

胡正阳

香港大学 · 博士研究生

u3010250@connect.hku.hk · +852 5493 3801 · datou30.github.io

研究方向：表格数据基础模型预训练 · 非线性相关性建模 · 大语言模型推理

教育背景

香港大学

2023.08—至今

电气与计算机工程 · 博士研究生 · GPA: 3.9 / 4.3

导师: Yanchao Yang 博士、Kaibin Huang 教授

中国科学技术大学 · 专业 GPA: 3.87 / 4.3

2019.09—2023.06

数学与应用数学 · 理学学士 · 少年班学院

科研与业界经历

明汭投资

— 机器学习量化研究实习生

2026.08—2026.10

- 研究依赖性度量能否作为训练损失。在固定 MLP 架构下预测 A 股未来五个交易日收益，设计结合依赖性损失的训练后重拟合策略，将**测试集平均横截面 IC 从 0.105 提升至 0.12**；对照为 MSE 基线，两组均采用 10 个随机种子的模型集成。

华为高性能计算团队 (RAFT)

— 外部驻场学生研究员

2024.09—2026.01

- 开发基于表征的模型训练诊断方法，并在 NanoGPT、ViT 和 Pythia 上验证；相关研究已投稿 ICLR。

诚奇资产

— 量化研究实习生

2022.08—2022.11

- 基于多类数据源研究 Alpha 策略，**5 项策略**被纳入公司 Alpha 策略库。

代表性研究与论文

1. InfoNet: Neural Estimation of Mutual Information without Test-Time Optimization

Z. Hu et al. · ICML 2024 (Oral)

开发用于随机变量间依赖性估计的表格基础模型，在低维基准测试中实现 **1,000 倍**加速，并取得更高估计精度。下游任务涵盖特征选择、图像分割和关键帧提取。

2. InfoAtlas: A Foundation Model for Zero-Shot Statistical Dependence Estimate

Z. Hu et al. · ICML 2026

通过更丰富的合成数据与模型架构设计，将 InfoNet 扩展至高维统计依赖估计。搭建数据生成与分布式预训练流程，用于在 **32 张 H200 GPU** 上训练 **6 亿参数模型**。

3. From Anisotropy to Anomaly: An Online Geometric Diagnostics Framework for Transformer Training

Z. Hu et al. · ICLR 审稿中

基于隐藏表征间的统计依赖关系构建在线诊断方法，可用来识别 Transformer 训练中的异常，包括过拟合、数据异常和损失突增。

4. DensityFM: Foundation-Style Pretraining for Density Estimation

Z. Hu et al. · ICLR 审稿中

预训练用于密度估计的表格基础模型，在合成与真实数据基准上优于所有已评估的经典基线方法。应用包括尾部风险估计与异常值检测。

补充经历

Optiver Quant Focus — 为期三天的博士生项目。

2026.09

团队获 Quant Challenge 第一名；个人独立解出全部 12 道数学与智力题中的 **5 道**。

Susquehanna (SIG) PhD Brainteaser Battle

2026.10

团队获第二名；个人独立解出赛事中最难的谜题。

考试成绩: TOEFL iBT 105 · GRE 322 (数学 170) | 技能: Python · PyTorch · pandas