

朱晗

- 学校: 清华大学
 - 地址: 北京市 海淀区 清华大学 东配楼 11-419
 - 移动电话: +86 15201519155
 - 邮箱: zhuhan10@gmail.com
-

教育经历

- 2010.8-2014.7 于清华大学软件学院攻读学士学位
 - 2014.8-2017.7 于清华大学软件学院信息系统与工程研究所攻读硕士学位
-

项目经历

万达电影票房预测

在这个项目中, 我们需要预测电影上映首周可能取得的票房成绩。数据方面, 我们能够拿到历史上在万达影院上映过的所有电影的基本信息以及购票数据等。本人在项目中主要负责数据整理与清洗、特征选择、算法设计三个部分。最终, 在测试集上进行测试时, 预测值和真实值的绝对误差(APE)是12%。这一结果对今后影院的排片有很大的指导意义。

海洋浮游生物识别

这是Kaggle平台上的"National Data Science Bowl"竞赛项目, 任务是对拍摄得到的海洋浮游生物进行分类。官方给出了分属于121类的3万多张浮游生物的图片。在这一竞赛中, 由于图片数据不够多, 本人首先对训练数据进行了增广, 然后设计了多个层次不同的神经网络独立完成分类训练, 最后进行模型组合给出分类结果。其中, 单个模型的top-1分类准确率最高达到了88%。

HSQLDB的事务机制优化

HSQLDB是一个用Java编写关系型数据库开源项目。在其早期版本中, 仅支持"SERIALIZABLE"和"READ COMMITTED"两种事务级别。所以, 我和我的团队对其事务机制进行了优化, 增加了"REPEATABLE READ"和"SNAPSHOT"两种新的事务级别。另外, 我们对HSQLDB的表级锁进行了拓展, 实现了行级锁, 以此来提高高并发情况下的查询效率。

研究工作

研究领域

机器学习, 数据质量, 计算机视觉, 深度学习

发表文章

- Deep Hashing Network for Efficient Similarity Retrieval. **Han Zhu**, Mingsheng Long, Jianmin Wang and Yue Cao (AAAI 16)
 - Deep Quantization Network for Efficient Image Retrieval. Yue Cao, Mingsheng Long, Jianmin Wang, **Han Zhu**, Qingfu Wen (AAAI 16)
 - Constraint-Variance Tolerant Data Repairing. Shaoxu Song, **Han Zhu**, Jianmin Wang (SIGMOD 16)
 - Probabilistic Correlation-based Similarity measure on text records. Shaoxu Song, **Han Zhu**, Lei Chen (Information Science 289, 8-24)
-

掌握技能

- 数据分析
- 机器学习, 深度学习
- 计算机视觉
- C/C++, Java, Python, JS, Matlab