

何智星

电话: 17710177926 邮箱: hzx957@sina.com
微信: hezhixing957 主页: [hzxx957.github.io](https://github.com/hzxx957)



工作经历

| | | |
|------|-------|-------------------|
| 百度 | 研发工程师 | 2021.10 – 至今 |
| 深兰科技 | 算法工程师 | 2020.09 – 2021.10 |

教育背景

| | | |
|------------|------|-------------------|
| 美国弗吉尼亚理工大学 | 物理博士 | 2013.08 – 2020.05 |
| 浙江大学 | 物理学士 | 2009.08 – 2013.05 |

项目成果

- 百度搜索垂类扶持的开发** 2021.10 – 至今
- 负责百度自然搜索小说垂类的排序的算法研发和业务拓展。小说项目用户覆盖广，浏览量高，业务重要而复杂。
 - 通过 NLP 模型 Ernie 实现小说的意图，领域分类；通过树模型实现统一的小说资源排序；通过章节拉链等自研算法实现高时效性的业务场景。
 - 对复杂业务下框架同时有整体把控和细节钻研。包括各个网站的内容抓取解析，大数据处理和分析，算法召回和排序等。
- 数据挖掘方向自动化机器学习平台的研发** 2020.09 – 2021.10
- 完成分类、回归、时序预测等数据挖掘方向的自动化机器学习 Pipeline 的开发建设
 - 研发了高灵活性的自动特征生成、模型选择和超参搜索模块，优化集成学习算法 (Bagging, Stacking)
 - 在结构化数据上引入深度学习模型 (TabTransformer)，在二分类任务上 AUC 平均提升 3%
 - 总结时序结构化数据 AutoML 开发和研究经验，撰写论文被 SIGKDD2021 录用^[1]
- 开发高泛化性时序异常检测的框架** 2021.04 – 2021.08
- 利用自研残差融合策略，集成 TCN、VAE、FFT、Matrix Profile 等多种算法，准确率达到 88%
 - 通过优化样本生成策略，解决回归算法的检测偏移问题
 - 获得 KDD CUP 2021 Multi-dataset Time Series Anomaly Detection 冠军

相关技能

编程工具: Python, C++, Bash, Docker, Git, Flask, Scikit-learn, PyTorch
机器学习/深度学习: GBDT, LR, SVM, CNN, LSTM, VAE, Transformer, Ernie, GBrank
软技能: 熟练的英语口语和写作，良好的沟通能力，富有团队合作精神和创造力

竞赛/论文

冠军: *KDD Cup 2021 Multi-dataset time series anomaly detection* -- https://youtu.be/1v64_tzJPTE
冠军: *SIGIR 2021 eCom'21 Competition* -- <https://sigir-ecom.github.io/data-task.html>
季军: *PKDD 2021 Ariel ML Data Challenge* -- <https://www.ariel-datachallenge.space>

[1] AutoSmart: An Efficient and Automatic Machine Learning Framework for Temporal Relational Data, SIGKDD 2021
[2] Interferometric spectroscopy and high-speed orientation detection of individual gold nanorods, Nanoscale 2020
[3] Assembly and disorder dissipation in superparamagnetic nanoparticle chains in a rotating magnetic field, arXiv2020