



李鸿敏

17/10/1992

(+81)070-3668-3456

lihongmin@mma.cs.tsukuba.ac.jp

日本茨城县つくば市桜 2-26E 棟 105

教育背景

筑波大学 计算机科学 (博士) - 信息系统工程系 2019.04 - 2022.03

- GPA : 4.1 / 4.3
- 在校荣誉 : 筑波奖学金 (每月 80000 日元)。

筑波大学 计算机科学 (硕士) - 信息系统工程系 2017.04 - 2019.03

- GPA : 4.0 / 4.3
- 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 2011.09 - 2015.07

- 主修课程 : 线性代数, 信息论, 概率论与数理统计。

学术研究

博士阶段 - 研究课题 : 关于提高谱聚类算法的效率和鲁棒性的研究 2019.04 - 至今

研究内容 :

- 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 k -means 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。
- 谱聚类集成学习算法 : 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, 对拉普拉斯图进行一致性优化, 较现有算法提高了 3~20% 的精度。
- 对大规模的 Nystrom 谱聚类的采样算法 : 创新利用局部 Hubness 值进行快速采样, 提高了 1~5% 的精度。

论文成果 :

- 谱聚类集成学习算法成功发表在数据挖掘顶会 ICDM 2020, 接受率为 19.7% (2020.11.17) ;
- Nystrom 谱聚类的采样算法成功发表在 AI 国际会议 IJCNN 2020 (2020.07.19)。

硕士阶段 - 研究课题 : 机器学习算法在多种场景中的优化 2017.04 - 2019.03

研究内容 :

- 对不平衡数据过采样方法框架 : 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。
- 二次复发的癌症预测 : 负责算法开发和实验验证工作, 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。
- 保护隐私的分布式特征选择算法 : 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, 并对原本数据进行特征选择。负责公式验证, 代码实现和实验框架的开发, 指出并顺利修正了原理上隐私泄露问题。
- 大规模的谱聚类算法 : 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%。

论文成果 :

- 基于拉普拉斯特征映射的过采样算法框架发表在期刊 Neurocomputing (2020.07) ;
- 保护隐私的分布式特征选择算法成功发表在 AI 顶会 IJCAI2019, 接收率 19.7% (2019.08) ;
- 二次复发的癌症预测成功发表在国际医学期刊 International journal of medical sciences (2019.06) ;
- 大规模的谱聚类算法成功发表在 CBDCom 2018, 荣获最佳论文奖 (2018.10)。

科研经历

筑波大学 AI 中心的合作研究

2018.10 - 至今

研究助理

- **AI+肾脏病数据**：独立对肾脏病人的诊断数据进行数据清洗、缺失值补全、归一化操作，熟练运用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。
- **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。
- **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。

研讨会经历

- ✓ 国际学习表征会议(ICLR) 2021.05
- ✓ IEEE 数据挖掘国际会议(ICDM) 口头报告 2020.11
- ✓ 神经网络国际联合会议(IJCNN) 口头报告 2020.07
- ✓ 2019 智慧医疗研讨会 2019.09
- ✓ G20 贸易和数字经济部长级会议 “数字化科学” 联合国际研讨会 海报报告 2019.06
- ✓ IEEE 大数据与云计算国际会议(IEEE CBDCOM) 口头报告 2018.09

技能爱好

专业技能：精通 Matlab，Python，R 语言，Linux 等软件，及熟练掌握 Numpy、Pandas、Sklearn 和 TensorFlow 等框架构建。

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、谱方法等算法。

语言技能：中文（普通话流利）、英语（精通，TOEIC 795，3 次国际会议全英文发表）、日语（专业二级，可作为工作语言）。

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）。

参赛奖项

- ✓ 第一届 AETA 地震预测 AI 算法大赛二等奖（北京大学深圳研究生院主办） 2020.11
- ✓ 第三届易观算法大赛-PV、UV 流量预测大赛特别奖 2019.10
- ✓ IEEE CBDCOM 最佳论文奖 2018.09

自我评价

- **专业能力**：在机器学习和数据挖掘领域成功发表 6 篇的学术论文，两篇入选顶级会议，多次参加顶流专业学术研讨会，拥有丰富的科研项目经验，能独立进行项目开发等一系列工作，具有团队精神，能指导和培养团队成员跟进项目，进行相关研究计划的制定和实施；善于演讲展示，能进行全英文或日文的学术报告，曾受邀在 G20 峰会上学术海报展示并与参会的 20 国官员以及科学家进行交流，具有全面的专业素养和科研能力。
- **综合能力**：性格稳重，处事严谨，善于理解和沟通组织协调，具备团队协作能力及问题解决能力，有良好的职业道德、有较强的责任心和敬业精神，能承受工作压力；对自己已掌握的技术敢于创新，精益求精，能够妥善的处理周围的人际关系；学习能力和逻辑性强，在严密分析的基础上结论，在全面考虑的前提下决断。