

李鸿敏

电话: 070-3668-3456 | 邮箱: lihongmin@mma.cs.tsukuba.ac.jp
现居城市: 日本茨城县つくば市桜2-26E棟105



教育经历

筑波大学 - 计算机科学 博士 信息工程系 GPA : 4.1 / 4.3	2019年03月 - 2022年03月
筑波大学 - 计算机科学 硕士 信息工程系 GPA换算 : 4.0 / 4.3	2017年04月 - 2019年03月
全英语教学课程: 机器学习 (A+), 计算机科学实验设计 (A+), 进化计算 (A+), 学术写作 (A)。	
宁夏大学 - 电子信息工程 本科 物理电气信息学院 相关课程: 线性代数, 信息论, 概率论与数理统计。	2011年09月 - 2015年07月

研究经历 (6篇论文, 2篇顶会)

博士研究: 关于提高谱聚类算法的效率和鲁棒性的研究	2019年04月 - 至今
1. 大规模谱聚类与集成学习算法 (准备发表): 提出新的分而治之的kmeans算法进行典型性数据采样, 改进近似估算数据 KNN稀疏矩阵, 简化二项图分割计算公式, 较现有解决方案加速10~120%并保持高精度。	
2. 谱聚类集成学习算法 (数据挖掘顶会ICDM 2020, 接受率19.7%): 定位当前集成聚类算法中的误差来源于基础聚类的随机性, 引进对同一数据的多个谱嵌入, 进行对拉普拉斯矩阵一致性优化, 较现有算法提高了3~20%的精度。	
3. 对大规模的Nystrom谱聚类的采样算法 (IJCNN 2020): 创新利用局部Hubness值进行快速采样, 提高了1~5%的精度。	
硕士研究: 机器学习算法在多种场景中的优化	2017年04月 - 2019年03月
1. 对不平衡数据过采样方法框架 (期刊Neurocomputing): 设计开发基于拉普拉斯特征映射的过采样算法, 提升了对分类器的AUC达0.1~0.4。	
2. 二次复发的癌症预测 (期刊IJMS): 设计和开发结合过采样的集成特征选择算法, 提高了对二次复发的癌症预测AUC超过0.2。	
3. 保护隐私的分布式特征选择算法 (AI顶会IJCAI2019, 接收率 19.7%): 开发基于中间表现的分布式算法, 在执行算法时的保护隐私, 并对原始数据进行特征选择。负责公式验证, 代码实现和实验框架的开发。指出并协助修正了原理上隐私泄露问题。	
4. 大规模的谱聚类算法 (CBDCOM 2018, 最佳论文奖): 通过数据间拓扑关系实现采样构造稀疏特征矩阵, 较当前方案提升精度达2~10%。	

科研项目经历

"AI+" 合作研究 - 研究助理: 负责主要开发工作, 同时指导和培养组员跟进项目。	2018年10月 - 至今
1. AI+ 肾脏病数据	
◦ 对肾脏病人诊断数据进行数据清洗、缺失值补全、归一化操作, 简历逻辑回归模型预测了未来是否恶化, 评定特征重要度, 推导指标以指示未来恶化可能性。目前的三变量指标的AUC已达到0.7以上。	
◦ 因数据显示控制组 (介入式治疗) 和对照组在病情发展上不能符合统计学有差异假设, 研究一度停滞。通过统计检验查明控制组和对照组在特征重要度上存在较大差异, 做出有多个人群子集存在的假设, 初步确定了介入式治疗后好转的人群特点。	
2. AI+ 基因数据	
◦ 使用Linux生物学软件Plink对全基因数据进行预处理, 应用UAMP和t-SNE等降维算法对数据进行数据可视化, 成功展示出不同人种间的差异, 实现人口分层。	
◦ 使用R语言的生物信息的库Seurat对患癌症的小鼠血细胞RNA-seq数据进行质量控制、降维、可视化、聚类等操作, 成功识别出不同种类的细胞, 并观察到癌症组织内的NK细胞明显增加的现象。	
3. AI+ 零售数据	
◦ 利用滑动窗口将时序预测转换为回归问题, 进行特征工程, 分别建立Lasso, LSTM和LightGBM模型进行预测, 其日均误差百分比2%低于该企业委托微软开发的预测模型。	

技能

- **掌握技术:** Matlab (精通, 论文1~7)、Python (熟练, 项目1、2, 竞赛获奖1、2)、R语言, Linux。Numpy、Pandas、Sklearn和TensorFlow等框架经验。
- **熟悉算法:** 熟悉机器学习、聚类、特征提取、稀疏学习、维度降低、核方法、谱方法、特征提取等算法。
- **语言:** 普通话 (母语), 英语 (TOEIC 795, 3次国际会议发表), 日语能力二级 (学习环境为日语)。
- **兴趣爱好:** 中国哲学经典: 参加筑波大学的中国哲学读书会和东方哲学演习课, 并协助了《朱子语类》的日文译注。

获奖经历

1. 北京大学深圳研究生院主办 第一届AETA地震预测AI算法大赛 二等奖	2020年11月
2. 第三届易观算法大赛-PV、UV流量预测大赛 特别奖	2019年10月
3. 2018 IEEE CBDCOM 最佳论文奖	2018年9月

国际活动

G20部长级会议贸易和数字科学论坛 (学术海报展示)	2019年06月 - 2019年06月
展示自身研究内容, 与参会的20国官员以及科学家进行交流。	