

Master thesis

University of Tartu

March 22, 2023

1 Meeting notes 6

1. I need to do write up about methods I'm using (prefix-tuning, prompt-tuning, lora, adapters, etc). I will write an article
2. Yova also pointed out that it is ok to mix models i.e. some method would be based of T5-Large, and another based on T5-Large-LM-adapt. This is needed because prompt-tuning is working for me only with T5-Large-LM-adapt. We just need to add accuracy on T5-Large in the appendix. We will also report that in [Lester et al., 2021] they also hypothesized that it would be much harder for prompt-tuning on T5 with span corruption objection to unlearn such an objective.
3. Stick with conventional research steps i.e. show test-set accuracy for the best checkpoint chosen based on validation-set. In this way it will keep me less tempting to report results based test-set accuracy alone.
4. In terms of experiments in total we are going to have four of them i.e. four tables where each table would use common training dataset(f or f+cf or f+a or f+cf+a).
 - (a) All methods trained on **factual** dataset, and evaluation presented as:
 - i. exact match on **factual** test-set
 - ii. exact match on **counterfactual** test-set
 - iii. exact match on **empty** test-se
 - iv. exact match on **random** test-set
 - (b) ...
 - (c) ...

- (d) All methods trained on **f+cf+a** dataset, and evaluation presented as:
- i. exact match on **factual** test-set
 - ii. exact match on **counterfactual** test-set
 - iii. exact match on **empty** test-se
 - iv. exact match on **random** test-se

Here is the experiments table template.

Methods(datasets)	f	cf	empty	random
adversarial-training(f)	-	-	-	-
prompt-tuning(f)	-	-	-	-
adapters(f)	-	-	-	-
lora(f)	-	-	-	-
fine-tuning(f)	-	-	-	-
DisentQA fine-tuning(f)	-	-	-	-

We don't care about predicting multiple answers because it isn't a point that we try to make. In [Neeman et al., 2022] it was core of their paper, because they were studying whether it is possible to disentangle data in case of knowledge conflicts.

We stick with predicting single answer, and our point is to force T5 predict an answer based on contextual data i.e. giving higher priority to contextual rather than parametric data. And thus we want to avoid memorization, but rather expect algorithm to emerge.

5. We could also in all our experiments use any T5 variation we like(T5 or T5-LM-adapt), and just in the appendix show that they are on par with baseline model. My assumption is that T5-LM-adapt converges faster than T5, but performance would be the same.
6. In [Neeman et al., 2022] they don't provide all the results for T5-Large thus for missing results we will fine-tune ourselves, and show our fine-tuning matches theirs(when they provided them).
7. I need to improve my presentations i.e. explain a graph where they should pay attention to on so on. Otherwise, it is hard to understand me. And also I should have good visuals.
8. It is good to prepare for the presentation. Yeah, next time I would create slides because I think they are easy to comprehend.

Todos:

1. Write article about LoRA
2. Add table in wandb instead of graphs because it is more appropriate
3. Find a way to change wandb graphs because lines are too thin
4. Instructive in-context-learning for T5-FLAN-XXL(no weights updates)
5. Prepare slides for my next presentation
6. Read HPC documentation, to understand allocation of two machines.
7. **Investigate:** Conduct additional experiments with adversarial-training like:
 - (a) use $\lambda = 1$, to see whether it would match with fine-tuning that was training on mix of factual and counterfactual knowledge.
 - (b) combine factual and counterfactual into single batch - investigate whether such data splitting affect training somehow.
 - (c) penalize factual loss - to explore the impact of penalization if such exist.
 - (d) beside penalizing counterfactual I could also penalize empty and random contexts - extension for other datasets.

Desirable outcome:

1. Write article about our methodologies(like LoRA, etc).
2. Finish the first experiment
3. Improved presentation of my experiments
4. T5-FLAN-XXL inference for instructive in-context-learning

References

- [Lester et al., 2021] Lester, B., Al-Rfou, R., and Constant, N. (2021). The power of scale for parameter-efficient prompt tuning.
- [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.