

倪嘉志

学校: 中国科学院软件研究所 手机: (+86) 186-8674-6125 邮箱: nijiazhi14@otcaix.iscas.ac.cn

教育背景

- | | | | | |
|-------------------|-----------------|----------|------|----------|
| ➤ 2014.09-至今 | 中国科学院软件研究所 (保送) | 计算机软件与理论 | 工学硕士 | 排名 2/150 |
| ➤ 2010.09-2014.07 | 哈尔滨工程大学 | 计算机科学与技术 | 工学学士 | 排名 2/181 |

获奖、论文情况

- | | |
|--|---------------|
| ➤ 阿里-天池大数据竞赛季军 (3/2734) | 2015 年 |
| ➤ “共享杯” 科技创新大赛二等奖、宁波国际机器人大赛水中鱼项目二等奖 | 2013, 2015 年 |
| ➤ 黑龙江省三好学生、省优秀毕业生 | 2014 年 |
| ➤ 中国计算机学会 CCF 优秀大学生奖 (受邀参会, 全国 100 名) | 2013 年 |
| ➤ 本科、研究生国家级奖学金 (3 次), 中科院学业奖学金一等奖, 校级一等奖学金 (7 次) | 2011 年-2016 年 |
| ➤ 《基于多段间隔监督度量学习的病人相似度算法》 计算机系统应用 | 2015 年 |
| ➤ 《基于迁移学习的病人相似性度量算法》 第四届 CCF 大数据学术会议 (报告论文) | 2016 年 |
| ➤ 《度量学习在医疗领域的应用研究》中国科学院大学第 13 届研究生学术研讨会 (报告论文) | 2016 年 |

项目经历

- | | |
|---|--------------------------|
| ➤ 基于相似度的健康数据检索 : | 2016.01 - 今 |
| ■ <u>项目描述</u> : 旨在利用历史健康数据 (门诊记录, 住院记录等) 对健康主体进行建模, 结合领域专家信息实现有监督的度量学习进而产生有效的相似度度量方式, 利用其进行检索; 在不同疾病领域通过迁移学习手段完成知识共享 | |
| ■ <u>主要贡献</u> : 健康数据模型设计、特征工程、相似度算法 (LSML) 迁移学习算法 (TML) 研究及实现、多标签分类算法及测试 | |
| ■ <u>主要技术</u> : 特征提取、R(caret)、multi-label classify、自动编码器、回归预测、信息检索、数据索引与缓存 | |
| ➤ 广东-公交线路客流预测 (独立完成, 阿里竞赛): | 2015.10 - 2015.11 |
| ■ <u>项目描述</u> : 根据广州市 2014 年 8 月 1 日至 2014 年 12 月 31 日五个月的公交刷卡数据 (6 时-21 时) 对公交行为进行建模, 数据涉及 200 万用户、2 条线路, 约 800 万条, 同时提供天气状况, 预测次年 1.1-1.7 的分时段流量数据(含假日) | |
| ■ <u>主要贡献</u> : 公交模型设计、特征工程、回归算法的研究及使用、时序性数据的研究 | |
| ■ <u>主要技术</u> : 数据清洗、特征提取、Python(sklearn)、GBDT、RF、线性模型、 | |
| ➤ Haflow : 大数据分析平台 (国家 863 项目) | 2015.06 - 2015.10 |
| ■ <u>项目描述</u> : 一站式大数据分析应用开发平台, 支持多种大数据处理分析框架的接入, 并借助其分布式存储和计算能力, 加快大数据分析的速度。构建基于 Web 的拖曳式的数据分析应用开发界面, 同时将用户从平台搭建, 工具集成, 接口学习, 算法管理, 实验脚本编写和多任务调度等繁杂工作中抽离出来 | |
| ■ <u>主要贡献</u> : 设计并实现系统基础架构; 基于 Hbase 的中间数据管理; 设计基于组件以及模型驱动的数据分析业务流程处理框架; 设计与实现基于嵌套图识别与分层的模型转换算法; 大数据平台 MR 作业分析及优化 | |
| ■ <u>主要技术</u> : Hadoop, MapReduce 优化, Hbase, Hibernate, java 注解, 数据流模型转换、Spring MVC | |

实习经历

- | |
|--|
| ➤ 腾讯 MIG 事业群 --- 定位组 --- 定位算法基础研究 (2016.07 - 至今) |
| ■ 实习内容: 1、使用 Pig/MapReduce 对海量用户 WiFi 定位数据进行处理, 并利用 Hbase 进行数据持久化; |
| 2、处理公交线路数据, 提取公交线路特征, 对公交路线进行建模; |
| 3、设计并实现基于 MapReduce 的用户 WiFi 定位数据挖掘算法, 用于实现用户与公交线路的匹配; |

专业技术

- | |
|--|
| ➤ 熟悉 Java 语言; 了解并可运用 Python、R 等主流语言; 掌握常用的算法及数据结构; CET-6 |
| ➤ 熟悉 Hadoop 生态圈及 MapReduce 实现原理, 了解大数据相关工具如 HBase, Pig, Yarn, Hive, Zookeeper 等; |
| ➤ 掌握常用的数据挖掘和机器学习方法, 了解 ML 相关工具如 sklearn、XGboost, 及语义网相关技术, 如 Jena、sparql 等; |