Chenghui Li

cli539@wisc.edu · (+1) 608-358-2030 · in chenghui-li-a9a29817b

Education

University of Wisconsin-Madison, Wisconsin, The United States — Aug. 2018 – Present

M.S. in Data Science; Expected to May. 2020; *GPA: 4.00/4.00*

Zhejiang University, Zhejiang, China — 2015 – 2019

B.S. in Mathematics and Applied Mathematics; *GPA: 3.77/4.00; Major GPA: 3.95/4.00*

Projects & Experience

Consistency of an algorithm for manifold clustering — Jan. 2019 – Present

Programming, Algorithm design and Theoretical proof Professor: Nicolas Garcia Trillos

- **Brief Introduction:** We proposed a technique in recovering a low-dimensional representation of the data coming from several intersected structures. Therefore, the problem was multi-manifold cultering, which aimed to label each data point according to the surface it came from.
- **Outcome**: We proposed a methodology for manifold clustering for which we can develop rigorous mathematical theory establishing its consistency, by using ideas from geometry, probability and calculus of variations.
- **Keywords:** Manifold Clustering, Discrete to continuum limit, Gamma-convergence, Point cloud, Dijkstra Algorithm

FFI Algorithm Performance in bin-packing MinSum Problem — Feb. 2018 – June 2019

Theoretical proof Professor: Zhiyi Tan

- **Brief Introduction:** MinSum problem is NP hard, and the heuristic algorithm FFI was not studied well. The past upper bound was 1.83, and the lower bound was 1.25. There was a big gap between the bounds.
- **Outcome**: I used the construction method to improve the limiting upper bound to 1.5 and the lower bound to 1.35.
- **Keywords:** Combinatorial Optimization, bin packing, MinSum Problem, Limiting bound, FFI Algorithm

Lagrangian Flux Calculation Via Donating Volumes In 3 Dimension — Oct. 2017 – Dec. 2017

Directed Study: Multigrid Method on Irregular Domains Professor: Qinghai Zhang

- **Brief Introduction:** When it came to donating volumes, it might be inconvenient to compute the lagrangian flux. So, the scalar conservation laws for donating volumes should be necessary.
- **Outcome**: I found the generalization problem's gaps in the proof and then fixed them.
- **Keywords:** Flux set, Generalized donating regions, Streakline, Winding number, Hopf theorem, Cycle decomposition

Yelp Business Analysis — Oct. 2019 – Nov. 2019

STAT 628: Data science Practicum, Course project Professor: Hyunseung Kang

- **Brief Introduction**: We need use millions of data with serious missing problem to give advice to business in order to improve their scores on Yelp.
- **Repsonsibility**: I used NLP to split texts into different parts, used LSTM to predict, then combined the statistics tests and prediction to give the feedback. Finally, I used Shiny with html to present real-time response.

High Dimensional Simulation in SPCA, PCA, LASSO and PLS — Apr. 2019 – May 2019

Stat 471: Introduction to computational statistics, Course project Professor: Fangfang Wang

- **Brief Introduction**: SPCA is a generalization of PCA, and we tried to figure out how its performance on multicorrelated data comparing with other algorithms.
- **Responsibility**: I used Latent model to give some solid theory on SPCA, and used R to generate the toy data, program SPCA algorithm and then test the algorithm.

Modeling for hematopoiesis prediction Dec. 2018

STAT 601: Statistical Inference, Course project Professor: Zhengjun Zhang

- **Brief Introduction**: Hematopoiese cellular data that we used had a serious multicollinearity issue, so we need to portray its nonlinear part. GMC can be used as a complement of linearity, so we also tested GMC behavior in selection of models.
- **Modeling**: Fitted linear models using LASSO, PCA to reduce data dimension that was capable for high-dimensional data, then used GMC to evaluate the model. We selected the principal components or genes that contributed most to the response and interpret. Then, selected best models using different standards.

Artificial Intelligence and Industry Program in Oxford Aug. 2018

- **Brief Introduction**: This is a two-week program to go to G5 universities to meet with professor and listen to their lectures about artificial intelligence.

International Teaching Volunteer in Indonesia Feb. 2017

Teaching Volunteer in Yunnan province in China$\times 2$ June 2016 & June 2017

♡ Honors and Awards

- Visiting International Student Academic Excellence Award $\times 2$ at Madison — Mar. 2019 & July 2019
- Chu Kochen Honors Program at Zhejiang University — June 2019
- Qiushi Pursuit Science Class(Major in Mathematics) at Zhejiang University — June 2019
- Putnam Exam Top 200 — 2019
- First Class Scholarship for Elite Students in Basic Science at Zhejiang University — Jan. 2017

⚙ Skills

- **Programming Languages**: Proficient in R, Python and Matlab, experienced in Latex, C and HTML
- **Platform**: Linux
- **Development**: SQL

ℹ Miscellaneous

- **GitHub**: https://github.com/chl781
- **Personal website**: https://chl781.github.io/
- **Relevant courses**: Mathematical statistics, Real Variable analysis, Topics in probability, Multi-variate statistical analysis, Statistical Methods, Statistical learning theory, Optimal Transport Machine Learning, Stochastic Processes, Regression analysis, Point set topology