Project 2: 实验报告

花叶果，骆浩然

2025 年 5 月 31 日

Part A: Implementing Language Models

在本部分，我们手写实现了 RNN 模型，调包实现了 LSTM, Transformer 模型。我们对每一个模型均进行了超参搜索，使它们能达到良好的训练效果，避免过拟合与欠拟合的发生。图 1是各个模型的 loss 曲线。

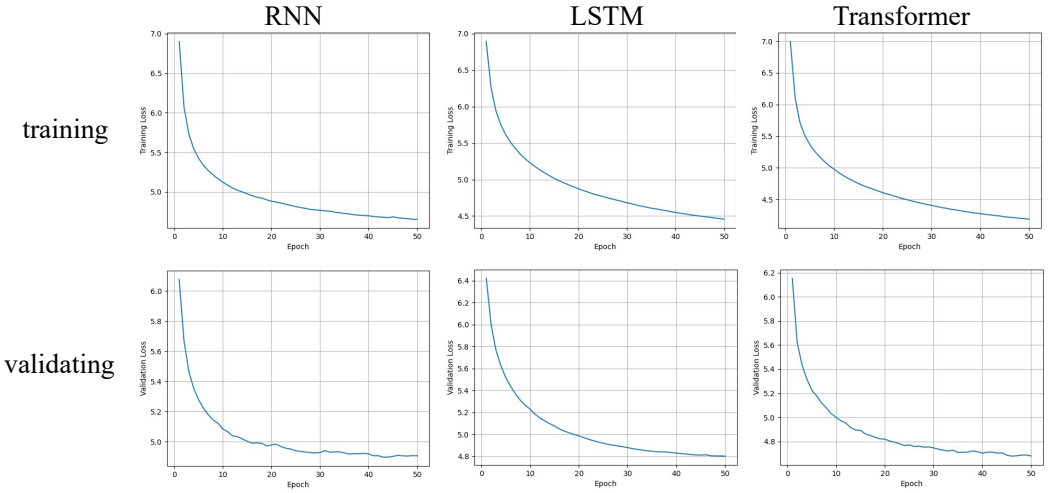

图 1: 三个模型的 loss 曲线

Part B: Comparing Across Models and Domains

在本部分，我们在训练数据所处领域（金融）以及其他领域（例如气候）上评估了三个模型的困惑度（perplexity），并尝试从不同数据集的词频分布差异的角度来解释困惑度的变化。最后我们使用续写任务进行了模型的采样测试。

图 2比较了三个模型在特定的训练阶段的困惑度。表格 1以气候相关的维基百科文章为例，比较了三个模型的泛化能力，其中 Transformer 模型的泛化能力最强。

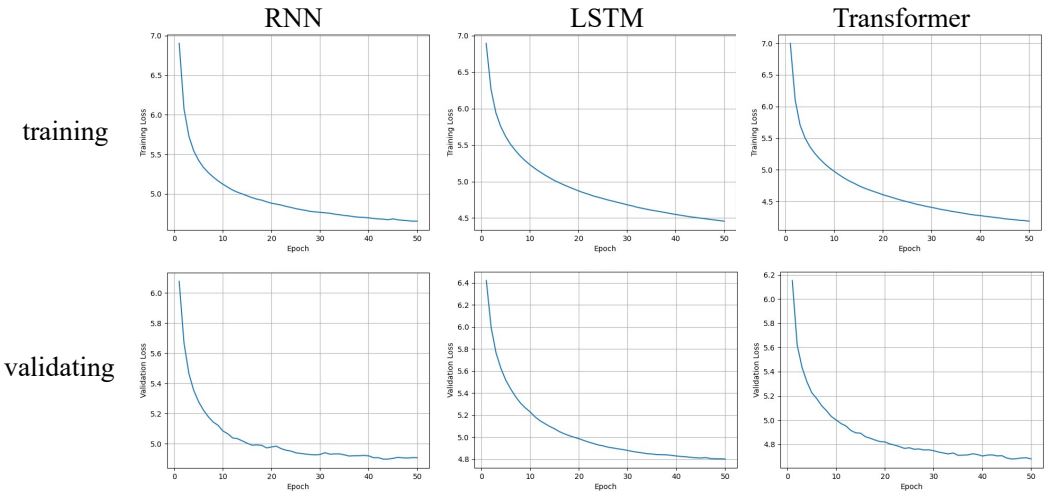

图 2: 三个模型的困惑度比较

wiki-sentence	RNN	LSTM	Transformer
This article is about the present-day human-induced rise in global temperatures	335.21	285.82	231.15
Present-day climate change includes both global warming—the ongoing increase in global average temperature—and its wider effects on Earthś climate system	1296.15	3007.36	1388.18
Earthś average surface air temperature has increased almost 1.5 °C (about 2.5 °F) since the Industrial Revolution	117.71	171.27	168.93
...	...	...	...
Average	827.3	705.7	**657.9**

表 1: 三个模型的泛化能力比较

在图 3中我们比较了训练集与其他领域数据的词频分布，并以一些有代表性的词为例展示了分布的差异。由于词频分布的差异较大，模型对于新词汇分布的判断变得不确定，导致了困惑度的提升。

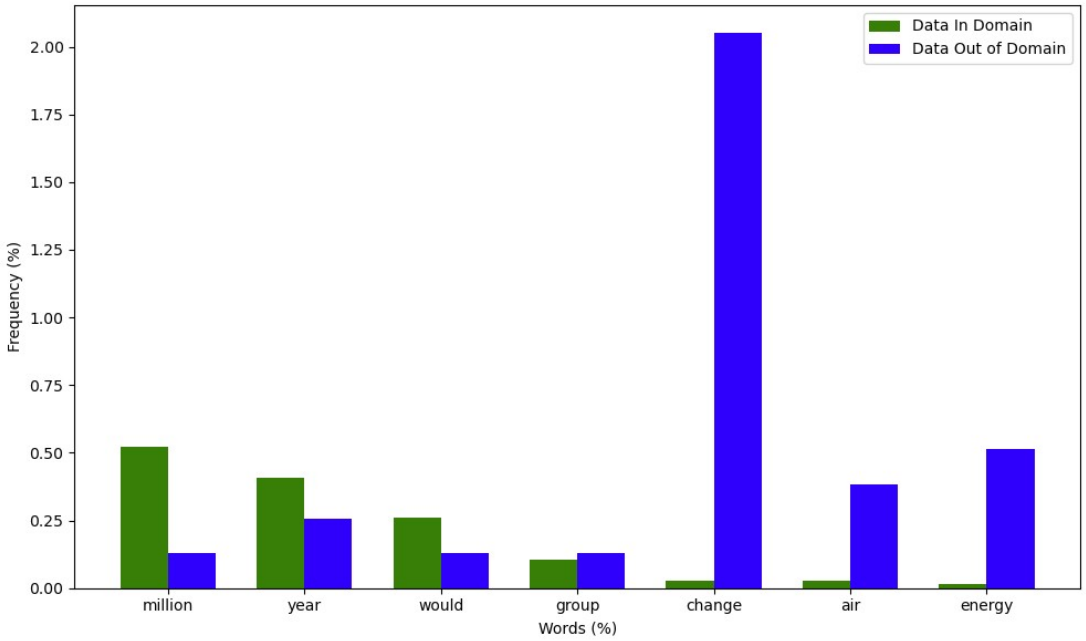

图 3: PTB 数据与维基百科文章数据的词频分布比较

在采样测试中，我们以 “the meaning of life is” 作为开头，要求模型进行续写。我们使用了温度控制的采样策略，并在表格 2 中呈现了低温度值和高温度值下模型的输出。可以看出，RNN 的输出内容比较简单，高温下可能产生语义分散；LSTM 的输出句子结构完整，但信息量有限；Transformer 的输出内容最丰富，语法基本正确，但句子过长时语义可能不连贯。综合而言，Transformer 表现出最佳的连贯性和流畅度。

RNN	the meaning of life is now <eos>
LSTM	the meaning of life is the biggest market <eos>
Transformer	the meaning of life is putting what is able to a specialty and about fiscal N million from other defense <eos>

(a) 低温采样结果

RNN	the meaning of life is new days forced easier change because that attacks <eos>
LSTM	the meaning of life is obviously seeing government <unk> now accountants sought production it reached managers ' <unk> tools using rumors ranging money but healthy billions china looks featuring glass work <unk> affecting television deductions activities slashing property movements papers last icahn in last aggressive <eos>
Transformer	the meaning of life is buying george and economics instead as movie plus one imported bags a five-year compensation of bribery totaling million works from campeau sells discussions on electronics services disposable items next fiscal in highly court-appointed colorado stepped worrying passed male in rate of securities agency yen because patients tissue segment operating

(b) 高温采样结果

表 2: 三个模型的采样结果

Part C: Fine-Tuning a Pretrained Language Model

在本部分，我们使用 TinyStories 数据集对 Qwen2-0.5B 模型进行了全量参数微调，并分别用定量和定性的方式评估了微调结果。除此以外，我们还测试了微调的时间、微调使用的数据量对于结果的影响。我们在图 4中展示了微调过程的 loss 曲线，在表格 3中定量测试了模型在微调前后的困惑度变化。

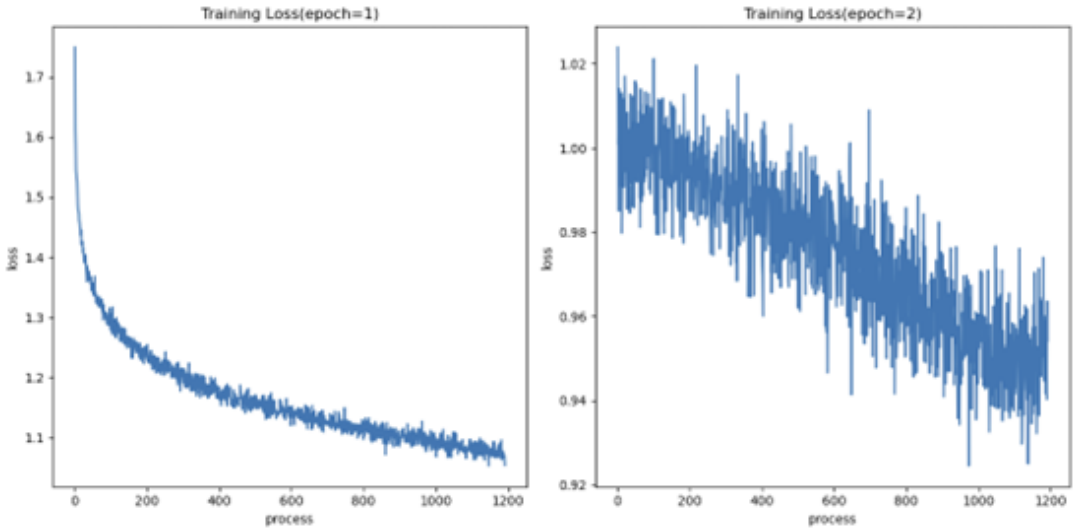

图 4: 微调过程的 loss 曲线

TinyStories-sentence	PPL Before fine-tuning	PPL After fine-tuning
At home, Lily served the bird some water and food. The bird was very happy and started to feel better. Lily was happy too because she helped the bird. From that day on, Lily and Max went on walks and helped other animals they saw on the trail	81.13	**26.16**
When they got to the store, they gave the bottle of wine to the nice lady at the counter. They told her that they wanted to return the wine because it was not what they wanted. The lady smiled and gave them their money back. The family was happy	138.74	**49.39**
...	...	...
Average	120.18	**36.23**

表 3: 微调前后的困惑度比较

在定性测试中，我们以 “A brave knight sets out on a journey to rescue a princess.

Write the next sentence in TinyStories style.” 作为提示，让模型进行故事续写，输出的结果呈现在表格 4中，可以发现模型在微调后的输出更符合 TinyStories 风格，在微调前的输出虽然语句通顺但与提示不相关。进一步地，为了验证模型在微调过程中是否保留了原来的语言能力，以及是否学习了一些微调数据的深层结构特征，我们以与微调数据集无关的 “Climate change is causing glaciers to melt, sea levels to rise, and extreme weather events to become more frequent, posing a threat to human survival. Protecting the environment is urgent. Please continue writing.” 作为提示，让模型进行续写，输出的结果在表格 5中。模型在微调后良好地保留了原来的语言能力。有趣的是，微调后的输出往往擅长构造事件和结局，而微调前的输出只是复述和解释提示中的内容，并不会构造事件。这种能力也许是在 TinyStories 中提取并学习的。

Before fine-tuning	Once he has arrived at the castle, what do he do? What happens? The knight then goes to the palace and asks for the princess's hand in marriage. Once she accepts, what happens? What do the knights and princess do to celebrate? I was a knight in my youth. I had a reputation for being a brave and honorable man. I had the courage to face off against the evil lord of the castle and his army. But I was also a man with a heart of gold
After fine-tuning	As the knight journeys further and further, the princess finds a dangerous path. The knight comes across a dragon. The dragon is huge and angry, and it wants to hurt the princess. The knight is scared but he knows he must rescue the princess. He quickly grabs a sword and gets ready to fight. The knight and the dragon fight for a long time. The knight is brave and strong, but the dragon is very powerful. In the end, the knight is able to rescue the princess from

表 4: 以”A brave knight...” 为提示的续写结果

Before fine-tuning	The planet's climate is rapidly changing, causing significant ecological impacts. As temperatures rise, glaciers are melting at an unprecedented rate, leading to rising sea levels and more frequent extreme weather events. These changes threaten the survival of many species, including humans and their
After fine-tuning	Everyone needs to take action and work together to protect the environment. So the people all work hard to clean up the oceans, stop melting them, and make sure the sea levels stayed low. Everyone work together and they made a difference. One

表 5: 以"Climate change is..." 为提示的续写结果

在表格 6中，我们比较了不同微调时间、不同数据量下的结果。微调时间以 epoch(s) 数来衡量，所用的数据量以占总数据量的百分比 (%) 来衡量，采样的提示使用与之前相同的"A brave knight..."，并用灰色标出微调前的结果。可以看出，在足够的训练时长下，更大的数据量有助于提升模型的泛化能力、叙述逻辑和语言连贯性，使模型生成的故事更为自然；在数据固定的情况下，更长的训练周期有助于模型从浅层的语言模仿逐步提升到深层次的语义理解与故事结构把握。

Epochs	Data(%)	Overall PPL	Output
2	90	**36.23**	As shown above
2	60	36.62	The knight arrived at the castle. He heard the princess crying. The knight knocked on the door. “Who is it?” asked the princess. ” It’ s me, the knight. I’ m here to rescue you!” The princess opened the door. She was very happy. The knight helped her to her feet. He hugged her and kissed her. ” Thank you for rescuing me! I was
1	90	41.37	The knight is riding along, and the princess is hiding in a tree. She says, ”I love you, Lord.” The knight’s heart swells with joy. He quickly climbs up the tree and finds the princess hiding. He approaches her and says, ”I love you, Princess.” The princess smiles and says, ”I love you, Lord.” The knight feels a sense of happiness and relief. He tells her that he will always be there for her, no matter what challenges
None	None	120.18	As shown above

表 6: 微调时间、数据量对结果的影响

Bonus: RLHF

在本部分，我们采用 PPO（近端策略优化）对 Qwen2-0.5B 模型进行优化。我们的优化目标是让模型输出带有积极情感的回复内容。我们所使用的奖励模型是一个在 IMDB 数据集上训练的打分器模型，它能够对含有积极情感的语句赋予高分。我们所使用的价值模型是建立在原模型隐空间上的一个 ValueHead，它在训练过程中不断更新。图 5中的红色线是我们的训练过程中奖励均值、方差的变化曲线，强化学习训练使模型输出的

奖励均值从 0.943 变为 2.480。表格 7展示了训练前后模型的采样结果。可以看出，经过强化学习训练后的模型能输出更加积极的回复。

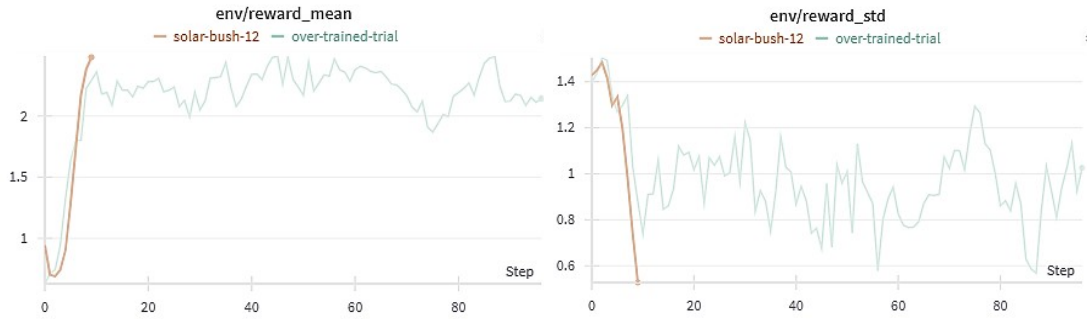

图 5: 强化学习过程的奖励变化曲线

但与此同时，强化学习训练后的模型偶尔也会产生一些非预期的输出，如表格 8所示。模型可能错误地学习了通过堆砌正面词汇来盲目赞美某物的模式，这导致了重复或与提示不相关的输出。我们推测可能的改进点是：调整更优的 PPO 相关超参数；修正奖励函数使之惩罚重复语义；交替进行监督微调和强化学习等。

Prompt	Output	Reward
Along with Darkwing Duck this is	a very goodgood piece of art. It was a great piece, but so great, the art was set for reference and be love this with	2.78
This movie was horrible. I	love the movie. It was so funny. I love it. It's a great movies good. You the be love	1.84

表 8: 强化学习导致的偶然非预期输出

小组分工

花叶果：代码编写；训练，采样和分析；制作实验报告；方案讨论
骆浩然：制作 PPT；汇报；方案讨论

以上是我们小组第二次大作业实验报告的全部内容。感谢老师和助教!

Prompt	Before RL	Reward	After RL	Reward
Life Begins	On, The Life Begins On. There is no life without suffering. No one can be without suffering	1.71	, I was one very happy and I would hearted them with my friends and my husband and I	**2.48**
Yes, as the other reviewers have	noted, the current version of the game has a problem with an extra character, a problem I had	-1.11	written, I've been very lucky to read this, and am very fond of this read.	**1.96**
For your own good, it would	be better to live in a place that's comfortable to you than one that's not, even if it's the same area	0.69	help you, and will keep you, a wonderful friend. The great friends which were in love, on its bump is good on the	**2.58**
Now I don't	know about you, but I'm a sucker for a good classic story. I can't say I read	0.52	know what to talk about, but I'm a very lucky guy and I thank the book and a book	**2.18**
There are no people like	you, but you have to take care of them. You have to make sure that they are not	0.71	you, I'm very pleased that you are so very lucky and wise with the help that. help other People who	**2.27**

表 7: 强化学习前后的采样测试