**小明**

电话1894879XXXX 邮箱 daxuelu@foxmail.com

# IT技能

* 能够熟练使用Python语言进行开发，有良好的PEP 8编码习惯，掌握Python字典、列表、集合，元组，Python异常捕捉，Python面向对象编程、封装、继承与重写（这里类属性，类方法，静态方法一定要搞清楚，内置方法\_\_del\_\_ \_\_str\_\_有哪些，都有什么作用，私有方法，私有属性作用，多继承继承顺序，Property属性，with等）。
* 掌握常用数据结构与算法，如栈，队列，链表，红黑树，位图，二分及哈希查找，八大排序算法，贪心，能够熟练使用Python原生sort等。
* 能够熟练使用常用Shell命令，VIM编辑器，查看Linux日志，无秘钥登录，定时任务，虚拟环境配置等
* 掌握单例模式，工厂模式，装饰器模式，生成器的使用，会使用面向接口编程，面向切面编程
* 掌握MySQL数据库的增删查改操作，视图，事务的使用，常用索引的使用，主从备份机制的基本使用。
* 掌握 Python多进程、多线程编程，会使用进程池，线程池。掌握进程间通信，epoll 等多路 IO 模型。掌握 socket 网络编程， 熟悉 TCP/IP，HTTP 协议，会使用NAT穿透。
* 熟练使用Numpy科学计算工具、Pandas数据分析包、Matplotlib数据可视化工具，熟悉各种常用业务指标含义，如何进行埋点采集，能够多角度分析业务；
* 熟悉数据集的基本处理方法(空值处理, 数据归一化和标准化)，jieba分词，欠拟合、过拟合产生的原因及解决办法，模型超参数的调优，常用损失函数和梯度下降算法的优化
* 熟悉Scikit-learn机器学习框架，熟练掌握KNN、LinearRegression、Ridge Regression，LogisticsRegression, Decision Tree、随机森林， K-Means等经典机器学习算法
* 掌握Tensorflow 2.0 keras使用，熟练使用dataset及tfrecord，能够搭建多层神经网络解决分类及回归问题，会使用批归一化，dropout，网格搜索，自定义损失函数，自定义层，自定义求导对模型进行调优
* 掌握图片预处理，能够使用CNN实现图片分类，掌握深度可分离卷积，ResNet模型，掌握文本预处理方法，能够使用RNN（lstm，gru）实现文本分类及文本生成，同时熟悉如何调整参数对模型进行优化
* 掌握Git版本控制系统及Gitee/GitHub代码仓库的使用，基本的单元测试技巧。

# 教育及工作经历

**中南大学** 2014年9月 - 2018年7月

计算机专业（不是计算机相关专业这里可以不写专业） 本科 长沙

* 荣誉/奖项：三级奖学金、数模竞赛三等奖，优秀学生干部

# 项目经历

**基于用户定位数据的商圈分析**

项目描述：该项目利用用户手机的定位数据，计算不同商圈的人均停留时间和日均人流量，分析商圈类别，挖掘人流量的分布规律，有助于商场开发潜在客户。

职责描述：

1、提取数据，通过解析、过滤，从外部数据源（移动运营商）得到用户定位数据。

2、数据预处理，计算出人均停留时间、日均人流量等特征数据。

3、模型构建与参数调优，属于层次聚类算法对用户数据进行分析，划分商圈类别。

技术描述：pandas(数据清洗)，scipy(确定聚类中心个数K值)，Scikit-learn(构建层次聚类模型),

Matplotlib(数据可视化)

**二手车价格预测(可以不写)**

使用工具：Scikit-learn,Numpy,Pandas

* 数据来源于阿里天池学习赛， 加载后进行特征提取，缺失值处理以及标准化，PCA降维等特征工程。
* 分别使用正规方程，随机梯度下降以及岭回归等算法进行训练和预测。
* 使用L2正则化来解决过拟合问题，并计算模型的精确率和召回率以及f1\_score。

**CIFAR-10图片分类**

使用工具：TensorFlow-keras，Pandas，Matplotlib，卷积神经网络CNN

* 数据来源于kaggle竞赛，训练集样本数为50000，测试集样本数为300000，图片标签共十个类别。
* 读取数据将其转化为TFRecord格式(便于加载),并对图像进行缩放、翻转、剪切等预处理。
* 利用深度可分离卷积搭建神经网络模型，模型层数超过20层，参数共计超过870万，训练批次为50次；
* 训练后模型在测试集上的准确率为85%以上。
* 利用残差网络ResNet50对模型进行改进，模型参数达到了2300万，在测试集上的准确率达到了93%。

**海洋生物术语翻译助手**

开发工具：TensorFlow-keras，Python，Pandas，Numpy，Scikit-learn，Matplotlib

项目简介：该项目是一个用于快速翻译海洋生物专业术语的机器翻译项目，采用Transformer和缩放点积注意力机制，通过所给的语料集和输入的名词术语可自动生成解释术语的文本。

* 使用Tokenizer将术语名词和翻译内容进行subword-level分词，并进行padding和shuffle等预处理。
* 通过位置编码，给每个词标记位置，实现单词前后位置的联系。
* 使用全连接层和缩放点积搭建Attention层，encoder和decoder层使用多头注意力、feed-forward、
* Add&Normalize和dropout搭建。
* 使用Adam优化器，自定义学习率和损失函数，使用梯度下降对参数进行更新。
* 通过模型进行术语翻译，并绘制翻译的矩阵热度图。

**机器翻译**

使用框架：TensorFlow 2.0，sk-learn 技术：seq2seq，注意力机制

* 该项目是一个用于西班牙语翻译为英语的机器翻译，将源语言和目标语言通过Tokenizer进行word-level的分词
* 使用karas Embedding对数据进行降维，使用gru搭建Encoder层，在decoder中使用BahdanauAttention，将context\_vector合并后经过gru得到output
* 使用自定义损失函数计算损失
* 使用Adam优化器，自定义学习率

**小黄鸡智能聊天机器人**

使用框架：TensorFlow 2.0 seq2seq，注意力机制，tornado，JavaScript，html

项目简介：该项目是B/S模式，前端通过浏览器输入聊天语句，通过http协议将数据传输给后端tornado服务，tornado拿到数据后，将对应数据喂给训练好的模型，得到模型输出后，回复给前端

* 模型训练数据集是50万条的小黄鸡语料集
* 聊天机器人采用的模型是seq2seq，调整的超参数有训练轮数，batch样本数，训练时dropout的保留比例，hidden\_size，rnn神经元单元类型，编码器和解码器的层数，词嵌入的维度，编码器和解码器是否共用词嵌入，解码允许的最大步数，梯度裁剪的阈值，学习率初始值，编码器是否使用双向rnn

# 其他

**语言能力：** 英语：CET6 524分 CET4 588分

**兴趣爱好：** 游泳，跑步，喜欢阅读技术博客及看一些技术书籍