

地址
江苏省南通市港闸区怡
园北村 16 幢 401 室

电话和 Skype
+33 664 294 183(fr)
158 513 90 734(cn)

Mail

hsz1273327@gmail.com

Git

github.com/hsz1273327

自我评价



编程语言

C ★★★★★
Python ★★★★★
java ★★★★★
javascript ★★★★★
scala ★★★★★
Go ★★★★★
R ★★★★★

数据库

RDB ★★★★★
Casandra ★★★★★
MangoDb ★★★★★

大数据

hadoop ★★★★★
Spark ★★★★★

操作系统偏好

Debian ★★★★★
MacOS ★★★★★
Windows ★★★★★

语言

英语 ★★★★★
法语 ★★★★★

黄思喆

数据科学工程师

Occupation Experience

- 01/12 - 06/12 控制工程师 南通科技, 南通
1. 研发机床控制系统
 2. 控制电路设计
 3. 研究多伺服电机同步控制
- 06/11 - 11/11 运营管理助理 平安人寿, 南通
1. 会议和活动组织管理
 2. 区办业务管理支持及业绩统计
 3. 对区办业绩汇总分析直接向内勤总经理提供报告

Education Experience

- 2013 - 现在 计算机数学硕士工程师 (BAC+5) EISTI, 法国
- 研究方向: 数据科学
主要研究领域: 数据挖掘, 社交网络分析, 大数据, 分布式计算, 复杂数据
- 2007 - 2011 自动化学士 重庆大学, 中国
- 自动化专业
主要学习科目: 控制理论, 系统论, 数据结构和算法, 嵌入式系统, 汇编语言.
- 2005 - 2007 高中毕业 江苏省南通第一中学, 中国
- 理科

Achievement

- 2009 全国大学生数学建模竞赛 重庆赛区一等奖
- 选题: 制动试验台的控制方法分析

Web技术

- Web 前端 熟悉 html5,css,javascript, 熟悉 bootsrap 框架, 熟悉 Ajax 技术, 对 react 有一定了解.
- web 后端 熟悉 python 的 Flask,Django 框架, 理解 http 协议

Date Science

- 数据收集 熟悉 python 爬虫技术, 熟练掌握 beautiful soup, 了解 scrapy 框架
- 数据存储 熟悉非关系数据库 Mongodb, 熟悉关系数据库 mysql 和 Sqlite. 了解 HDFS
- 数据分析 熟悉 python 常用数据分析包, 如 numpy,pandas,sicpy,scikit-learn, 熟悉 jieba 中文分词, 熟悉 Spark 及其标准库. 了解基本的机器学习及统计学模型, 如随机森林,SVM,PCA, 贝叶斯网络, 神经网络等
- 数据可视化 熟悉 matplotlib 和 D3.js

Project Experience

05/15 - 11/15 推荐系统

工具: Python,Mongodb

主要算法: 贝叶斯分类器,pagerank 等

指导老师: Maria Malek

说明: 利用协同过滤算法以及基于社会化标签的贝叶斯网络算法构建推荐系统, 数据是爬取的 coursera 和 afpas 上的课程. 我负责数据库构建, 数据爬取和推荐算法研究部分. 前端后端是其他人.

10/14 - 04/15 标签传递法 **mapreduce** 框架下的改进与实现

工具: Spark,GraphX,R,python2.7,igraph.

指导老师: Jean-Philippe Attal, Maria Malek

说明: 该项目服务于一个保险公司的用户社群分类项目, 目的是利用 LPA 算法实现社群划分, 用来挑选出适合推广他们公司的保险产品的人群. 我负责利用 spark 实现人群划分, 并利用 igraph 出图.

02/14 - 04/14 人类面部特征数据分析

工具: scilab,weka

主要算法: 决策树,Logistic regression,KNN,SVM

指导老师: Rachid Chelouah

说明: 该项目是数据分析课的学期作业, 要用以上所有算法对人类面部特征数据进行分析, 建立分类器, 并评价各个分类器在该组数据上的效果. 最后 svm 算法以 64.7% 的正确率胜出.

02/14 - 04/14 用 **MaxDiff** 工具分析市场对手机特征的偏好

工具: C,.net,access

主要算法: maxdiff

指导老师: Maria Malek

说明: 该项目是一个英国公司给的市场调查项目, 要求我们设计一个投票网站收集用户对手机不同特征的偏好 (如价格, 是否可换电池, 摄像头像素高低, 音质, 屏幕大小, 屏幕质量等). 然后实现 maxdiff 算法以分析市场需求. 项目小组由 4 人组成, 因为不会 C # 所以我负责数据库设计部分

Small projects

简单的搜索引擎

python 写的简单搜索引擎, 数据源是百度百科

工具: Python,SQLite

主要算法: pagerank,knn

说明: 分词用的 jieba,web 界面用的 flask, 主要为了体会 < 数学之美 > 中所说的搜索引擎的工作原理:

数据爬取 ==> 索引 ==> 排名, 还在开发中

俄罗斯方块

js 写的俄罗斯方块, 纯粹照着书做

工具: lufylegend

说明: 练手小项目, 主要是为了熟悉 js 的语法

steamAPI 封装

python 写的一个小命令行工具, 可以用来查看目标最近的 **dota** 战绩

工具: python

说明: 目的是为了写个对应功能的 webapp 玩, 同时也是为了以后分析 dota2 用户群体做准备. 目前还没有实现全部功能, 但基本功能已经实现

一些数据分析的小项目利用 **jupyter notebook** 为平台, 写的一些数据分析的小脚本

工具: jupyter,python,spark

说明: 这些项目有的是股市分析, 有的是图像分析, 有的是分析社交网络图.