

# 云计算与数据科学研究所

**团队负责人：孙永奇**

交通大数据与人工智能教育部重点实验室

计算机科学与技术学院 北京交通大学



# 主要内容



**第一部分** 研究所概况

**第二部分** 主要成员介绍

**第三部分** 主要科研项目

**第四部分** 优秀研究生成果

# 1

## 研究所概況

- **中国工程院院士、** 顾问教授1人；  
教授3人，副教授1人，讲师1人；  
在读硕士、博士研究生50余名。
- 研究**云计算与云原生、分布式人工智能、数据挖掘与知识图谱、深度学习与强化学习**等理论及应用技术。
- 负责北京市**大数据人才培养**示范基地的建设。
- 负责**BJTU-AWS**云创学院的建设。
- 负责《数据结构》、《大数据技术》、《并行与分布式计算》、《计算智能》等课程建设

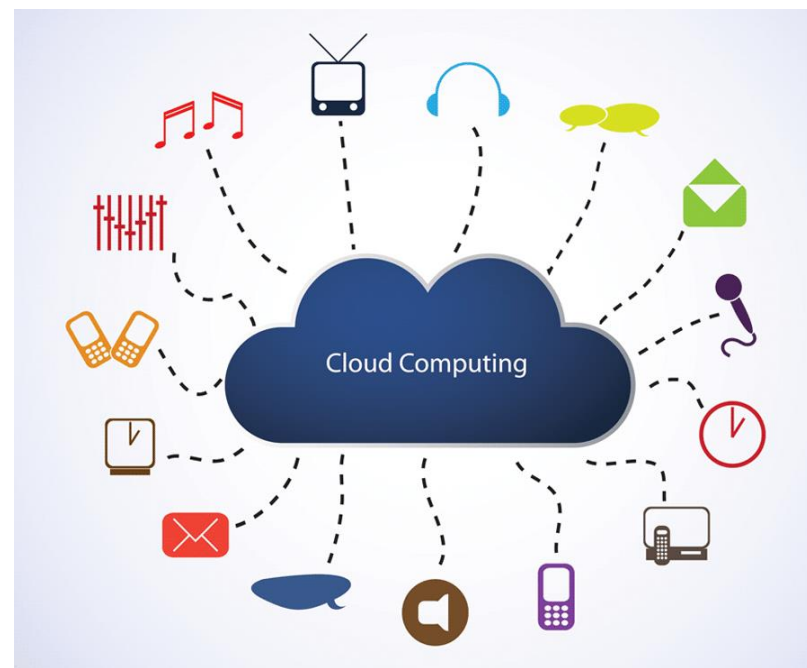
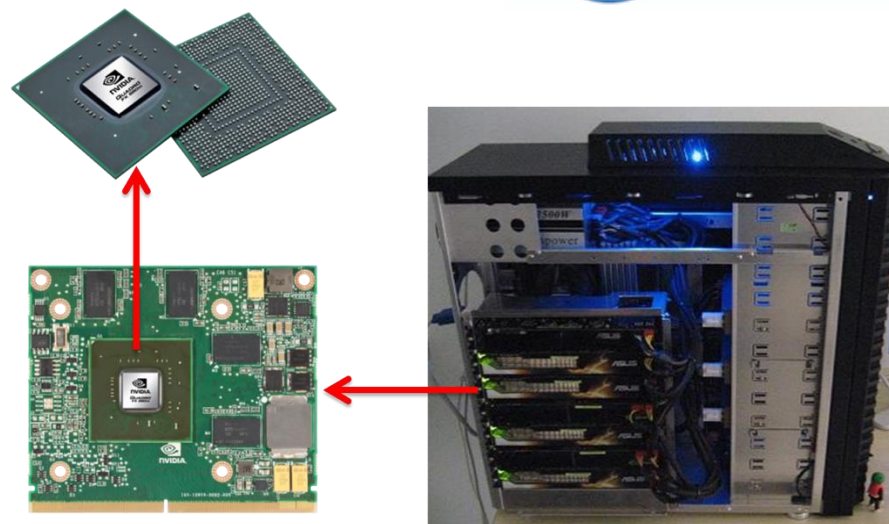
# 云计算与分布式计算



北京交通大学

高性能计算的范围很广，包括：

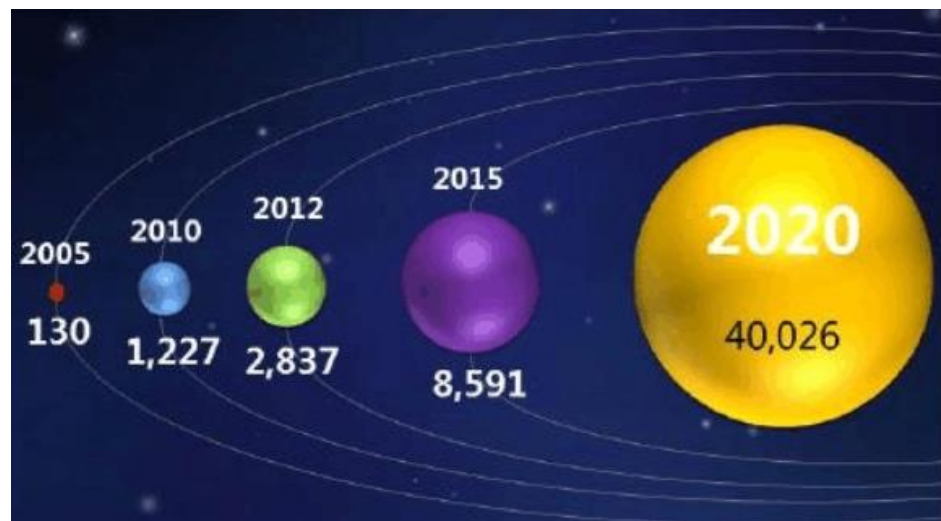
- 并行计算（大型机，MPI）
- 多核计算（OpenMP）
- 分布式计算（**云计算**）
- 基于GPU的并行计算（CUDA架构）
- 众核计算（混合架构， OpenMP+ MPI）



# 数据的发展

数据正以每年超过 50%的速度爆发式增长。IDC 最新的研究报告显示，到 2020 年，全球的数据资料存储量将超过 40ZB。

大数据的特点(4V): 量大 (Volume)、种类多 (Variety)、产生和处理速度快 (Velocity)、价值密度低 (Value)



# 数据的发展



股票交易



卫星勘测



超市交易

**时间序列数据**



实时车载信息



Internet通信



大数据为体  
云计算为术

# 2

## 主要成员介绍

# 主要成员介绍

## ■ 赵春江 院士

- 研究所学术带头人，北京交通大学顾问教授
- 中国工程院院士
- 研究方向：
  - 人工智能
  - 数据挖掘与数据库技术
  - 智能化专家系统
  - 模型与决策支持系统



# 主要成员介绍

## ■ 王志海 教授

- 博导，北京市教学名师

- 研究方向：  
数据挖掘与商务智能  
数据库与数据仓库  
机器学习与计算智能



➤ zhhwang@bjtu.edu.cn

## ■ 孙永奇 教授

- 博导，研究所所长，人工智能专业责任教授

- 研究方向：  
大数据与云计算  
人工智能  
图表示学习  
联邦学习



➤ yqsun@bjtu.edu.cn

# 主要成员介绍

## ■ 段莉 教授

- 博导，国家级青年人才



- 研究方向：

区块链：区块链安全及其应用、跨链安全等

隐私计算：联邦学习、安全多方计算等

服务安全与数据隐私

➤ [duanli@bjtu.edu.cn](mailto:duanli@bjtu.edu.cn)

## ■ 原继东 副教授

- 博导，研究所副所长
- 北京市优秀人才
- 毕业生“我最敬爱的老师”



- 研究方向：

时间序列数据挖掘

知识图谱

时间序列模型对抗攻击

深度学习、强化学习

➤ [yuanjd@bjtu.edu.cn](mailto:yuanjd@bjtu.edu.cn)

## ■ 刘海洋 讲师

- 博士，硕导
- 毕业生“我最敬爱的老师”



- 研究方向：

时间序列数据挖掘

多标记学习

图数据挖掘

联邦学习

➤ [haiyangliu@bjtu.edu.cn](mailto:haiyangliu@bjtu.edu.cn)

# 3

## 主要科研项目

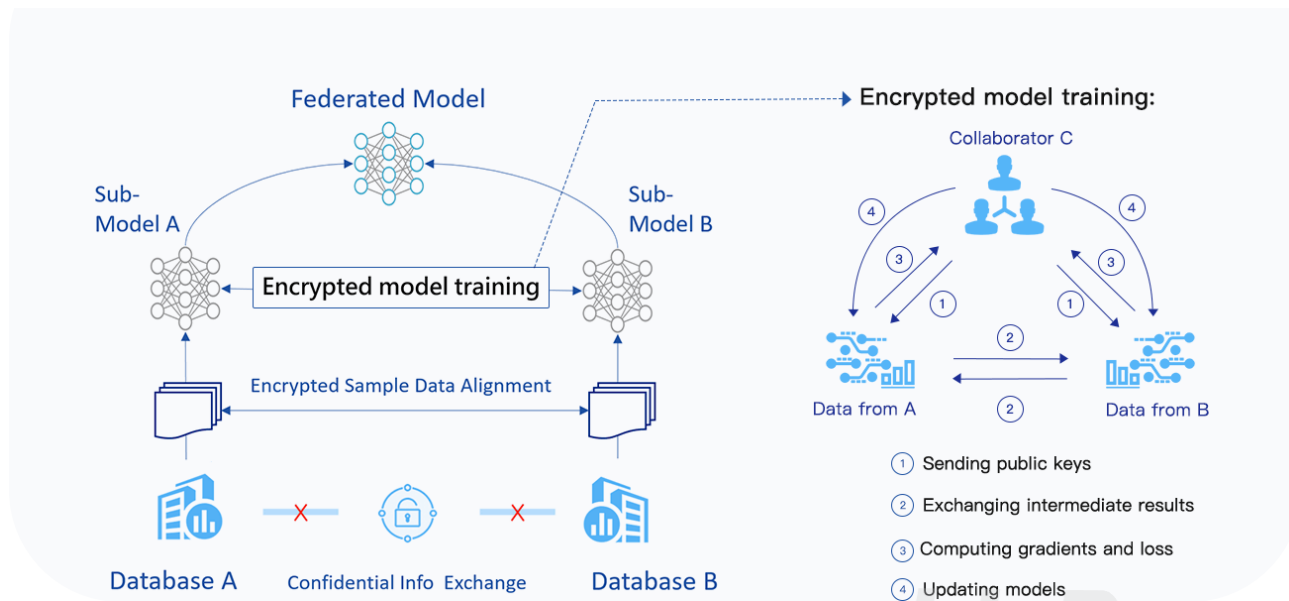
# 3.1-分布式人工智能与联邦学习

(孙永奇、刘海洋、原继东)

- 个性化联邦学习
- 联邦推荐系统
- 联邦学习节点大规模扩展技术
- 自适应、高效的隐私保护技术
- 落地平台：PaddlePaddle（百度）、MindSpore（华为）、Tensorflow、PyTorch

## ➤ 联邦学习系统架构

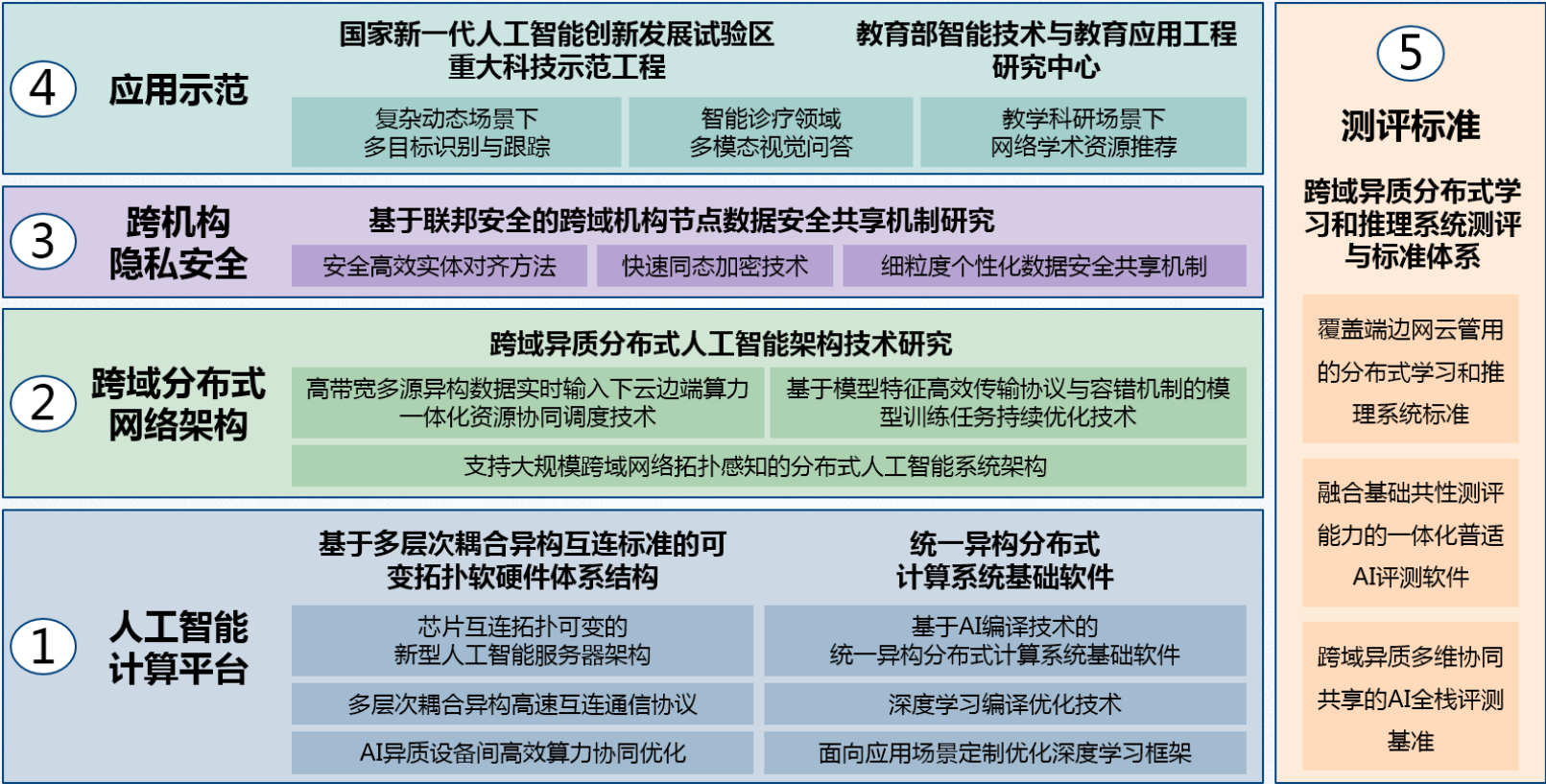
- 联邦学习：分布式计算+深度学习/机器学习，分横向、纵向、迁移三种
- 解决数据孤岛问题、保护用户的数据隐私



# 3.1-分布式人工智能与联邦学习

## ➤ 跨域异质分布式学习和推理系统（与浪潮、百度合作）

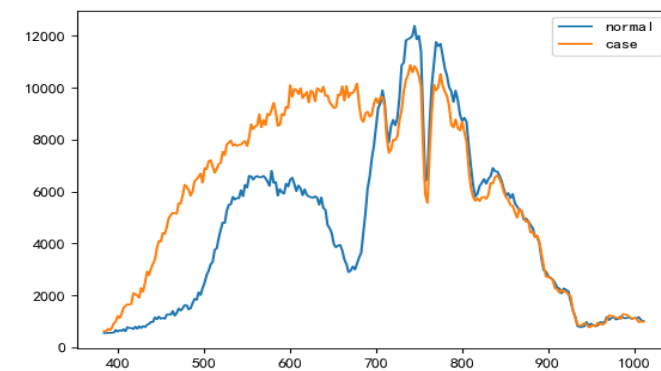
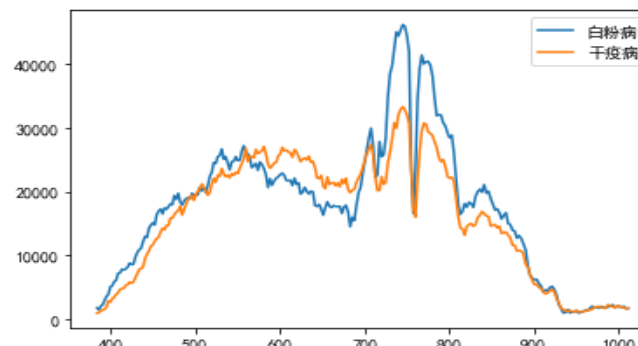
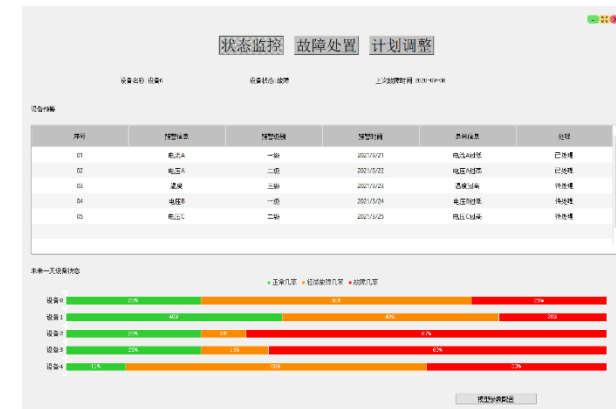
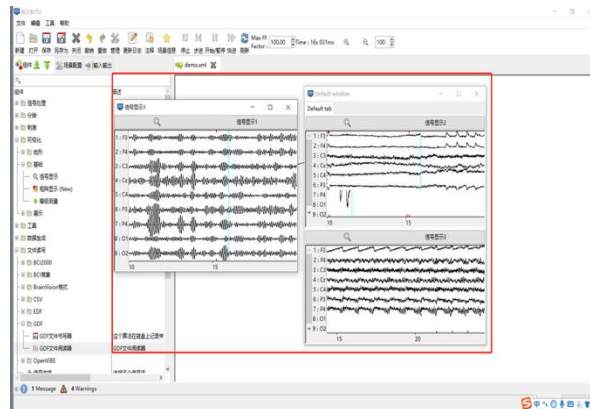
■科技创新2030 “新一代人工智能” 重大项目，研究跨域异质分布式学习和推理系统面临的大规模跨域网络复杂化、跨机构数据隐私保护、AI异质设备多样化等重大科学难题。



# 3.2-大规模时间序列分析与预测

(原继东、刘海洋、孙永奇、王志海)

- 时间序列表示学习
- 时间序列辨别性模式发现
- 深度学习模型的对攻击与防御
- 时间序列分类与预测
- 数据流概念漂移检测
- 序列推荐与因果推荐系统



应用场景：基于高光谱时间序列的病虫害检测、可穿戴数据的快速识别与分类、用户特征分析、手写汉字识别、在线学习行为预测与分析以及交通流量、股票趋势、设备故障的预测等。

## 3.2-大规模时间序列分析与预测

### ➤ 掌子面围岩条件实时智能感知理论-国家重点研发项目

■挑战：基于历史工程建立的机器学习模型不适用于新建工程；新建工程数据量少，深度学习模型效果差

■对策：基于**TBM破岩机理**的数据转换关系 + 数据**迁移学习**方法

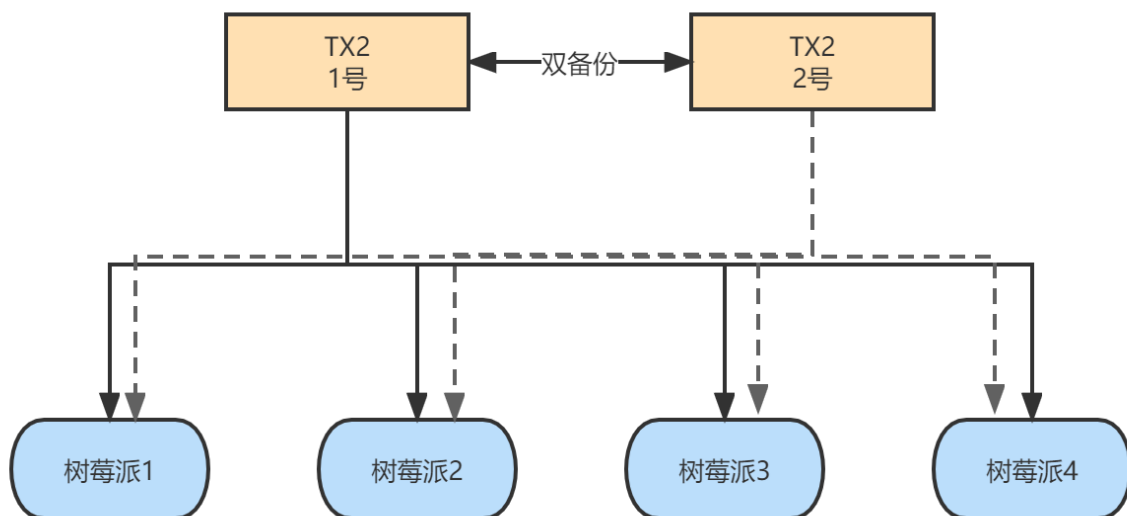
■意义：有效**利用历史工程的数据知识**指导新建工程的TBM施工



# 3.3-分布式对抗博弈与多智能体仿真技术

(刘海洋、原继东、孙永奇、徐保民)

- 将深度强化学习算法构建为微服务，基于分布式硬件资源的虚拟化完成统一调度与管理。

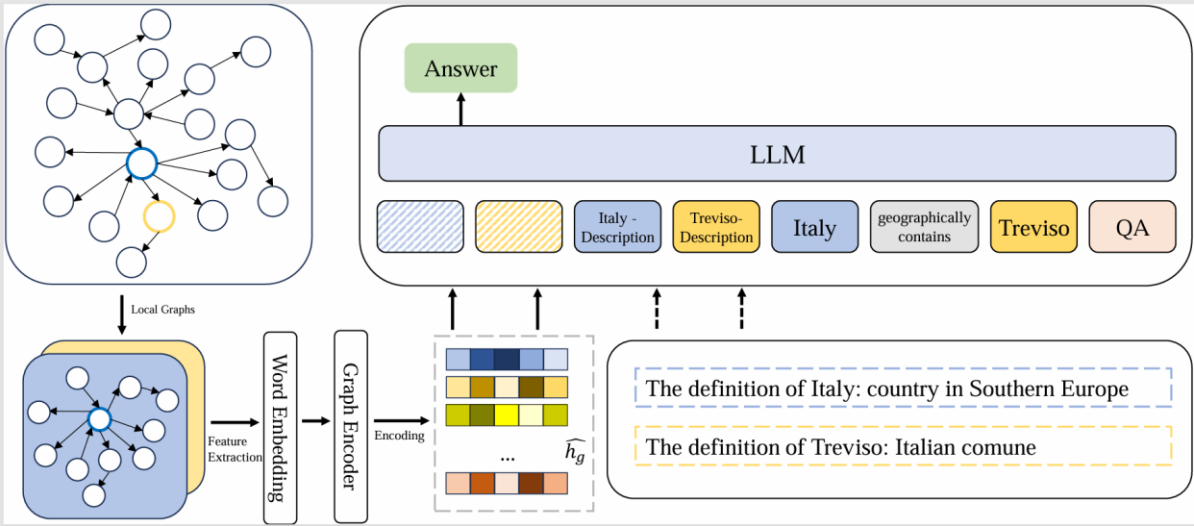


- 研究数字孪生与并行仿真技术，利用多智能体控制技术保证仿真环境真实性



# 3.4-大模型应用与系统平台

## ■ 基于LLM的知识图谱补全方法



三元组分类结果

| Model    | FB15K237-N |       | Wiki27k |       | WN18RR |       |
|----------|------------|-------|---------|-------|--------|-------|
|          | Acc        | F1    | Acc     | F1    | Acc    | F1    |
| TransE   | 0.802      | 0.800 | 0.616   | 0.554 | 0.565  | 0.474 |
| TransH   | 0.786      | 0.790 | 0.614   | 0.556 | 0.566  | 0.508 |
| TransD   | 0.796      | 0.797 | 0.608   | 0.547 | 0.562  | 0.524 |
| TransR   | 0.804      | 0.811 | 0.611   | 0.585 | 0.590  | 0.561 |
| RotatE   | 0.868      | 0.871 | 0.555   | 0.580 | 0.557  | 0.517 |
| ComplEX  | 0.734      | 0.728 | 0.583   | 0.474 | 0.550  | 0.517 |
| DistMult | 0.698      | 0.703 | 0.591   | 0.478 | 0.555  | 0.535 |
| KG-BERT  | 0.901      | 0.907 | 0.757   | 0.794 | 0.773  | 0.666 |
| PKG      | 0.902      | 0.920 | 0.878   | 0.880 | 0.676  | 0.736 |
| KoPA     | 0.800      | 0.854 | 0.852   | 0.848 | 0.863  | 0.897 |
| KG-LLAMA | 0.824      | 0.805 | 0.690   | 0.647 | 0.851  | 0.875 |
| LG-KGC   | 0.965      | 0.978 | 0.895   | 0.944 | 0.960  | 0.975 |

专利：一种基于大语言模型和图神经网络的知识图谱补全方法(已公开)

## ■ 基于LLM+RAG的文档审查与信息抽取

(含技术培训、技术中介)

项目名称: 江西省国土空间规划 (2020-2035 年)

委托人 (甲方): 江西省自然资源厅

受托人 (乙方一、联合体牵头单位): 中国城市规划设计研究院

受托人 (乙方二): XXXXX

受托人 (乙方三): XXXXX

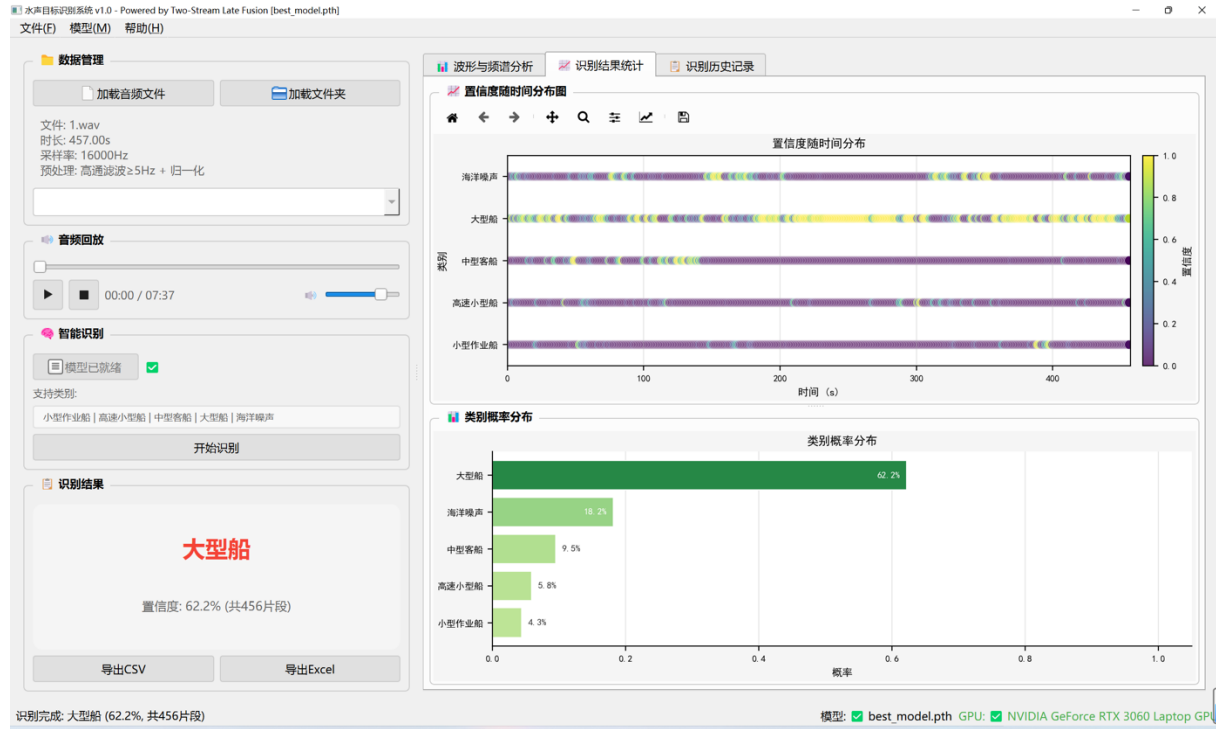
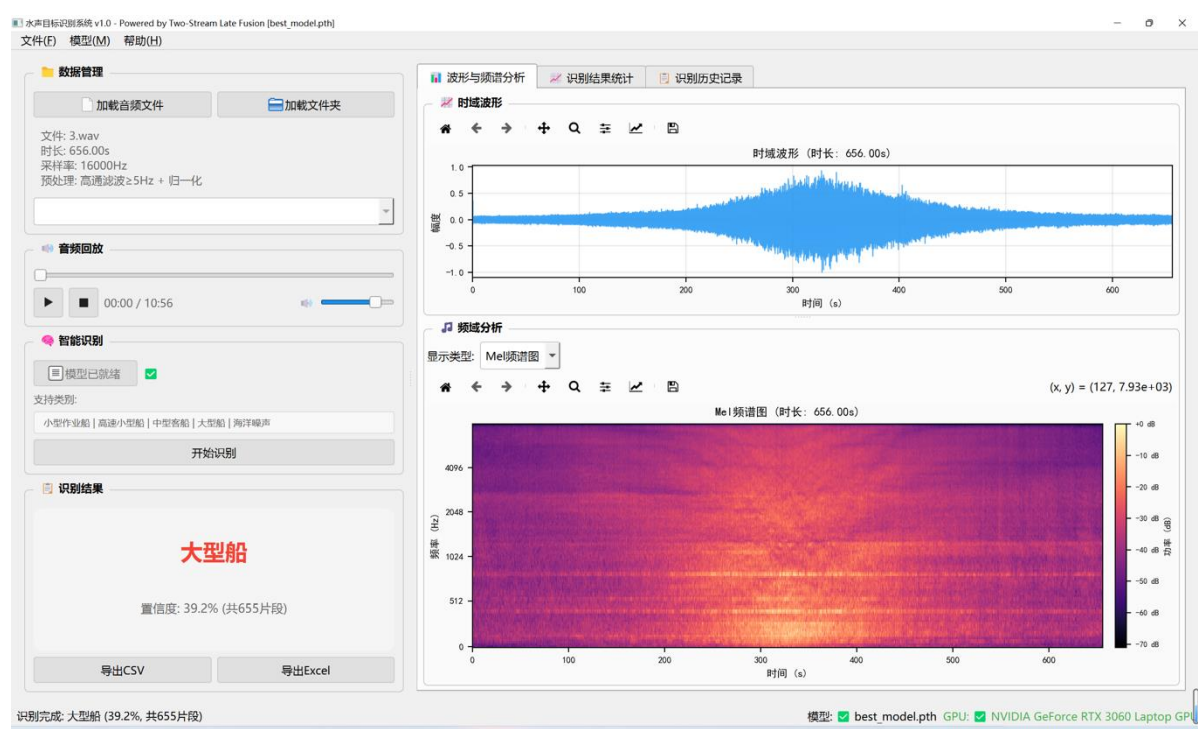
|       |     |      |           |
|-------|-----|------|-----------|
| 签订地点: | 江西省 | 省(市) | 南昌市、县 (区) |
|-------|-----|------|-----------|

```
{  
  "contractNameAndBackground": {  
    "name": "技术服务合同",  
    "backgroundInfo": "项目名称: 江西省国土空间规划 (2020-2035 年); 委托人 (甲方): 江西省自然资源厅; 受托人 (乙方一、联合体牵头单位): 中国城市规划设计研究院; 受托人 (乙方二): XXXXX; 受托人 (乙方三): XXXXX。",  
  },  
  "contractCoverInformation": {  
    "name": "合同封面信息",  
    "checkFlag": "不符合",  
    "checkInfo": "合同封面内容不符合初审要求: 缺失“签订日期”、“有效期限”,  
    "originalText": "签订地点: 江西省南昌市; 其他信息未明确。",  
  },  
  "others": {  
    "name": "其他",  
    "checkFlag": "符合",  
    "checkInfo": "达到规定合同份数。",  
    "originalText": "本合同一式拾叁份, 甲方肆份, 乙方一伍份, 乙方二叁份, 招标代理机构壹份",  
  }  
}
```

相关功能：实现合同内容的审查，返回各项审查结果

# 高质量数据集建设项目

■ 面向大规模船舶噪声数据，实现数据清洗、高质量数据集构建、目标识别与轨迹预测等



# 智慧缆索项目

■ 花江峡谷大桥是以智慧主缆监测为核心，融合实时感知、数据采集、可视化分析和智能运维的世界级峡谷桥梁智慧管养示范工程。



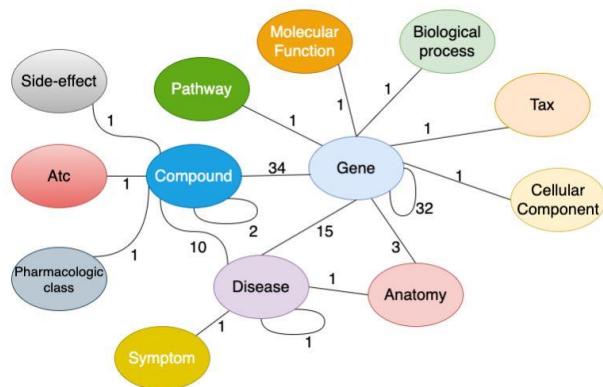
# PIS智慧运维平台

- 面向城市轨道交通的智能运维平台，利用智能化、信息化、可视化和大数据等手段，实时监控设备的运行状态、采集和下载PIS系统的运行数据。



# 3.4-大模型应用与系统平台

- 基于知识图谱的**药物重定位**，XX方案推荐（原继东，刘海洋等）
- 基于深度学习的**图像文本、手写票据内容识别**（孙永奇等）
- 基于文本挖掘的**舆情分析系统**（王志海、原继东等）
- 轨道交通车辆中心作业及设备**运维一体化平台**（徐保民等）

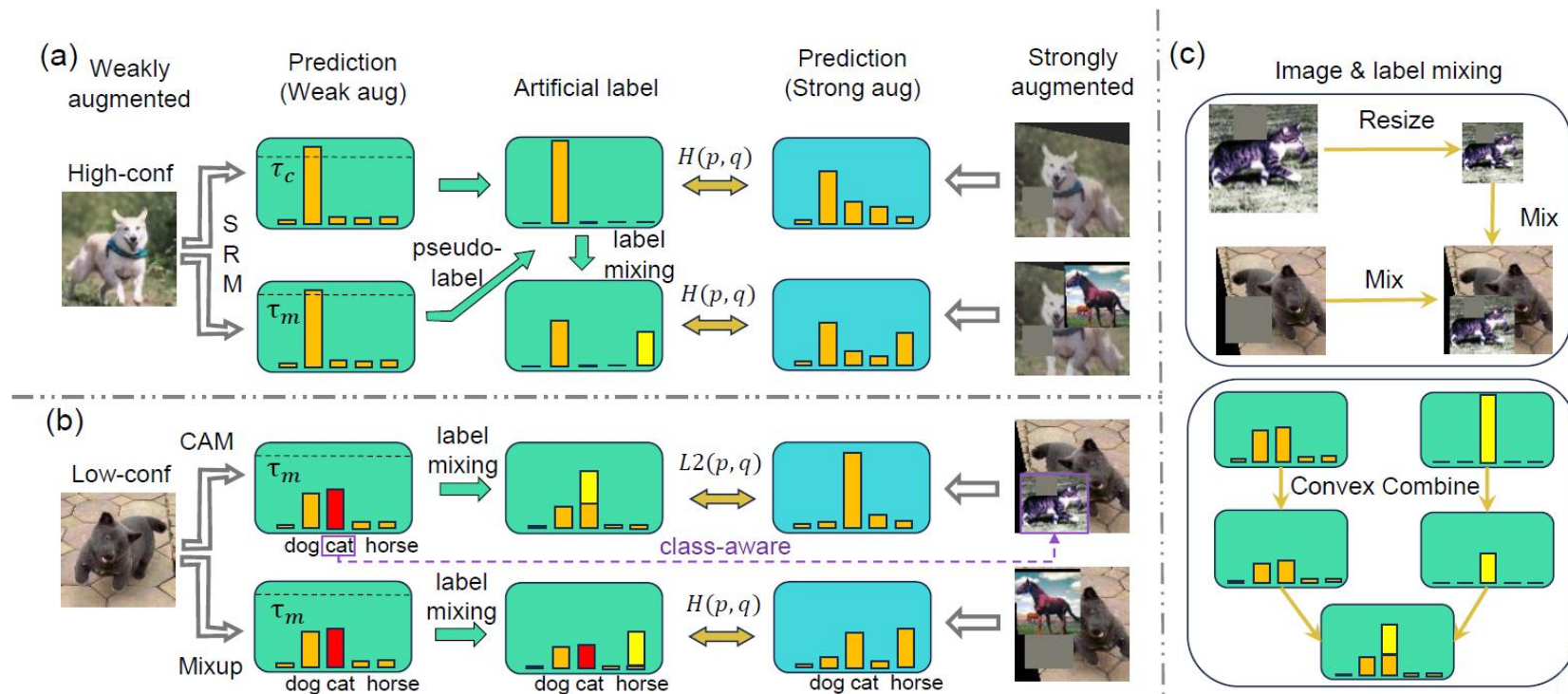


# 4

## 优秀研究生成果

# 优秀研究生成果

## ■ 韩浩荣 (半监督小样本学习, 2026年优秀硕士毕业论文)



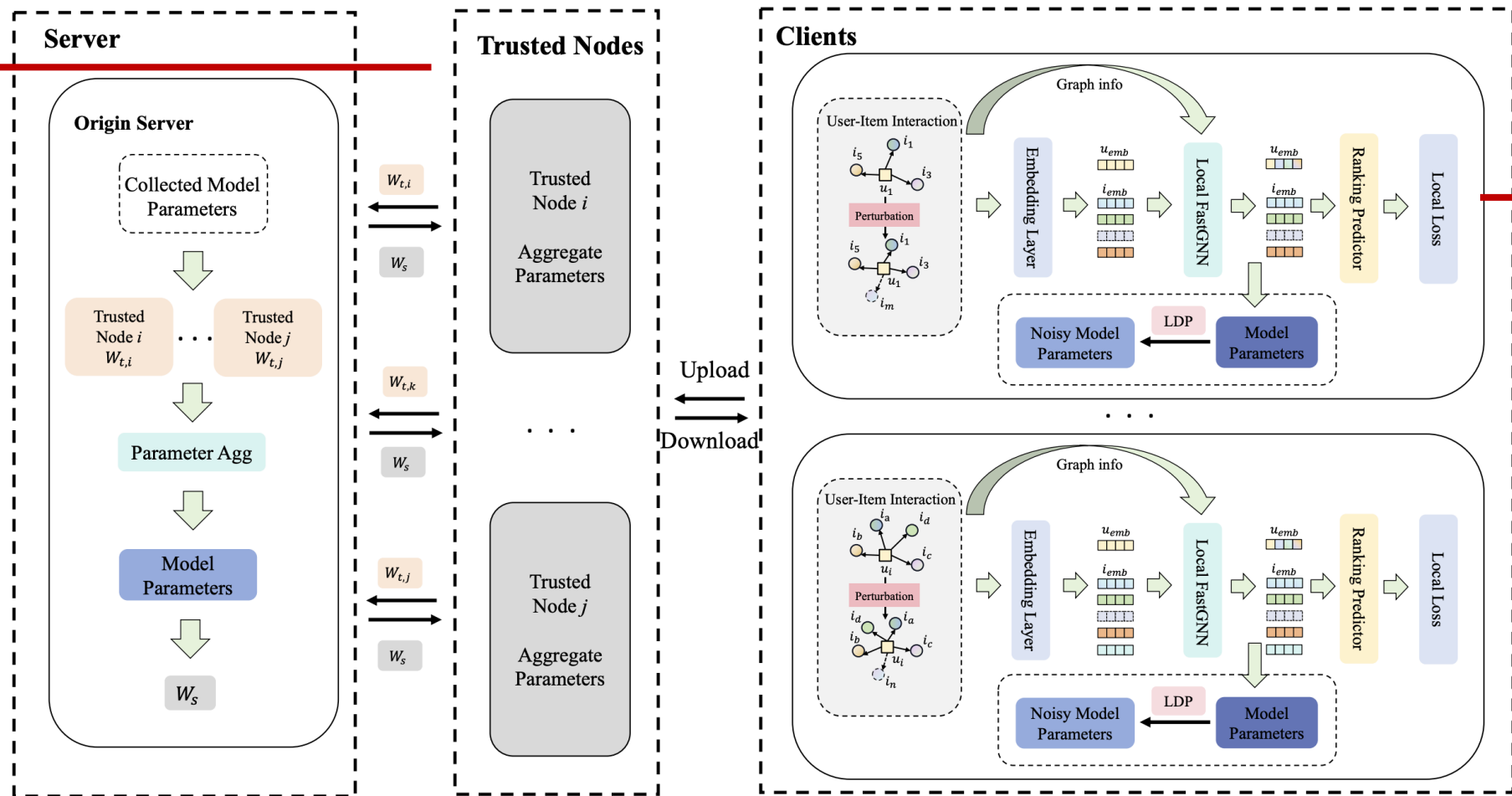
只有1%样本有标签的情况下, 取得了比全标注更好或接近的效果

Haorong Han, Jidong Yuan\*, Chixuan Wei, Zhongyang Yu. RegMixMatch: Optimizing mixup utilization in semi-supervised learning. AAAI 2025. (CCF A)

# 优秀研究生成果

## ■ 闫振兴 (联邦推荐, 2026年优秀硕士毕业论文)

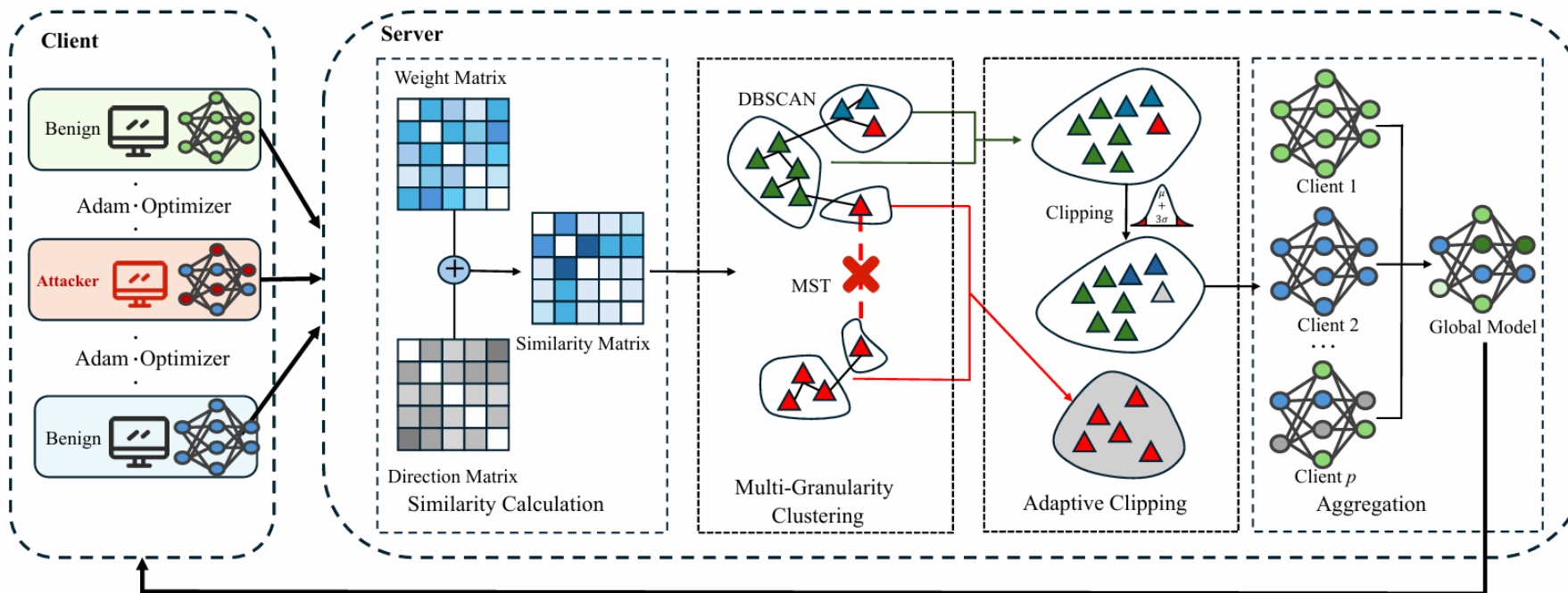
解决梯度上传不安全问题



解决GNN收敛速度慢问题

# 优秀研究生成果

## ■ 张起航（联邦环境下的攻击和防御算法研究，硕博连读）

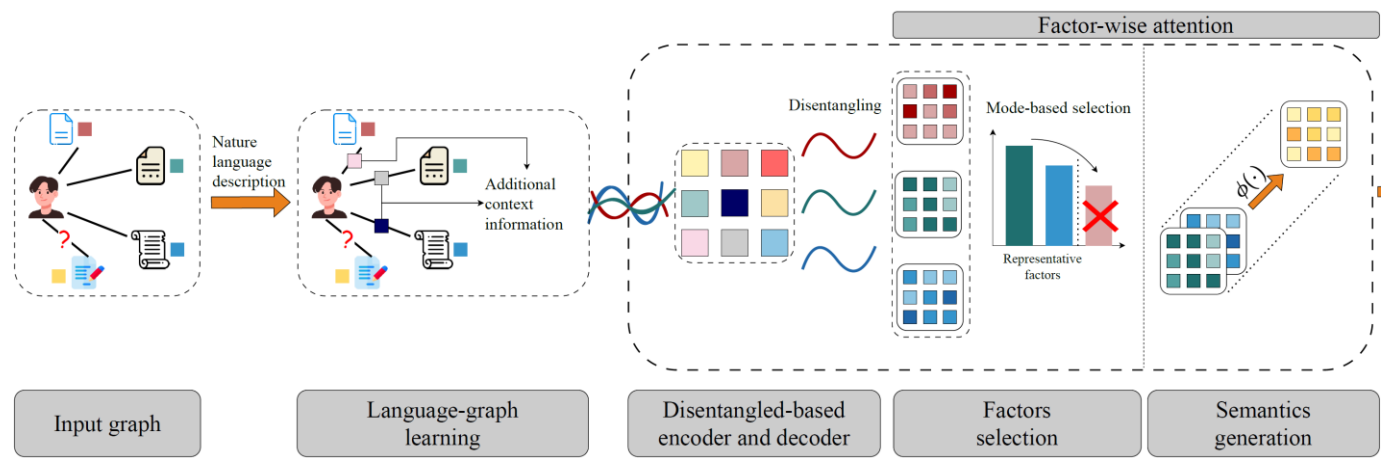


通过多粒度聚类的方式有效防御分布式联邦环境下的潜在攻击

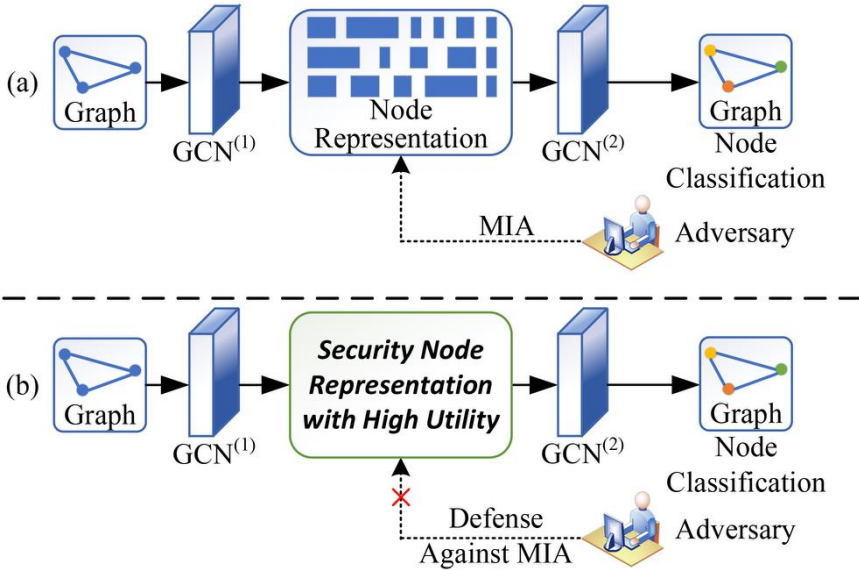
Jidong Yuan, Qihang Zhang, Naiyue Chen, Shengbo Chen, Baomin Xu. A multi-granularity clustering approach for federated backdoor defense with the Adam optimizer. IJCAI 2025. (CCF A)

# 优秀研究生成果

## ■ 方荣强（语言模型辅助与语义感知的异构图链接预测方法）



结合语言模型学习节点和边的上下文信息和丰富的语义信息提高  
异质图（知识图谱）的连边预测准确率



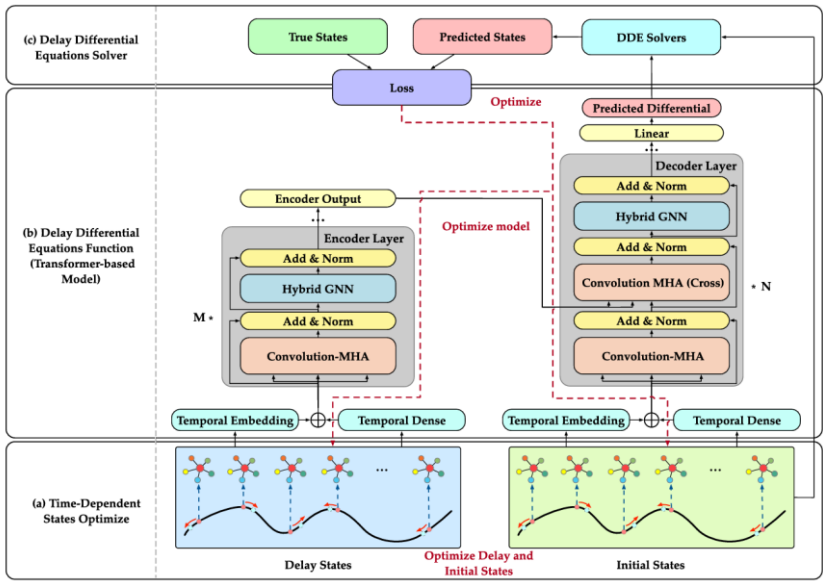
基于解耦的抗成员推理攻击方法

Rongqiang Fang, Yongqi Sun, Jidong Yuan\*, hongbo Cao, Jinkun Dong. A Language-Assisted Semantic-Aware Disentangled Method for Link Prediction on Heterogeneous Graphs. ACM MM 2025. (CCF A)  
Rongqiang Fang, Yongqi Sun, Jidong Yuan\*, Hongbo Cao. DMIA: A disentangled-based method for graph convolutional network against membership inference attack. IEEE Transactions on Dependable and Secure Computing. 2025, 22(6): 7956-7970. (CCF A, An1)

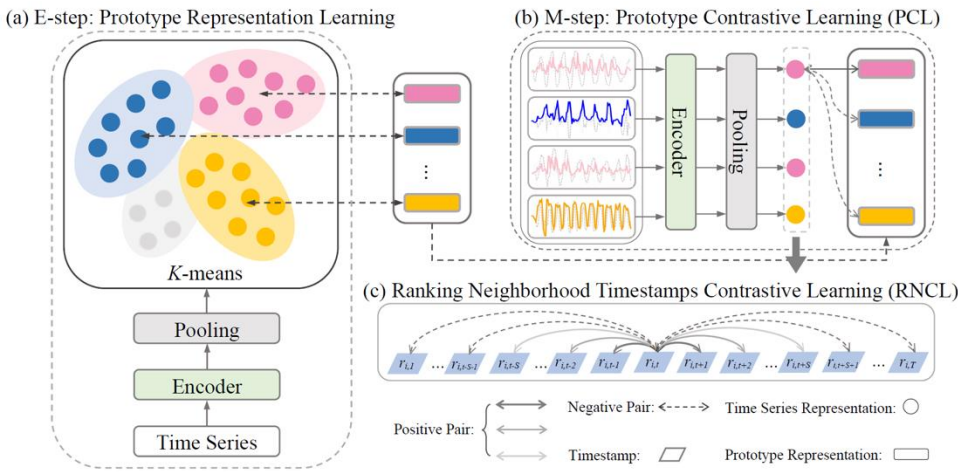
# 优秀研究生成果

## ■ 于钟洋、朱卫国、魏池璇等（时间序列分类与预测）

相关论文发表在人工智能领域顶级IEEE汇刊, 如IEEE TITS, TNNLS, TII, TBD, TCSS, TETCI, TAI 等



基于微分方程的时空交通流预测



通过EM聚类实现时间序列的原型学习

Chixuan Wei, Zhihai Wang, Jidong Yuan, Xiaokang Wang, Haiyang Liu\*, Qiyang Zhao. SemiHAR: Improving Semi-supervised Human Activity Recognition via Multi-task Learning[J]. IEEE Transactions on Neural Networks and Learning Systems. 2025, 36(1): 1884-1897.

Jidong Yuan, Bingyu Gao, Xiaokang Wang\*, Haiyang Liu, Lingyin Zhang. Counterfactual music recommendation for mitigating popularity bias[J]. IEEE Transactions on Computational Social Systems. 2025, 12(2): 851-861.

Yingying Zhao, Yongqi Sun, Naiyue Chen\*, Yifan Sui, Hongbo Cao. FedFOC: Personalized federated learning via fine-grained one-shot clustering. IEEE Transactions on Industrial Informatics, 2025.

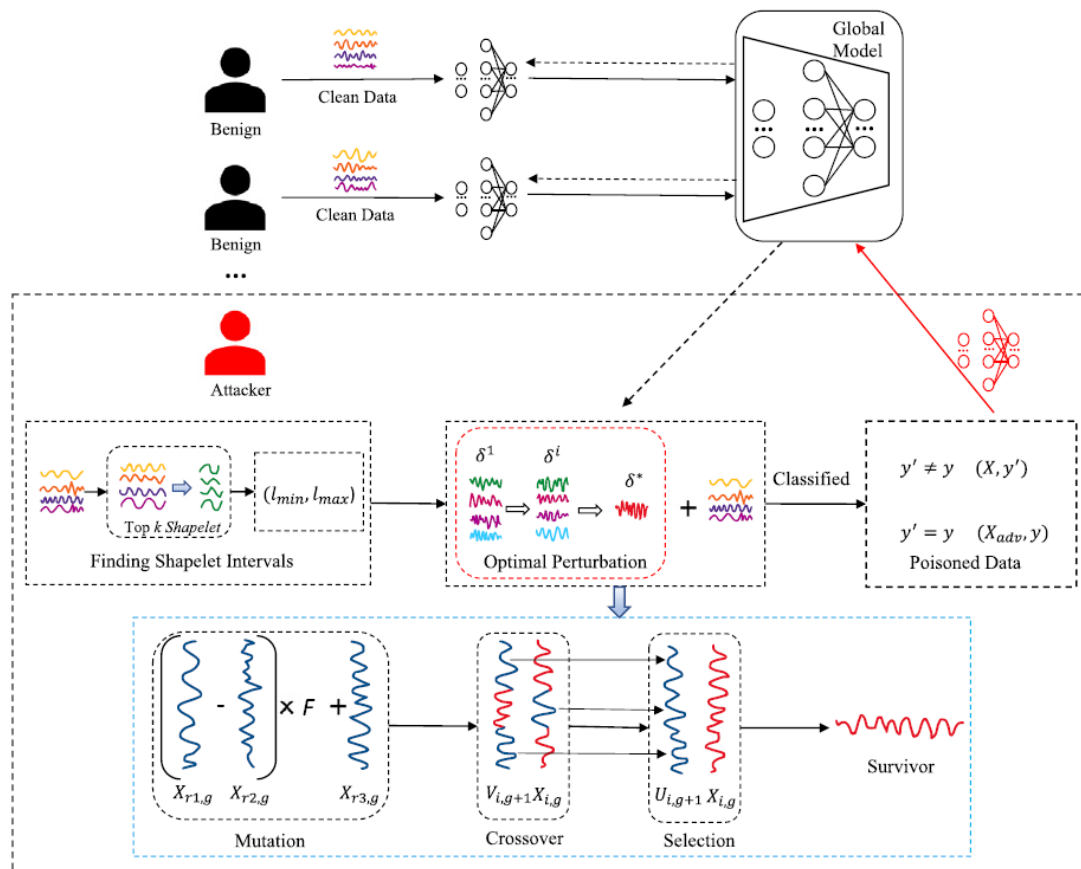
Zhongyang Yu, Zhihai Wang, Jidong Yuan\*, Huiting Pei, Yuxuan Fan, Shijiang Li. DTWformer: A DTW-based Transformer for multivariate time series forecasting[J]. IEEE Transactions on Big Data. 2025.

Jidong Yuan, Lingyin Zhang, Weiqiang Jia, Jia Guo, Haiyang Liu\*, and Jinfeng Wang. Ranking time-frequency contrastive learning for multivariate time series classification. IEEE Transactions on Artificial Intelligence. 2026.

Chixuan Wei, Jidong Yuan\*, Yi Zhang, Zhongyang Yu, Yanze Liu, Haiyang Liu. Ranking neighborhood and class prototype contrastive learning for time series[J]. IEEE Transactions on Big Data. 2025.

# 优秀研究生成果

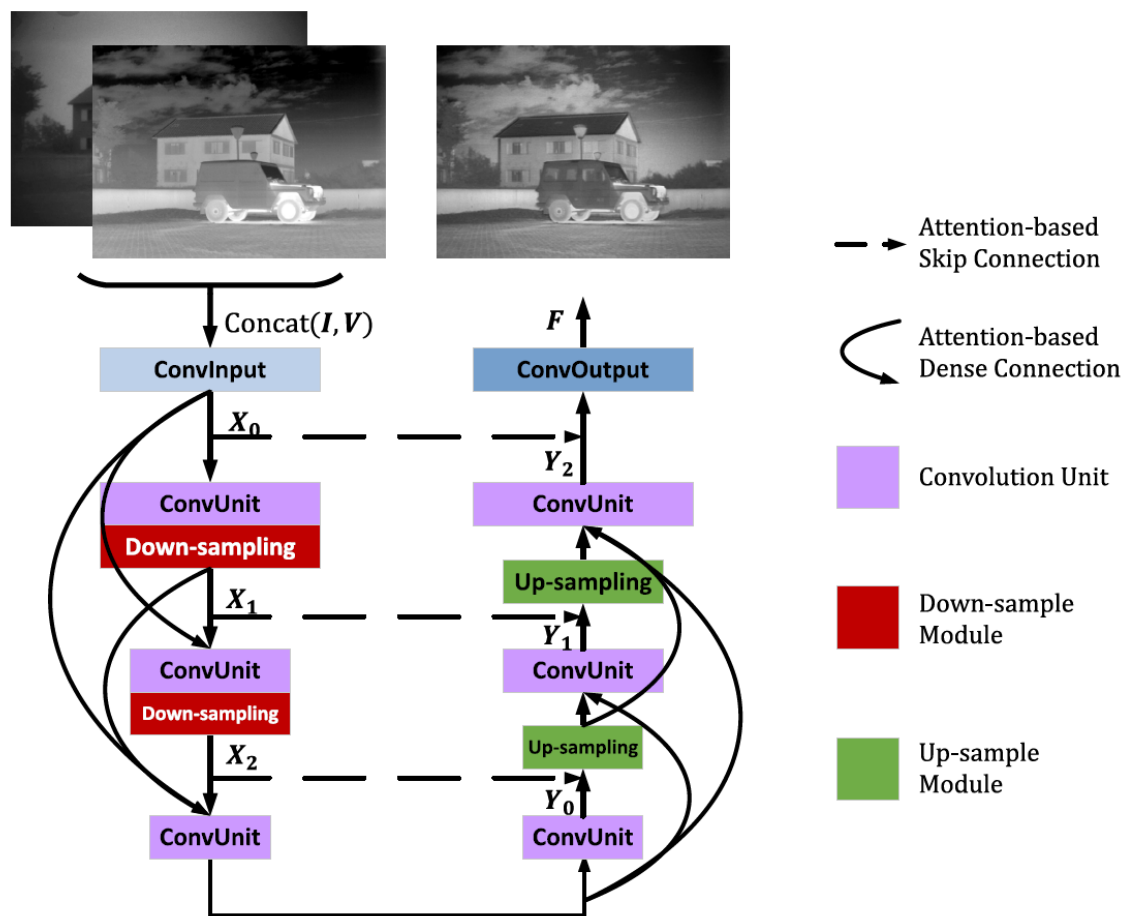
## ■ 陈升波（联邦环境下的时间序列攻击算法研究，2024届优秀毕业生）



Shengbo Chen, Jidong Yuan\*, Zhihai Wang, Yongqi Sun. Local perturbation-based black-box federated learning attack for time series classification[J]. Future Generation Computer Systems, 2024, 158: 488-500. (SCI, An1 Top)

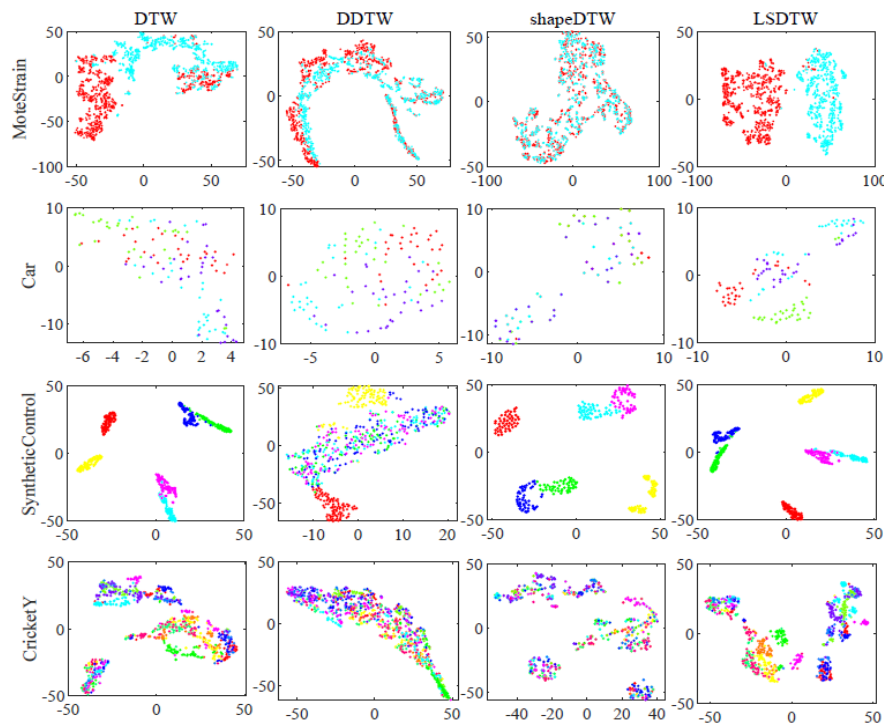
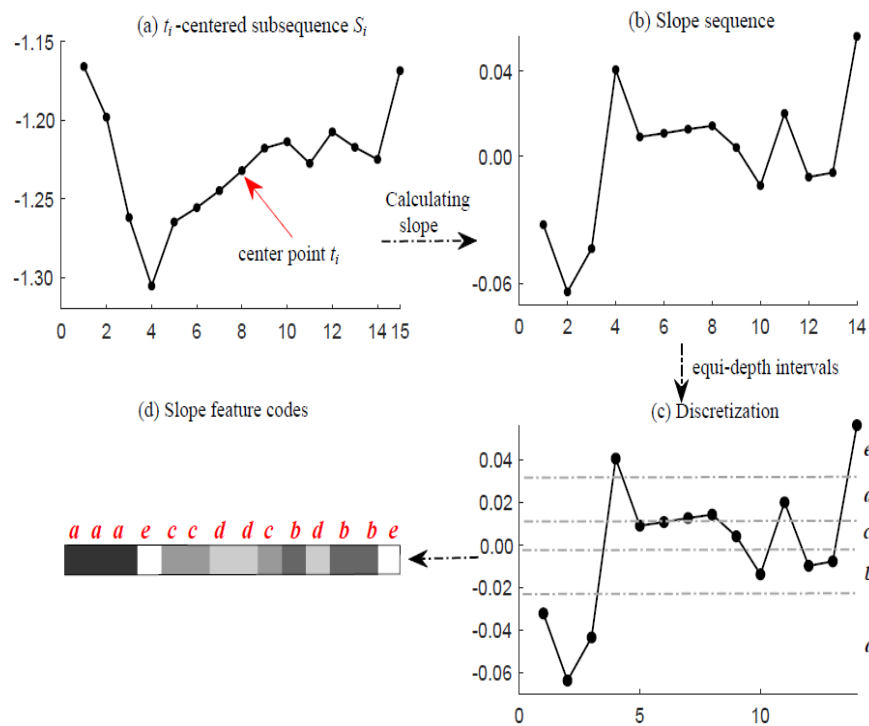
# 优秀研究生成果

## ■ 程辰（基于风格损失的可见光和红外图像融合模型，2023届优秀毕业生）



# 优秀研究生成果

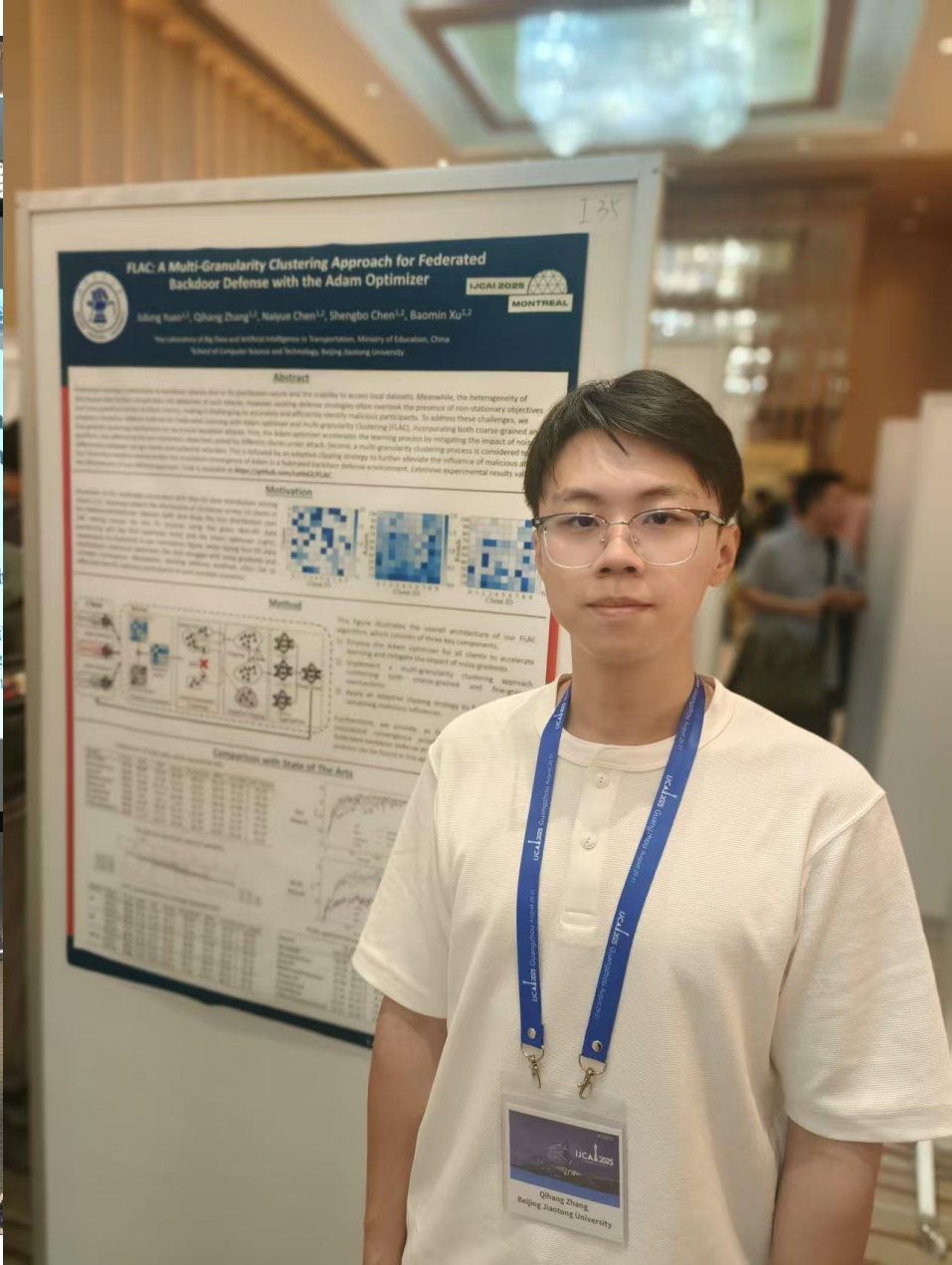
## ■ 林钱洪（基于形状特征的时间序列分类算法，2020届优秀毕业生）



Yuan J, Lin Q, Zhang W, et al. Locally Slope-based Dynamic Time Warping for Time Series Classification[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management(CIKM 2019). 2019: 1713-1722. (CCF-B)

魏池璇, 王志海, 原继东\*, 林钱洪. 时间序列可变尺度的时域与频域特征求解及其分类算法[J]. 软件学报, 2022. (CCF A中文)

# 国际交流



# 毕业生的去向

---

## 1. 硕士研究生

阿里、百度、快手、字节跳动、京东、银行、移动、联通、城铁、中铁通号公司等

## 2. 出国深造

日本大阪大学、澳大利亚悉尼科技大学、美国达特茅斯学院、美国贝勒医学院、英国兰卡斯特大学、法国约瑟夫傅里叶大学

## 3. 博士研究生

公安部、中国船舶集团、中国电子科技集团、中国医学科学院、北京交通大学、中国传媒大学、郑州大学、桂林电子科技大学、北京建筑大学等科研院所与高校

# 校企合作

---



# 团队建设



# 目前研究方向总结

---

- ① 时间序列分类、预测（交通流）、异常检测：Transformer、AI Agent
- ② 车辆路径规划（VRP、TSP）：深度强化学习技术
- ③ 生成式人工智能与安全：LLM、GAN、扩散模型
- ④ 图数据表示与学习（graph）：结合预训练模型/LLM
- ⑤ 联邦学习：个性化联邦、联邦环境下攻击与防御
- ⑥ 知识图谱：结合大模型的知识图谱补全、面向铁路大模型的数据平台构建关键技术



2026夏令营

**欢迎加入我们团队**  
**掌握云计算、探索大数据!**  
**<http://ccd.bjtu.edu.cn>**