

乔永超

邮箱: yongchaoqiao9@163.com Github: <https://github.com/YongchaoQiao>

教育背景:

乔治华盛顿大学, 统计学硕士 (绩点: 3.84/4.00)

华盛顿特区, 美国; 2018.08-2020.05

中国地质大学 (武汉), 统计学学士

湖北武汉, 中国; 2014.09-2018.06

出版物及项目经历:

Co-Author, "An experimental investigation on the influence of coal brittleness on dust generation," *Powder Technology*, 364: 457-466. 2020.

机器学习: 细胞显微图片单一标签及多标签分类

2020.01-2020.05

- 采用 python 依托 GCP 对从美国疾控中心获取的 8600 张细胞图片及所属四类细胞文本信息进行配对, 获得四类观测细胞比为 53: 20: 6: 1, 对图片进行旋转, 翻转以扩充训练集样本, 保证各类别观测数量基本平衡; 构建基于 keras 的 MLP 模型, 采用 Macro F1 和 Kappa 值的算数平均数为模型评价指标, 最终获得班级最高测试集得分 0.8042。
- 采用 python 依托 AWS 对从美国疾控中心获取的 2 万张单张包含多类细胞的图片进行旋转, 翻转等进行数据增强, 构建基于 pytorch 的 CNN 多标签分类器, 采用 Binary Cross-Entropy loss 为模型评价指标, 获得班级最优测试集得分 0.2048。
- 采用 python 依托 AWS 对从 Kaggle 获取的 8 千张分别属于四种健康状况的树叶照片进行旋转, 翻转等进行数据增强, 构建基于 pytorch 的包含分支的 CNN 分类器, 最终模型 AUC 得分为 0.95。

自然语言处理: 文本数据分析

2019.09-2019.11

- 利用 python 为公司的社交平台设计评论过滤器, 通过对评论进行词嵌套处理并采用基于 pytorch 框架的 LSTM 识别恶意评论, 基于其结果采用改进的 Naïve Bayes 检测恶意词汇及结构, 最终获得含有建议信息但不含恶意内容的评论。

高级时间序列分析: 汽车与航空产业链公司股票协整分析及预测

2019.09-2019.10

- 采用 R 对来自波音航空工业链和特斯拉产业链的共 7 组近 6 年股票数据进行协整分析, 发现 VECM 模型比 VAR 模型在可解释性和预测性上表现更突出。

数据挖掘: 微软恶意软件预测

2019.03-2019.05

- 依据方差大小, 相关系数, 卡方检验等方法对特征进行筛选, 对具有高数量类别特征的变量进行基于频率的编码。
- 在 R 及 Google Cloud 上对来自 Kaggle 的 890 万的观测值基于 52 个特征并采用 LDA、QDA、Logistic Regression, Decision Tree, SVM 等作为单模型对目标值进行预测, 并通过 xgBoost 构建集成模型, 获得 Kaggle 公共分数为 0.68, 第一名为 0.71。

线性回归分析: 美国大都市区人口死亡率影响因素研究

2018.09-2018.12

- 利用 SAS 应用多元线性回归模型以美国大都市区人口死亡率为因变量进行分析, 模型参数通过 F-test, t-test, 基于 VIF 的多重共线性检验等, 最终发现死亡率很大程度上取决于温度、降雨量和教育背景。

工作经历:

风脉能源有限公司, 数据分析师 (实习)

湖北武汉, 2018.01-2018.03

- 在 C# 里控制 DataGridView 设计交互界面, 内嵌读入与输出函数, 对风机测得的以 txt 文件储存的数据进行读入与输出, 并在主窗口与子窗口进行传值。在 C# 中调用 R 来清洗数据并进行描述性统计, 包括按组计算风速的最大值、最小值、均值和频率, 并绘制的风玫瑰图, 获得数据的可视化。
- 在 R 中对原始数据进行有效范围检验, 通过风速插补法和风向插补法, 单塔插补法和异塔插补法, 线性插补法和切变插补法等方法对缺失部分进行插值, 并采用 ARIMA 模型对超短期 (4 小时) 风速进行预测。

中国工商银行长江支行, 信贷助理 (实习)

湖北黄冈, 2016.07-2016.08

- 在 Excel 中对获得的初步社保卡数据进行重复剔除、错误修正、筛选漏填、检测不完整数据并做备份。
- 在 SPSS 中采用线性回归模型分析该银行在第一、二季度的存款客户层次结构, 并预测第三季度的存款量。

荣誉:

中国地质大学 (武汉) 优秀毕业生

2018.06

技术技能:

- R; Python; AWS; GCP; SAS; SPSS; Eviews; Matlab; Excel; Word; Power Point