东北师范大学

研究生学位论文开题报告

论 文 题 目：Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do.

报 告 人 姓 名：张三

研 究 方 向：深度学习理论

学 科 专 业：计算机科学与技术

年　　级：2024 级

学 历 层 次：博士生 □ 硕士生 ✓

学 位 类 型：学术学位 ✓ 专业学位 □

指 导 教 师：Lorem ipsum dolor.

培 养 单 位：信息科学与技术学院

撰写说明

1.文献综述应基于选题领域内具有代表性的文献进行，需满足一定的字数要求。博士生：文科不得少于 10000 字，理科不得少于 6000 字。

2.参考文献是指在开题报告中实际引用的文献。博士生实际引用文献须不少于 50 篇，硕士生实际引用文献须不少于 30 篇。参考文献格式参照学位论文格式要求，建议文中引用文献以脚注形式标注，并在文末按照字母顺序列出所有引用文献。

3.博士生论文开题时间与学位论文通讯评阅时间间隔原则上不少于 1.5 年，硕士生论文开题时间与学位论文通讯评阅时间间隔原则上不少于 8 个月。 开题报告审查小组根据开题报告情况，在相应的 □ 内打号。合格的开题报告，由学院存档并作为毕业审核材料之一。

4.开题报告中的字体字号均用宋体小四，页边距上下 20MM,左右 25MM，用 A4 纸打印，于左侧装订成册。 开题结束后，研究生需针对开题中所提问题与建议进行修改，并向学院提交开题报告修订花脸稿。

一、 研究背景（分析本选题范畴内尚未得到较好解决的学术或实践难题，阐述选题的缘起与依据）

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 postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题，

二、 文献综述（系统梳理本选题相关的具有代表性的文献，分析相关研究的发展脉络与进展，评述已有研究存在的问题与不足）

2.1. Transformer 的表达能力与通用近似能力

2.2. Transformer 的泛化能力

2.2.1. 泛化理论

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 postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题， 分析本选题范畴内尚未得到较好解决的学术或实践难题，

2.2.2. 大模型的泛化现象

2.3. Transformer 的训练

三、 研究问题（提出本论文拟回答的核心问题及具体研究问题）

四、 研究意义（阐述本研究可能的理论贡献与实践价值）

五、 研究设计（针对研究问题，详细阐述本选题的研究内容、基本思路或总体框架、理论基础、具体研究方案等）

六、 进度安排（按照时间顺序，就研究的进度做出具体的规划）

2024 年 11 月-2025 年 1 月：代码编写，实验验证

2025 年 2 月-2025 年 3 月：证明 D-TF 的 token-coherence 与收敛和泛化的关系

2025 年 4 月-2025 年 5 月：撰写论文

2025 年 6 月：论文修改，准备答辩

七、 参考文献

[1] S. Chatterjee and P. Zielinski, “On the generalization mystery in deep learning,” *arXiv preprint arXiv:2203.10036*, 2022.

[2] A. Vaswani, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.

[3] A. Kazemnejad, I. Padhi, K. Natesan Ramamurthy, P. Das, and S. Reddy, “The impact of positional encoding on length generalization in transformers,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[4] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, “Ammus: A survey of transformer-based pretrained models in natural language processing,” *arXiv preprint arXiv:2108.05542*, 2021.

[5] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in vision: A survey,” *ACM computing surveys (CSUR)*, vol. 54, no. 10s, pp. 1–41, 2022.

[6] L. Dong, S. Xu, and B. Xu, “Speech-transformer: a no-recurrence sequence-to-sequence model for speech recognition,” in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2018, pp. 5884–5888.

[7] A. Radford, “Improving language understanding by generative pre-training,” 2018.

[8] H. Touvron *et al.*, “Llama 2: Open foundation and fine-tuned chat models,” *arXiv preprint arXiv:2307.09288*, 2023.

[9] J. Pérez, P. Barceló, and J. Marinkovic, “Attention is turing-complete,” *Journal of Machine Learning Research*, vol. 22, no. 75, pp. 1–35, 2021.

[10] C. Yun, S. Bhojanapalli, A. S. Rawat, S. J. Reddi, and S. Kumar, “Are transformers universal approximators of sequence-to-sequence functions?,” *arXiv preprint arXiv:1912.10077*, 2019.

[11] A. Petrov, P. H. Torr, and A. Bibi, "Prompting a pretrained transformer can be a universal approximator," *arXiv preprint arXiv:2402.14753*, 2024.

[12] V. Nagarajan and J. Z. Kolter, "Uniform convergence may be unable to explain generalization in deep learning," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[13] O. Bousquet and A. Elisseeff, "Stability and generalization," *The Journal of Machine Learning Research*, vol. 2, pp. 499–526, 2002.

[14] D. A. McAllester, "Some pac-bayesian theorems," in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 230–234.

[15] P. Foret, A. Kleiner, H. Mobahi, and B. Neyshabur, "Sharpness-aware minimization for efficiently improving generalization," *arXiv preprint arXiv:2010.01412*, 2020.

[16] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] Z. Li, H. Liu, D. Zhou, and T. Ma, "Chain of thought empowers transformers to solve inherently serial problems," *arXiv preprint arXiv:2402.12875*, 2024.

[18] Y. Pan and Y. Li, "Toward understanding why adam converges faster than sgd for transformers," *arXiv preprint arXiv:2306.00204*, 2023.

[19] L. Bottou and others, "Stochastic gradient learning in neural networks," *Proceedings of Neuro-Nımes*, vol. 91, no. 8, p. 12, 1991.

[20] Y. Yu *et al.*, "Low-rank adaptation of large language model rescoring for parameter-efficient speech recognition," in *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2023, pp. 1–8.

[21] F. Meng, Z. Wang, and M. Zhang, "Pissa: Principal singular values and singular vectors adaptation of large language models," *arXiv preprint arXiv:2404.02948*, 2024.

[22] Q. Zhang *et al.*, "AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning," *arXiv preprint arXiv:2303.10512*, 2023.

[23] S. Wang, L. Yu, and J. Li, “LoRA-GA: Low-Rank Adaptation with Gradient Approximation,” *arXiv preprint arXiv:2407.05000*, 2024.

[24] J. Wei *et al.*, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in neural information processing systems*, vol. 35, pp. 24824–24837, 2022.

[25] Q. Dong *et al.*, “A survey on in-context learning,” *arXiv preprint arXiv:2301.00234*, 2022.

[26] B. Neyshabur, S. Bhojanapalli, and N. Srebro, “A pac-bayesian approach to spectrally-normalized margin bounds for neural networks,” *arXiv preprint arXiv:1707.09564*, 2017.

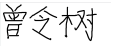
[27] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, “The curious case of neural text degeneration,” *arXiv preprint arXiv:1904.09751*, 2019.

[28] M. E. Ildiz, Y. Huang, Y. Li, A. S. Rawat, and S. Oymak, “From Self-Attention to Markov Models: Unveiling the Dynamics of Generative Transformers,” *arXiv preprint arXiv:2402.13512*, 2024.

[29] A. Krogh and J. Hertz, “A simple weight decay can improve generalization,” *Advances in neural information processing systems*, vol. 4, 1991.

[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[31] M. Hardt, B. Recht, and Y. Singer, “Train faster, generalize better: Stability of stochastic gradient descent,” in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger, Eds., in Proceedings of Machine Learning Research, vol. 48. New York, New York, USA: PMLR, 2016, pp. 1225–1234. [Online]. Available: https://proceedings.mlr.press/v48/hardt16.html

指导教师签字： 

2024 年 11 月 5 日

审查小组意见		
开题报告审查小组成员名单		
张三	讲师	东北师范大学
李四	教授	北京大学
王五	教授	复旦大学
赵六	副教授	南京大学
孙七	讲师	浙江大学

审查结论

合格，修改后可以进入学位论文写作阶段 □

不合格，需再次进行学位论文开题报告 □

组长签字：曾令树

单位公章： 2024 年 11 月 5 日