佘帅杰

shesj@smail.nju.edu.cn · (+86) 188-5116-0388 ·

教育背景

南京大学 2022 – 至今

在读直博 计算机科学与技术系，导师黄书剑, 预计 2027 年 6 月毕业

南京大学 2018 – 2022

学士 计算机科学与技术

科研经历

生成式摘要的事实一致性评估 一作，发表于 AAAI-23

本研究提出了一种利用 prompt 来控制模型偏好来检测生成式摘要中的事实不一致框架 (CoP)。该方法通过分离无关偏好，不需要训练就可以精确的检测出事实不一致。此外还可以衡量特定类型的偏好以检测具体不一致类型。我们还探索了结合 prompt tuning 来强化偏好，高效的从少量真实数据中学习。我们在三个不一致检测任务上取得了 SOTA 结果，证明了方法的有效性。该论文发表于 AAAI2023 (CCF-A)，并在大会上做 Oral presentation。论文地址：https://arxiv.org/abs/2212.01611

探究在对话摘要事实性中的大模型 一作，2023 年

本研究提出了一个新的数据集，用于评估主流大模型的对话摘要能力。分析结果表明现有的大模型的对话摘要一致性依旧较差。我们进一步的尝试用大模型评估这些错误，初步实验发现直接应用现有的大模型并不能解决事实性的评估问题。为了解决这个问题，我们提出了一种多任务伪数据构建策略，并对大模型进行微调，有效提升了模型的主客体理解能力，论文将于近期公开于 arxiv。

大模型数据处理与训练以及思维链 (CoT) 能力研究 算法实习生，2023 年 4 月

在第四范式实习期间，参加式说 (SageGPT) 研发。主要工作有两个方面, 一方面负责清洗和筛选 TB 级别无监督语料。另一个方面主要负责大模型的 CoT 能力以及 SageGPT 数理逻辑学科能力提升的学术研究，有使用 DeepSpeed，FSDP，PEFT 等相关工具训练多个 7B 和 13B 中文、英文模型的经验。

基于人称指代消解的生成式对话摘要改进方法研究 一作，2022 年

本研究提出了一种基于人称指代消解的生成式对话摘要改进框架 (WHORU)，通过向现有的模型注入额外的指代消解信息，帮助模型理解复杂的指代问题，提升生成质量。进一步的针对对话摘要提出了一个启发式的指代消解策略，计算速度快，达到 SpanBERT 的 11 倍推理速度。大量实验表明了 WHORU 在多个基座模型上获得了大幅增长，达到了新的 SOTA 水平。

项目经历

开源和学术社区贡献

维护多个仓库，积极在开源社区贡献自然语言处理相关的代码，github 主页：Ricardokevins。目前共计在开源社区 github 获得了 565 个 stars。公开的训练模型权重获得超过 2000 次下载。担任 EACL23, ACL23 的审稿人，参加了 NIPS22, ICLR23, IJCAI23 的审稿工作。

2022 阿里天池全球人工智能技术创新大赛

阿里天池全球人工智能技术创新大赛赛道三: 短文本语义匹配比赛。在团队中主要承担算法研发和实现，使用了多种预训练语言模型。采用 MLM 预训练，知识蒸馏，自蒸馏，对抗训练，多模多折融合等技术。最终排名 44/5345 (Top 1%)

其他

英语四级 615 分，六级 594 分

2022 美国大学生数学建模 M 奖，2022 全国大学生数学建模江苏省一等奖