马浩杰

18151681382 · 1137633684@qq.com · angelhunt ·

个人信息

男，1994 年出生　　求职意向：算法工程师　　工作经验：约1年

教育经历

- 硕士，南京大学，计算机科学与技术专业，2017.9~2020.7
- 学士，南京大学，计算机科学与技术专业，2013.9~2017.7

工作经历

- **微软中国 公司，m365 部门，2020.6~2021.4**
  大数据平台用户/集群异常检测
  基于scope语法树静态分析实现的data lineage系统

科研经历

- 两篇CCF A类一作会议论文
- Unsupervised Human Activity Representation Learning with Multi-task Deep Clustering (**UbiComp 2020**)
  - 构建了多任务无监督深度聚类，用于人类活动识别
  - 模型通过对聚类，分类和自编码三个任务的联合学习，获取了高效的隐层表征
  - 模型在无标签的情况下能超过主流无监督方法，并且通过少量标签的fine-tuning可以逼近监督算法性能
- AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition (**Ijcai 2019**)
  - 多模态场景下人类的行为识别
  - 通过self attention模型进行了多模态传感器信号的dynamic fusion

项目经历

- 大数据集群异常检测
  - 基于策略的检测和报警，报警合并机制
  - 基于PCA的用户行为异常检测与拉黑
  - 集群节点异常检测，例如任务阻塞，job失败率过高
- 基于Intel MKL加速spark MLlib算法性能（本科毕设）
  - 通过MKL实现了PCA, ALS, Logistic Regression等算法
- NJU-ProjectN
  - 实现了完整的流水线CPU，c语言子集编译器，MIPS模拟器和简易操作系统
  - 在FPGA上运行自己实现的MIPS CPU, x86硬件模拟器，操作系统，编译器，并编译运行仙剑奇侠传
- 造的一些轮子
  - python实现的机器学习算法，例如决策树，朴素贝叶斯
  - c++, MPI, cuda等库实现的并行机器学习算法
  - 高性能KV存储，lsm tree based存储引擎
  - 使用Nlp技术分析金庸小说，实现多标签聚类算法，展示出金庸小说人物关系图

获奖经历

- 2015年-2017年 南京大学人民奖学金
- 2016年-2017年 南京大学第二届程序设计竞赛二等奖
- 2019年-2020年 国家奖学金

Haojie Ma

18151681382 · 1137633684@qq.com · angelhunt ·

Information

1994/07/09 Job intention: Algorithm engineer Work experience: near 1 year

Education

- Master, Nanjing University, Computer Science and Technology, 2017.9~2020.7
- Bachelor, Nanjing University, Computer Science and Technology, 2013.9~2017.7

Experience

- **Microsoft China, m365 department, 2020.6~2021.4**
  Big data platform user/cluster anomaly detection
  Data lineage system based on static analysis of scope syntax tree

Research

- Two CCF-A conference papers
- Unsupervised Human Activity Representation Learning with Multi-task Deep Clustering (**UbiComp 2020**)
  - Constructed multi-task unsupervised deep clustering for human activity recognition
  - The model obtains hidden layer representations through joint learning of the three tasks of clustering, classification and auto-encoding
  - The model can surpass mainstream unsupervised methods without labels, and the performance of supervised algorithms can be approximated by fine-tuning with a small number of labels
- AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition (**Ijcai 2019**)
  - Human behavior recognition in multi-modal scenes
  - Dynamic fusion of multi-modal sensor signals is carried out through the self attention model

Project

- Big data cluster anomaly detection
  - Policy-based detection and alarm, alarm merging mechanism
  - PCA-based user behavior anomaly detection and blackout
  - Cluster node abnormality detection, such as task blocking, job failure rate is too high
- Accelerate the performance of spark MLlib algorithm based on Intel MKL (undergraduate completion)
  - Implemented PCA, ALS, Logistic Regression and other algorithms through MKL
- NJU-ProjectN
  - Implement a complete pipeline CPU, c language subset compiler, MIPS simulator and simple operating system
  - Run the MIPS CPU, x86 hardware simulator, operating system, and compiler implemented by yourself on FPGA, and compile and run Legend of Sword and Fairy
- Some small project
  - Machine learning algorithms implemented in python, such as decision trees, naive Bayes
  - Parallel machine learning algorithms implemented by libraries such as MPI and cuda
  - High-performance KV storage, lsm tree based storage engine
  - Use Nlp technology to analyze Jin Yong's novels, implement a multi-label clustering algorithm, and display the relationship diagram of Jin Yong's novels

Award

- 2015-2017 Nanjing University People's Scholarship
- 2016-2017 Second Prize of the Second Program Design Competition of Nanjing University
- 2019-2020 National Scholarship