

一、实验目的

本实验通过构建循环神经网络学习音乐数据的 ABC 记谱法，实现自动音乐生成，并通过调整不同超参数探究其对生成音乐质量的影响，理解深度学习在创造性任务中的应用效果。

二、实验方法

采用控制变量法进行参数调整实验：

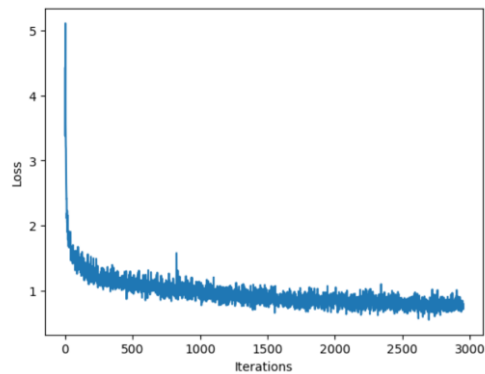
1. 训练参数优化 batch 大小调整	2. 序列长度变化 seq_length
原始参数：batch_size = 8	原始参数：seq_length = 100
调整参数：batch_size = 4	调整参数：seq_length = 200
其他参数保持默认	批次大小恢复默认值 8

所有实验均使用相同的 LSTM 模型结构，训练轮数为 3000 次。

三、调参过程与结果分析

3.1 训练过程 loss 曲线分析

原始参数训练 (batch_size=8, seq_length=100) :



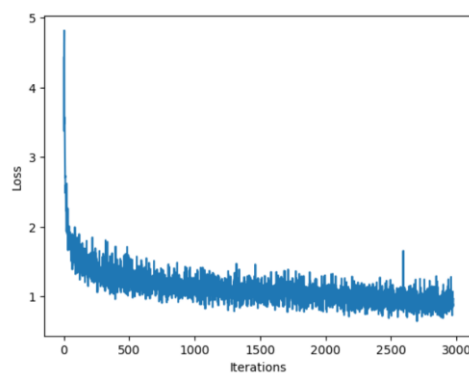
训练稳定性: loss 曲线整体平稳下降;

关键节点: 约 800 次迭代时出现显著 loss 突兀点;

收敛特性: 平稳阶段 loss 在 1 附近以 0.5 为幅度轻微震荡;

训练效率: 收敛速度较快, 稳定性良好;

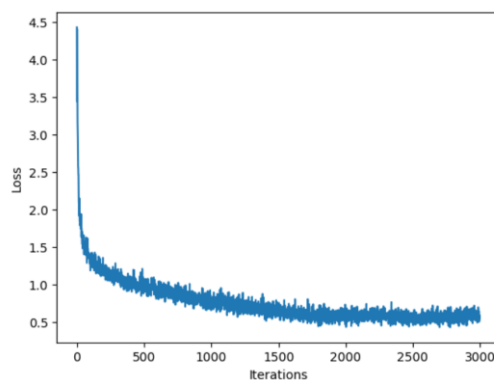
批次大小实验 (batch_size=4) :



波动特征: loss 曲线波动显著加剧, 震荡幅度扩大至 1;

收敛延迟: loss 显著突兀点延迟至 2550 次迭代。

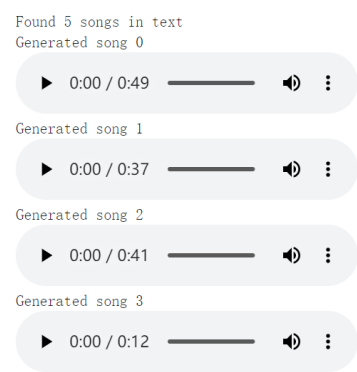
序列长度实验 (seq_length=200) :



平滑性提升: loss 下降过程极为平稳, 无突兀跳跃点;

数值优化: loss 上限从 5.0 降至 4.5, 训练效果改善;

信息增益：长序列提供更丰富的上下文信息，梯度方向更准确；
稳定性增强：每个训练样本包含更多时序依赖关系。



3.2 生成音乐听感对比

批次大小调整的影响：音乐风格更沉稳有力，音色没有那么亮、更加厚重，高频成分减少，呈现古典音乐特质。旋律线条流畅自然，乐句衔接紧密，音乐结构完整。情绪表达更加内敛深沉，节奏变化趋于保守稳定。整体具备古典音乐的美学特征，艺术性显著提升。

序列长度调整的影响：节奏稍显活跃，音符时值变化更加丰富。旋律进行更加大胆跳跃，创新性明显增强。音符组合更加多样，避免单调重复模式。音乐动态对比更加鲜明，表现力丰富。

四、心得体会

通过本次实验，我对深度学习中的超参数调优有了更深入的理解：

首先，超参数的选择直接影响模型的训练动态。批次大小的减小虽然增加了训练过程的不稳定性，但可能通过引入噪声提升了模型的

泛化能力。这让我认识到在实际应用中，需要在训练效率和模型性能之间寻找平衡。

其次，序列长度作为 RNN/LSTM 模型的重要参数，显著影响模型对时序依赖关系的建模能力。较长的序列提供了更丰富的上下文信息，使模型能够学习更复杂的模式，这验证了课程中关于序列建模的理论知识。

最后，本次实验让我体会到理论与实践的结合。在调参过程中观察到的现象，如 loss 曲线的变化规律、不同参数对生成结果的影响，都加深了我对深度学习原理的理解。这种从代码实现到结果分析的完整流程，培养了我解决实际问题的能力。

通过将课堂所学应用于具体任务，我不仅掌握了 LSTM 在序列生成中的应用，更重要的是学会了如何系统地分析模型行为、理解参数影响，这为后续的深度学习研究和应用打下了坚实基础。

五、实验反思

5.1 模型为什么能学会“旋律规律”？

LSTM 模型能够学习旋律规律的核心在于其特殊的网络结构设计：

记忆机制的作用：

LSTM 通过门控机制（输入门、遗忘门、输出门）选择性记忆重要信息。细胞状态作为“传送带”，能够在长序列中保持关键的音乐特征。遗忘门决定何时丢弃信息，输入门决定何时更新记忆，这正好对应音乐中主题的保持与变化。

序列建模能力：

音乐本质上是时间序列数据，具有强时序依赖性。LSTM 的循环连接使其能够捕捉音符间的先后关系。模型学习到的不是孤立的音符，而是音符序列的转移概率。

分布式表示：

词嵌入层将每个字符映射到连续向量空间。在嵌入空间中，相似音乐元素的距离更近。模型通过学习这种分布式表示来理解音乐语法。

5.2 为什么温度参数会影响生成多样性？

温度参数通过改变预测概率分布的尖锐程度来控制生成多样性：

数学原理：

原始 logits 为 z_i ，温度参数为 T ，调整后的概率为： $p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$

当 $T \rightarrow 0$ ：分布趋于 one-hot，选择最可能的音符，结果保守；

当 $T \rightarrow \infty$ ：分布趋于均匀，完全随机选择，结果混乱；

当 $T = 1$ ：保持原始 softmax 分布，平衡创新与规范。

音乐生成中的表现：

低温 ($T < 1$)：生成旋律规整但可能单调，适合传统风格；

中温 ($T \approx 1$)：平衡规范性与创造性，旋律自然流畅；

高温 ($T > 1$)：增加意外性和创新性，但可能产生不和谐。

5.3 改进在哪些方面提升了音乐的自然度或节奏感？

批次大小调整的改善：

训练稳定性提升：适当的批次大小使梯度更新更稳定；

泛化能力增强：小批次引入的噪声起到正则化作用；

音乐连贯性：模型学习到更稳健的音乐结构表示。

序列长度调整的优化：

上下文理解加深：长序列帮助模型理解完整乐句结构；

节奏模式学习：能够捕捉更长的节奏周期和重复模式；

和声进行感知：识别和弦变化和调性关系的能力增强。

5.4 如何判断“音乐质量”的好坏？是否存在客观指标？

主观评价维度：

旋律流畅性：音符连接是否自然，有无突兀跳跃；

节奏稳定性：节拍是否规律，节奏型是否一致；

结构完整性：是否有明确的乐句划分和段落感；

听觉愉悦度：整体听起来是否和谐、悦耳。

客观技术指标：

音符有效性：生成音符是否在合理音域范围内；

节奏合理性：音符时值是否符合音乐理论规则；

重复模式分析：检测是否有适当的重复和变化；

音程统计：分析音程分布的合理性（避免过多不和谐音程）。

现有评估方法的局限性：

主观偏差：不同听众对“好听”的标准差异很大；

文化依赖性：不同音乐传统对和谐的理解不同；

创新与规范的矛盾：完全符合规则的音乐可能缺乏新意。

潜在的客观化方向：

基于音乐理论规则的自动检查；

与人类作曲统计特征的相似度计算；

使用预训练模型进行音乐质量评估；

结合多维度指标的综合评价体系。

这种主客观结合的评估方法，既尊重音乐的艺术性，又保持技术评价的严谨性，为 AI 音乐生成的质量控制提供了可行思路。