

# 齐建鹏

大数据 · 数据挖掘等

烟台大学 计算机与控制工程学院, 山东 · 烟台

☎ (+86) 156-1535-0639 | ✉ jianpengqi@126.com | 📱 qijianpeng | 🌐 qijianpeng

*"One person's noise is another person's signal."*

## 教育经历

硕士研究生, 计算机与控制工程学院, 烟台大学 (导师: 于彦伟)

山东 · 烟台

2015.09 - 今

- 学业奖学金一等、二等奖; 优秀研究生; 优秀共产党员; 研究生优秀科技创新成果奖二等奖

大学本科, 计算机与控制工程学院, 烟台大学

山东 · 烟台

2011.09 - 2015.06

- 优秀毕业生; 优秀学生干部; 优秀团员干部; 齐鲁软件大赛三等奖

## 培训经历

2016 可视化暑期学校, 北京大学

北京

学员

2016.06

甲骨文工程教育实践中心 (Oracle WDP)

山东 · 烟台

学员, 软件开发

2013.09 - 2014.06

## 职业技能

大数据 数据挖掘, 实时流数据分析, Storm, Hadoop, etc.

编程 J2EE 相关, C/C++, R, Python, Matlab, LaTeX, C#.

其它 **OS:** Linux; **J2EE:** Struts2, Spring, Hibernate, iBatis/MyBatis; **Database:** Oracle, MySql, SQL Server; **HTML:** JavaScript, CSS3, HTML5; etc.

## 研究经历

大数据流分布式下基于距离的异常检测

硕士研究课题

论文撰写中...

### • Jobs:

理论分析, 建模, 算法优化以及在 Storm 平台上进行实现. 包括 Storm 集群环境的搭建等.

### • Highlights:

Apache Storm, 网格划分下的负载均衡, 数据流, 空间索引.

### An Effective and Efficient Hierarchical K-means Clustering Algorithm

Journal

International Journal of Distributed Sensor Networks(SCI 检索); Jianpeng Qi, Yanwei Yu\*, Lihong Wang,

2017

Jinglei Liu, Yingjie Wang

- 提出了一种类似于层次聚类的优化的  $k$ -means 聚类算法, 由三种优化原则组成.

### • Jobs:

- 1). 增量式更新 Mean & SSE(Sum of Square Errors);
- 2). 通过数学推导公式对簇集进行合并, 降低随机选择初始点的敏感性以及提升效率;
- 3). 通过推导, 发现了可以对搜索空间进行剪枝的距离半径, 大大提升了迭代速度.
- 4). 大量实验对算法效率与效果进行了验证.

- SocialCom2016 会议论文扩展版本

- Highlights: 聚类, K-means, 算法优化.

## 基于多线程的不确定移动对象连续 $k$ 近邻查询

Journal

浙江大学学报(工学版)(EI 检索); 齐建鹏, 于彦伟\*, 王创存, 曹磊, 宋鹏

2017.07.09

- 提出了一种高效的基于多核多线程的并行查询处理框架.
- **Jobs:**
  - 1). 提出了一种针对  $k$  近邻可以进行查询复用的方法;
  - 2). 使用一种基于密度网格的多线程数据分发方法, 解决了多线程并行条件下的负载均衡问题;
  - 3). 通过实现多线程间的内存共享机制, 对已计算过的移动对象的预测区域通过建立共享区来实现计算复用, 以避免重复计算.
- 已录用
- **Highlights:** 多核多线程, PR-Tree 空间索引, 数据流, 查询复用.

## 移动社交网络异常签到在线检测算法研究

中国·昆明

中国计算机学会人工智能会议 (CCFAI2017); 赵冠哲, 齐建鹏, 于彦伟\*, 刘兆伟, 宋鹏

Aug. 4-6, 2017

- 针对移动社交网络中用户异常签到位置检测问题, 提出了一类基于用户移动行为特征的异常签到在线检测方法.
- **Jobs:**
  - 1). 算法讨论与研究;
  - 2). 论文修改等.
- 已被《智能系统学报》录用
- **Highlights:** 数据流, 基于距离的异常检测, 好友圈.

## 面向不确定移动对象的连续 $K$ 近邻查询算法

中国·桂林

第 6 届中国数据挖掘大会 (CCDM2016); 于彦伟\*, 齐建鹏, 宋鹏, 张永刚

May 20 - 22, 2016

- 提出了一种高效的针对不确定移动对象的连续  $k$  近邻查询算法.
- **Jobs:**
  - 1). 利用最近一段时间窗口内的位置采样、速度和方向预测移动对象在查询时刻到未来可能的位置区域;
  - 2). 使用最小距离与最大距离区间描述移动对象到查询对象的距离. 采用优化的基于模糊 (Vague) 可能度判定的排序方法查找查询对象的  $k$  近邻.
- 出版于《模式识别与人工智能》, Vol. 29, 2016.
- **Highlights:** 移动对象, 预测, 模糊决策, PR-Tree, 数据流.

## K\*-Means: An Effective and Efficient K-Means Clustering Algorithm

Atlanta, USA

The 9th IEEE International Conference on Social Computing and Networking(SocialCom2016)(EI 检索);

Oct. 8 - 10, 2016

Jianpeng Qi, Yanwei Yu\*, Lihong Wang, Jinglei Liu

- 提出了一种类似于层次聚类的优化的  $k$ -means 聚类算法, 由三种优化原则组成.
- **Jobs:**
  - 1). 增量式更新 Mean & SSE(Sum of Square Errors);
  - 2). 通过数学推导公式对簇集进行合并, 降低随机选择初始点的敏感性以及提升效率;
  - 3). 通过推导, 发现了可以对搜索空间进行剪枝的距离半径, 大大提升了迭代速度.
- **Highlights:** 聚类, K-means, 算法优化.

## MR-Swarm: Mining Swarms from Big Spatio-Temporal Trajectories Using MapReduce

中国·扬州

The 17th International Conference on Intelligent Data Engineering and Automated Learning(IDEAL' 2016)(EI 检

Oct. 12 - 14, 2016

索); Yanwei Yu\*, Jianpeng Qi, Yunhui Lu, Yonggang Zhang, Zhaowei Liu

- 针对时空数据提出了一种分布式挖掘 Swarm 模式的计算框架, 通过 MapReduce 链来实现并行挖掘.
- **Jobs:**
  - 1). Hadoop 集群维护及管理;
  - 2). 针对该算法在 MapReduce 下编程, 并考虑通信, 负载问题对算法进行优化;
  - 3). 在会议上进行了英文报告.
- 出版于 "Lecture Notes in Computer Science", vol 9937. Springer, 2016.
- **Highlights:** MapReduce 链, 时空数据, 负载均衡 (Partition), 通信优化 (Combiner), Linux.

- 提出了一种高效的基于 MapReduce 的分布式蜂群模式挖掘算法。
- Jobs:**
  - 利用最小时间支集优化了串行挖掘算法;
  - 提出了蜂群模式的并行化挖掘模型, 利用蜂群模式时间域无关性, 并行化了聚类与子时间域上的蜂群模式挖掘过程;
  - 设计了一个基于 MapReduce 链式架构的分布式并行挖掘算法, 通过四个阶段快速地实现了蜂群模式的并行挖掘。
- 出版于《计算机工程与科学》, Vol. 38, 2016.
- Highlights:** DBSCAN, 时空数据, MapReduce, Linux.

## 项目经历

### 研究型

#### 面向不确定数据的 KNN 高性能并行实时查询

烟台大学研究生科技创新基金

主持人, 提前完成, 已达到结题要求

2016.12 - 2017.10

- Keywords:** 并行, 连续  $k$  近邻查询, 不确定移动数据, 数据流

#### 流式大数据分布式异常检测算法研究

国家自然科学基金

参与人, 导师项目

2015.09 - 2017.7

- Keywords:** 大数据, 数据流挖掘, 分布式数据挖掘, 异常检测, 聚类分析

### 应用型

#### 企业人力资源管系统

Oracle WDP

成员

2013.09 - 2013.12

- 开发环境: Windows, Tomcat.
- 开发平台: MyEclipse, Oracle 11g, SVN.
- 技术: SSH2, JSP, JAVA, JS, Oracle, Ajax.
- 项目描述: 实现公司资源的在线管理, 提高办公效率.
- 贡献:
  - 培训管理模块实现, 针对各单位的培训需求调研, 计划制定阶段, 培训实施阶段, 培训效果评价阶段作统一的过程管理.
- 总结: 通过陪培训模块的实现, 掌握了审批流程的处理方法, 对 Ajax 使用, 富文本编辑器插件也有了很好的把握. 对 SSH 框架达到了可以熟练使用的目的. 对 SVN 达到了熟练使用的程度. 对于团队合作, 协调能力也有了很大的提升.

#### Online Mall

Oracle WDP

成员

2014.01 - 2014.04

- 开发环境: Windows, Tomcat.
- 开发平台: MyEclipse, Oracle 11g, SVN.
- 技术: JSP, JQuery, Servlet.
- 项目描述: 本项目为大型企业提供一个在线商品交易平台, 平台的主要目的是让企业在平台上发布商品及资讯, 用户可以在此平台上购买商品并参与商品的评论. 围绕这一目的, 系统需要实现登录注册, 产品前端显示, 产品全文搜索, 用户评论管理, 人工客户服务, 商品管理, 商品订购, 商品促销, 内容管理, 会员用户管理等功能.
- 贡献:
  - 安全登录: 使用 MD5 加密算法, 对用户登录数据进行保护.
  - 用户管理: 网站员工可以查看到注册用户的信息, 对恶意用户的账号可以实施禁用. 另外还可以为丢失密码的用户找回密码;
  - 员工管理: 维护员工的基本信息, 如姓名, 联系电话、电子邮件等信息. 对离职的员工可以实施账号停用;
  - 权限管理: 网站管理员可以添加权限组, 为员工分配权限. 产品管理: 产品具有类别、品牌、生产厂商、供应商、样式、尺码等属性, 产品类别可以实现无限级分类, 品牌具有中英文名称及 Logo. 网站员工在该模块中可以对产品信息进行管理. 通过更换模版商可以实现产品页面不同的显示风格;
  - 用户评论管理: 网站员工对用户的评论可以进行回复、删除操作.
- 总结: 通过本项目开发, 熟练掌握了运用 Ajax 进行记录管理的操作. 并掌握了在 SSH 框架下实现文件的上传功能. 更加深入的了解了角色权限的管理.

## 荣誉与经历

- **2015** 级研究生班长, 烟台大学, 计算机与控制工程学院 2015.09 - 今
- 研究生会主席, 烟台大学, 计算机与控制工程学院 2016.09 - 今
- 计算机学生党支部组织委员, 烟台大学, 计算机与控制工程学院 2016.06 - 2017.06
- **2011** 级班级组织委员, 烟台大学, 计算机与控制工程学院 2012 - 2015