Information Retrieval Project Proposal Domain Adaptation using Parameter Efficient Fine Tuning (PEFT)

Janneke Nouwen (s1101750), Daan Brugmans (s1080742), Julian Roddeman (s1080491)

I. Introduction

According to this study [8], sparse and dense LLM-based retrieval models like BERT (Bidirectional Encoder Representations) are less capable of making predictions in unseen domains. BERT is a complex model, fine-tuning all 110 million trainable parameters of the model for an unseen domain is highly expensive, and the overcomplexity of fine-tuning on this number of parameters can lead to overfitting [1].
Parameter Efficient Fine Tuning (PEFT) approaches only fine-tune a small, new set of parameters while keeping the original parameters of the models frozen. This enables rapid adaptations of pre-trained language models to varied domains without fine-tuning the original parameters of the model, lowering computational and storage costs significantly. Because only a small number of parameters is used, this method of fine-tuning is less susceptible to overfitting [7]. In this research we will compare the capability of plain BERT to classify unseen documents to a fine-tuned approach using Parameter Efficient Fine Tuning (PEFT) on BERT.
We will use the TREC-COVID dataset from the BEIR framework [8, 9] to determine the model's capability to classify documents in an unseen domain. This decision is based on the fact that BERT was pre-trained solely on an unlabeled, plain text corpus [2] and has no exposure to literature pertaining to the COVID-19 virus, given that its training data predates the pandemic. As a result, the specificity and freshness of the COVID-19 topic certainly designates it as an unseen domain for the model, establishing a valid dataset to assess PEFTs potential to boost the performance of classifying documents of an unseen domain.

II. Research Question

How can a Parameter Efficient Fine Tuning (PEFT) method improve the performance of BERT for classifying documents in unseen domains, when we compare it to a plain BERT setup?

III. Resources

Python 3.10 will be used to test plain BERT and BERT fine-tuned with PEFT. To apply different PEFT methods to BERT, the 'adapter-hub' library will be used [6]. We will compare the results using the Scikit-learn package's precision, F1-score and recall [5]. For data preprocessing, Numpy and Pandas will be used [3, 4].

IV. Experimental Design

Objective: Our objective is to compare the capability of BERT to classify documents in an unseen domain, comparing a non-fine-tuned plain version of BERT with a dense layer, against a fine-tuned approach with PEFT. We will do this in an unseen domain, specifically in a Covid-19 domain. In figure 1, a flowchart shows the experimental setup. **Hypothesis:** We expect that PEFT will enable BERT to effectively classify documents in an unseen (COVID-19) domain with improved performance compared to a plain BERT model. **Independent Variable:** Fine-tuning approach (plain BERT vs. fine-tuned BERT using PEFT) **Dependent Variable:** Model's performance in classifying relevancy labels of COVID-19 documents using the precision, recall and F1-score.

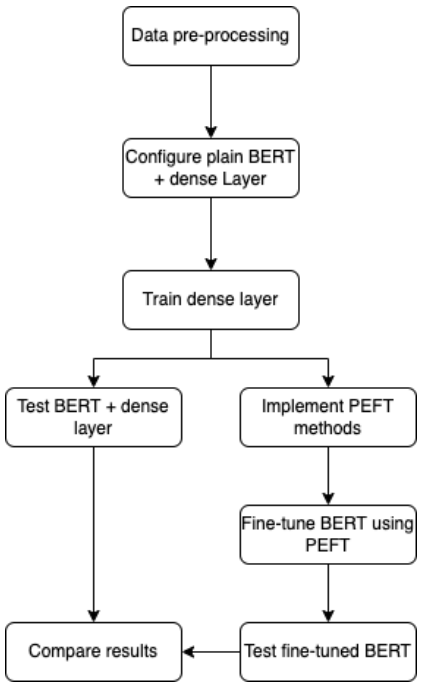

Fig. 1. Flowchart of the experiment setup

References

[1] Mohammad Mahdi Bejani and Mehdi Ghatee. "A systematic review on overfitting control in shallow and deep neural networks". In: *Artificial Intelligence Review* (2021), pp. 1–48.

[2] Jacob Devlin et al. “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805* (2018).

[3] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

[4] Wes McKinney et al. “Data structures for statistical computing in python”. In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX. 2010, pp. 51–56.

[5] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[6] Jonas Pfeiffer et al. “AdapterHub: A Framework for Adapting Transformers”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Oct. 2020, pp. 46–54. DOI: 10.18653/v1/2020.emnlp-demos.7. URL: https://aclanthology.org/2020.emnlp-demos.7.

[7] Deepak Soekhoe, Peter Putten, and Aske Plaat. “On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks”. In: Oct. 2016, pp. 50–60. ISBN: 978-3-319-46348-3. DOI: 10.1007/978-3-319-46349-0_5.

[8] Nandan Thakur et al. “Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models”. In: *arXiv preprint arXiv:2104.08663* (2021).

[9] Ellen Voorhees et al. “TREC-COVID: constructing a pandemic information retrieval test collection”. In: *ACM SIGIR Forum*. Vol. 54. 1. ACM New York, NY, USA. 2021, pp. 1–12.