程锦国

18835438251 1261430883@qq.com github.com/Chengyui

教育背景

至今 2023.09 **四川大学 (985)** • 计算机学院
计算机科学与技术 • 硕士

2023.09 2019.09 **四川大学 (985)** • 计算机学院
计算机科学与技术 • 学士

荣誉

- 2024 年英特尔奖学金
- 四川大学 2023-2024 学年优秀研究生干部
- 四川大学 2023-2024 学年优秀研究生
- 四川大学研究生一等奖学金
- 四川大学优秀学生

实习经历 (AI 业务 + 工程落地)

2025.4 2024.12 **百度智能云 (成都)-AI 能力引擎研发组-后端研发工程师**

- 项目背景: TextMind 智能文档分析平台以 AI 大模型（NLP、多模态）为模型算法底座，基于平台提供场景化预置模型，同时建设从数据标注到模型生产、以及模型效果进化端到端的能力，同时建设了以文档内容审查、文档结构化抽取、智慧招采等重点场景化应用，帮助客户提升海量文档处理效率，加固风险防控体系。
- 应用中心模块，参与**文档抽取**，**智慧招采**业务的 LLM 工作流具体设计和实现。
  - 基于 **Asynq** 分布式任务队列异步处理框架实现串行 workflow 业务解耦。对 Asynq 并发分配方式由一个协程串行打平为每个不同的 worker 协程**并发分配分离**，可灵活进行并发控制，响应时间减少 20% 以上，解决实例扩容时，资源利用率较低的情况。
  - 使用 OpenTelemetry 和 ThreadLocal，收集分布式追踪数据，利用 Jaeger UI 可视化界面定位链路性能瓶颈。
  - 为了解决系统吞吐量过小导致的任务超时问题，微调文心 Speed 实现**多切片分类**，分类 F1score 与单切片效果持平，任务处理速度提升了 3.3 倍。
  - 使用 Redis 分布式锁及 Lua 脚本，通过轮询管理 Redis 中的 LLM API 账号池，防止账号 QPS 超限，并通过实时更新账号 TPM 额度，进行动态限流，防止 TPM 超限。
  - 引入 LightRAG 技术，对用户上传的文档进行知识图谱构建、管理、检索增强推理，提升招标文件生成的效果。
- 数据中心模块，参与**智慧招采**业务数据飞轮，用户通过标注数据维护数据集，联动模型中心提升效果。
  - 用户复制或新增数据集版本时，发起后台异步任务复制数据集版本-异步任务信息-用户标注记录，实现版本间解耦。
  - 大数据集统计标注信息时，通过分页查询 Mysql 进行内存优化，并使用游标法，防止深分页问题。

2024.11 2024.8 **曦谋决策 (杭州) 智能科技有限责任公司-研发部-算法开发工程师**

- **国家电网光明电力大模型-分布式光伏时序预测大模型：** 1. 基于时序大模型Time-LLM架构，开发自然语言模型与时序模态对齐算法，结合高精度天气预报数据，构建以百度千帆大模型为核心的分布式光伏电力预测系统，完成客户电站功率建模及超长期预报。
- 2. 基于 BentoML 搭建时序大模型在线推理服务，针对用户数据特征进行个性化微调以提升预测精度，设计微调参数存储策略，并实现在华为昇腾 910A 计算卡上的推理部署。
- 3. 在河北地区实现户级（5 万户）48 小时及县区级（109 个县区）240 小时功率预测，准确率超过 93%，系统部署于中国电力科学研究院电力自动化所。

2024.08 2024.06 **清华大学启元实验室**

- 1. 基于 SAM 模型处理任务数据集，融合 Refcoco 通用场景数据生成面向特定任务的空间感知 SFT 数据集，利用 GPT4o 生产决策链路数据，加入 SFT 数据集。采用 DeepSpeed 加速训练 MiniCPM-V2.5 的空间视觉理解能力。
- 2. 基于 LangChain 实现多模态模型工具调用与任务规划，使具身智能机器人能根据自然语言指令输出目标物体 2D 坐标，完成物品抓取及状态驱动的动作指令生成。

主要开源项目 (通用时序基座模型)

- `UnifiedTSLib`：通用时间序列基座模型预训练及微调框架（暂未开源）
  - 与字节跳动合作的TimeMixer++时序开源预训练及微调框架。支持数据并行训练，模型保存为 huggingface 风格。
  - 第一个通道混合的时序预训练框架。对于不同数据集，在批次大小和通道数量上取得平衡，提升计算稳定性，降低 padding 填充导致的带宽浪费。
  - 在大规模数据训练任务时采用磁盘批量读取方式，规避内存读取导致大数据集内存溢出（300B token）。
- `NuwaTS`：基于大语言模型的缺失时序数据补全基础模型（62 stars）
  - 为论文 2 的开源的项目。NuwaTS 基于 LLaMA，GPT-2，BERT 预训练语言大模型，采用先进设计的语言-时序对齐设计特征空间嵌入技术，具备强大的 zero-shot 能力和先进的微调模块设计，服务于工业场景数据收集和建模时面临的数据丢失困境，并实现跨域预测。
  - 在公众号集智俱乐部分享最新大语言模型数据分析最新研究进展，收听人数达到 2300 人次 +，具有较大学术影响力,链接直达。
- `SUMformer`：城市时空流量预测方法（20 stars）
  - 为论文 1 的开源的项目。与成都交通管理局合作，项目中引入了线性注意力机制解决细粒度的城市交通流量空间管理建模造成的较大计算开销、高频特征过滤模块去除冗余噪声，提升峰值流量预测准确度。
  - 提出 SUMformer 模型进行时空建模，在北京，成都与纽约三个城市，五个数据集上的预测精度达到学界最高水平

科研成果 (时间序列 + 大模型)

- （论文 1：第一作者）Rethinking Urban Mobility Prediction: A Super-Multivariate Time Series Forecasting Approach `TITS(SCI 一区，CCF-B)`
- （论文 2：第一作者）NuwaTS: a Foundation Model Mending Every Incomplete Time Series `TNNLS 在投`
- （论文 3:第二作者)Evaluating Temporal Plasticity in Foundation Time Series Models for Incremental Fine-tuning(`IJCNN CCF-C`)
- （论文 4：第二作者）Integrating Future Exogenous Information into Multi-mode Travel Demand Forecasting at Gateway Hubs(`ICONIP CCF-C`)

技能

Go	Go 基础、协程、Gin、Gorm
其他语言	Python、Java（了解）、C++（了解）
数据库	MySQL 、Redis、ElasticSearch
消息中间件	Asynq
工具、部署	Jaeger、k8s、Docker、Nginx
AI 工具	Pytorch 、Ollama、LangChain
人工智能	多模态模型，时间序列-时空数据神经网络
语言	**英语** — CET6（516）