



学历

2017–2020 计算机技术专业（数据科学方向），硕士，都柏林理工大学绩点：3.6/4.0

专业课程：

- 数据分析编程（SQL, R 编程）(A), 机器学习 (A-), 深度学习 (A), 数据挖掘 (B+)

2013–2017 应用化学专业，学士，北京化工大学

项目经验

2019.09 毕业论文, A Comparison Study on State-of-the-art Minority Class Data Oversampling Techniques for Imbalanced Learning .

2020.01 项目内容和总结：

- 研究多种应对不均衡数据分布时所采用的过采样方法（以 SMOTE 及其改进算法为主），分析不同过采样方法和数据分布形式之间的关系。
- 论文根据数据的统计学分布特征和数据本身的性质，推荐相匹配的数据过采样方法，避免过采样所导致的数据干扰和训练模型过拟合的问题。
- 阅读文献，选取十种常用的数据过采样技术，进行深入研究。
- 在多个数据集上计算选取的过采样算法的表现，比较它们表现的差别并进行统计学检验。
- 根据算法原理，分析改进算法的表现和数据分布形式之间的关系。
- [项目链接: Msc Dissertation Han Tang](#)

2019.08 实习经历, 中国邮政储蓄银行数据分析实习.

- 通过贷中客户数据，构建客户信贷逾期预警模型，分析预测客户下个季度信贷违约情况。

2019.09 ○ 针对数据中的连续变量和分类变量，以及数据不均衡的问题，进行数据准备和描述性分析。

- 使用 Python 分析构建关联矩阵特征相关关系，筛选多重共线变量，进行特征工程操作。
- 通过多种机器学习算法，包括 Xgboost 和 Random Forest（随机森林）构建预测模型，并且通过模型调参进一步优化模型。
- 预测模型在测试数据中召回率达到 0.82。

2019.04 项目经历, 街景路牌数字识别.

- 构建卷积神经网络模型识别图片中的号码牌数字，并判断数字的值。数据包含 10 万张街景图像，其中各包含一个或多个数字。

2019.06

- 设计卷积神经网络架构，并通过图像处理算法将彩色街景图像处理成灰度图，节省计算资源。
- 使用 Python 中的库 TensorFlow 和 Keras 构建多个不同架构的卷积神经网络模型，其中最优模型在测试集中预测 F1 值为 0.9438。
- [项目链接: Steet View House Number 数字图像识别](#)

专业能力

熟练 Python, R, SAS, LaTeX, 机器学习算法, 数据分析和可视化, 数据科学相关的常用库 (Scikit-learn, Pandas, Numpy, Scipy 等)

熟悉 SQL, Web 开发, Java, 深度学习算法 (CNN, RNN), 深度学习常用库 (Pytorch, TensorFlow), 网页爬虫

语言

英文 熟练的听说读写能力

TOEFL : 90

课外活动

足球 参加多支足球队。

- 北京化工大学理学院足球队。
- 北京化工大学足球队。
- 参加北京市足球业余联赛。
- 参加爱尔兰地区级足球业余联赛 UCD Super League。