Master thesis

University of Tartu

February 27, 2023

1 Meeting notes 3

1. Finetune T5-XL model. Since T5-XXL might be not feasible to finetune. I hypothesize that if finetuning T5-Large is taking 30 hours and T5-Large is 15 times smaller than T5-XXL, thus compute time should grow linearly i.e. $(15 * 30)/24 = 18.75$ or 19 days to finetune. T5-XL(3B), on the other hand, $(3 * 30)/24 = 3.75$ or 4 days, which is acceptable.

2. While we are waiting for the data from [Neeman et al., 2022]. I should start implementing a) **prefix tuning**([Li and Liang, 2021]), b) **adaptors**([Houlsby et al., 2019]). But I'm not sure about **in_context_learning** ([Brown et al., 2020]), because the maximum input for T5 is 512 tokens, and considering that NQ dataset, has quite long passages. Such circumstances make it infeasible to apply in context learning. On top of that it is usually requires to provide several example for the model, which makes situation even worse.

3. Read again [Neeman et al., 2022] and analyse thoroughly their results for T5-Large, and contrast them with results for T5-XXL. The main goal is to understand whether usage a smaller model also proves their points. Maybe it is the case, then I should try next big model i.e. T5-XL. Hopefully, we won't need to train T5-XXL, to prove our point.

4. My question was: What are the most important parts in the thesis?

   - Methodology
   - Results
   - Discussion

5. Since thesis report is also really important I need to concentrate my efforts 50/50 for report and research correspondingly.

6. For Yova good thesis would be the one that you could send to the conference and receive conference score 3 or higher. But I'm not sure what this conference score or impact factor represents. So I need to find out about the process of submitting the paper.

Todos:

1. Read [Neeman et al., 2022] [Li et al., 2022], and understand their results in relation with model size.
2. Ask Maarja: 1) about thesis requirement? 2) who will be in the committee? and 3) how thesis is graded? whether is presentation part of grading or only report is matter.
3. Read HPC documentation, to understand allocation of two machines.
4. Find out about process of submitting paper to the conference.
5. Start fine-tuning T5-XL
6. Start implementation prefix-tuning and adaptors.
7. Setup wandb
8. Submit code for finetuning t5

References

[Brown et al., 2020] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners.

[Houlsby et al., 2019] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. (2019). Parameter-efficient transfer learning for nlp.

[Li et al., 2022] Li, D., Rawat, A. S., Zaheer, M., Wang, X., Lukasik, M., Veit, A., Yu, F., and Kumar, S. (2022). Large language models with controllable working memory.

[Li and Liang, 2021] Li, X. L. and Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation.

[Neeman et al., 2022] Neeman, E., Aharoni, R., Honovich, O., Choshen, L., Szpektor, I., and Abend, O. (2022). Disentqa: Disentangling parametric and contextual knowledge with counterfactual question answering.