陈英发 Yingfa Chen

chenyingfa1999@qq.com | 188 0117 9013 | www.github.com/chen-yingfa

北京 | 研究方向: NLP、LLM、长文本、知识更新

教育经历 Education

清华大学 博士 | 计算机科学与技术，自然语言处理 2024 年 8 月 - 现在
导师: 刘知远 北京

清华大学 硕士 | 计算机科学与技术，自然语言处理 2022 年 9 月 - 2024 年 7 月
导师: 刘知远，GPA: 3.9/4.0 北京

清华大学 本科 | 计算机科学与技术 2018 年 8 月 - 2022 年 7 月
- GPA: 3.4/4.0 北京
- 二学位: 数字媒体设计

发表文章 Publications

Cost-Optimal Grouped-Query Attention for Long-Context LLMs
Yingfa Chen, Yutong Wu et al.

- 从 FLOPs 和 memory 开销的角度，对比不同 GQA 头数量的配置，在不同序列长度下最小化开销。
- 发发现 context 较长时，应使用更少的 attention 头且更大的模型，可保证能力不下降的情况下，减少约 50% 的开销。

Stuffed Mamba: State Collapse and State Capacity of RNN-Based Long-Context Modeling
Yingfa Chen et al.

- 发现 Mamba 模型长度泛化能力差，且将其归因于模型没有学会健壮的遗忘机制。
- 发现存在一个跟 state 大小有关的阈值，当且仅当训练长度超过该阈值，Mamba 才能学会遗忘并且能够长度泛化。

∞-Bench: Extending Long Context Evaluation Beyond 100K Tokens (ACL 2024)
Xinrong Zhang, Yingfa Chen et al.

- 构造首个超过 100K 长度的大模型评测集。
- 包含不同语言（中英）和不同领域（数学，代码，自然文本），同时包含合成任务和真实任务。

Robust and Scalable Model Editing for Large Language Models (COLING 2024)
Yingfa Chen et al.

- 提出一种基于检索的大模型知识更新的框架，可以同时处理串行和并行的更新操作。
- 构造一个更具有挑战性的知识更新评测数据集。

CFDBench: A Large-Scale Benchmark for Machine Learning Methods in Fluid Dynamics (preprint)
Yining Luo, Yingfa Chen et al.

- 构造了首个针对机器学习模型的，包含多种边界条件、几何形状和流体物性的流体力学评测集。

Sub-Character Tokenization for Chinese Pretrained Language Models (TACL 2023)
Yingfa Chen, Chenglei Si, Zhengyan Zhang et al.

- 将汉字根据字形或者发音转换为 sub-character 序列。
- 此分词器可以获得更短的编码从而提高训练速度，以及对同音字导致的噪声更鲁棒，且它没有牺牲准确率。

其他

- 编程: PyTorch、Python、Huggingface、BMTrain、C++。
- 语言: 普通话、粤语、英语（流利）、挪威语（近乎母语）。
- 兴趣: 羽毛球（系队）。
- 竞赛经历: 数学和信息学（挪威国家队）。
- 个人背景: 本人是挪威籍三代华裔，父母分别是越南和柬埔寨华裔，本人在挪威出生。