



李威岳 (Larry)

+1 (970) 290-4669 | wel019@ucsd.edu
WeChat: GSWSCurry | https://weiyueli7.github.io/



教育经历

加州大学圣地亚哥分校 - 数据科学、应用数学、经济 三专业 本科

2020年06月 - 2024年03月

海外QS前100

GPA: 3.99/4.0.

相关课程: 机器学习、深度学习、数据挖掘、推荐系统、探索性数据分析和推断、数据可视化、数据库、算法、统计方法、线性代数、概率论、实分析、计量经济学、风险与决策、微观/宏观经济学等。

工作与实习经历

HDSI & Intel Corporation - Student Researcher DCA

2023年01月 - 至今

- 在两位教授的带领下, 与两位成员在英特尔的DCA数据库进行了探索性数据分析。应用曼-惠特尼U检验研究了用户的设备、充电方式和屏幕大小等因素对充电时长的潜在显著差异。通过对DCA数据库的深入研究, 成功估算了能量消耗值, 并运用线性模型, 根据各个网站类别中的时间花费数据, 建立了高度精准的能量损耗预测模型。该模型在自定义的损失函数下, 达到了70%以上的预测准确率, 为英特尔的能耗管理研究提供了有力的支持。同时, 通过对每个网站类别的线性模型系数进行深入分析, 为英特尔研发团队提供了优化芯片和硬件设计的重要方向。
- 个人还负责协助英特尔数据部门对DCA数据库进行了严格的数据验证工作, 并运用马尔可夫链等概率模型进行数据建模, 成功侦测到了5处存在于DCA数据库的标记错误, 保障了数据的质量和准确性。

Qualcomm Institute - Data Science Intern

2022年09月 - 至今

- 在一个五人团队中, 与哈佛、MIT以及UCSD的医生合作, 从UC San Diego Health的内部数据库中收集到COVID-19患者数据和相关信息, 并通过数据清理、聚合和验证的方式确保数据的准确性和可靠性。随后, 利用AWS构建了高效的机器学习管道, 进行数据分析和统计建模, 并开发了可视化交互式仪表盘, 向医生呈现具有统计学意义的结果。这一工作帮助医生基于数据做出决策, 推动了数据驱动的医疗决策, 并为COVID-19特效药的药效监测提供了有力的支持。
- 在一个两人团队中, 利用ECG数据进行心血管疾病预测。通过将ECG信号转换为频谱图图像, 配合由LSTM和AlexNet+SVM搭建的组合型深度学习模型与传统机器学习模型如SVM、KNN、MLP和Random Forest进行对比。研究结果表明, AlexNet+SVM的组合型模型在保持较高的F1 score水平下, 能够达到96%的预测准确率, 显著超过传统机器学习模型的预测准确率。这一研究为心血管疾病预测提供了高度精准的解决方案。

Institute for Neural Computation - Data Science Research Assistant

2022年08月 - 至今

- 在Terry Sejnowski教授的实验室与他的博士生一起利用功能性磁共振成像 (fMRI) 和结构性磁共振成像 (sMRI) 技术进行青少年脑认知发展 (ABCD) 的研究。
- 独自负责通过查询和网络爬虫收集来自351个外部数据库的数据, 并创建了高效的数据查找表, 以确保数据的完整性和准确性。
- 与两名团队成员一同合作, 开发了一种结合了3D-CNN、sMRI和fMRI多模态数据进行抑郁症检测的深度学习管道, 并通过能量模型对数据的可靠性进行检测, 从而解决了不平衡数据集可能导致的偏见问题。同时, 作为对比, 还建立了基于Spike Neural Network的fMRI和3D-CNN的sMRI模型。通过对比实验结果, 证明了深度学习管道在抑郁症检测方面的高准确性和可靠性。

相关项目

深度学习图像分类与配文 - 项目负责人

2022年01月 - 2022年04月

- 从零开始使用PyTorch搭建了LSTM和其他自定义RNN模型, 对COCO数据集的30K+张图像执行了自动图像注释 (AIA) 任务。通过调试, 确定了最佳模型为LSTM, 在AIA任务中取得了显著的成果, 测试损失为1.4, BLEU-1评分为68.3, BLEU-4评分为8.9, 表明LSTM在AIA任务中能够生成具有逻辑性的配文。
- 同时, 实施了自定义和预训练的CNN模型, 并采用带有动量和正则化的多层神经网络, 对食物图像和平假名笔迹进行图像分类任务。经过优化, 这些模型能够高精度地 (>80%正确率) 对10-20类图像进行分类。

用 COVID-19 统计数据反映各国幸福指数 - 项目组员

2021年10月 - 2021年12月

- 研究了 Covid-19 疫苗接种率、死亡率以及疫苗接种率较低的国家的幸福指数之间的关系。收集了有关 Covid-19 疫苗接种记录、按国家/地区划分的幸福指数和人口报告的数据集, 并进行了数据清洗和探索性数据分析, 以发现有意义的模式。
- 利用统计模型 (如OLS和潜在变量模型) 以及机器学习方法 (如PCA、SVR和RandomForestRegressor), 进一步分析问题。
- 确定了较早获得 Covid-19 疫苗的国家与较高幸福指数之间存在正相关关系, 并将研究结果报告给大学防疫部门。

犯罪指数和社区财富分析 - 项目负责人

2020年09月 - 2020年12月

- 使用ArcGIS从圣地亚哥周边社区收集和清洗数据, 进行探索性数据分析, 研究社区的平均财富水平及其相应的犯罪指数。
- 创建数据可视化并执行A/B测试, 以更好地解释变量关系, 构建用于预测分析的机器学习管道, 并应用地理空间分析技术来支持具有统计学意义的发现, 例如发现平均家庭收入较高的地区通常更容易出现犯罪现象。

专业技能

编程技能: 熟练掌握Python编程, 并熟悉常用库如Pandas、Numpy、sklearn、seaborn、matplotlib、plotly、SciPy、statsmodels、nlTK、bs4等, 能够进行高级数据分析和可视化。精通PyTorch深度学习框架, 具备Java、R (ggplot2)、SQL、HTML、JavaScript、Spark、Shell等编程语言的扎实基础。熟练运用Tableau等数据可视化工具。

技术技能: 熟悉AWS云计算平台, 具有Databricks和BigQuery等大数据处理经验。熟练操作Microsoft Office (Excel), 能够进行复杂数据处理和分析。

软技能: 注重细节, 具备出色的公开演讲和学术写作能力, 能够有效沟通和解决问题。擅长多任务处理和DEI (多元化、公平性和包容性) 原则的应用, 能够在跨团队合作中展现卓越的协调和领导能力。

认证: 获得谷歌数据分析专业证书。