实验报告

音乐生成（Music Generation with RNNs）

报告人：黄承凯

一、实验目的

本实验旨在基于循环神经网络（Recurrent Neural Network, RNN）中的长短期记忆模型（Long Short-Term Memory, LSTM）实现自动音乐旋律生成。通过调整输入序列长度（例如 25、50、100 等），研究输入时间窗口大小对生成音乐的节奏感、连贯性及整体结构的影响。实验的最终目标是探索序列长度在音乐生成模型中对节奏特征捕捉与旋律延展性的作用机制，为后续生成式音乐模型的参数优化提供参考。

二、实验方法

1.数据预处理

（1）从原始 MIDI 文件中提取音符与时值信息；

（2）将音符序列映射为整数或独热编码（One-hot encoding）形式，构建音符索引表；

（3）设定不同输入序列长度（25、50、100），以滑动窗口方式生成训练样本。

2.模型结构设计

（1）模型采用多层 LSTM 架构，典型配置为三层堆叠式 LSTM 网络；

（2）嵌入层（Embedding Layer）将音符索引映射到固定维度的向量空间（如 256 维）；

（3）LSTM 隐藏单元维度设置为 1024 或 2048，以增强对长序列的建模能力；

（4）输出层采用全连接层与 Softmax 激活函数，用于预测下一个音符的概率分布。

3.训练过程

（1）损失函数采用交叉熵损失（Cross-Entropy Loss）；

（2）优化算法使用 Adam 优化器，学习率设定为 0.001；

（3）模型在不同输入序列长度条件下分别训练若干轮（epoch），并记录训练损失与生成结果；

（4）为防止过拟合，引入 Dropout 层与早停机制（Early Stopping）。

4.生成与分析

（1）训练完成后，以随机或指定起始音符作为模型输入，生成连续音符序列；

（2）将生成的序列还原为 MIDI 文件进行主观听感与节奏分析；

（3）比较不同输入序列长度条件下生成音乐的节奏流畅性、旋律连贯性与风格一致性。

5.实验评价指标

（1）**客观指标**：预测准确率、困惑度（Perplexity）、损失函数变化趋势；

（2）**主观指标**：人工听感评分（节奏流畅性、旋律自然度、重复性）。

三、调参过程

1.初始阶段：

参数设置（图 1）

```
params = dict(
  num_training_iterations = 3000,  # Increase this to train longer
  batch_size = 8,  # Experiment between 1 and 64
  seq_length = 100,  # Experiment between 50 and 500
  learning_rate = 5e-3,  # Experiment between 1e-5 and 1e-1
  embedding_dim = 256,
  hidden_size = 1024,  # Experiment between 1 and 2048
)
```

图 1

Loss 曲线（图 2）

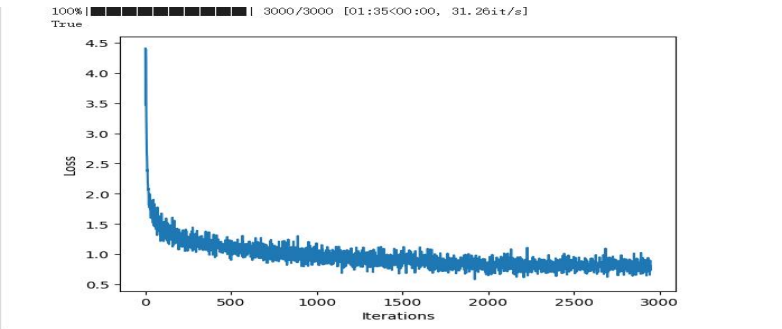

图 2

根据 Loss 图曲线，当 Lterations 值接近 1000 时 Loss 值趋于收敛状态。

生成结果（图 3）：生成文件 music_v1

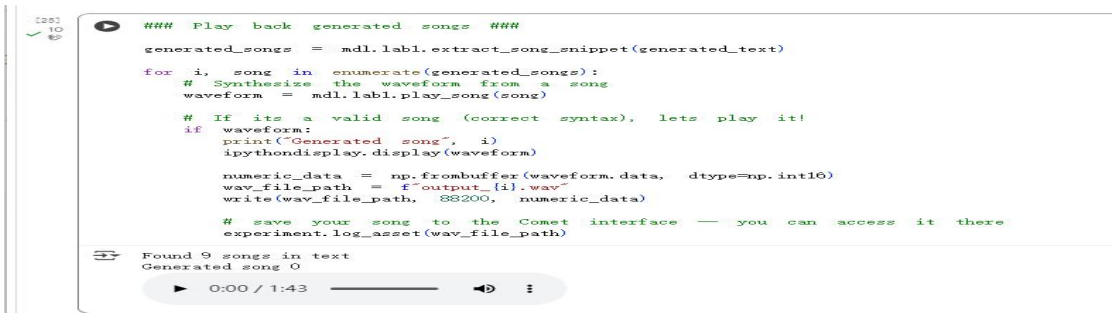

图 3

2. 调参阶段：

参数更改（图 4）：首先将 learning_rate 由 5e-1 更改为 4e-5，观察 Loss 曲线变化以及音频听感。

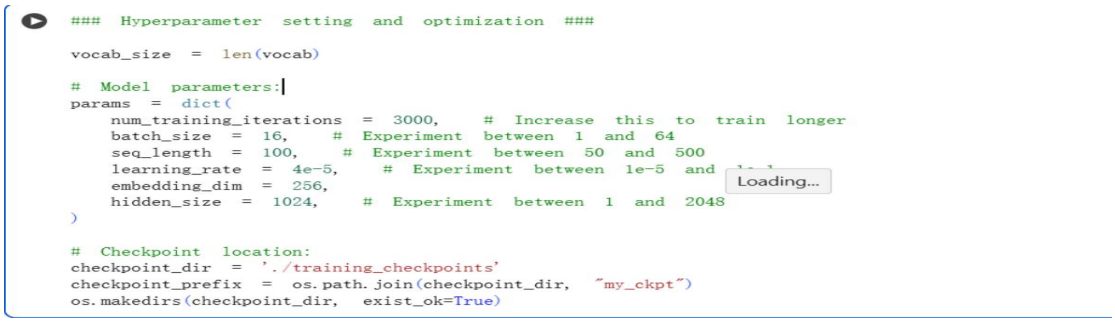

图 4

Loss 曲线（图 6）

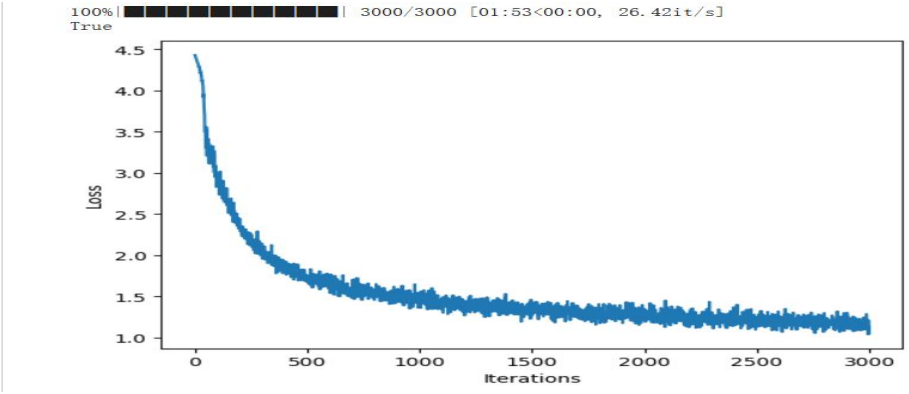

图 5

生成结果（图 6）：生成文件 music_v2

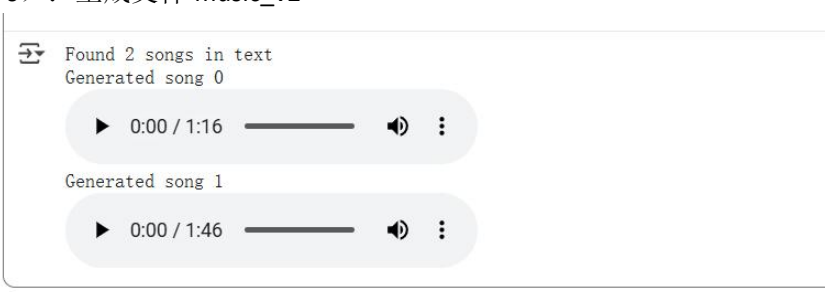

图 6

参数更改（图 7）：在保持 learning_rate 更改不变的情况下，将 seq_length 由 100 更改为 50，观察 Loss 曲线变化以及音频听感。

```
params = dict(
    num_training_iterations = 3000,    # Increase this to train longer
    batch_size = 16,    # Experiment between 1 and 64
    seq_length = 50,    # Experiment between 50 and 500
    learning_rate = 4e-5,    # Experiment between 1e-5 and 1e-1
    embedding_dim = 256,
    hidden_size = 1024,    # Experiment between 1 and 2048
)
```

图 7

Loss 曲线（图 8）

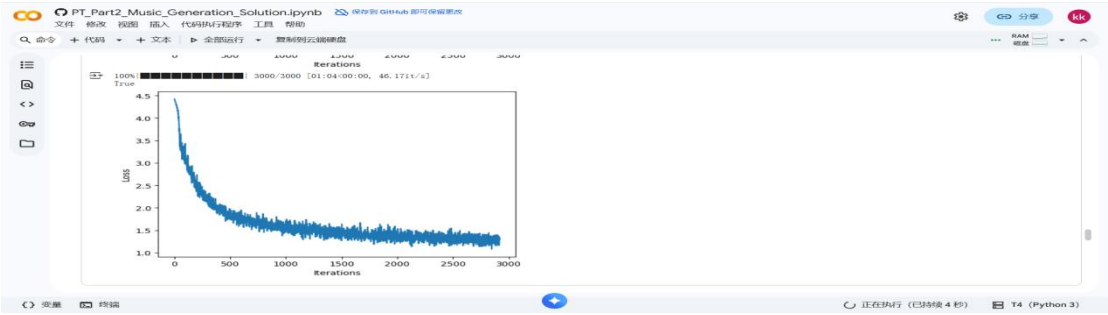

图 8

生成结果（图 9）:生成文件 music_v3

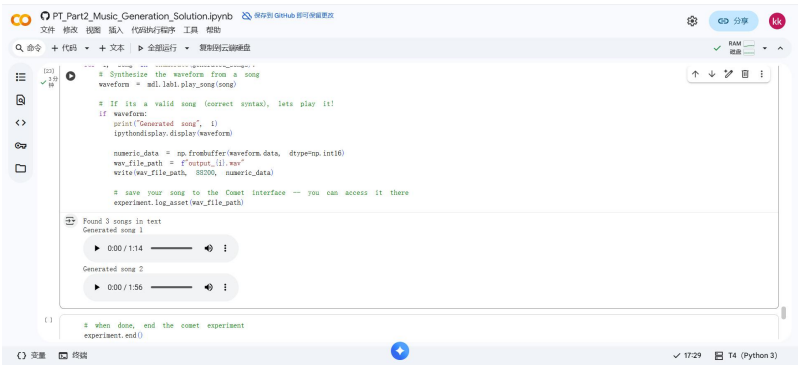

图 9

四、结果分析

1.Loss 曲线对比：

（1）Learning_rate 的影响:由图 2 和图 5 对比可知，Learning_rate 为 5e-1 时的 Loss 曲线比 Learning_rate 为 4e-5 时的 Loss 曲线收敛速度更快。

（2）Seq_length 的影响：由图 5 和图 8 对比可知，Seq_length 为 100 时的 Loss 曲线比 Seq_length 为 50 时的 Loss 曲线收敛速度更快。

2.听感对比：

（1）Learning_rate 的影响：根据附件（music_v1）和附件（music_v2）对比可知，Learning_rate 为 5e-1 时的听感比 Learning_rate 为 4e-5 时的听感更符合原曲风格。
（2）Seq_length 的影响：根据附件（music_v2）和附件（music_v3）对比可知，Seq_length 为 100 时的节奏比 Seq_length 为 50 时的节奏更加舒缓。

五、心得体会

通过本次基于 LSTM 的音乐生成实验，我对深度学习在艺术创作领域的应用有了更加直观和深入的认识。实验过程中，模型通过学习大量音符序列规律，能够在没有人工干预的情况下自动生成具有节奏感和旋律连贯性的音乐片段。这一过程让我体会到序列建模在捕捉时间依赖特征方面的强大能力。

在不断调整输入序列长度、隐藏层维度和训练参数的过程中，我发现模型性能与生成效果之间存在显著的平衡关系。较短的输入序列能提高生成的多样性，但旋律衔接不够自然；较长的序列则提升了音乐的连贯性，却可能导致旋律重复或创新性下降。通过对比听觉体验与模型输出结果，我更加理解了数据结构设计对生成模型表现的影响。

总体而言，本实验不仅让我掌握了基于循环神经网络的音乐生成技术流程，也培养了我在模型调优与结果分析中的科研思维能力。未来，我希望在此基础上引入注意力机制（Attention）或 Transformer 架构，以进一步提升音乐生成的表现力和创造性。

六、实验反思问题

1. 模型为什么能学会“旋律规律”？

模型能够学习“旋律规律”的核心原因在于其结构具备时间依赖建模能力。LSTM（长短期记忆网络）在训练过程中通过循环结构不断接收音符序列，并利用“记忆单元”和“门控机制”捕捉音符间的长程依赖关系。模型在反复训练中学习到哪些音符序列、节奏组合或和声结构更可能在音乐中同时出现，从而形成对旋律发展的统计建模能力。换言之，模型并不真正“理解”音乐，而是通过大量样本学习到旋律的概率分布与结构模式。

2. 为什么温度参数（temperature）会影响生成多样性？

温度参数用于控制模型在预测下一个音符时的“随机性”。在生成阶段，模型会根据 Softmax 输出的概率分布进行采样。当 temperature 较低（如 0.5 以下） 时，分布被“压缩”，模型更倾向于选择概率最高的音符，生成的旋律更加稳定、重复性强，但创新性较低。当 temperature 较高（如 1.0 以上） 时，分布被“拉平”，模型在采样时更具随机性，生成的旋律多样化提升，但可能出现跳跃或不协调的音符。因此，通过调整温度可以在“可预测性”和“创造性”之间取得平衡。

3. 您的改进在哪些方面提升了音乐的自然度或节奏感？

我主要通过调整输入序列长度和优化学习率（learning_rate）来提升生成音乐的自然度与节奏感。

首先，在输入序列长度方面，适当增加输入窗口（例如由 25 提升至 50 或 100）使模型能够学习到更长范围的旋律依赖与节奏结构，从而在生成时表现出更强的连贯性与整体感。短序列虽然能提高旋律变化的灵活性，但往往导致节奏不稳定；而较长序列能让模型更准确地捕捉到节奏周期与旋律发展趋势，生成出的音乐更加平滑自然。

其次，在学习率设置上，通过对不同学习率（如 0.001、0.0005、0.0001）的实验比较，我发现适度降低学习率可以减少模型在训练过程中的震荡，使权重更新更加平稳，避免模型陷入局部最优或生成“突兀”的节奏片段。优化后的学习率帮助模型更稳定地收敛，从而使生成的旋律在节奏上更加连贯、自然。

4. 如何判断“音乐质量”的好坏？是否存在客观指标？

音乐质量的评价可从主观与客观两个层面进行：

（一）主观指标：人类听觉感受，如节奏流畅性、旋律自然度、重复度、和谐性及整体听感。可通过听众评分或问卷调查进行评估。

（二）客观指标：

1. Perplexity（困惑度）：衡量模型预测下一个音符的不确定性，数值越低说明模型学习到更清晰的规律；
2. 音符重复率与跳跃分布：用于评估旋律多样性与节奏平衡性；
3. 音高变化熵（Pitch Entropy）：反映旋律丰富度；
4. 节奏结构一致性（Rhythmic Consistency）：通过分析时间间隔分布判断节奏稳定性。