

WANG Xiasi

☎ +86 15156017806 • ✉ wxs119226@gmail.com

Education

HongKong University of Science and Technology

PhD in Data Science and Analytics

2020.09 – 2024.06

GPA: 3.931/4.30

Related Courses: Statistical Machine Learning, Deep Learning, Data Mining

Research interests: Machine Learning, Learning with Noisy Label, Self-Supervised Learning

University of Science and Technology of China

Bachelor of Science in Statistics

2016.09 – 2020.06

GPA: 3.45/4.30

Related Courses: Probability, Mathematical Statistics, Regression Analysis, Multivariate Analysis, Time Series Analysis, Real Analysis, Functional Analysis, Statistical Software, Mathematical Modeling

Research Projects

Research on learning with noisy label data

Guide: Prof. Yuan Yao, Department of Math, HKUST

2021.09 – Present

- Systematically investigate literature about learning with noisy label.
- Propose an algorithm dubbed as S3C: Self-Screening and Self-Correction for noisy label data. S3C adopts an iterative idea to train the classifier and select a new training dataset with higher quality alternatively. Discover a problem worth of noting: hard samples can be confused with noisy samples since they share similar characteristics such as large losses during training.
- Algorithm achieves good performance on artificially created noisy CIFAR10 dataset.
- Discover new problems after analyzing: Noisy label hampers the model training at the very beginning. However, most of existing methods need a warm up training stage which uses the whole training set regardless of their labels' correctness.
- Propose to use self-supervised learning, more specifically, contrastive learning for the warm up stage of model training.

Nexperia semi-conductor image classification

Guide: Prof. Yuan Yao, Department of Math, HKUST

2021.03 – Present

- Explore the industrial data and discover problems – class imbalance and noisy label.
- Focus on noisy label problem. Try some methods such as “self-adaptive training”, “robust loss function”, and “Mix-up”'s methods on our industrial data.

Research on systemic risk contagion in financial networks

Guide: Prof. Yu Chen, Department of Statistics and Finance, USTC

2019.07 – 2020.01

- Learn a risk contagion model for financial networks which triggers the contagion before default.
- Construct a liability matrix of a Chinese bank network by assigning prior distributions to banks' connection probabilities and weights of their connections for the Gibbs Sampler.
- Explore the model's sensitivity to its parameters. Impose stress tests to the Chinese bank network and discover the main causes of changes of contagion losses of the past ten years.
- **Publication:** Solvency contagion risk in the Chinese commercial bank network (*Physica A*, 2021).

Honors and Scholarships

- Full Postgraduate Scholarship, HKUST 2020 – 2024
- Outstanding graduate student, USTC 2020
- Outstanding Undergraduate Scholarship, USTC 2017, 2018, 2019

Skills

- Computer Languages & Tools: Python, R, C, MATLAB, Mathematica, \LaTeX
- English: TOEFL: 98 (R27 L22 S23 W26) GRE: 322+4.0 (V152 Q170)

教育经历

香港科技大学

Ph.D. | 数据科学 Data Science

2020.09 – 2024.06

GPA: 3.931/4.30

相关课程: 统计机器学习, 深度学习, 数据挖掘

研究兴趣: 机器学习, 噪声标签, 自监督学习

中国科学技术大学

B.S. | 统计学 Statistics

2016.09 – 2020.06

GPA: 3.45/4.30

相关课程: 概率论, 数理统计, 回归分析, 多元统计分析, 时间序列分析, 实分析, 泛函分析, 统计计算软件, 数学建模

研究经历

深度学习中的噪声标签研究

指导老师: 姚远, 香港科技大学数学系

2021.09 – 现在

- 系统研究了“噪声标签”相关的文献
- 提出了一个算法 S3C: 自筛选和自改正模型。S3C 采用一种迭代训练的思想: 交替地训练分类器和筛选并且更正数据标签得到一个数据质量更高的训练集。关注了一个新问题: 在训练过程中, 噪声数据和难样本难以分开, 因为两者具有一些相同的特性, 比如训练过程中有较大的训练损失
- 算法在有噪声标签的 CIFAR10 取得的不错的效果, 能够有效提高模型抵抗噪声标签的能力
- 在对算法结果进行仔细分析后发现了新的问题: 噪声标签会在模型训练的最开始就影响模型的训练。然而目前绝大多数的噪声标签学习算法都需要一个“warm up”阶段, 这一阶段是使用全部数据集来进行训练的
- 提出可以使用自监督学习 (对比学习) 来代替一般算法的“warm up”阶段

Nexperia 半导体图像分类项目

指导老师: 姚远, 香港科技大学数学系

2021.03 – 现在

- 仔细分析了数据并且发现问题 – 类别不平衡和噪声标签
- 关注于噪声标签问题。在 Nexperia 数据集上尝试了“SAT”, 鲁棒损失函数, mixup, manifold mixup 和 patch up 等方法
- 代码: <https://github.com/wangdaha1/Nexperia>

金融网络系统性风险研究

指导老师: 陈昱, 中国科学技术大学统计与金融系

2019.07 – 2020.01

- 研究了一个在金融机构违约前触发风险传播机制的模型
- 通过给银行间的借贷连接概率和分布指定先验分布从而创建银行间借贷矩阵网络, 并且用吉布斯取样的方法生成样本矩阵用于实证分析
- 在压力测试中研究了模型关于参数的敏感性, 创建移植资产表发现造成传播风险变化的原因
- 发表作品: Solvency contagion risk in the Chinese commercial bank network (*Physica A*, 2021)

获奖情况

- 全额博士生奖学金, 香港科技大学 2020 – 2024
- 优秀毕业生, 中国科学技术大学 2020
- 优秀本科生学业奖学金, 中国科学技术大学 2017, 2018, 2019

专业技能

- 计算机语言: Python, R, C, MATLAB, Mathematica, L^AT_EX
- 英语: TOEFL: 98 (R27 L22 S23 W26) GRE: 322+4.0 (V152 Q170)