未来工作

1. **实时交互**：将 Melody 作曲助手打造成为实时“对话式”创作工具，用户可随时输入修改点并即时听取变化。
2. **更智能的评价体系**：引入更多音乐美学评价指标或 AI 评审，以自动判断音乐的创意性和商业化潜力。
3. **跨文化与多语言适配**：扩展对各类语言歌词的理解与生成能力，包括中文、英文、日语、韩语等多语种。
4. **与 VR/AR 融合**：在虚拟现实和增强现实环境下，音乐的生成与演唱可与视觉和交互结合，进一步拓展艺术表现方式。

7. 结论

本文提出了一种基于 DeepseekR1 大模型的 Melody 作曲助手，通过提示词工程实现对歌词、风格、速度、和弦及伴奏形式等多维度指令的理解和生成，同时结合个性化演唱模型，将用户的声音特征注入音乐作品之中。通过迭代式 Midi 优化（“千锤百炼”）方式，极大减轻了后期人工编辑压力，提升了音乐创作效率和质量。实验结果显示，Melody 作曲助手能够在流行、摇滚、民谣等不同风格中生成高质量的旋律与编配。随着深度学习和大模型技术的进一步演进，我们有理由相信，Melody 作曲助手将在音乐教育、商业制作、个人创作等领域产生深远影响。

参考文献（示例）

1. Huang, C. Z. A., Vaswani, A., & Uszkoreit, J. (2018). Music Transformer: Generating Music with Long-term Structure. *International Conference on Learning Representations (ICLR)*.
2. Payne, C. & Tyack, G. (2019). MuseNet: A Large-Scale Generative Model of Music. *OpenAI Blog*.
3. Lu, S. (2021). Prompt Engineering for Large Language Models. *IEEE Transactions on Neural Networks & Learning Systems*.

4. Lee, J. et al. (2022). Singing Voice Synthesis with Personalized Voice Characteristics. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

作者: XXX, 2025 年