

实验报告：基于 LSTM 的音乐生成实验

一、实验目标

1. 理解序列建模（Sequence Modeling）与循环神经网络（RNN/LSTM）的基本原理；
2. 掌握音乐数据的数值化（MIDI → 序列 → 模型输入）的过程；
3. 熟悉 Notebook 中的模型结构与训练流程；
4. 尝试通过调整模型结构或参数，提升音乐生成的质量；
5. 输出一段可播放的旋律文件（.mid）。

二、实验方法

通过使用 PyTorch 实现基于 LSTM 的音乐生成模型。

首先，利用 MIT Deep Learning Lab 提供的 `mitdeeplearning.lab1` 数据集加载数百首训练乐曲，将所有曲谱文本拼接后建立字符到索引和索引到字符的映射，以便将字符序列转换为可输入模型的张量。

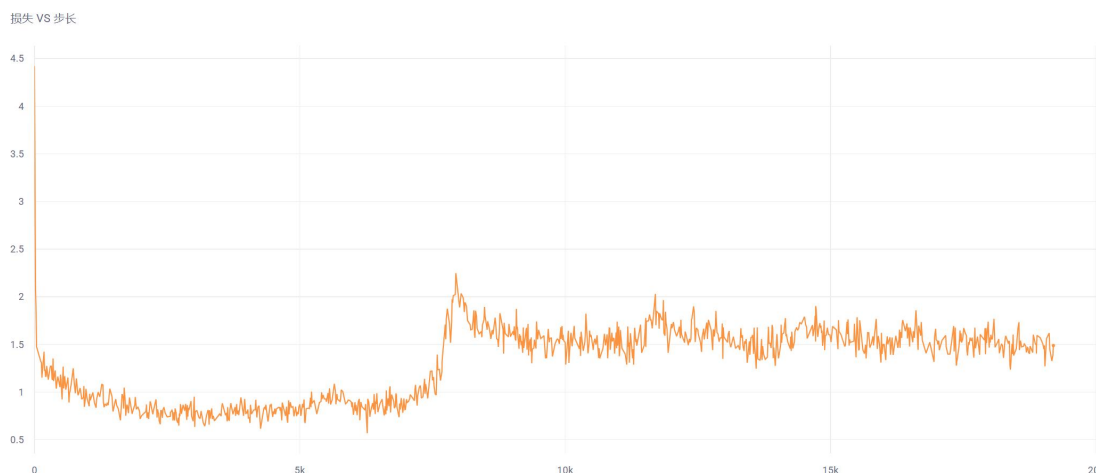
接着，设计训练模型，模型结构包括嵌入层、单层 LSTM 和全连接层输出到字符空间。训练过程中，输入序列长度设为 100，批大小为 8，使用 Adam 优化器和交叉熵损失函数进行反向传播。每个训练批次随机截取连续的片段，以学习序列间的上下文关系。训练共进行 3000 步，每 10 次输出一损失，并将指标通过 Comet.ml 记录。

最后，利用 `abc2midi.exe` 与 `timidity.exe` 将生成的乐谱文本转换为 wav 音频文件，实现可听化输出。

三、调参过程

在初始实验中，embedding 维度为 128，hidden size 为 256，训练收敛速度较慢且生成结果结构松散。后续将隐藏层规模扩大至 1024 并提高嵌入维度至 256，模型对长序列依赖的捕捉能力显著增强，损失下降更平稳。

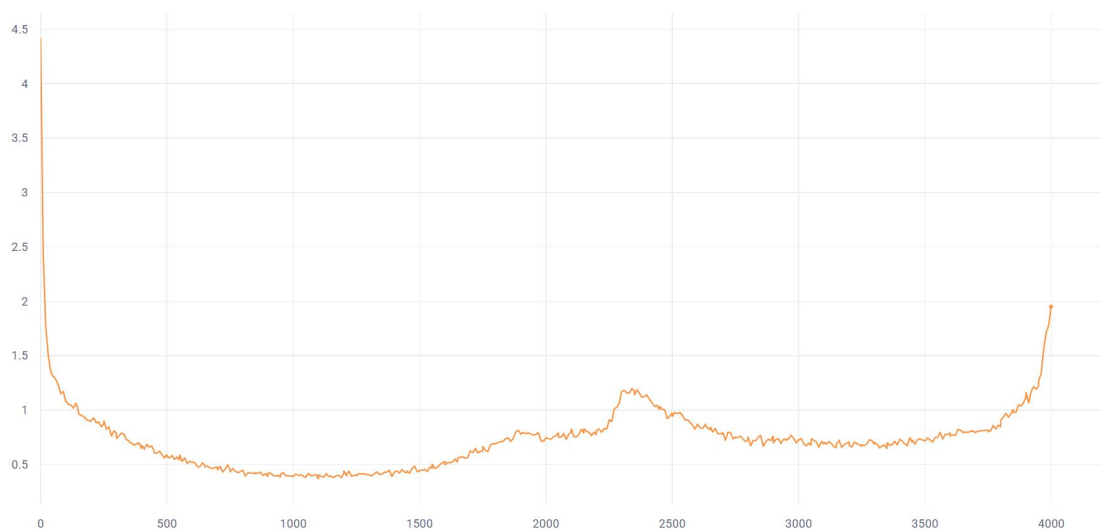
训练阶段，我尝试将训练次数 `num_training_iterations` 设置为 19200，观察不同批次大小 `batch_size` 的 Loss 曲线情况，当 `batch_size` 为默认的 8 时，其训练的 Loss 曲线在 2000-3000 之间收敛，Loss 值大约为 0.7，之后 Loss 值开始上升，



损失函数曲线图

当我将批次大小 `batch_size` 设置为 64 时, Loss 曲线在 1000 左右收敛, Loss 值大约为 0.36。

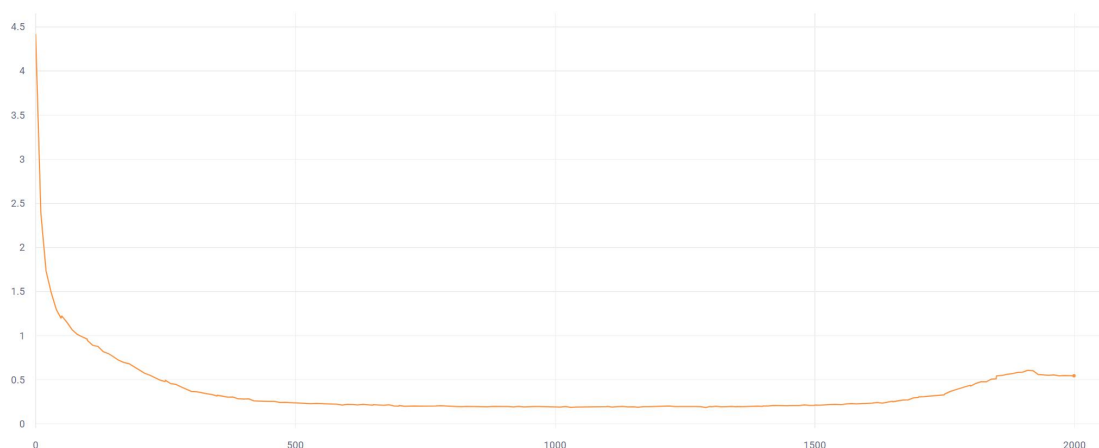
损失 VS 步长



损失函数曲线图

而当我将 `batch_size` 设置为 256 时, Loss 曲线也在 1000 左右收敛。

损失 VS 步长



不同的 `batch_size` 对生成的音乐也有一定的影响, 以 `batch_size=8` 和 `batch_size=256` 作为对比, 其他参数不变, 使用前者参数生成的歌曲, 单首歌内节奏变化较多, 但旋律不够明显, 不同歌曲之间差异较大; 使用后者参数生成的歌曲, 节奏变化较少, 整体较平稳, 但旋律更加清晰, 但是不同歌曲之间差异较小。

生成阶段还尝试了不同的温度参数 `temperature`, 为确保可以清晰的看出不同温度参数下, 歌曲的区别, 我将 `batch_size` 设置为 64, 当 `temperature=1.5` 时生成的音乐节奏变化较多不同歌曲之间存在明显的区别, 当 `temperature=0.5` 时, 生成的音乐节奏变化较小不同歌曲之间没有明显的区别。

四、结果分析

训练完成后, 模型成功生成多段音乐片段。通过 Comet.ml 的可视化界面可以观察到损失曲线平稳下降, 说明模型已有效学习到音乐结构。生成的 wav 文件

在听感上具备一定节奏性与旋律走向，部分片段甚至能呈现出类似人类创作的旋律结构。

在实验结果中，模型能生成不同长度、节奏变化的片段，显示其具备一定的音乐语法学习能力。但仍存在部分不协调音符与过早重复模式的问题，表明模型尚未充分理解全局音乐结构。

五、心得体会

本实验让我深入理解了 LSTM 在序列数据建模中的优势，尤其是在捕获长期依赖关系方面的表现。同时，实验让我意识到音乐生成与自然语言建模在原理上的相似性，二者均依赖序列上下文建模。

实践过程中也发现深度模型对计算资源依赖较强，在 GPU 上训练可显著加快收敛速度。

六、实验反思问题

1. 模型为什么能学会“旋律规律”？

模型能够学会“旋律规律”，主要是因为 LSTM 模型可以捕捉序列数据中的时间依赖性，从训练数据中学习音符的前后关系和节奏模式。

2. 为什么温度参数（temperature）会影响生成多样性？

较低温度倾向于选择概率较高的音符，使生成的旋律更稳定但重复性高；较高温度增加了低概率音符被选择的可能性，使旋律更加多样但可能出现不协调的音符。

3. 您的改进在哪些方面提升了音乐的自然度或节奏感？

通过调整模型超参数和生成策略，本实验在音乐的自然度和节奏感上得到了提升，例如在单首歌内适度出现节奏变化，使旋律更丰富且连贯。

4. 如何判断“音乐质量”的好坏？是否存在客观指标？

对于“音乐质量”的好坏，目前仍以人的主观判断为主，但是歌曲的旋律连贯性、节奏变化、和谐度等等可以作为“音乐质量”的辅助判断。