

# 苏云祥

手机：(+86) 18811382629  
邮箱：suyx21@mails.tsinghua.edu.cn



## 教育背景

清华大学，软件工程，在读博士研究生	2021/08 - 2026/06
清华大学，数理基础科学，理学学士	2017/08 - 2021/06

## 研究方向

主要研究方向为 DB4AI，探索数据库能力驱动的机器学习，尤其关注时间序列数据库的存储特性，从而提升数据处理和机器学习效率。相关成果以独立第一作者身份在数据库领域 A 类会议 SIGMOD、KDD、VLDB 上发表论文 4 篇，并在开源时间序列数据库 Apache IoTDB 中实现系统部署和应用验证。

## 研究成果

- (1) **Yunxiang Su**, Yikun Gong, Shaoxu Song. Time Series Data Validity. SIGMOD 2023. (独立一作, CCF-A 类会议)
  - 贡献：首次定义时序数据的有效性，并提出度量方法，基于时序数据库存储特性设计高效算法。
  - 效果：相比现有关系型数据有效性度量，取得更准确的度量结果，同时显著降低计算时间开销。
- (2) **Yunxiang Su**, Wenxuan Ma, Shaoxu Song. Learning Autoregressive Model in LSM-Tree based Store. KDD 2023. (独立一作, CCF-A 类会议)
  - 贡献：针对时序数据库内机器学习，改进传统自回归模型，通过时序数据库预计算信息加速学习。
  - 效果：保证模型训练结果完全一致的同时，显著降低时序数据库内模型训练的时间开销。
- (3) **Yunxiang Su**, Shaoxu Song, Xiangdong Huang, Chen Wang, Jianmin Wang: Distance-based Outlier Query Optimization in Apache IoTDB. VLDB 2024. (独立一作, CCF-A 类会议)
  - 贡献：提出时序数据库内的基于距离的异常检测算法，实现数据库内的高效异常检测。
  - 效果：相比其他先进基线方法，在结果相同的情况下，时间开销降低一个数量级。
- (4) **Yunxiang Su**, Kenny Ye Liang, Shaoxu Song: In-Database Time Series Clustering. SIGMOD 2025. (独立一作, CCF-A 类会议)
  - 贡献：提出基于形状的高效时序数据聚类方法，能够高效处理长序列，并针对数据库内优化。
  - 效果：相比其他基于形状的基线方法，取得显著准确的聚类结果，同时显著降低计算时间开销。
- (5) Kenny Ye Liang, **Yunxiang Su**, Shaoxu Song, Chunping Li: Turn Waste Into Wealth: On Efficient Clustering and Cleaning Over Dirty Data. TKDE 2025. (第二作者, CCF-A 类期刊)

## 获奖情况

(1) 国家奖学金	2023/11
(2) 清华大学综合奖学金（一等）	2024/11

## 工程项目成果

---

- (1) 国家重点研发计划：制造大数据驱动的预测运行与精准服务技术及系统 2021/03 - 2023/11
- (2) 中冶赛迪信息技术（重庆）有限公司：时序数据库及数据质量研究及开发 2023/05 - 2024/11
- (3) Apache IoTDB： *Apache* 顶级开源物联网数据库项目 2020/12 - 2024/11
- (4) IoTDB-Quality： *Apache* 顶级开源项目 *IoTDB* 下二级项目 2020/12 - 至今