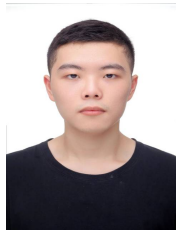




赵福方宸

☎ 13700903801 ✉ zhaofufangchen@bupt.edu.cn



🎓 教育经历

北京邮电大学	2023.09 – 至 今
计算机科学与技术 博士 计算机学院	北京
• 硕博连读	
北京邮电大学	2021.09 – 2023.06
智能科学与技术 硕士 人工智能学院	北京
海南大学	2017.9 – 2021.06
计算机科学与技术 本科 网络空间安全学院	海口
• GPA: 3.82/4.00	
• 全国建模大赛（国二2019，国二2020）	
• 一等奖学金（2019，2020）	
• HnuSec网络安全技术挑战赛省一等奖（2019）	

📁 实习经历

商汤科技-商汤研究院	2023.04 – 2024.07
NLP见习研究员	北京
• 项目1：工业大模型跨团队研究中一致性检测工具的开发。	
– 独立设计和开发（从数据集到整体工具实现）一套工业大规模语言模型跨团队研究一致性检测工具，用于对工业研究中大模型跨团队交付进行一致性检测。	
– 目前该工具已经在商汤研究院内部所有大语言模型研发部门推广。	
– 目前成果也额外形成一篇论文，已经投稿NAACL2025。	
• 项目2：基于Sensechat的金融领域RAG能力的Embedding模型研发。	
– 使用金融领域对比学习训练数据，使用句子最后的EOS token的hidden embedding作为句子embedding，直接微调mistral-7b作为embedding模型。	
– 完成有监督对比学习训练数据的hard negative pairs的挖掘代码构建以及整个有监督数据的构造pipelines。	
– 配合GPT-4和自研的有监督数据构造pipeline，生成高质量金融领域有监督数据100W条。	
– 成果：在mteb leaderboard的中文榜单中刷到第一（2024/3月）	
• 为sensechat研发提供数据方面支持	
– 完成构建训练数据和benchmark的构建pipeline，生成300B tokens的高质量预训练语料和20B tokens的高质量sft标注语料，为团队模型训练和评测提供了坚实的保障。	
• 大模型端侧加强研究	
– 配合正式员工，利用progressive wanda中的剪枝方法，配合高质量的calibration sample，在保证特定任务（实际应用场景中的高频任务）能力基本不变的前提下将模型参数量降低75%。	
悉之智能	2021.12 – 2023.03
见习研究员	北京
• 重写simpy包，并基于改写后的simpy进行推理节点的开发	
• 负责基于G2T+BERT模型的数学应用题答案校验的相关工作，并将模型转为onnx，提高推理速度	
• 基于COT(chain-of-thought)，分别从原文的8个prompt模板以及自己设计的prompt模板对chatgpt进行prompt，使得chatgpt对数学问题生成合理的分步求解步骤。	

- 从Auto-COT(Automatic Chain-of-Thought Prompting)角度对chatgpt进行了和上一项工作类似的相关实验工作
- 利用chatgpt和GPT4生成代码，解决绘图问题，主要从zero-shot和few-shot两个方向出发

科研经历

SimCT: 大模型研发周期的一致性检测的简单框架

第一作者 北邮 & 商汤研究院

设计一套一致性测评框架，主要服务于工业界大规模语言模型研究中的跨团队交付难以保证一致无误的问题。目前该论文已经投NAACL2025。

reCSE:基于特征重塑的可移植无监督句子表征对比学习框架

第一作者 北京邮电大学

通过特征重塑的方式，加强对比学习句子表征框架的能力，展现出了强大的可移植性，在非离散化增强的对比学习框架中取得了Sota的水平。目前论文已经形成并已投NAACL2025。.

BSCSE: 基于软样本的对比学习无监督句子表征任务

第一作者 北京邮电大学

以SimCSE为基本框架，从prompt-tuning和prefix-tuning角度，同时引入软样本和新的损失增强训练过程，并在语义相似度任务上达到了Sota水平，目前已经形成论文预投ICASSP2025。

公开成果

F. Zhao, G. Jian, and D. Yan, "Recse: Portable reshaping features for sentence embedding in self-supervised contrastive learning," arXiv preprint arXiv:2408.04975 (NAACL 2025 under review), 2024.

[\[paper\]](#)

F. Zhao, G. Jin, R. Zhao, J. Huang, and F. Tan, "Simct: A simple consistency test protocol in llms development lifecycle," arXiv preprint arXiv:2407.17150 (NAACL 2025 under review), 2024.

[\[paper\]](#)

专业技能

- 技能: python, NLP, pytorch, git,
- 语言: 英语 (CET-6)
- 兴趣爱好: 篮球 (院队)