

路海阔 Ezra

·28岁·18811338441·ezraluuu@gmail.com·上海·

具备统计建模与系统优化双重背景，专注AI算法实现与大模型工程应用。

熟悉非线性模型、Agent框架、文档解析等场景，擅长用C++/Python构建高性能系统并推动实际落地。

教育经历

Boston University

Master of Arts in Statistics

College of Math and Statistics

2018.09 – 2020.01

- 核心课程: Statistical Learning, General Linear Regression, Bayesian Statistics, Computational Statistics, Stochastic Process

Boston, MA

GPA: 3.58/4.0

北京理工大学

统计学学士

数学与统计学院

2014.09 - 2018.06

- 核心课程: 数学分析, 高等代数, 实分析, 概率论, 数理统计, C语言等

北京

GPA 3.3/4.0

工作经历

博佳医药科技有限公司

中级算法研究员 | 上海 | 2023.08 –至今

非线性混合效应模型算法开发:

- 复现前沿论文算法 (论文 → Python Demo → C++), 结果与行业金标准软件一致
- 设计测试用例, 进行数据模拟与采集, 验证算法准确性
- 优化计算性能, 减少重复运算, 提升运行效率 (如通过BLAS/LAPACK替换Eigen, 加速70%+)

浙江来未来科技有限公司

数据科学家 | 数据智能部 | 2021.07 –2023.08

- 医疗AI项目研发:** 主导医疗科研合作项目的AI算法设计与实现, 涵盖数据采集、模型研发、算法部署等全流程, 服务医院临床科研与管理需求。
- 数据搜集及清理流水线搭建:** 根据医院临床科室需求, 综合整理不同平台患者数据, 根据多样数据格式搭建对应结构化数据流水线。
- 多模态数据处理:** 基于ResNet, 卷积网络等, 处理时间序列 (如心率波形)、医学图像等多模态医疗数据, 完成特征提取与融合。

项目经历

FOCE(First Order Condition Estimation)与SAEM 算法开发

- 主导从零实现FOCEI与SAEM两种主流非线性混合效应模型(NLME)参数估计算法, 涵盖Hessian

近似、梯度计算与优化器集成。

- 参考前沿文献对梯度求解流程进行数值与解析混合优化, 显著减少迭代耗时, 整体加速约30%。
- 构建了可复用的矩阵与三维数组结构, 采用内存池管理, 减少Eigen内部内存分配瓶颈, 在数万重复模拟任务中提升内存利用率与运行效率。
- 在多种模型结构与真实数据模拟下, 系统性对比输出与NONMEM、Monolix等商业软件的一致性, 验证算法精度与稳定性。
- 完善了模型比较流程、VPC预测检验图输出与Bootstrap置信区间估计, 构建完整的建模与评估工作流。

基于大模型的PDF文档结构化信息抽取与自动标注系统

- 项目背景:** 公司内部大量医疗数据以PDF格式存储, 涉及临床研究方案、统计报告等, 且因数据涉密, 需使用私有化部署的大语言模型。统计部门频繁需要从PDF文档中提取结构化信息, 自动化需求强烈。
- 系统设计与搭建:** 使用Dolphin及Deepseek R1构建从PDF到结构化数据的端到端流水线, 利用Dolphin工具完成版面解析与格式标准化, 输出结构化JSON。
- 大模型标注接口:** 将结构化JSON输入私有部署的大模型, 结合prompt编排与Agent架构实现实体识别、字段提取与标签生成, 自动完成数据标注。
- 实际成效:** 系统已在多个临床统计项目中落地, 显著减少人工抽取工作量, 提高数据提取速度与准确率。

个人总结

- 编程语言:** Python, C++, R, Bash, SQL
- 常用框架:** PyTorch, NumPy, Pandas, Polars, Scikit-learn, Eigen, SymPy, OpenBLAS/LAPACK
- 建模方法:** FOCEI, SAEM, Bootstrap, Nonlinear Mixed Effects Models, Bayesian Inference
- 发表:** PAGE 31 (2023) Abstr 10676 [www.page-meeting.org/?abstract=10676]
- 语言能力:** 中文(母语), 英文(TOEFL 110)