

HW3

Transformer-based music generation

Name: 汪宣甫
Student id: R11944074

Outline

- Method
- Experiments
- Conclusion
- Appendix
- Reference

Method

- I trained two types of transformer-based language model, GPT2 [1] and llama [2].
- For each model, I fix the loss scale i.e. all models are with the same losses on training data, and tried three decoding strategies: greedy, beam search (number of beams=3) and top-p sampling ($p = 0.95$, temperature=0.95).

Δ : absolute difference with the training dataset's score.

Experiments

Method	H1	H4	GS
Compute directly from the dataset			
Train dataset	1.94	2.53	0.78
Generated from GPT2			
Greedy	1.84 ($\Delta=0.1$)	2.42 ($\Delta=0.1$)	0.85 ($\Delta=0.07$)
Top-p w/ temperature=0.95	1.79 ($\Delta=0.15$)	2.40 ($\Delta=0.13$)	0.86 ($\Delta=0.08$)
Beam search (num_beams=3)	1.90 ($\Delta=0.03$)	2.40 ($\Delta=0.12$)	0.84 ($\Delta=0.06$)
Generated from LLama			
Greedy	1.93 ($\Delta=0.00$)	2.50 ($\Delta=0.03$)	0.78 ($\Delta=0.00$)
Top-p w/ temperature=0.95	0.03 ($\Delta=0.03$)	0.03 ($\Delta=0.03$)	0.03 ($\Delta=0.03$)
Beam search (num_beams=3)	1.96 ($\Delta=0.02$)	2.45 ($\Delta=0.08$)	0.87 ($\Delta=0.09$)

Conclusion

- Since the ranking depends on the “closeness” to the training dataset, so I list the delta value (the absolute diff.) in page 4.
- From page 4. we can see that generally greedy and beam search are better strategies than top-p sampling if we consider the delta value.
 - The result is quite intuitive since those strategies will choose the overall high probabilities combinations.
- From page 4. we can also see that llama generally outperforms gpt2.
- Also, I have tried qualitative analysis by myself, and I found that sometimes model will output exactly same song as the human player.
- I choose the best setting: llama + greedy decoding as the submission.

Appendix: generated examples

- You can check the examples here:

https://drive.google.com/drive/folders/1byOQ4b760ZqVCIr8OtLETD4GKogQKxe9?usp=drive_link

Reference

- [1] Radford, Alec, et al. "Language models are unsupervised multitask learners." OpenAI blog 1.8 (2019): 9.
- [2] Touvron, Hugo, et al. "Llama: Open and efficient foundation language models." arXiv preprint arXiv:2302.13971 (2023).