

李博怡

+86-151-5810-2539 ◊ boyizjuer2017@gmail.com ◊ <https://clamli.github.io> ◊  ◊ 

教育经历

德克萨斯州大学奥斯汀分校

硕士学位，主修计算机科学 | GPA: 4.0/4.0

奥斯汀，德克萨斯州，美国

2018.08 – 至今

浙江大学

学士学位，主修计算机科学与技术 | GPA: 3.91/4.00 (10%)

杭州，浙江，中国

2014.08 – 2018.06

科研/实习

基于协同主动学习树的冷启动物品推荐

2017.10 - 2018.02

西蒙菲莎大学，加拿大，指导教授：Martin Ester 教授，Chung Fu-Lai 教授

- 提出协同主动学习树模型 (CALT)，解决冷启动电影的推荐问题。
- 在基于树的学习算法中结合电影内容信息，对未获得用户评分的电影，利用用户评分电影与当前电影的相似度预测用户评分，缓解已有方法存在的低用户反馈度的问题。
- 采用动态划分方法划分树节点，将节点中电影依评分预测值均匀分入子节点，以解决传统方法不平衡树的问题。
- 提出以相同偏好的用户集合作为预测对象，调用 Spark 库函数对用户集合实现矩阵分解 (MF)，降低评分预测复杂度。
- 实现两种树结构评分预测模型作为对照，在 MovieLens 电影数据集设计实验，将对冷启动电影的评分预测准确度提高 5%。

用于健康记录和食物推荐的 Android 程序开发

2017.07 - 2017.09

Ubiquitous Computing 实验室，新加坡国立大学，新加坡

- 利用 Android 的 AsyncTask 机制解决 APP 内打开相机时导致的界面阻塞问题。
- 设计满足 RESTful 风格的接口，连接服务端和移动端 APP，获取、提交用户拍摄图片并返回图片中食物营养信息。
- 利用遗传算法实现基于两个约束条件的食物排序，定义损失函数，随机交叉适应度高的食物序列，迭代直至获得满意解。

项目经历

Topic-aware Seq2Seq Model for Text Summarization

基于 Seq2Seq 模型的文本摘要生成 (Python)

- 实现基于 Attention 的 Encoder-Decoder 模型，利用 Pointer-generator 网络决策从文本中提取单词或从词库中生成单词。
- 在 Beam search 时将重复出现的相同词组的概率置 0，缓解生成摘要中出现重复词的问题。
- 利用 LDA 生成文本的主题模型，在 Encoder 阶段对输入文本中的主题词计算与其主题相关的 Embedding。

WatchOut!

基于声音的无线入侵检测系统 (C++/Python)

- 调用 PortAudio 获取 Raspberry Pi 麦克风矩阵的声音数据，将其转化为二进制数据并保存至远端数据库。
- 实现基于声音的事件分析器，利用能量信息检测声音发生与否，利用振幅变化信息降低噪音干扰，利用 MUSIC 算法分析声音发生方位。
- 利用 Dash 建立监视网页，实时显示声音数值信息和波形图，并调用事件分析器分析入侵状况。

MiniSQL

类 MySQL 的数据库管理系统 (Java)

- 实现数据库中表的增/删，记录的选择/插入/删除/更新，索引的创建/删除。
- 基于 B+ tree 实现索引管理模块，将所有数据指针存储在树的叶子节点，提高数据查询效率和稳定性。
- 基于 LRU 内存调度算法实现缓存块管理模块，在内存满时移出最近最少使用的数据块，减少磁盘重复访问次数。

Mini-C Compiler

类 C 语言小型编译器 (C)

- 使用 Flex/Bison 工具实现词法分析和语法分析，生成抽象语法树。
- 使用动态规划算法实现指令选择，每次选择使当前代码段执行开销最小的指令集合，提高目标代码的执行效率。
- 利用图着色算法实现寄存器分配，只对溢出变量实现内存存取，降低内存读写次数和使用率。

获奖经历

校学业优秀一等奖学金 (2015) | 校学业优秀三等奖学金 (2016, 2017) | 校优秀学生三等奖学金 (2016, 2017) | 浙江大学 ACM 程序设计竞赛三等奖 (2016) | 美国大学生数学建模竞赛二等奖 (2017)

技能

Java, Python, C/C++, Matlab, R, HTML, Verilog, SQL

NLP, RecSys, Data Analytics, Database, Wireless