

宋思远

<https://github.com/siyuan-song> | 867086542@qq.com | 手机/微信: 13651601370 | 出生年月: 1992.09



自我评价

- 六年数据科学与四年机器学习、深度学习、神经网络开发经验，涉及自然语言处理（NLP），计算机视觉（CV）等领域，一年半的项目管理经验。在环保，金融，互联网和智能汽车等多个行业积累了丰富的经验，参与并成功交付了多个大型项目。
- 专业负责，独立完成项目。在美邦独立搭建和部署针对用户售后评论的自然语言分析模型。
- 勤于思考，提升工作效率。在华为担任软件开发工程师期间主导模型转代码的流水线的优化，及开发自测试平台的搭建。
- 结果导向，善于解决问题。在华为担任项目经理期间对外沟通客户诉求，对内协调各领域资源，保障酷户需求的高效传递和顺利交付。
- 英语可作为工作语言。有统计学与应用数学等领域专业留学背景。

相关技能

- 编程语言及可视化工具：Python3, MS SQL Server, Tableau, MatLab
- 框架和库：NumPy, Pandas, Matplotlib, Seaborn, OpenCV, Sklearn, NLTK, TensorFlow, Keras, SciPy, Plotly, XGBoost 等
- 机器/深度学习模型：线性回归，逻辑回归，K 近邻，支持向量机，自然语言处理，K 均值聚类，主成分分析，朴素贝叶斯，梯度提升决策树，随机森林，推荐系统，人工神经网络，循环神经网络，卷积神经网络等

工作经历

华为技术有限公司（委派自北京外企德科有限公司） 智能汽车解决方案部 2022 年 8 月 – 至今 上海市

软件开发工程师（车控运动域） 2022 年 8 月 – 2023 年 10 月

工作内容：负责汽车轮胎六分力模型算法的预研、模型转代码流水线脚本的开发和维护、模型自测试 MIL 平台的开发和优化。

项目 1: 汽车轮胎六分力模型算法预研

项目描述：作为和汽车运动状态联系最为紧密的轮胎状态表征之一，轮胎六分力的精准估计对于整车动力控制系统的算法起着关键的作用。

在传统的轮胎六分力估算基础上，结合机器学习算法，预研更高精度和鲁棒性的轮胎六分力估算模型。

- 个人贡献：
1. 负责试验车上千公里路试数据的整理、清洗和预处理。
 2. 开发基于传统汽车动力学和机器学习/神经网络相结合的轮胎六分力估计模型。
 3. 拟合回归不同工况下基于魔术公式的公式系数分布。
 4. 预研极限状态外的高置信度模型外推方案。
 5. 路试回归验证模型，持续迭代，提高模型精度和鲁棒性。

项目 2: 模型转代码流水线脚本的开发和维护

项目描述：在已有版本构建流水线的基础上，优化模型转代码脚本逻辑，提升版本构建速度和效率，实现本地测试版本的可追溯性。

- 个人贡献：
1. 负责落地车控部门基于标定代码和逻辑代码分离的共主干方案模型仓/代码仓 UCI 工程。
 2. 协助开发增量模型转代码的 UCI 工程。
 3. 编写、适配、测试、部署、维护工程脚本，提升 70%模型转代码效率，提升 30%端到端版本构建速度。
 4. 负责本地测试版本迭代管理，做到本地版本 100%历史可追溯。

项目 3: 模型自测试 MIL 平台的开发和优化

项目描述：搭建和优化 MIL 自测试平台，为应用层开发提供平台化的模型自测试方案，提高开发的自验证效率。

- 个人贡献：
1. 负责 MIL 自测试管道流程雏形的设计和搭建。
 2. 测试管道流程的运行能力，优化管道流程的运行速度，提高管道流程运行的泛用性，维护管道流程的运行稳定性。
 3. 部门内串讲推广试用版 MIL 自测试管道使用方法，收集管道流程优化意见。

项目交付经理（车控热力域）

2023 年 10 月 - 至今

工作内容：负责某主机厂客户项目群的热管理产品的客户面交付管理统筹工作。作为客户面需求输入及软硬件产品交付的第一责任人，对外统筹管理需求接纳流程及交付节奏安排，合理管控客户面预期；对内协调包括软硬件系统、设计、开发、标定、测试、采购、制造、供应、质量、维护、服务、商务、法务等在内的各领域资源组的协同工作，保障项目按时高质量的交付。

项目 1: 某主机厂已量产项目的改款子项目及出海子项目

项目描述：在某主机厂已成熟的量产车型基础上，交付改款子项目及出海子项目的热管理产品。

- 个人贡献：
1. 对接客户侧针对产品硬件的设计变更需求，拉通硬件研发及下层供应商评估及锁定设计变更方案，推进变更流程。
 2. 对接客户侧针对产品软件的功能变更需求，拉通软件设计、开发、测试等资源评估变更方案，迭代软件版本。
 3. 根据项目开发节奏和对应质量闸点，协调内部资源，准时高质量交付软件版本和硬件样机。
 4. 作为热管理产品部门首个出海项目，针对全新场景，组织法务、商务等相关部门，梳理产品交付新风险，端到端打通产品交付的新模式。为后续同类场景项目的落地提供了丰富的实战经验。所在团队收获当年度部长激励奖项。

项目 2: 某主机厂在研新项目

项目描述：新一代热管理产品平台化项目下，首个立项交付的客户化项目。

- 个人贡献：
1. 对接客户侧针对产品硬件的设计变更需求，拉通硬件研发及下层供应商评估及锁定设计变更方案，推进变更流程。
 2. 对接客户侧针对产品软件的功能变更需求，拉通软件设计、开发、测试等资源评估变更方案，迭代软件版本。
 3. 根据项目开发节奏和对应质量阀点，协调内部资源，准时高质量交付软件版本和硬件样机。
 4. 作为新一代热管理产品平台化项目下的首个立项交付的客户化项目交付负责人，提前识别项目交付风险，积极协调各领域协同攻关解决交付难点和堵点，全力推进项目进展，为后续各客户化项目起到积极的表率作用。

上海美特斯邦威服饰有限公司 战略大数据分析部

2020 年 12 月 - 2022 年 7 月 上海市

数据分析师

2020 年 12 月 - 2022 年 7 月

工作内容：负责对公司及竞店结构化与非结构化产品数据的抓取、处理、分析与报告。独立负责机器学习相关独立项目，包括自产品评论数据的深度学习分析模型的搭建、产品畅滞销监测模型构建等。独立负责自动化网页信息管理工具的模块适配和应用，极大提升业务部门数据获取效率。主导公司中台评论分析模块搭建，支持面向各业务部门及领导层的评论数据分析可视化展现。

项目 1: 自动化网页信息管理工具的模块适配和应用

项目描述：利用自动化网页信息管理工具（RPA），抓取各相关网页结构化与非结构化产品数据。

- 个人贡献：
1. 独立负责完成公司平台化 RPA 工具的整体流程设计与管道搭建。
 2. 针对每个业务部门的差异化需求，在平台化 RPA 工具的基础上，嵌入不同 Python 模块，支持各部门日常工作。
 3. 整体上提升 300% 业务部门数据采集效率，大大降低人工使用成本。

项目 2: 产品评论数据的深度学习分析模型的搭建

项目描述：基于天猫、淘宝等网络销售平台产品售后评论，搭建深度学习分析模型。

- 个人贡献：
1. 负责组织对线上产品售后评论等非结构化数据做分类、标注等预处理工作。
 2. 独立负责完成利用 NLP（自然语言处理）BERT 模型对预处理数据做下游训练及调参的工作并搭建售后评论分析模型。
 3. 模型用于完成日常分析评论、拆解评论主题、判断评论情感度、整理评论热词等售后评论相关分析任务。
 4. 编写 Python 程序，定时自动化输出评论周报。

项目 3: 公司中台评论分析板块的搭建

项目描述：搭建公司中台评论分析板块，汇总公司及竞店评论分析数据，支持面向各业务部门及领导层的可视化展现。

- 个人贡献：
1. 撰写 SQL 代码，将售后评论分析模型输出数据上传至数据库。
 2. 建立数据库格式表格，设计中台前端展示页面，配合 IT 部门落地中台评论分析板块。

项目 4: 产品畅滞销监测模型的构建

项目描述：通过产品销售数据，构建产品畅滞销监测模型，为产品管理决策提供支持。

- 个人贡献：
1. 通过商品流量数据、销售数据及市场大盘数据等，主导负责构建以 K 均值聚类算法为基础的畅滞销监测模型。
 2. 编写 Python 程序，定时自动化输出产品畅滞销监测周报。

Global AI

美国纽约市

数据科学家 实习

2019 年 5 月 - 2020 年 10 月

- 调用联合国 SDG（可持续发展目标）API 爬取与 SDG 相关的每日新闻情感值，并通过 Quandl API 获取超 500 家公司每日股价数据。
- 利用 Python 对新闻情感值和每日股价进行探索性数据分析，包含描述性统计数据分析 and 可视化分析。
- 利用 Python 对数据进行预处理操作，包括去重，缺省值处理，独热编码，标准化，归一化以及特征工程。
- 调用多种机器学习算法构建并拟合模型，包括使用网格搜索调整超参数，K 折交叉验证监测模型鲁棒性以及模型融合以整合预测模型。
- 利用 Tableau 创建数据仪表盘和报告，为公司建立与 SDG 相关的基准投资组合决策提供数据支持。

Ever Cleaning Environmental, Inc.

美国费城

数据分析师 实习

2018 年 12 月 - 2019 年 4 月

- 将数据集图像分成训练集与验证集。对图像做做注释和标签，并将其保存为 Python 字典形式的 json 文件。
- 应用 Mask R-CNN 卷积神经网络框架对目标容器实现对象实例分割及分类（如可乐瓶等不同类型的容器）。
- 基于 COCO 数据集调用预训练模型权重对模型进行初始化与下游训练，监测模型损失值以防止过拟合及欠拟合，经过 3000 次网络结构参数更新迭代后得到最终模型。
- 编写 Python 代码对模型进行输出测试并做性能评估来检验模型的有效性。评估模型的性能指标（平均精度均值 mAP）达到 0.87。

教育经历

南卡罗莱纳大学 统计学 硕士

2016 年 8 月 - 2020 年 5 月 美国

纽约州立大学布法罗学院 应用数学 硕士

2015 年 1 月 - 2016 年 5 月 美国

复旦大学 物理学 本科

2010 年 9 月 - 2014 年 7 月 上海市