

冯兆鹏

手机：13733562319

邮箱：zhaopengfeng424@gmail.com

个人主页：<https://fzp0424.github.io>

出生年月：2001. 04

Github：<https://github.com/fzp0424>

研究领域：大模型 RL/Post-train/Agent/机器翻译

教育经历

浙江大学	硕士	计算机技术	2023. 09-至今
哈尔滨工业大学（深圳）	本科	自动化	2019. 08-2023. 06

个人自述

- 有极强的自驱力，持续的学习吸收能力，良好的沟通合作能力，扎实的工程实践能力，能够在合适的环境中发挥巨大能量。
- 未来感兴趣的研究方向涉及强化学习，多模态，目前对于大模型发展的认知：框架 > 模型 。希望能做出更普适更有影响力的工作并且相信自己能做到。

科研经历

- Zhaopeng Feng***, Jiayuan Su*, Jiamei Zheng, Jiahao Ren, Yan Zhang, Jian Wu, Hongwei Wang, Zuozhu Liu, “[M-MAD: Multidimensional Multi-Agent Debate for Advanced Machine Translation Evaluation](#)”, *ACL 2025 main*.
 - 应对现有机器翻译评估方法在鲁棒性、精确性和可解释性上的挑战，提出 M-MAD 框架。通过维度划分（解耦 MQM）、多智能体正反方辩论（LLM Debate）及最终综合判断，系统化提升 LLM 翻译评估的质量。
 - **成果**：在片段级评估上显著超越现有 LLM-as-a-judge 方法，系统级性能保持顶尖，整体表现优于多数先进的自动评估指标，且无需额外训练。有效提升了错误定位的准确性。
- Zhaopeng Feng***, Ruizhe Chen*, Yan Zhang, Zijie Meng, Zuozhu Liu, “[Ladder: A Model-Agnostic Framework Boosting LLM-based Machine Translation to the Next Level](#)” *EMNLP 2024 main*.
 - 针对通用大模型训练成本高昂及专用翻译模型依赖大量标注数据的痛点，提出 Ladder，一个模型无关、经济高效的训练范式。利用现有 LLM 生成伪改进三元组，并通过阶梯微调策略（HFT），实现从易到难逐步提升翻译改进性能。
 - **成果**：MT-Ladder-7B 的翻译改进效果媲美 GPT-4o mini，显著提升 GPT-3.5、ALMA-13B 等模型的性能；MT-Ladder-2B 使小型通用模型性能接近甚至超越专业翻译大模型。
- Zhaopeng Feng***, Yan Zhang*, Hao Li, Wenqiang Liu, Jun Lang, Yang Feng, Jian Wu, Zuozhu Liu, “[TEaR: Improving LLM-based Machine Translation with Systematic Self-Refinement](#)”, *NAACL 2025 Findings*. ([Github 40+ stars](#))
 - 针对顶尖 LLM 在翻译中仍存在错误的问题，提出 TEaR 框架，通过“翻译-评估-修正”三模块实现系统化自我修正。核心在于引入独立的、模仿人类 MQM 的评估模块，为 LLM 提供清晰、可操作的内部反馈以指导优化。
 - **成果**：有效提升了 GPT-4o mini 等模型在多种语言对上的翻译质量，在自动指标和 MQM 分数上均超越现有内部修正及外部反馈方法。证明迭代修正能持续减少翻译错误，评估模块是关键。Similar and Recognized by Andrew Ng’s [translation-agent](#).
- Zhaopeng Feng**, Shaosheng Cao, Jiahao Ren, et al. “[MT-R1-Zero: Advancing LLM-based Machine Translation via R1-Zero-like Reinforcement Learning](#)”, *EMNLP 2025 Findings*. ([Github 50+ stars](#))
 - 探索将 R1-Zero 范式应用于纯文本机器翻译，旨在通过强化学习激发大模型的“翻译推理能力”，无需监督微调或 CoT 冷启动。创新性提出规则-度量混合奖励机制（Rule-Metric Mixed Reward）和 GRPO 算法进行端到端优化。
 - **成果**：首个 MT 领域 R1-RL 范式开源实现。7B 模型在中英翻译任务上达到 SOTA 闭源模型水平，3B 模型在语义指标上超越远大于其容量的 70B+模型。实验证明奖励信号可塑造翻译风格，并观察到模型思考过程的动态演变。
- Zhaopeng Feng***, Jiahao Ren*, Jiayuan Su, et al. “[MT-RewardTree: A Comprehensive Framework for Advancing LLM-Based Machine Translation via Reward Modeling](#)”, *EMNLP 2025 Findings*.
 - 针对机器翻译领域过程奖励模型（PRM）构建与评估方法缺失，特别是词元级标注成本高昂的问题，提出 MT-RewardTree 框架。采用近似蒙特卡洛树搜索（MCTS）自动生成词元级偏好数据，用于训练隐式 PRM。
 - **成果**：同时构建了 MT-RewardBench 基准。实验证明，基于 MCTS 生成的词元级偏好数据训练的 PRM 在判别准确率上显著优于序列级偏好方法，并成功应用于测试时对齐（TTA）和词元级信用分配（Credit Assignment）。

⑥ Yan Zhang*, Zhaopeng Feng*, Zhiyang Teng, Zuozhu Liu, Haizhou Li, “[How Well Do Text Embedding Models Understand Syntax?](#)”, *EMNLP 2023 Findings*.

- 目前的文本表征模型在语义理解, MTEB 评测上取得了卓越的成绩, 但我们发现其在句法理解能力上的评估不足, 我们开发了 SR 基准。通过结构启发式 (主被动转换) 和关系理解 (概念顺序扰动) 探究模型对句法变化的敏感度。
- **成果:** 发现先进的文本表征模型 (如 OpenAI-Ada-002, SimCSE) 在 SR 基准上表现不佳, 难以捕捉句法结构变化对语义的影响。提出的针对性数据增强策略能有效提升模型的句法理解能力, 且不损害其在传统 STS 任务上的性能。

⑦ Zhaopeng Feng*, Yupu Liang*, Shaosheng Cao, et al. “[MT³: Scaling MLLM-based Text Image Machine Translation via Multi-Task Reinforcement Learning](#)”, *Submitted to NeurIPS 2025*.

- 针对文本图像机器翻译 (TIMT) 中精确 OCR、视觉推理与高质量翻译的挑战, 提出 MT³ 框架, 将多任务强化学习应用于 MLLM 进行端到端 TIMT。通过分解子技能并设计多混合奖励机制进行协同优化。
- **成果:** MT³-7B-Zero 在 MIT-10M 基准上显著优于现有系统, 甚至超越更大规模的 MLLM, 并展现强大领域外泛化能力。同时构建了首个真实社交媒体 TIMT 测试集 XHSPost。

⑧ Jiayuan Su*, Fulin Lin*, Zhaopeng Feng*, et al. “[CP-Router: An Uncertainty-Aware Router Between LLM and LRM](#)”, *Submitted to AAAI 2026*.

- 针对推理模型 (LRM) 在简单问题上“过度思考”导致的效率低下问题, 提出 CP-Router 框架。这是一个免训练、模型无关的动态路由机制, 基于一致性预测 (CP) 估计的不确定性, 在标准 LLM 和 LRM 间智能分配任务。
- **成果:** 在保持甚至提升推理准确率的同时, 显著降低了 token 消耗和对昂贵 LRM 的调用频率。在多个数学、逻辑推理基准上验证了其有效性和泛化能力。

⑨ Zhaopeng Feng*, Zijie Meng*, Zuozhu Liu, “[EC-Guide: A Comprehensive E-Commerce Guide for Instruction Tuning and Quantization](#)”, *KDD Workshop, 2024*.

⑩ Zhaopeng Feng, Keyang Zhang, Shuyue Jia, Baoliang Chen, Shiqi Wang, “[Learning from mixed datasets: A monotonic image quality assessment model](#)”, *Electronics Letters 2023*.

实习经历

阿里巴巴 - 智能信息事业群 - 夸克基础模型 RLHF

mentor: 曹宇

2024. 11 - 2025. 03

- 探索双轮训练范式 (轮次范式 vs. 续写范式) 对激发模型 Self-Reflection 能力的影响。探索合成数据 (自激发 vs. 蒸馏) 对模型数学推理能力的起始与后续 RL 训练的影响。
- **产出:** ① AC 数据合成框架: 设计 solution-level 和 step-level 的多轮反思交互范式, 结构化地让模型自我蒸馏生成包含推理、评估与自主纠正过程的训练数据。② 双轮训练探索结论: 通过实验验证了直接对答案进行奖惩的“轮次范式”RL 在小模型上难以激发有效反思, 且方法严重依赖基础模型。提出了“续写范式”作为更优解, 它能更稳定地引导模型进行探索性重答 (Exploration), 为后续结合 Verifier 提升多轮推理能力探索出一条有效路径。

小红书 - 社区 NLP 算法 - 大模型 Post-training

mentor: 曹绍升

2025. 03 - 至今

- 探索将前沿强化学习范式 (如 R1-Zero) 应用于纯文本及多模态图文机器翻译任务, 旨在通过设计精细化的混合奖励机制与优化算法, 端到端地提升模型的翻译推理与多技能协同能力。
- **产出:** ① MT-R1-Zero 框架: 主导设计并开源了首个将 R1 强化学习范式应用于机器翻译的框架。该框架使 7B 模型在中英互译任务上性能媲美 SOTA 闭源模型。② MT³ 多任务 RL 框架: 针对图文翻译, 提出 MT³ 框架并构建了首个真实场景社交媒体测试集 (XHSPost)。该方法使 7B MLLM 在 TIMT 基准上的性能显著超越了更大规模的模型。

其他经历

- KDD Cup 2024 知识推理赛道 Top 2 (Best Student Team Award, TEAM LEAD), 全赛道 Top 5
- 曾运营公众号“深度学习自然语言处理”(10w 订阅)
- 创建小红书论文分享账号“nlper 今天读 paper 了吗”(1.3w 关注)
- 深圳高校赛艇联赛男子八人艇季军, 篮球国家二级裁判员, 半程马拉松 150
- ROBOCON 全国大学生机器人大赛国家三等奖 (电控组)