

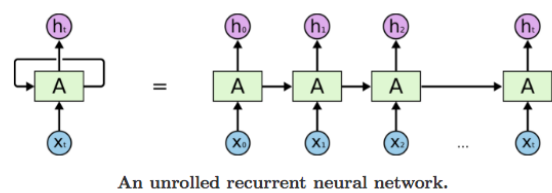
神经网络与深度学习第 3 次作业 实验报告

2251924 晏景豪

1. 模型解释

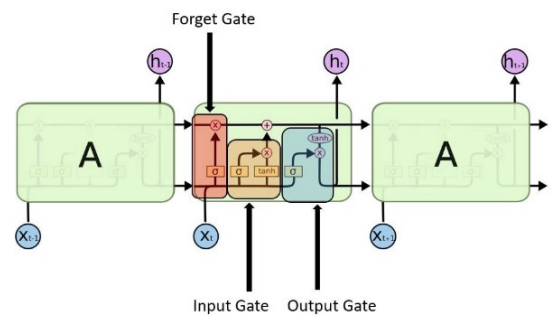
1.1. 循环神经网络 (RNN)

循环神经网络 (RNN) 是一种专门用于处理序列数据的神经网络，通过在网络中引入循环结构，使每个时间步的输出不仅依赖于当前输入，还受先前状态的影响，从而捕捉数据的时间依赖性。其结构使得 RNN 在自然语言处理、语音识别等任务中能够有效建模上下文信息，但同时也面临梯度消失或爆炸的挑战，通常通过引入长短时记忆 (LSTM) 或门控循环单元 (GRU) 等变体来缓解这一问题。



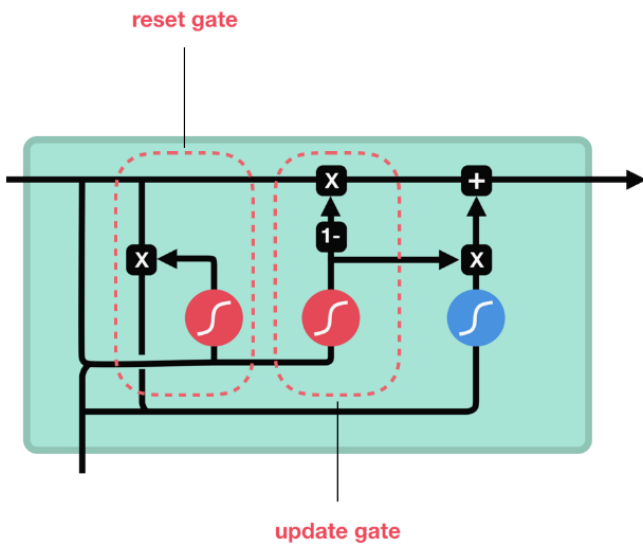
1.2. 长短期记忆网络 (LSTM)

长短期记忆网络 (Long Short-Term Memory, LSTM) 是一种特殊的 RNN，通过引入细胞状态 (cell state) 和门控机制 (输入门、遗忘门、输出门) 解决传统 RNN 的梯度消失和长期依赖问题。其核心在于细胞状态沿时间步传递关键信息，门控单元通过 Sigmoid 函数和逐元素相乘操作，动态调控信息的存储、遗忘与输出，从而在时序数据 (如文本、语音) 中高效捕捉长期依赖关系，广泛应用于机器翻译、时间序列预测等领域。



1.3. 门控循环单元（GRU）

门控循环单元（Gated Recurrent Unit, GRU）是 RNN 的轻量化变体，通过整合长短期记忆网络（LSTM）的门控机制并简化结构，解决梯度消失和长期依赖问题。其核心包含更新门（调控历史状态保留比例）和重置门（控制历史状态信息遗忘程度），利用 Sigmoid 函数与 Hadamard 积实现动态门控，将隐藏状态与细胞状态合并为单一状态传递，在减少参数冗余的同时保持对时序特征的建模能力，广泛用于文本生成、语音识别等序列任务。



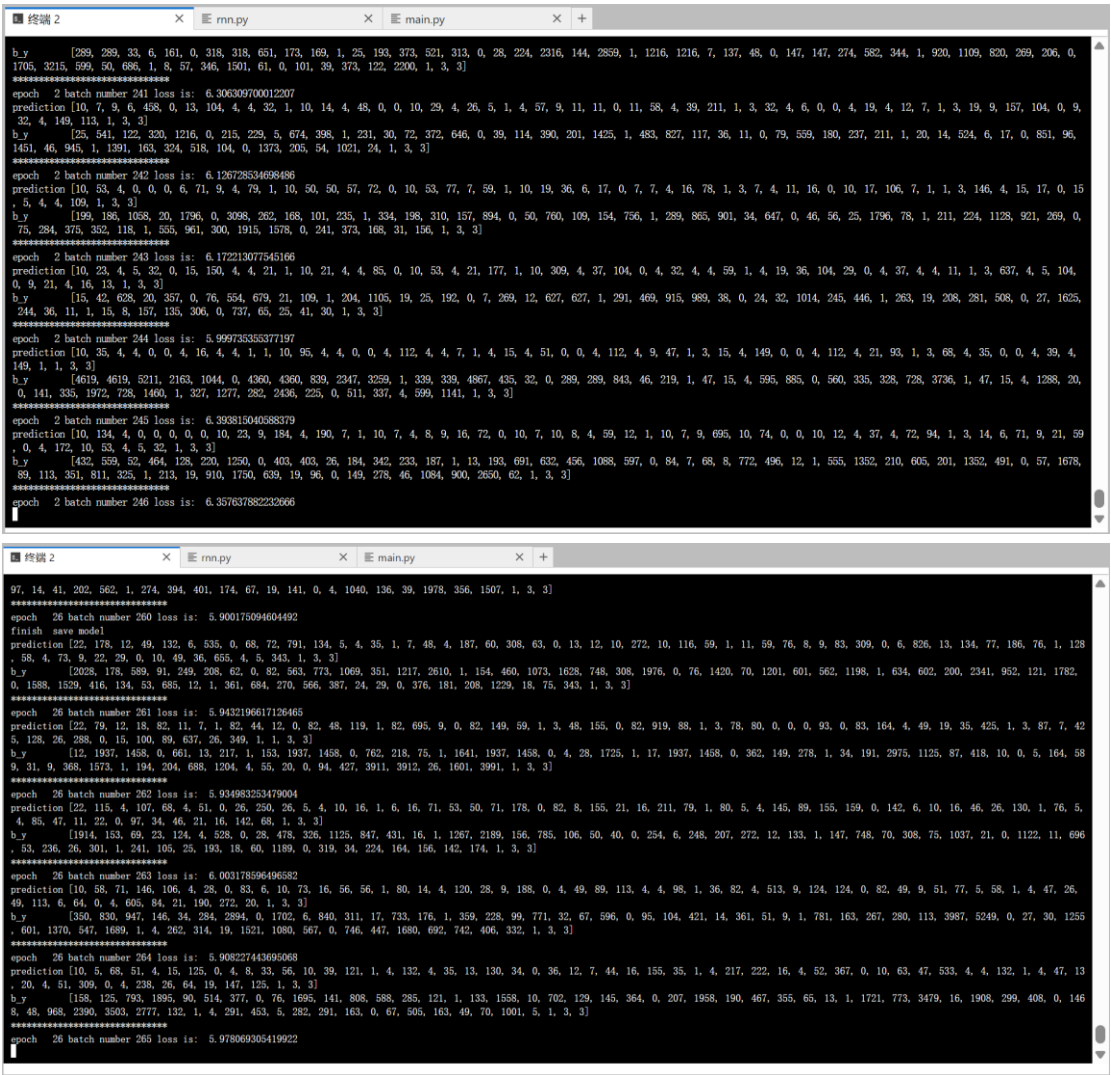
2. 生成过程

本诗歌生成模型基于双层 LSTM，通过序列到序列的深度学习框架实现古诗创作。模型架构包含词嵌入层、循环神经网络和全连接分类器，利用唐诗语料进行有监督训练，最终通过自回归生成机制实现给定首字的诗歌续写。

具体实现中，系统首先对原始语料进行字符级向量化处理，构建字词映射表；训练阶段采用滑动窗口构建训练对（ X_t 为当前字符， Y_t 为下一字符），使用 NLLLoss 计算损失并通过 RMSprop 优化；生成阶段通过 argmax 策略递归选择概率最高的下一个字符，结合起止符控制和生成长度限制，确保输出符合诗歌格式要求。

3. 实验结果

选择 pytorch 版本进行实现，训练过程中的截图如下：



训练过程中的 GPU 使用率如下：



生成结果如下：

A terminal window titled '终端 2' with tabs for 'rnn.py' and 'main.py'. The command '(base) root@autodl-container-5f994a86fd-bac6393a:~/autodl-tmp/tangshi_for_pytorch# python main.py' is executed. The output shows a series of 'error' messages and a 'UserWarning' about deprecated log_softmax usage. The generated text is: '日下无忧，诗人不能能不能。', '云外不能开地尽，地平无复有风尘。', '金章不可问，不是有时情。', '何人不相别，一笑不可寻。', '夜多长有，玉中风月不如霜。', '不能天地升春去，不得人间不得人。', '为看山色好，不似白头看。', '海昔明月夜，月明风雨滴无声。', '不须不得无人问，不是人间有意还。', '清风一一色，一树不离襟。'. The prompt '(base) root@autodl-container-5f994a86fd-bac6393a:~/autodl-tmp/tangshi_for_pytorch#' is shown at the bottom.

4. 实验总结

本次实验基于 PyTorch 框架实现了双层 LSTM 网络，重点探索其在序列建模中的核心特性。模型通过词嵌入层将字符映射为 100 维语义向量，利用 LSTM 的遗忘门、输入门和输出门机制，在 128 维隐藏状态中动态维护时序记忆。训练采用 RMSprop 优化器配合梯度裁剪策略，有效平衡了参数更新速度与稳定性，经过多轮迭代后损失函数呈现稳定收敛趋势。

实验验证了 LSTM 处理长序列依赖的优越性，其门控结构可自适应选择记忆与遗忘节点。但在生成测试中观察到局部重复现象，反映出传统自回归解码策略的局限性。本次实验深化了我对循环神经网络内部状态传递机制的理解，为后续学习打下了坚实的基础。