UNIVERSITY OF TARTU

Institute of Computer Science

Data Science Curriculum

Mart-Mihkel Aun

Prefix Tuning Applicability In Language Models

Master's Thesis (15 ECTS)

Supervisor:

Sven Laur, DSc

Tartu 2026

Prefix Tuning Applicability In Language Models

Abstract:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

Keywords: LLM, fine-tuning

CERCS: P176

Prefix Tuning Applicability In Language Models

Lühikokkuvõte:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

Võtmesõnad: LLM, fine-tuning

CERCS: P176

Contents

1. Introduction

$$h_i = \mathrm{LM}_\phi(z_i, h_{<i})$$

2. Related Work

2.1. Context based parameter efficient fine-tuning

Prefix tuning [1], P-Tuning [2], Prompt tuning [3].

2.2. Weight based parameter efficient fine-tuning

LoRA [4].

2.3. Language models

BERT [5], Llama [6].

Few-shot learning [7].

3. Conclusion

Bibliography

[1] X. L. Li and P. Liang, “Prefix-tuning: Optimizing continuous prompts for generation,” *arXiv preprint arXiv:2101.00190*, 2021.

[2] X. Liu *et al.*, “GPT understands, too,” *AI Open*, vol. 5, pp. 208–215, 2024.

[3] B. Lester, R. Al-Rfou, and N. Constant, “The power of scale for parameter-efficient prompt tuning,” *arXiv preprint arXiv:2104.08691*, 2021.

[4] E. J. Hu *et al.*, “Lora: Low-rank adaptation of large language models.,” *ICLR*, vol. 1, no. 2, p. 3, 2022.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[6] H. Touvron *et al.*, “Llama: Open and efficient foundation language models,” *arXiv preprint arXiv:2302.13971*, 2023.

[7] T. Brown *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

4. Appendices

5. License