



韩泽熙

24 岁, 波士顿/北京, +1(617)816-9210, zexihan@outlook.com

github.com/zexihan, [linkedin.com/in/zexihan](https://www.linkedin.com/in/zexihan), zexihan.com

教育背景

| | | |
|---|-----------|-----------|
| 2017.01 至今 | 东北大学, 波士顿 | 数据科学 (硕士) |
| 主修课程: 算法, 机器学习, 数据挖掘, 大规模并行数据处理, 计算机视觉; GPA: 3.8/4.0; 机器学习助教 | | |
| 2012.09-2016.06 | 北京邮电大学 | 电信工程 (学士) |
| 主修课程: 线性代数, 概率论与随机过程, 数字信号处理, 通信原理, 数字系统设计, 数据结构, 软件工程, 高级网络编程; GPA: 3.5/4.0; 国家级创新项目二等奖; 北邮优秀毕业论文 (12/680) | | |

技术掌握

编程语言: Python, Java, Scala, SQL, R, C, MATLAB
框架工具: Hadoop, Spark, AWS, Tensorflow, Caffe, OpenCV, Docker, Git, Linux

实习经历

2018.01-2018.06

Rue Gilt Groupe（奢侈品闪购网站）

Data Scientist

从网购用户中识别转销商

从购买记录、网站行为记录等数据中构建了 127 个转销商的网购特征，并迭代地标记数据、训练调整 XGBoost 模型以从两百万余网购用户中识别转销商，准确率达到 90%，用以改进网站的推荐引擎；

开发 Docker App 实现特征提取、模型训练和推断，将其部署到了亚马逊 ECS 和 Airflow 上对模型和标记数据进行周期性的自动维护更新。

2016.06-2016.08

清华大学人机交互与媒体集成研究所

研究助理

智能手表边缘感知

通过在不同场景从 4/6/8 个不同方向轻敲手表的边缘收集并预处理的数据，探索智能手表加速度传感器的数据变化模式；

训练逻辑回归模型对传感器数据进行分类，并开发 Android Wear App 实现对智能手表边缘敲击动作方位的实时感知。

2015.08-2016.05

中科院自动化所模式识别国重实验室

研究助理

基于深度神经网络的图像特征表示与快速检索

构建一种基于深度卷积神经网络的三级混合图像检索框架（分类，识别，匹配），并应用到同款商品图片检索的任务中；

使用 Caffe 框架将多种 CNN 模型应用到阿里五百万商品图像数据集进行测试，数据集分为 10 个大类和 676 个小类；

开发安卓 App 的后端程序实现了实时的同款商品图像检索，在线检索总时间限制在每张图片 552.49 毫秒；

mAP 最高在化妆品类别上达到了 57.5 %，在上衣、零食、饮料等类别上亦表现优异。

项目经历

2017.09-2017.12

在 Yelp 及人口普查数据上的联合数据挖掘

从 Yelp（类似大众点评）数据集中提取街区级别的商业动态特征；

采用不同的聚类算法来探究商业动态和街区动态之间的关系。

2017.02-2017.04

基于 YouTube-8M 数据集的大规模多标签视频分类

用 Tensorflow 训练一种根据 Youtube-8M 数据集中提供的视频深度特征进行网络视频分类的分类器；

对比了逻辑回归、支持向量机、人工神经网络三种机器学习模型在训练以及准确度上的表现。

2017.02-2017.04

GitHub 优秀开源项目的生命周期

用 R 实现 GitHub 版本资料库中活动数据（Issues & Commits）的收集、清理以及预处理；

分析不同受欢迎度的项目的活动模式，并识别与某种模式最直接相关的指标与重要要素。