Modern NLP M1 Project Plan

Team UniGPT

May 2023

1 Introduction

In this report, we present a plan for the project to develop an effective educational assistant using techniques similar to those employed in ChatGPT. We are building an assistant for a neuroscience course. Our project will be divided into the following key stages:

- Collection of the dataset
- Training the reward model
- Fine-tuning the base generative model
- Evaluating the performance of the fine-tuned model
- If time permits, fine-tuning the generative model using RLHF

2 Dataset

To train a robust reward model, we need to first prepare a suitable dataset. We propose the following strategies for dataset construction:

- Combine prompts and responses generated with ChatGPT for each Multiple Choice Question from the EPFL neuroscience course, created by our team members as well as those produced by other teams.
- Clean and preprocess the data obtained from other teams. In this step, we will thoroughly process the collected data to ensure its quality and compatibility.
  1. Separate the data into Conversations (inputs) and Confidence/label (outputs).
  2. Partition the dataset into training, validation, and test subsets. In this step, we will carefully distribute confidence score labels approximately equally among the subset.
  3. Ensure consistency in formatting, before training and evaluation of the model.
  4. To further augment the training set, we may ask ChatGPT to generate additional questions to the original ones and provide answers to them. These 100-200 data points will be then assigned with a confidence score by us.

3 Reward model

For the reward model, we plan to fine-tune an existing pre-trained model, such as BioBERT [1], which has been trained on PubMed and PMC papers.

Initially, we will fine-tune the pre-trained classification model with 5 classes corresponding to the confidence scores. Then if time permits, and we are able to figure out how to implement this, we will formulate the task as a regression problem and train the model to produce a scalar output (for the confidence score) given the input.

In terms of model optimization, we will experiment with the optimizers, different learning rates, and other relevant parameters.

As for the loss function, we plan to use the standard cross-entropy loss for the classification task. Moreover, we will explore incorporating a KL-divergence term into our loss function, as it has been recommended in numerous recent publications, such as InstructGPT [2].

4 GPT-2 fine-tuning

We intend to fine-tune the GPT-2 [3] model using the dataset we annotated in milestone 1 and the data we get from other teams. We can then evaluate our fine-tuned model based on the reward model we trained in milestone 2.
For this part, we intend to use the huggingface transformers library. We do not need to write a lot of code in this section since the huggingface transformers library provides all the necessary API.
For the extra credit part, we plan to fine-tune our model using PPO with the help of TLR library.

5 Evaluation metrics

To evaluate a reward model, we plan to use a multiclass F1 score as a standard metric in classification tasks. For the final model, there are more options to choose from. We will use our reward model and, in addition, since factual consistency is important for the question-answering model, we plan to use the Q^2 metric [4] which was found to perform rather well for this task [5]. This metric is based on Natural Language Inference (NLI) and treats the ground truth text answer as a premise and the generated text answer as a hypothesis. An entailment in this case corresponds to factual consistency.

As a baseline model, we plan to use the initial GPT-2 before it is fine-tuned on our neuroscience datasets.

To qualitatively evaluate the generations of our final model, we aim to compare examples with high and low scores. We can also look for particular types of questions for which the model usually fails to give a correct answer.

6 Conclusion

By following this plan, we aim to create an educational assistant that can help users in various learning scenarios and provide valuable insights or support.

References

[1] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, sep 2019.

[2] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, "Training language models to follow instructions with human feedback," 2022.

[3] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[4] O. Honovich, L. Choshen, R. Aharoni, E. Neeman, I. Szpektor, and O. Abend, "q^2: Evaluating factual consistency in knowledge-grounded dialogues via question generation and question answering." *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.

[5] O. Honovich, R. Aharoni, J. Herzig, H. Taitelbaum, V. Cohen, D. Kukliansky, T. Scialom, I. Szpektor, A. Hassidim, and Y. Matias, "True: Re-evaluating factual consistency evaluation," 2022.