



曾展航

期望职位: 数据挖掘 / 机器学习 / 计算机视觉相关岗位

住址: 英国, 苏格兰, 爱丁堡, 5 Buccleuch Place (目前住址) / 广东省珠海市

☎ (+44) 07422544002 (英国) / (+86) 13168698208 | 📧 zengzhanhang | ✉ zengzhanhang@outlook.com

🏠 zengzhanhang.github.io/

Education 教育经历

爱丁堡大学 (University of Edinburgh)

2018年9月 - Exp. 2019年11月

人工智能 (M.Sc. in Artificial Intelligence 硕士在读) | 预计学位等级: Merit

英国, 爱丁堡

- 课程: 深度学习 (CNN/RNN/GAN 及优化器 (Adam/RMSProp) 等原理, 卷积/池化的 Numpy 实现), 机器学习与模式识别 (逻辑回归/贝叶斯回归/决策树/集成学习等), 数据挖掘 (EDA/预处理/降维 (PCA/kernelPCA/MDS 等) /建模和评估方法), 自然语言处理, 数据可视化

北京理工大学珠海学院

2014年9月 - 2018年6月

应用统计学 (B.Sc. 理学学士) | GPA: 3.93/5 (专业前 3%) | 奖项: 优秀毕业生、特等奖学金 (系第 1)

广东, 珠海

- 奖项: 美国大学生数学建模竞赛国际二等奖; 优秀毕业生; 特等奖学金 (系第 1); 4 次一等奖学金 (系前 5); 优秀毕业论文
- 课程: 应用回归分析、时间序列分析、多元统计分析、计量经济学、统计预测与决策、微/宏观经济学、利息理论、运筹学

Skills 技能证书

语言及工具 熟悉 Python (Keras, TensorFlow, scikit-learn, xgboost), R 语言 (可视化: ggplot2, plotly), SQL; 了解 Web: html/ php

软件及其他 熟悉 Git 及 GitHub 进行团队协作开发, 熟悉 Linux 下的开发; 软件: Tableau, Spss, LaTeX, Markdown 及 Office 办公套件

语言 英语: 英语四/六级 (能熟练阅读英文文献和技术文档); 中文 (普通话/粤语)

Experience 项目经验

Kaggle 竞赛: APTOS 2019 糖尿病型视网膜病变检测 (Private LB: Top15%)

2019年8月-2019年9月

- 图像分类是计算机视觉的一个基本任务之一, 该任务旨在建立模型以实现糖尿病型视网膜病变的图像分类检测。
- 对初始视网膜图像进行预处理 (数据增强) 以突出图像中的特征, 项目基于 Keras 实现 ResNet101, DenseNet121, EfficientNet-B5 三种网络, 对于 5 分类的序列分类 (Ordinal Classification) 任务, 采用了回归的策略进行训练和预测。
- 测试时采用了 TTA 的策略进行预测, 最后运用 blending 的方式对模型结果进行融合, 寻求在 Kappa 评价指标更好模型组合。

临床脑科学中心 (CCBS, UoE) 毕业项目: 阿尔兹海默症风险预测的可解释机器学习模型

2019年2月-2019年8月

- 在爱丁堡大学 Centre for Clinical Brain Science 老师指导下的毕业项目。项目基于 SHARE (Survey of Health, Ageing and Retirement in Europe) 数据旨在构建可解释的机器学习模型用于探索风险因子和阿尔兹海默症的关系, 提供风险概率的解释。
- 建模前将数据离散化, 使用关联规则 Apriori 分析特征与阿尔兹海默症的相关关系进行特征选择。同时分析数据特点, 利用过采样处理数据不平衡。分类器采用随机森林和 XgBoost, 最终 XgBoost 达到 0.88 的 geometric 准确度, 其中达到了 0.99 的特异度 (specificity) 和 0.80 的敏感度 (sensitivity), 探索了不同的方法 (如 Permutation Importance/ PDP/ Apriori/ SHAP Value) 以分析和解释模型结果。

深度学习项目: 基于生成对抗网络 (GAN) 的超分辨率的图像重建研究

2019年1月-2019年4月

- 项目旨在实现对低分辨率图像的分辨率重建。采用了生成对抗网络 (GAN), 复现了 SRGAN 模型, 在 Google Cloud 上进行实验探索。
- SRGAN 的判别器 (Discriminator) 采用 VGG 的思想构建, 生成器 (Generator) 采用了 ResNet 的结构堆叠残差块, 最后采用 Pixel Shuffle 卷积进行上采样生成图像。项目探索了 Batch Norm, Weight Norm, Instance Norm 等 normalization 方法的影响, 改用了相对平均 GAN 损失作为判别器的损失函数, 最终结果采用 PSNR 和 SSIM 指标进行评估。

文本挖掘项目: 基于 The 4 Universities Dataset 的文本数据挖掘

2018年11月-2018年12月

- 项目旨在用 5 所大学的网页数据建模以识别网页的所属部门 (7 分类), 数据包含超过 8000 个 HTML 页面储存在 7x5 个文件夹中。
- 进行数据处理等预处理, 使用 Python 中的 HTMLParser 删除无关信息 (如 HTML 标签), 结合 TF-IDF 对文本提取特征, 利用 EDA 分析数据特点进一步提取手工特征。尝试了 PCA 和 Autoencoder (Keras 框架) 等降维方法, 项目使用 scikit-learn 探索了多种机器学习模型: 逻辑回归, SVM, KNN, 决策树等, 其中最优模型 LR 在准确率上达到 80.78%, 在 F1-score 上达到 80.54%。

数据挖掘项目: 慕课用户翘课行为预测

2018年1月-2018年3月

- 项目旨在利用网页中的 log 数据预测潜在的翘课用户。原始数据为 '学堂在线' 提供的 KDD 2015 数据集, 记录了时间和操作信息, 采用特征工程进行特征提取解决特征不足, 基于 R 语言的 h2o 包构建 Stacking 集成模型, 集成 13 个不同的模型进行预测, 如随机森林, 逻辑回归。

Honors & Awards 奖项荣誉

2018 年 北京理工大学珠海学院优秀毕业生

2018 年 北京理工大学珠海学院优秀毕业论文

2018 年 北京理工大学珠海学院特等 (全系第 1) / 一等 (4 次; 全系前 5) / 二等奖学金 (2015 年-2018 年)

共六次

2016 年 美国大学生数学建模竞赛国际二等奖 (Honorable Mention)

2017 年 全国大学生统计建模竞赛优胜奖