Hongmin Li

Sakura 2-26 E Building 105 • Tsukuba • Ibaraki, 305-0003, Japan • li.hongmin.xa@alumni.tsukuba.ac.jp • 070-3668-3456

EDUCATION

University of Tsukuba Tsukuba, Japan

Ph.D., Computer Science. March 2022

- GPA: 4.1/4.3
- Tsukuba Scholarship recipient(80,000¥ per month)

M.A., Computer Science. March 2019

- GPA: 4.0/4.3
- Coursework in Machine Learning (A+), Evolution Algorithm (A +), Academic Writing (A).

University of Ningxia Yinchuan, China

B.A., Electronic Information Engineering June 2015

- Major Course: Linear Algebra, Information Theory, Probability Theory & Mathematical Statistics.

ACADEMIC RESEARCH EXPERIENCE

University of Tsukuba Tsukuba, Japan

PhD Researcher April 2019 – Now

Research Subject: Research on Improving the Efficiency and Robustness of Spectral Clustering Algorithm

- **Large-scale Spectral Clustering and Ensemble Algorithm (ready for publication):** Proposed a divide-and-conquer kmeans algorithm for representative data sampling, improved KNN sparse matrix for approximate estimation data, simplified binomial graph partitioning formulation, accelerated 10~120% compared with the current scheme and maintained high accuracy.
- **New Ensemble Clustering Framework:** Simultaneously optimized the base embeddings and consensus one to avoid the misclustering from -means, performed the optimization of Laplacian matrix consistency, and improved 3~20% accuracy.
- **A Quick Local Hubness Sampling Method for Nystrom Spectral Clustering:** Innovatively used the local Hubness values for fast sampling, and improved 1~5% clustering accuracy in our experiment.

Research Results:

- Spectral clustering and ensemble algorithm successfully published in the International Conference on Data Mining (ICDM), accepted rate 19.7% (17/11/2020).
- The sampling algorithm for Nystrom spectral clustering was successfully published in International Joint Conference on Neural Networks (IJCNN) (19/07/2020).

Master's Researcher April 2017– March 2019

Research Subject: Research on Optimization of Machine Learning Algorithms in Multiple Scenarios

- **An Oversampling Framework for Imbalanced Classification based on Laplacian Eigenmaps:** Designed and developed the oversampling algorithm based on Laplacian Eigenmap, improved 0.1~0.4 AUC of classifiers.
- **Ensemble Feature Learning to Identify Risk Factors for Predicting Secondary Cancer:** Responsible for algorithm development and experimental validation work, developed an ensemble feature learning framework and brought more than 0.2 AUC.
- **Distributed Collaborative Feature Selection Based on Intermediate Representation:** Developed distributed algorithm based on intermediate representation, protected privacy during algorithm execution, and feature selection of original data. Responsible for formula validation, code implementation and experimental framework bulding, pointed out and successfully corrected the privacy leakage problems in principle.
- **Large Scale Spectral Clustering Using Sparse Representation Based on Hubness:** Developed a sparse representative matrix by sampling through the topological relationship between the data, and the accuracy is improved by 2~10% compared with the current scheme.

Hongmin Li

Research Results:

- An Oversampling Framework for Imbalanced Classification based on Laplacian Eigenmaps was published in Neurocomputing (25/07/2020).
- Distributed Collaborative Feature Selection Based on Intermediate Representation was published in IJCAI2019, accepted rate 19.7% (08/2019).
- Ensemble Feature Learning to Identify Risk Factors for Predicting Secondary Cancer was successfully published in the International Journal of Medical Sciences (06/2019).
- Large Scale Spectral Clustering Using Sparse Representation Based on Hubness was successfully published in CBDCom 2018, with the Best Paper Award (08/10/2018.10.08).

TEAMWORK EXPERIENCE

University of Tsukuba, Center for Artificial Intelligence Research Tsukuba, Japan

Research assistant for Interdisciplinary Cooperation October 2018 – Now

- **AI+Nephrology Data:** Independently performed data cleaning, missing value completion, and normalization operations on diagnostic data of kidney patients, skillfully used logistic regression models to predict whether future deterioration, rated the importance of features, and derived indicators to predict the likelihood of future deterioration.
- **AI + Genetic Data:** Used Linux Plink to pre-process the whole genetic data, and applied dimensionality reduction algorithms such as UAMP and t-SNE to visualize the data, and successfully demonstrated the differences between different ethnic groups to achieve population stratification. Using Seurat, a library of bioinformatics in R language, to perform quality control, dimensionality reduction, visualization, and clustering on RNA-seq data from blood cells of mice suffering from cancer, successfully identified different kinds of cells and observed a significant increase of NK cells within cancer tissues.
- **AI+Retail Data:** Usedsliding windows to convert time-series forecasts into regression problems, Lasso, LSTM and LightGBM models are built for forecasting respectively, and their average daily error values of 2% were lower than those of the forecasting models developed by Microsoft commissioned by enterprises.

SEMINAR PRESENTATION EXPERIENCE

- 2020 IEEE International Conference on Data Mining (ICDM) 11/2020
- 2020 International Joint Conference on Neural Networks (IJCNN) 07/2020
- 2019 Quanta Smart Medicine Symposium 09/2019
- International Symposium on “Digital Science Now” in association with the G20 Ministerial Meeting on Trade and Digital Economy 06/2019
- 2018 IEEE Cloud & Big Data Computing (IEEE CBDCom) 09/2018
- English Presentations: 2 oral presentations in international conference, 1 post presentation in the meeting of G20 Trade and Digital Economy Ministers.

SKILLS & INTERESTS

Technical: Proficient in programming with MATLAB, Python, R; using Linux, Numpy, Pandas, Sklearn and TensorFlow.

Language: Chinese (Native), English (TOEIC 795), Japanese (N2)

Algorithms: Familiar with machine learning, Clustering, Feature Extraction, Sparse Learning, Dimensionality Reduction, Kernel Methods, Spectral Methods, etc.

Interests: Ancient Chinese history, Ancient Chinese book (participated in Chinese philosophy reading group and Oriental philosophy exercise class at Tsukuba University, and assisted in the translation and annotation of Zhu Zi Yixue in Japanese.)

AWARDS of AI COMPETITION

- The second prize of 1st AETA Earthquake Prediction AI Algorithm Competition (2020)
- Special award in the 3rd Analysys International Algorithm Competition -PV, UV Prediction Competition (2019)
- IEEE CBDCom Best Paper Award (2018)