Master thesis

University of Tartu

February 22, 2023

1 Meeting notes 2

1. Alright, due to our miscommunication. I started baseline implementation for the wrong paper([Welbl et al., 2020]). Instead, I need to implement baseline for [Neeman et al., 2022]. The idea is following, in [Neeman et al., 2022] in order to fight problem where model isn't paying attention to the context, authors used mainly data-augmentation(such as augmenting with counterfactual, and non-answerable entries). However, what we wanna do is to improve generalization on a training model, and data level. We will do comprehensive study of how individually those modification would improve model performance(i.e. not-paying attention), as their combinations.

   The model level involves following approaches, that could allow to improve generalization:

   (a) **prefix tuning**([Li and Liang, 2021])
   (b) **adaptors**([Houlsby et al., 2019])
   (c) **in_context_learning**([Brown et al., 2020])

   The data level involves:

   (a) Counterfactual examples([Neeman et al., 2022])
   (b) Non-answerable examples([Neeman et al., 2022])

   The baseline involve finetune T5 11B and 770M models on natural questions dataset.

2. The data from [Neeman et al., 2022] would be available in two weeks.

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 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.

[Welbl et al., 2020] Welbl, J., Minervini, P., Bartolo, M., Stenetorp, P., and Riedel, S. (2020). Undersensitivity in neural reading comprehension.