

杜亚磊

统计学研究生(在读)

电话: 18301928261
网址: badbye.github.io
邮箱: yaleidu@163.com

简介

上海理工大学, 理学院统计学在读研究生。对**数据挖掘**, **机器学习**和**数据可视化**感兴趣。数理统计基础扎实, 但不喜欢纯理论。

技能

数据处理

熟悉各种类型的数据格式: csv, json, xml; 能运用正则表达式处理非结构化数据, 如文本字符

数据分析

熟练基本的描述统计和常用的统计推断算法(回归分析, 方差分析, 聚类, 分类等)

数据展示

撰写html5的分析报告, 混合交互式的动态图形

软件

R

Excel

Git

SQL

Matlab/Octaves

Linux

博客

郑州市日历热图

通过API获取2012年郑州市的每日温度数据, 从xml(json)格式转化成表格, 并绘制日历热图。

研究生英语试卷的词频分析

搜集2003到2013的研究生考研英语试卷, 统计各个单词出现的频率, 并绘制词云展示出现频率最高的词汇; 以统计出的词频为特征构造支持向量机, 对阅读部分的文章划分类别, 如经济, 人文, 自然科学等(这部分未在博客中展示)。

更多

- R语言里的主成分分析
- R语言中的非参数检验

教育经历

2009 - 2013:

- 郑州大学, 商学院, 经济学学士

2013 - Present:

- 上海理工大学, 理学院, 统计学硕士