RELATION BETWEEN INTRINSIC DIMENSION AND THE RANK OF LoRA ADAPTORS

Prasanth YSS

1 BACKGROUND AND PROBLEM STATEMENT

Aghajanyan et al. (2020) try to explain the effectiveness of finetuning in large language models through the lens of ***intrinsic dimension*** [Li et al. (2018)]. *Intrinsic dimension* of a objective function is defined to be the minimum dimension needed to solve the original problem to a reasonable precision. They argue that pre-training a language model reduces the minimum description length of various downstream NLP tasks in the representation space of the pre-trained model. They go on to show that the *intrinsic dimension* for a given task is inversely related to the number of parameters. These ideas inspired the creation of Low Rank Adaptors (LoRA) [Hu et al. (2021)] for parameter efficient fine-tuning.

We define the *intrinsic dimension* (d_{90}) for a given pair of model and task to be the minimum dimension required to achieve 90% of the accuracy of full parameter fine-tuning. For this project, we wish to analyse the relation between the *intrinsic dimension* and the *rank* of LoRA for a given model and task pair. We want to ask the following questions

- For a given model, is the *intrinsic dimension* directly related to the rank of LoRA adaptors across tasks?
- For a given task, is the minimum LoRA rank r_{90} inversely related to the number of model parameters across models?

We will compare d_{90} to the minimum rank r_{90} of a LoRA adaptor to achieve the same 90% accuracy. It would also be interesting to observe the trend of *intrinsic dimension* with model parameters also apply to the rank of a LoRA adaptor, i.e. we wish to see if there is an inverse relation between r_{90} and number of parameters. These experiments could potentially indicate any shortcomings of LoRA adaptors and could lead to better approaches for parameter efficient fine-tuning.

2 EXPERIMENTS AND EVALUATION

We will take the evaluation approach given by Aghajanyan et al. (2020). For a given model and dataset, we run 10 subspace trainings with d ranging from 100 to 10000 on a log scale. The minimum d that achieves 90% accuracy gives us d_{90}. Similar evaluation will be done for r_{90} with r ranging from 1 to 1000. For every training run we might do a hyperparameter search across learning rates.

We evaluate the first set of experiments, i.e. comparing *intrinsic dimension* with rank of LoRA adaptor, on RoBERTa-Large [Liu et al. (2019)] across MRPC [Dolan & Brockett (2005)], SST [Socher et al. (2013)], ANLI [Nie et al. (2020)] and QQP [Chen et al. (2017)] datasets. Microsoft Research Paraphrase Corpus (MRPC) is a corpus consisting of 5,801 sentence pairs collected from newswire articles. Each pair is labelled if it is a paraphrase or not by human annotators. The Stanford Sentiment Treebank (SST) is a corpus consisting of 11,855 single sentences extracted from movie reviews. The Adversarial Natural Language Inference (ANLI) is a large-scale NLI benchmark dataset, the data is selected to be difficult to BERT and RoBERTa models. Quora Question Pairs (QQP) dataset consists of over 400,000 question pairs, and each question pair is annotated with a binary value indicating whether the two questions are paraphrase of each other.

The second set of experiments to evaluate the trend between r_{90} and number of model parameters will be done on the ANLI dataset across BERT-Base, BERT-Large, RoBERTa-Base, and RoBERTa-Large models. Bert-Base and Large models consist of $110M$ and $340M$ parameters respectively. Whereas the RoBERTa versions consist of $125M$ and $355M$ parameters. All the datasets and models are open-sourced and publicly available.

References

Armen Aghajanyan, Luke Zettlemoyer, and Sonal Gupta. Intrinsic dimensionality explains the effectiveness of language model fine-tuning, 2020.

Zihang Chen, Hongbo Zhang, Xiaoji Zhang, and Leqi Zhao. Quora question pairs. 2017. URL https://api.semanticscholar.org/CorpusID:233225749.

William B. Dolan and Chris Brockett. Automatically constructing a corpus of sentential paraphrases. In *International Joint Conference on Natural Language Processing*, 2005. URL https://api.semanticscholar.org/CorpusID:16639476.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021.

Chunyuan Li, Heerad Farkhoor, Rosanne Liu, and Jason Yosinski. Measuring the intrinsic dimension of objective landscapes, 2018.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. Adversarial nli: A new benchmark for natural language understanding, 2020.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1631–1642. ACL, 2013. URL https://aclanthology.org/D13-1170/.