

# 褚征

男 30岁 党员 (+86)185-9917-2150 [375257209@qq.com](mailto:375257209@qq.com) [个人网站](#) 新疆大学(双一流, 211)



## 教育背景

2018.09-2021.06	<b>新疆大学, 信息科学与工程学院, 计算机应用技术专业 (统招, 全日制)</b> 博士 年级排名: 1/15 研究方向: AI for System (流处理任务, 性能评估与预测, 机器学习, 深度学习) 重要奖项: 一等学业奖学金; 新疆维吾尔自治区奖学金; 优秀研究生标兵; 优秀学位论文; 优秀毕业生
2015.08-2018.06	<b>新疆大学, 软件学院, 软件工程专业 (统招, 全日制)</b> 硕士 年级排名: 2/30 研究方向: 分布式计算 (内存云(RamCloud)) 重要奖项: 一等学业奖学金; 两度新疆维吾尔自治区奖学金, 校级优秀学生干部, 院级优秀学生干部
2010.09-2014.06	<b>新疆大学, 软件学院, 软件工程专业 (统招, 全日制)</b> 学士 年级排名: 14/130, 软件设计师(中级) 研究方向: JAVA CS系统研发 重要奖项: 2012中国成都国际软件设计与应用大赛个人组优胜奖; 2011 JAVA组优秀奖

## 工作经历

2016.07-2018.08	<b>广州汇智通信技术有限公司-研发三部</b> 算法工程师 在民族关系紧张、外部势力渗透和国内公共安全技术匮乏背景下, 为保障地区安全稳定, 实现特殊人员及群体识别、预警和管控等目标, 采取LR, RF等算法挖掘异常样本, 利用Kafka、Flink等实时计算框架提升异常样本的预警效率。针对样本不平衡、数据规模大等难点, 使用大数据框架和机器学习算法, 最终实现了特殊人员及群体的建模、识别、预警和管控等任务。(行业涉密, 无法过多描述)
2014.07-2015.08	<b>中国联通乌鲁木齐分公司-信息化/市场/互联网发展部</b> 数据挖掘 在通信行业竞争加剧和(移动)互联网的巨大冲击下, 为了满足用户的多样需求, 利用数据挖掘、机器学习和深度学习技术挖掘利润增长点以提升公司营业收入。具体而言, 使用BPNN预测离网用户并维系挽留; 使用RF识别僵尸卡并关联真实用户; 根据补卡行为识别潜在终端购买人群并进行推荐等。通过应用上述方法, 单月营业收入提升约1% (30W左右)。此外, 异业联盟等其它项目不再过多描述。
2012.07-2014.08	<b>乌鲁木齐掌新动力信息技术有限公司</b> 软件工程师(实习) 在学籍管理信息化程度低等背景下, 为满足主管部门管理辖区学校, 实现标准统一的中小学学籍管理, 研发乌鲁木齐中小学学籍管理系统。针对业务建模困难, 开发效率低等问题, 使用Sprig MVC、Hibernate和模板技术, 最终实现用户需求。此外, 其它JAVA CS系统不再一一赘述 [1],[2].

## 科研经历

2018.09-2021.06	<b>面向流处理任务的性能评估与预测研究</b> 面向流处理任务的性能评估与预测技术以其对性能的评估能力和对性能的预测能力构成了对复杂流处理环境的良好支撑。为了解决评估手段匮乏和性能预测复杂等诸多不足, 实现对性能的准确评估与预测, 采取了自适应的性能评估方法、实时延迟预测方法、基于LPG-model的吞吐量预测方法和检查点执行时间预测方法。针对遇到的性能波动大、特征识别困难、冗余信息多、经验知识局限等问题, 使用跳跃滑动窗口、专家经验、LightGBM、PCA、随机森林等方法, 最终实现了准确且高效的性能评估与预测 ([1],[2],[3],[4], 2020年新疆维吾尔自治区科学技术二等奖)。 <b>图数据库查询性能预测</b> 由于查询计划和系统的复杂性, 图数据库中查询任务的执行时间预测已变得困难和挑战。为满足DBA和DBMS在查询任务执行期间和执行之前捕获准确的执行时间, 达到准确和高效地预测执行时间的目标, 采取了人工智能技术和图数据库理论相结合方法。通过使用RF、PCA和LSTM对特征进行筛选、降维和执行时间预测, 最终实现了81.34%的平均准确率[5]。 <b>端到端的水稻产量预测</b> 水稻产量对世界上半以上的人口是至关重要。因此, 准确预测水稻产量对贸易, 政策, 人道主义援助, 决策者都至关重要。为了克服传统机械模型和统计机器学习模型费力费时的缺点, 提出了一种新型端到端的预测模型。该模型融合BPNN和IndRNN网络以学习样本的时空特征, 最终实现了稳定的预测性能和较低的预测误差[6]。
2015.09-2016.06	<b>基于内存云的大块数据对象并行存取策略</b> 内存云 (RAMCloud) 是一种将所有数据存储已在部署服务器内存中的新型存储系统上, 具有高速访问和低延迟的优点。由于内存云仅支持存取小块数据对象 (1MB) 限制, 为支持大块数据对象存取, 提出了大块数据对象存取策略。利用数据分割思想和并行数据存取算法, 克服了内存云的存取限制, 最终实现了大块数据对象与小块数据对象相同的存取性能并能够应用于实际生产 [7],[8],[9],[10]。

## 知技能

编程技能	擅长Python、JAVA、SQL语言；熟练使用Linux系统，熟悉Linux环境开发。
分布式计算	熟悉HDFS、Hive、RamCloud、Flink分布式系统；擅长大数据存储、处理、分析及建模。
机器学习	熟悉Sklearn框架；擅长数据预处理、升降维、回归、分类等算法(PCA、LR、SVM、RF、LightGBM等)。
深度学习	熟悉PyTorch、Keras框架，擅长性能建模及预测，有丰富项目经验；擅长模型融合，了解模型优化方法。 熟悉BPNN、RNN(LSTM/GRU)、CNN等常见网络，了解DenseNet等模型，常见网络结构与训练技巧。

## 出版物

- [1] **Chu Z**, Yu J , Hamdulla A . LPG-model: A novel model for throughput prediction in stream processing, using a light gradient boosting machine, incremental principal component analysis, and deep gated recurrent unit network[J]. *Information Sciences*, 2020, 535:107-129. (SCI, JCR Q1, IF 5.910, 中科院一区, **Top Journal**, CCF B). [Link](#)
- [2] **Chu Z**, Yu J , Hamdulla A . Maximum Sustainable Throughput Evaluation Using an Adaptive Method for Stream Processing Platforms[J]. *IEEE Access*, 2020, 8(99):40977-40988. (SCI, JCR Q1, IF 4.098, 中科院二区). [Link](#)
- [3] 褚征, 于炯. 基于随机森林的流处理检查点性能预测[J]. *电子与信息学报*, 2020, 42(6): 1452-1459. (EI, ESCI). [Link](#)
- [4] **Chu Z**, Yu J , Hamdulla A . Throughput prediction based on ExtraTree for stream processing tasks[J]. *Computer Science and Information Systems*, 2020, 18:31-31. (SCI, JCR Q2, IF 0.742, 中科院四区). [Link](#)
- [5] **Chu Z**, Yu J and Hamdulla A. A novel deep learning method for query task execution time prediction in graph database[J]. *Future Generation Computer Systems*, 2020, 112 : 534-548. (SCI, JCR Q1, IF 6.125, 中科院一区, **Top Journal**, CCF C). [Link](#)
- [6] **Chu Z**, Yu J . An end-to-end model for rice yield prediction using deep learning fusion[J]. *Computers and Electronics in Agriculture*, 2020, 174:105471. (SCI, JCR Q1, IF 3.858, 中科院一区, **Top Journal**). [Link](#)
- [7] 褚征, 于炯, 鲁亮,等. 基于内存云的大块数据对象并行存取策略[J]. *计算机应用*, 2016, 36(6):1526-1532. [Link](#)
- [8] 褚征, 于炯, 王佳玉,等. 基于LDA主题模型的移动应用相似度构建方法[J]. *计算机应用*, 2017, 37(4):1075-1082. [Link](#)
- [9] 王佳玉, 张振宇, 褚征等. 一种基于轨迹数据密度分区的分布式并行聚类方法. *中国科学技术大学学报*, 2018,48(01):47-56. [Link](#)
- [10] 褚征. 基于内存云的大块数据对象并行存取策略[D]. *新疆大学*, 2018.

## 软件著作权

- [1] 基于EXT的毕业生就业后台管理系统V1.0 (编号: 2016SR241820)
- [2] 中学生综合素质管理信息系统V1.0 (编号: 2017SR445164)

## 主持项目

- [1] 大数据流式计算平台下配置项优化方法研究（XJUBSCX-201901），2019年新疆大学博士生科技创新项目，2019-2021，主持

## 自我评价

成熟稳重，善于团队合作，积极进取，拥有创新和独立完成工作的能力。热爱生活，户外运动。

# Zheng Chu

Email: [375257209@qq.com](mailto:375257209@qq.com) Xinjiang University (XJU)

Web: <https://chuzheng88.github.io/zhengchu.github.io/>

## Education

Ph.D. in Computer Application Technology, Xinjiang University, Urumqi, China	2018-2021
M.S. in Software Engineering, Xinjiang University, Urumqi, China	2015-2018
B.S. in Software Engineering, Xinjiang University, Urumqi, China	2010-2014

## Research Interests

**AI:** machine learning, deep learning

**AI for System:** (i.e., performance prediction for stream processing task, execution time prediction for query task in graph database)

**AI for Agriculture:** (i.e., rice yield prediction)

## Work Experience

<b>Machina Learning Algorithm Engineer</b> , Guangzhou Huizhi Communication Technology Co., Ltd, Guangzhou, China	2017-2018
<b>Software Development Engineer</b> , Guangzhou Huizhi Communication Technology Co., Ltd, Guangzhou, China	2016-2017
<b>Data Analyst</b> , China Unicom, Urumqi, China	2014-2015
<b>Software Development Engineer (Intern)</b> , Urumqi PalmSoft Electronic Information Technology Co., Ltd., Urumqi, China	2012-2014

## Research Experience

<b>Xinjiang University, Prof. Yu Jiong</b>	2018-2021
Research on Performance Evaluation and Prediction for Stream Processing Tasks [1], [2], [3], [4]	
Execution Time Prediction for Query Task in Graph Database using Deep Learning [5]	
Rice Yield Prediction with End-to-end Model using Deep Learning Fusion [6]	
<b>Xinjiang University, Prof. Yu Jiong</b>	2015-2016
Parallel Access Strategy for Large Data Objects based on RamCloud [7], [8], [9], [10]	

## Skills

**Programming:** Python, JAVA, SQL, LaTeX

**Distributed system:** Hadoop ecosystems (i.e., HDFS, Hive, Hbase, etc.), RamCloud, Apache Flink, Kafka

**Frameworks:** PyTorch, Keras, Scikit-learn, Pandas, Spring, Struts, Hibernate

## Awards & Honors

Outstanding Doctoral Dissertation Award	2021
Outstanding Graduate Model Award	2021
Outstanding Graduate Award	2021
First-class Academic Scholarship	2020
Scholarship of Xinjiang Uygur Autonomous Region	2020
First-class Academic Scholarship	2019
Scholarship of Xinjiang Uygur Autonomous Region	2018
First-class Academic Scholarship	2018
Scholarship of Xinjiang Uygur Autonomous Region	2017
Second-class Academic Scholarship	2017
School-level Outstanding Student Leader Award	2017
First-class Academic Scholarship	2016
College-level Outstanding Student Leader Award	2016
First-class Academic Scholarship	2015
Second-class Academic Scholarship	2013

Second-class Academic Scholarship	2012
China Chengdu International Software Design and Application Competition Individual Group Excellence Award	2012
Second-class Academic Scholarship	2011
China Chengdu International Software Design and Application Competition Java Group Excellence Award	2011
Second-class Academic Scholarship	2010

## Publications

---

- [1] **Chu Z\***, Yu J, Hamdulla A. LPG-model: A novel model for throughput prediction in stream processing, using a light gradient boosting machine, incremental principal component analysis, and deep gated recurrent unit network[J]. *Information Sciences*, 2020, 535:107-129. (JCR Q1, IF 5.910, **Top Journal**, CCF B). URL: [Link](#)
- [2] **Chu Z\***, Yu J, Hamdulla A. Maximum Sustainable Throughput Evaluation Using an Adaptive Method for Stream Processing Platforms[J]. *IEEE Access*, 2020, 8(99):40977-40988. (JCR Q1, IF 4.098). URL: [Link](#)
- [3] **Chu Z\***, Yu J. Performance Prediction Based on Random Forest for the Stream Processing Checkpoint [J]. *Journal of Electronics and Information Technology*, 2020, 42(6): 1452-1459. (EI, ESCI). URL: [Link](#)
- [4] **Chu Z\***, Yu J, Hamdulla A. Throughput prediction based on ExtraTree for stream processing tasks[J]. *Computer Science and Information Systems*, 2020, 18:31-31. (JCR Q2, IF 0.742). URL: [Link](#)
- [5] **Chu Z\***, Yu J and Hamdulla A. A novel deep learning method for query task execution time prediction in graph database[J]. *Future Generation Computer Systems*, 2020, 112: 534-548. (JCR Q1, IF 6.125, **Top Journal**, CCF C). URL: [Link](#)
- [6] **Chu Z\***, Yu J. An end-to-end model for rice yield prediction using deep learning fusion[J]. *Computers and Electronics in Agriculture*, 2020, 174:105471. (JCR Q1, IF 3.858, **Top Journal**). URL: [Link](#)
- [7] **Chu Z**, Yu J\*, Lu L, etl. Parallel access strategy for big data objects based on RAMCloud [J]. *Journal of Computer Applications*, 2016, 36(6):1526-1532. URL: [Link](#)
- [8] **Chu Z**, Yu J\*, Wang J Y, etl. Construction method of mobile application similarity matrix based on latent Dirichlet allocation topic model [J]. *Journal of Computer Applications*, 2017, 37(4):1075-1082. URL: [Link](#)
- [9] Wang J Y, Zhang Z Y\*, **Chu Z**, etl. A trajectory data density partition based distributed parallel clustering method. *Journal of University of Science and Technology of China*, 2018,48(01):47-56. URL: [Link](#)
- [10] **Chu Z**. Parallel access strategy for big data objects based on RAMCloud [D]. *Xinjiang University*, 2018.

## Software Copyright

---

- [1] Management System based on EXT for Graduate V1.0 (No: 2016SR241820)
- [2] Quality Information System for Middle School Student V1.0 (No. 2017SR445164)

## Research Projects

---

- [1] Research on optimization for configurations on stream processing platform (No.XJUBSCX-201901), Science and Technology Innovation Project for Doctoral Students, Xinjiang University, 2019-2021, Host.