涂浩伟

电话: +86 16621555340 | 邮箱: haowei.tu@outlook.com

教育经历

纽约大学 2022年09月 - 2025年05月
信息系统(CS+MBA 方向) 硕士 纽约，美国
相关课程：基础算法、操作系统、数据库系统、编程语言、多核处理器：架构与编程、实时与大数据分析、DevOps与敏捷式开发、商业数据科学、金融基础、智能投顾与系统化交易

卡尔顿大学 2018年09月 - 2022年05月
计算机科学+统计辅修 本科 渥太华，加拿大

技能

- **技能：** Python (Pandas, NumPy, sklearn, Django, Flask), Go (Gin), JavaScript (React.js, Node.js, Express.js), Java, C/C++ (OpenMP), SQL, Scala, Dart/Flutter, Linux, DevOps, REST, AWS, PostgreSQL, MongoDB, GitHub, CI/CD, Jira, Hadoop, Hive
- **语言：** 普通话（母语），英语(无障碍)

工作经历

特斯拉 2024年07月 - 2024年09月
软件工程师实习生(Go, Angular, Gin, C#, .NET, SQL, Elasticsearch, Jenkins) 上海，中国
- 负责内部工厂微服务的前端和后端功能的调试、开发和部署，涉及多种技术栈，推动全自动驾驶（FSD）技术在中国的部署进程
- 主导关键数据跟踪功能的全栈开发，管理大规模数据处理并优化系统性能，全面负责软件开发生命周期（SDLC），从初始设计到部署后的支持
- 设计并开发后端数据迁移任务，将数据库中的数据同步到Elasticsearch，支持通过API调用自定义参数（时间范围、ID、数据表）以及通过Kubernetes部署的定期自动化任务
- 与产品经理和质量保证团队跨职能合作，确保项目顺利推进，提供经过深思熟虑的解决方案，考虑到开发过程中潜在的挑战和改进机会

NuEnergy.ai 2022年05月 - 2024年07月
初级软件开发 (全职至2022年9月，此后为兼职) 渥太华，加拿大
- 在敏捷开发环境中工作，与产品管理团队、数据团队及高级开发人员密切合作，为首要客户开发Python Django RESTful全栈组件，提高客户满意度25%
- 合作重新架构了一个从静态到动态的 Django 应用程序；设计了 50 多张表格的关系模式；开发并维护了 RESTful API，以实现强大的 SQL 数据库交互；集中数据存储和访问，以实现可扩展的未来开发
- 通过使用各种 CSS 框架在 TypeScript 中开发 React 组件，将用户界面原型转化为面向客户端的应用程序功能
- 使用 AWS Cognito 和 Django 管理工具管理客户和管理员的登录凭证并定制登录功能
- 使用 Jira 进行任务管理，并使用 Bitbucket 进行代码签入、审查、合并以及从主版本库和测试版本库划分分支

Forkaia 2021年10月 - 2021年12月
数据科学与分析实习生 欧文, 美国德州
- 使用numpy、pandas批量收集用户数据，将数据转换为CSV；执行自动化数据处理，包括标准化和清洗；设计结构良好的数据模式并实施到PostgreSQL数据库
- 开发基于用户与其他用户及先前喜欢产品的余弦相似度的协同过滤推荐系统，减少每例数据处理时间80%

Hirebeat 2021年05月 - 2021年07月
数据科学家实习生 纽约，美国
- 利用多维评级从标题、工作经历的持续时间和最新性、技能关键词开发简历推荐算法，简化HR筛选流程，提高运营效率30%
- 利用 Python 中的 NLP 构建自动简历解析器，将简历部分转化为Excel表格数据；根据工作经验部分的关键字，利用补全 Naive Bayes 算法对职位名称进行准确分类

项目经历

并行行程编码器 2023年02月 - 2023年03月
- 通过使用 pthreads 实现可靠的线程池，用 C 语言开发了高性能多线程运行长度编码程序；在 10GB 以上的大型数据集上，与单线程版本相比，运行时间最多缩短了 75%
- 利用任务队列的概念，设计了自定义任务数据结构，以确保任务工作量均衡和未来合并过程高效；采用动态调度，允许同时进行任务编码和合并，减少了处理延迟
- 利用互斥和条件变量来防止共享数据结构上的所有竞争条件、繁忙等待和死锁；使用细粒度和读写锁来最大限度地减少互斥锁的持有时间，减少线程之间的竞争

多线程数独求解器 2023年01月 - 2023年05月
- 协作设计并开发了一个高效的多线程数独求解算法，使用C++和OpenMP，有效地将工作负载分布在多个核心上，将执行时间与单线程实现相比减少了高达90%
- 领导了3轮优化迭代，分析性能瓶颈并进行改进以获得更好的加速和可扩展性，将求解器的可扩展性从9×9数独扩展到更大的25×25矩阵，同时保持可接受的求解时间
- 记录了开发过程，包括设计原则、实施细节、性能基准和安装说明，以便于理解和故障排除

酒店房间最佳超售策略 2023年01月 - 2023年05月
- 进行分析以确定超售房间数量，从而实现最大利润，并最终呈现利润曲线的可视化
- 使用相关性分析和热图来识别和可视化影响目标变量——预订行为的关键变量；通过去除不太相关的变量来优化数据集，以提高模型性能
- 训练了多个预测模型，包括随机森林和逻辑回归；通过GridSearch超参数调优确定最佳模型；绘制ROC曲线以利用真阳性率和假阳性率确定最佳决策阈值
- 按客户的未到概率进行排名，通过整合关键酒店指标和超售罚款构建利润曲线，展示趋势，并确定房间超售的最佳百分比以最大化收入

纽约FHV数据分析 2023年09月 - 2023年12月
- 对7年的纽约FHV（Uber/Lyft）数据进行了详细的大数据分析，揭示出行模式并为城市出行提供洞察
- 在Hadoop上进行数据清洗和分析，解决了缺失和冲突的记录问题，应用过滤器排除不合规的数据，标记数据属性并将结果数据集重组为优化的Parquet格式
- 利用Hive按特定的上下车坐标和时间维度（月/日/小时）提取和组织数据，随后将优化后的数据格式化为TXT文件进行进一步分析
- 在Tableau中生成动态热图，以可视化出行的空间和时间分布，突出高需求区域和高峰时间；与团队成员的发现合作，为城市交通动态提供全面视图

Haowei Tu

Seek for SWE NG Opportunity | haowei.tu@outlook.com | +86 166-2155-5340

PROFESSIONAL SUMMARY

Fast-learner with backgrounds in Computer Science, Data Science, Statistics and Finance; currently enrolled in a CS and MBA combined master's program; demonstrated work expertise in Python, SQL, Excel, Data Analytics and various Data Science techniques; adept at promoting teamwork by consistently communicating with clarity and empathy

EDUCATION

New York University **New York, NY**

Master of Information Systems (MBA/CS Focus) 2022.09 – 2025.05

- **GPA:** 3.7/4.0, **Relevant Coursework:** Foundation of Finance, Robo-Advisor & Systematic Trading, Data Science for Business, Fundamental Algorithms, Operating Systems, Database Systems, Real-time and Big Data Analytics

Carleton University **Ottawa, Canada**

Bachelor of Computer Science, Minor in Statistics 2018.09 – 2022.05

- **Major GPA:** 3.81/4.00; Dean's List (2020, 2021); Entrance/Harry S. Southam/Lester Bowles Pearson Scholarship (Top 10%)

SKILLS & TECHNICAL COMPETENCES

- Programming: Python(numpy, pandas, sklearn, Django), JavaScript(Angular, React.js, Node.js), Go(Gin), Java, C/C++, SQL, Scala, Dart
- Techs and Tools: Linux, DevOps, REST, AWS, PostgreSQL, MongoDB, Github, CI/CD, Jira, Jenkins, Hadoop, Hive, Flutter

EMPLOYMENT EXPERIENCE

Tesla **Shanghai, China**

Software Engineer Intern (Go, Angular, Gin, C#, .NET, SQL, Elasticsearch, Jenkins) 2024.07 – 2024.09

- Led the development and deployment of both front-end and back-end features for internal factory platforms, facilitating the deployment process of Full Self-Driving (FSD) technology in China
- Spearheaded the development of a critical data tracking feature, managing large-scale data handling and optimizing system performance while taking full ownership of the SDLC from design through to post-deployment support
- Designed and developed a backend data migration job to sync data from databases to Elasticsearch, supporting both API calls with customizable parameters (time range, ID, table) and automated periodic jobs deployed on Kubernetes
- Collaborated cross-functionally with Product Managers and QA to ensure smooth project progression, delivering well-considered solutions that accounted for potential challenges and opportunities for improvement throughout the development process

NuEnergy.ai(NuEnergy.ai) **Ottawa, Canada**

Junior Software Developer (Full-Time until Sept 2022, Part-Time thereafter) 2022.05 – 2024.07

- Developed **Python Django** Full-Stack components in **Agile**, collabåorating with product and data teams for top-tier clients
- Re-architected a Django app to dynamic design; created relational schemas with 50+ tables and built scalable **RESTful APIs** for SQL database interaction; centralized data storage and access for scalable future development
- Converted UI prototypes into client-facing features by building **React** components in **TypeScript** with modern CSS frameworks
- Managed client's and admin's login credentials by customizing login features using **AWS** Cognito and Django Admin Tool
- Used **Jira** for task management and Bitbucket for code check-in, review, merging and branching from main and testing repositories

Forkaia(forkaia.com) **Irving, TX**

Data Science & Analytics Intern (Data Cleaning, SQL, Machine Learning Research, Recommendation Systems) 2021.10 – 2021.12

- Batch-collected user data using numpy, pandas by transforming the data into CSV; performed automated data processing including standardization and cleaning; designed well-structured data schemas and implemented into a PostgreSQL database
- Developed a collaborative filtering recommendation system to personalize a top list for users based on their cosine similarity with other users and similarity of previous preferred products, reduced 80% data processing time per case

Hirebeat(hirebeat.co) **New York, NY**

Data Scientist Intern (NLP) 2021.05 – 2021.07

- Developed resume recommendation algorithm by leveraging multi-dimensional ratings from titles, duration, recency of work experience, and skills keywords; simplified HR screening processes and improved operational efficiency by 30%
- Utilized NLP in Python to construct an automated resume parser that transforms resume sections into tabular data; leveraged the Complement Naive Bayes Algorithm for accurate classification of job titles, based on keywords from the work experience section

PROGRAMMING PROJECTS

Optimal Hotel Room Overselling Strategy (Predictive Analytics, Classification Modeling, Pandas, NumPy, sklearn) 2023.01 – 2023.05

- Conducted an analysis to determine the number of oversold rooms leading to peak profit, culminating in a profit curve visualization
- Employed correlation analysis and heatmaps to identify and visualize key variables influencing the target variable — booking behaviors; refined the dataset by removing less pertinent variables to enhance model performance
- Trained multiple predictive models including Random Forest and Logistic Regression; determined the best model by GridSearch hyperparameter tuning; plotted the ROC curve to identify the best decision threshold leveraging true positive and false positive rates

- Ranked customers by their no-show probability and constructed a profit curve by integrating key hotel metrics and oversell penalties, visualizing the trend, and determining the optimal percentage for room overselling to maximize revenue

New York FHV Data Analysis (Big Data Analytics, Data Visualization, Hadoop, Hive, Tableau) report 2023.09 – 2023.12

- Conducted a detailed **big data** analysis project on 7 years of New York FHV (Uber/Lyft) data, uncovering ride patterns and contributing to urban mobility insights
- Performed **data cleaning** and **profiling** on **Hadoop** to resolve issues with missing and conflicting entries, applied filters to exclude non-compliant data, marked data attributes and restructured the resulting dataset into an optimized Parquet format
- Utilized **Hive** to extract and organize data by specific pickup/dropoff coordinates and time dimensions (month/day/hour), subsequently formatting the refined data into TXT files for further analysis
- Generated **dynamic heatmaps** in **Tableau** to visualize spatial and temporal distribution of rides to highlight high-demand areas and peak times; collaborated with team member's findings to provide a comprehensive view of urban transportation dynamics

Multi-Threaded Sudoku Solver (C++, Parallel Programming, Performance Analysis, OpenMP) Report 2023.01 – 2023.05

- Collaboratively designed and developed a highly efficient multithreaded Sudoku solver algorithm using C++ and OpenMP, effectively distributed workload across multiple cores, reduced execution time by up to 90% compared to single-threaded implementations
- Led 5 rounds of optimization iteration of analyzing the performance bottleneck and then refining for better speedup and scalability, expanded the solvers' scalability from 9*9 Sudoku to significantly larger 25*25 matrices while maintaining acceptable solving time
- Documented the development process, including design principles, implementation details, performance benchmarks, and installation instructions of the multithreaded Sudoku solver, facilitating comprehension and aiding in troubleshooting

Parallel Run-Length Encoder (C, Multi-threading, Interprocess Communication, Thread Pool) Code 2023.02 – 2023.03

- Developed a high-performance multi-threaded run-length encoding program in C by implementing a robust thread pool using pthreads; achieved up to 75% reduction in running time compared to the single-threaded version on large data sets over 10GB
- Leveraged the concept of task queues, designed custom task data structures to ensure even task workload and efficient future merging process; Employed dynamic scheduling to allow simultaneous task encoding and merging, reduced processing latency
- Utilized mutexes and condition variables to prevent all race conditions, busy waiting, and deadlocks on shared data structures; Used fine-grained and read-write locks to minimize the time spent holding mutex locks, reduced contention among threads