



多源异质大数据 联合分析

智 联 异 质 数 据 孤 岛 · 赋 能 精 准 分 析 决 策

作品类别 | 自然科学类学术论文

参赛成员 | 张云帆 陈欣禧 谭泽熙 甘国基
龙志全 高逸菲 蔡升宏 陈俊仰

指导老师 | 张逸群 姬玉柱 骆文君



大数据 已成为推动社会进步和经济发展的核心资源

中华人民共和国中央人民政府 [www.gov.cn](#)

首页 > 政策 > 最新

中共中央 国务院印发《数字中国建设整体布局规划》

2023-02-27 来源：新华社

新华社北京2月27日电 近日，中共中央、国务院印发了《数字中国建设整体布局规划》（以下简称《规划》），并发出通知，要求各地区各部门结合实际认真贯彻落实。

《规划》指出，建设数字中国是数字时代推进中国式现代化的重要引擎，是构筑国家竞争新优势的有力支撑。加快数字中国建设，

中华人民共和国中央人民政府 [www.gov.cn](#)

首页 > 政策 > 最新

中共中央办公厅 国务院办公厅关于加快公共数据资源开发利用的意见

2024-10-09 来源：新华社

新华社北京10月9日电

国家数据局

国家发展改革委 国家数据局关于印发《数字经济促进共同富裕实施方案》的通知

2024.01.05

中华人民共和国中央人民政府 [www.gov.cn](#)

国务院印发《关于深入实施“人工智能+”行动的意见》

2025-08-26 来源：新华社

新华社北京8月26日电 国务院日前印发《关于深入实施“人工智能+”行动的意见》（以下简称《意见》）。

《意见》坚持以习近平新时代中国特色社会主义思想为指导，完整准确全面贯彻新发展理念，坚持以人民为中心的发展思想，充分发展我国数据资源丰富、产业体系完备、应用场景广阔等优势，强化前瞻谋划、系统布局、分业施策、开放共享、安全可控，以科技、产业、消费、民生、治理、国际合作等领域为重点，深入实施“人工智能+”行动，涌现一批新基建设施、新技术体系、新产业链、新就业岗位等，加快培育发展新动能，使全体人民共享人工智能发展成果，更好服务中国式现代化建设。

《意见》提出加快实施6大重点行动。一是“人工智能+”科学技术，加速科学发现进程，驱动技术研发模式创新和效能提升，创新科学社会科学研究方法。二是“人工智能+”产业发展，培育智能原创新模式新业态，推进工业全要素智能化发展，加快农业数字化转型升级，创新服务业发展新模式。三是“人工智能+”消费提质，拓展服务消费新场景，培育产品消费新业态。四是“人工智能+”民生福祉，创造更加智能的工作生活方式，打造更有品质的美好生活。五是“人工智能+”治理能力，开创社会治理新路径，打造全国一体化政务服务平台，共建美丽中国生态治理新画卷。六是“人工智能+”全球合作，推动人工智能普惠共享，共建人工智能全球治理体系。

《意见》提出强化8项基础支撑，包括提升模型基础能力、加强数据供给创新、开源生态繁荣、加强人才队伍建设和强化政策法规保障、提升安全能力水平等。

中华人民共和国中央人民政府 [www.gov.cn](#)

《“数据要素×”三年行动计划(2024—2026年)》发布

2024-01-04 来源：新华社

新华社北京1月4日电（记者 严赋憬）国家数据局4日发布消息，国家数据局等17部门近日联合印发《“数据要素×”三年行动计划（2024—2026年）》，旨在充分发挥数据要素乘数效应，赋能经济社会发展。

行动计划强调坚持需求牵引、注重实效，试点先行、重点突破，有效市场、有为政府，开放融合、安全有序等4方面基本原则，明确了到2026年底的工作目标。行动计划选取工业制造、现代农业、商贸流通、交通运输、金融服务、科技创新、文化旅游、医疗健康、应急管理、气象服务、城市治理、推动发挥数据要素乘数效应，释放数据要素价值。

下一步，国家数据局将会同有关部门，加强组织领导，开展试点工作，推动以赛促用，加强资金支持，加强宣传推广，确保各项工作落到实处。

国家数据局

国家发展改革委 国家数据局关于印发《数字经济促进共同富裕实施方案》的通知

2024.01.05

国家发展改革委 国家数据局关于印发《数字经济促进共同富裕实施方案》的通知

发改数据〔2023〕1770号

各省、自治区、直辖市及计划单列市、新疆生产建设兵团发展改革委、数据局，中央网信办、教育部、工信部、民政部、人力资源社会保障部、农业农村部、商务部、文化和旅游部、国家卫生健康委、金融监管总局、广电总局：

通过数字化手段促进解决发展不平衡不充分问题，推进全体人民共享数字时代发展红利

同富裕，我们研究制定了《数字经济促进共同富裕实施方案》。现印发给你们，请认真组织实施，加快推进各项任务。

充分开发数据资源，推进智能分析决策，成为重要命题

▶ 技术路线图

项目简介 技术创新 成果展示 团队介绍 拓展延续



多源异质大数据联合分析

研究价值

✓ 丰富理论方法

统一距离度量 | 无监督联邦学习机制 | 持续联邦学习范式

✓ 疏通应用堵点

数据多源异质 | 真实环境标签匮乏 | 知识持续更新

创新架构

信息融合损耗 ↓

创新点一
异质特征数据距离新度量

自研统一度量+多层次耦合编码

知识嗅探精度 ↑

创新点二
多源数据联邦聚类新策略

多粒度层次脱敏+簇分布自组织

知识时效性 ↑

创新点三
动态环境持续学习新模型

异步权重更新+主动漂移感知

核心问题

信息融合损耗大

数据异质 → 对齐程度弱

年龄
10
20
30



联邦聚类精度低

信息非完备 → 聚类依据匮乏



知识更新时效差

数据动态 → 更新链路复杂



痛点一

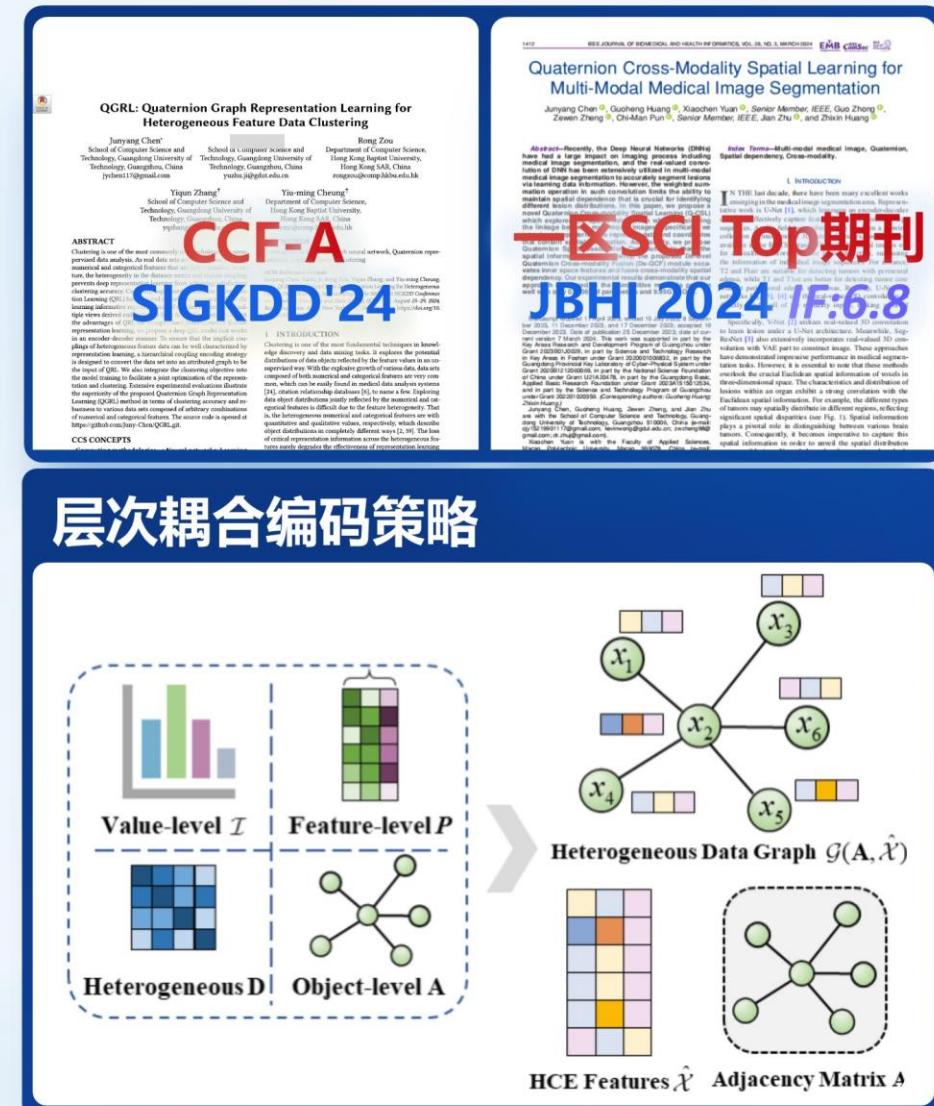
数据异质，对齐程度弱 → 信息融合损耗大

解决方法 → 自研统一度量 + 多层次耦合编码

- 使用统一度量方法衡量异质指标间距离，弥合信息差
 - 从“值-特征-异质-样本”四个层次进行编码，进一步提取隐式信息
 - 通过端到端的机器学习方法实现下游任务自适应

解决成效

信息融合损失降低37.49%



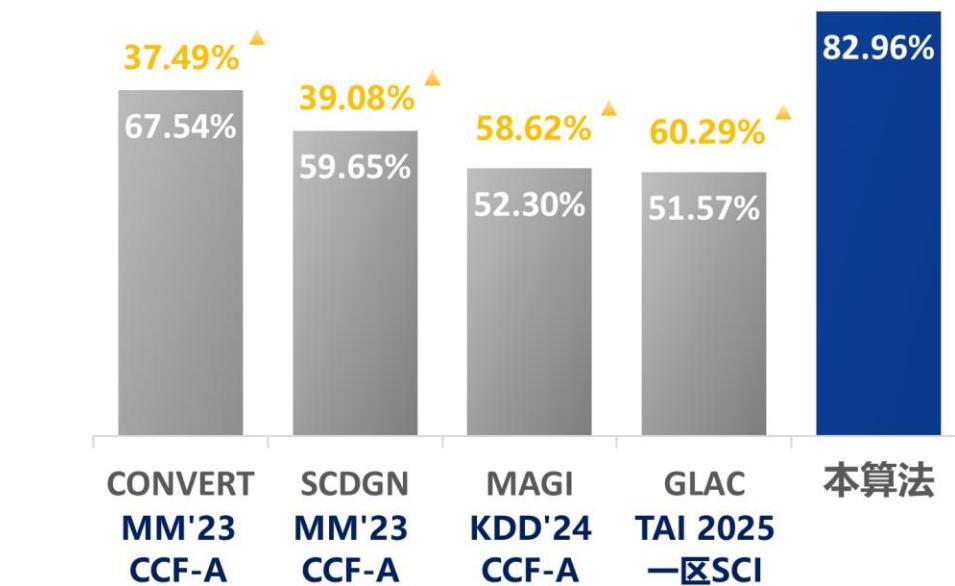
统一度量，精准融合，减少信息流失

▶ 实验结果

项目简介 技术创新 成果展示 团队介绍 拓展延续



当前领域先进算法: GLAC
(TAI 2025 一区SCI)



知识融合精度

51.57% → 82.96%

本算法: QGRL
(SIGKDD'24 CCF-A)



异质信息提取更充分, 分析决策精度高, 发表CCF-A类成果

痛点二 信息非完备，聚类依据匮乏 → 联邦聚类精度低

解决方法 → 多粒度层次脱敏 + 簇分布自组织

- 微簇竞争遵循“胜者通吃”原则，不显著微簇被淘汰
- 显著微簇进行层次融合，得到数据的多层次分布结构
- 基于多层次结构形成不同粒度的样本群体，避免“漏网之簇”

解决成效 知识嗅探精度高达90.50%

自动搜寻微簇，细查复杂分布，锁定关键少数

2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)

Robust Categorical Data Clustering Guided by Multi-Granular Competitive Learning

Shenglong Cai^{1,*}, Xiaopeng Liu², Yu-Ming Cheung³, Hong Jia⁴, Peng Liu¹
¹Guangdong University of Technology, Guangzhou, China
²Guangdong Institute of Intelligent Manufacturing, Shunde, China
³Hong Kong Baptist University, Hong Kong SAR, China
⁴Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shunde, China
312105074@stu.hkbu.edu.hk, hongjia1102@zjhu.edu.cn, lpcp@gsit.gdut.edu.cn
*Corresponding author

Abstract—Data set composed of categorical features is very common in big data analysis tasks. Since categorical features are discrete and non-metric, it is challenging to perform clustering on them. In this paper, we propose a multi-granular competitive learning (MGCL) algorithm to handle categorical data. MGCL first explores the nested granular cluster effect in the implicit distance space, which is induced by the categorical features frequently overlap in space or adaptive to the data distribution. However, the distance space cannot validate all the clusters in the data set. Therefore, MGCL also proposes a multi-granular Penetration Learning (MGPCL) algorithm to handle the clusters in the distance space. MGCL and MGPCL are integrated together to handle categorical data. MGCL can automatically explore the nested distribution of multi-granular clusters and tightly relate to categorical features. MGPCL can automatically handle the linear time complexity. MGCL is available to large-scale data sets and MGPCL is available to small data sets. Extensive experiments with various real public data sets show that MGCL and MGPCL outperform state-of-the-art counterparts on various real public data sets.

Keywords—Categorical data clustering, multi-granular competitive learning, number of clusters, cluster granularity

I. INTRODUCTION
Clustering is one of the most popular unsupervised learning techniques that divide objects into a certain number of clusters where each cluster is usually composed of similar objects. Quantitative values are often used to measure the similarity between objects. However, categorical values are qualitative values that cannot be directly compared. Therefore, how to encode the qualitative values into quantitative numerical values. How to measure the distance between categorical values is another challenge. In this paper, we propose a multi-granular competitive learning (MGCL) directly performs clustering by adopting categorical distance metrics that are specifically defined for categorical data. MGCL can automatically explore the nested distribution of categorical features from various domains, which brings great challenges to the clustering process. MGCL also proposes a multi-granular penetration learning (MGPCL) that overcomes a dangerous deficiency. MGCL can automatically explore the nested distribution of categorical data, the construction of

the unified heterogeneous attribute distance metric, and the hierarchical merging of micro-clusters.

CCF-B
ICDCS'24

2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)

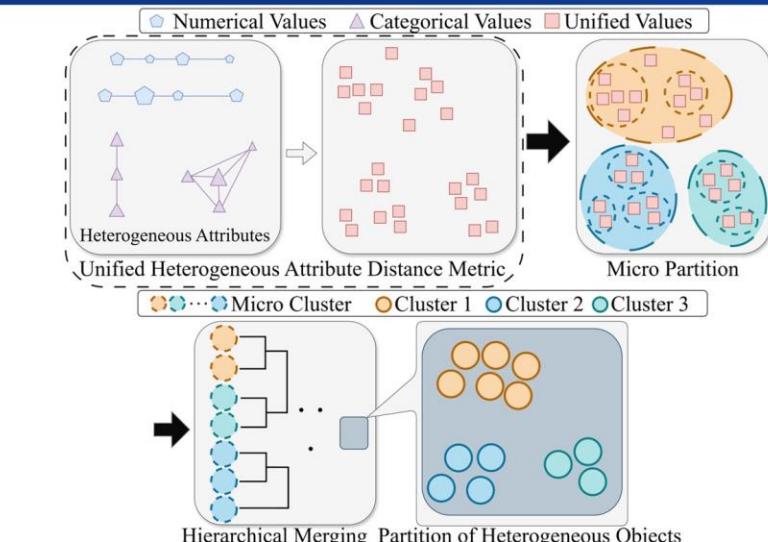
Federated Clustering with Unknown Number of Clusters

Rong Zou^{1,*}, Yanlin Zhang^{2,3}, Yiqian Zhang^{2,3}, Yang Lu^{1,3,4}, Mingke Li^{1,4*}, Yu-ming Cheung¹
¹Department of Computer Science, Hong Kong Baptist University, Hong Kong, S.A.R., China
²School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
³Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shunde, China
⁴312105074@stu.hkbu.edu.hk, hongjia1102@zjhu.edu.cn, lpcp@gsit.gdut.edu.cn
*Corresponding author

Best Paper Award
IEEE最佳论文奖

DOCS'24

微簇划分和分层合并策略



▶ 实验结果

项目简介 技术创新 成果展示 团队介绍 拓展延续



当前领域先进算法: GUDMM
(PR 2025 一区SCI)

知识嗅探精度
82.80% → 90.50%

本算法: MCDC
(ICDCS'24 CCF-B)

ACC (分类精度)



FM (Fowlkes-Mallows 指数)



ARI (调整兰德系数)



AMI (调整互信息)



分布检测更细致, 知识粒度自适应, 获IEEE最佳论文奖

痛点三

数据动态，更新链路复杂 → 知识更新时效差

解决方法 → 异步权重更新+主动漂移感知

- 对各节点群体检测数据进行知识概括，提取关键信息
- 对各节点数据实施脱敏处理以确保隐私安全，并通过跨节点的知识互补与簇分布推理，实现整合分析
- 动态接收并更新各节点数据

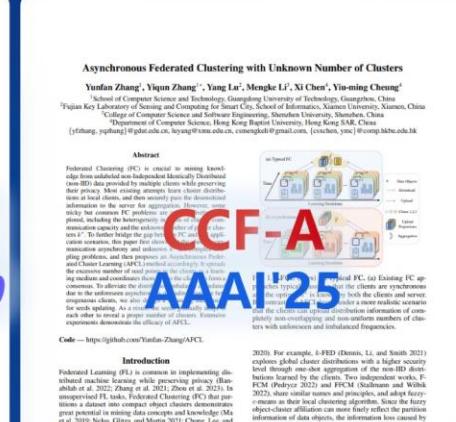
解决成效

知识更新时效提升高达 74.28%

异步联邦新架构，打通隐私壁垒，提升分析效能

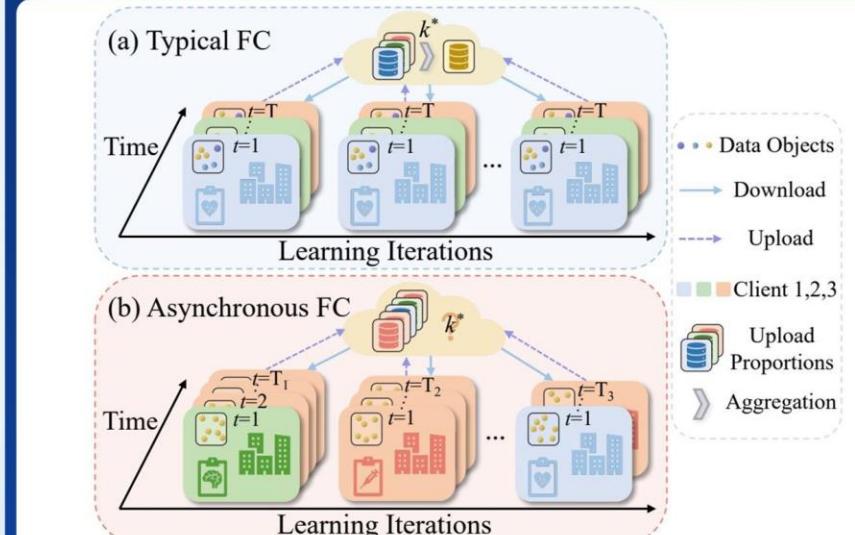


SCI Top期刊
TNNLS 2025 IF: 8.9



CCF-A
AAAI'25

自组织联邦学习+自适应异步更新



▶ 实验结果

项目简介 技术创新 成果展示 团队介绍 拓展延续



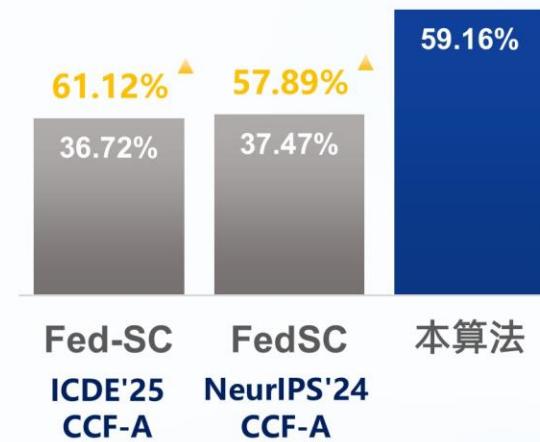
当前领域先进算法: FedSC
(AAAI'25 CCF-A)

全局知识嗅探精度
82.52% → 89.67%

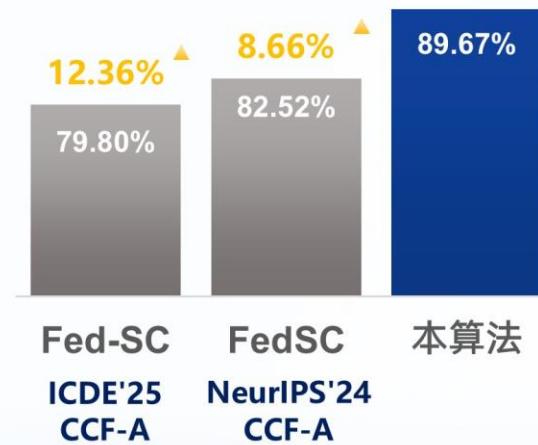
本算法: AFCL
(AAAI'25 CCF-A)

考虑更具挑战性场景: ✦ 异步通讯 (收发不同步) ✦ 先验知识不可知 (真实类别数未知)

在数据集Breast Cancer上对比



在数据集Live上对比



分析鲁棒性强, 决策精度再提升, 发表CCF-A类成果



1篇

CCF-B类

2篇

CCF-A类

4篇

一区SCI

1篇

IEEE最佳论文奖



**TNNLS 2025
IF:8.9**

Learning Self-Growth Maps Imbalanced Streaming

Yiqun Zhang, Senior Member, IEEE; Sen Feng, Peng Yuha, Li Member, IEEE, Rong Zou, and Yiqing Cheung

Abstract—Streaming data clustering is a popular research topic in data mining and machine learning. Since streaming data is often imbalanced and contains many outliers, it is difficult to evaluate the clustering performance. In this paper, we propose a learning self-growth map (SGM) that can automatically arrange various data objects in a semantic order. The SGM is based on the concept of “self-growth” and “self-improvement”. It can automatically learn the semantic order of data objects by mining their local features and then use this order to guide the clustering process. The proposed SGM can effectively handle the imbalanced streaming data and achieve better clustering performance.

CCF-C类

**ICPR'24
本科生一作**

Towards Unbiased Minimally Categorical-and-Numerical Clustering

Yiqun Zhang¹, Xinyang Li², Qianyu Zhou³, Yunfan Zhang⁴, Yiqun Zhang^{5*}

¹School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China; ²School of Computer Engineering, Guangzhou University, Guangzhou, China; ³Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China; ⁴Guangzhou Institute of Advanced Technology, Chinese Academy of Sciences, Guangzhou, China; ⁵College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Data analysis of categorical and numerical data is a challenging task. Most existing clustering methods are designed for categorical data or numerical data, which leads to bias in either the accuracy or the efficiency of clustering. In this paper, we propose a learning-based clustering method that can handle both categorical and numerical data simultaneously. The proposed method can automatically learn the semantic order of data objects by mining their local features and then use this order to guide the clustering process. The proposed method can effectively handle the imbalanced streaming data and achieve better clustering performance.

CCF-C类

**ICDCS'24
本科生一作**

Clustering by Learning the Qualitative Attributes

Shengkai Wang¹, Yiqun Zhang², Yiqun Zhang³*

¹School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China; ²School of Computer Engineering, Guangzhou University, Guangzhou, China; ³College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Data analysis of categorical and numerical data is a challenging task. Most existing clustering methods are designed for categorical data or numerical data, which leads to bias in either the accuracy or the efficiency of clustering. In this paper, we propose a learning-based clustering method that can handle both categorical and numerical data simultaneously. The proposed method can automatically learn the semantic order of data objects by mining their local features and then use this order to guide the clustering process. The proposed method can effectively handle the imbalanced streaming data and achieve better clustering performance.

CCF-C类

**IJCNN'24
本科生一作**

Federated Learning (FL) algorithm is often applied in potential clustering process. In addition to this, we propose a learning-based clustering method for federated learning tasks, which can handle both categorical and numerical data simultaneously. The proposed method can automatically learn the semantic order of data objects by mining their local features and then use this order to guide the clustering process. The proposed method can effectively handle the imbalanced streaming data and achieve better clustering performance.

CCF-C类

Robust Categorical Data Multi-Granular Con

Yunfan Zhang¹, Yiqun Zhang², Yiqun Zhang³*

¹School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China; ²School of Computer Engineering, Guangzhou University, Guangzhou, China; ³College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Data analysis of categorical features is very common in big data analysis tasks. Since categorical features are usually discrete and have no clear order, it is difficult to mine their inherent properties for clustering. However, some categorical features are always globally interdependent with a certain clustering task. A learning-based clustering method can be explored, including the heterogeneous nature of communication capacity and the heterogeneity of communication latency. In this paper, we propose a learning-based clustering method for categorical data analysis and unknown k. Existing FC approaches usually assume that the clients are synchronous and the server is located at the same place. The proposed method can handle both the clients and server in asynchronous mode. The proposed method can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed method can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

CCF-A类

**AAAI'25
本科生一作**

Asynchronous Federated Clustering with Unknown Number of Clusters

Yunfan Zhang¹, Yiqun Zhang², Yang Lu³, Mengke Li⁴, Xi Chen⁵, Yiqing Cheung⁶

¹Fujian Key Laboratory of Sensing and Computing for Smart City, School of Mathematics, Xiamen University, Xiamen, China; ²College of Computer Science and Technology, Xiamen University, Xiamen, China; ³Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China; ⁴College of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China; ⁵College of Computer Science, Xiamen University, Xiamen, China; ⁶College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Federated Clustering (FC) is crucial to mining knowledge from unlabeled non-independent identically Distributed (IID) data provided by multiple clients. In this paper, we propose a learning-based clustering method for asynchronous federated clustering (AFC). It spreads the excessive number of seed points among the clients and provides information to the servers. The proposed method is based on the Federated Learning (FL) algorithm. It can handle both the clients and server in asynchronous mode. The proposed method can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed method can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

CCF-A类

**SIGKDD'24
本科生一作**

Graph Representation Learning for Feature Data Clustering

Zuyu Jia¹, Rong Zhou², Junjie Jiang³, Yiqun Zhang⁴, Yiqing Cheung⁵

¹College of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China; ²College of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China; ³College of Computer Science, Xiamen University, Xiamen, China; ⁴College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁵College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Graph representation learning has been widely used in various applications, such as recommendation systems, social network analysis, and clustering. In this paper, we propose a learning-based graph representation learning framework for feature data clustering. The proposed framework can handle both the clients and server in asynchronous mode. The proposed framework can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed framework can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

CCF-A类

**CAIS 2024
本科生一作**

Heterarchical Merging for Accurate Mixed Spatial Dependency Learning

Zhang¹, Yiqun Zhang², Jian Zhu³, Zhixin Huang⁴

¹College of Computer Science and Technology, Xiamen University, Xiamen, China; ²College of Computer Science and Technology, Xiamen University, Xiamen, China; ³College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁴College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—In this paper, we propose a learning-based merging framework for mixed spatial dependency learning. The proposed framework can handle both the clients and server in asynchronous mode. The proposed framework can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed framework can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

二区SCI

**ICBHI 2024
本科生一作**

Multimodal Medical Image Segmentation with Cross-Spatial Dependency Learning

Yiqun Zhang¹, Senjian Zhang², Guo Zheng³, Jian Zhu⁴, Zhixin Huang⁵

¹College of Computer Science and Technology, Xiamen University, Xiamen, China; ²College of Computer Science and Technology, Xiamen University, Xiamen, China; ³College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁴College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁵College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—Medical image segmentation is a critical task in medical image processing. In this paper, we propose a learning-based cross-spatial dependency learning framework for multimodal medical image segmentation. The proposed framework can handle both the clients and server in asynchronous mode. The proposed framework can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed framework can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

一区SCI

**DOCS'24
本科生一作**

Multi-dimensional Optimal Goods Systems (DOCS) Clustering with Number of Clusters

Yiqun Zhang¹, Senjian Zhang², Guo Zheng³, Jian Zhu⁴, Zhixin Huang⁵

¹College of Computer Science and Technology, Xiamen University, Xiamen, China; ²College of Computer Science and Technology, Xiamen University, Xiamen, China; ³College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁴College of Computer Science and Technology, Xiamen University, Xiamen, China; ⁵College of Computer Science and Technology, Xiamen University, Xiamen, China

Abstract—In this paper, we propose a learning-based DOCS clustering framework for multi-dimensional goods systems. The proposed framework can handle both the clients and server in asynchronous mode. The proposed framework can also handle the client's local features and the server's global features for each update. As a result, the clients gradually adapt to each other to reveal a promising learning strategy for experiments. Experiments show that the proposed framework can achieve better clustering performance and highly outperform the state-of-the-art counterparts on various datasets.

IEEE 最佳论文奖 Best Paper Award

DOCS'24

团队取得论文成果17篇 其中10篇为本科生一作
另有高水平论文在投6篇

▶ 学术认可

项目简介 技术创新 成果展示 团队介绍 拓展延续



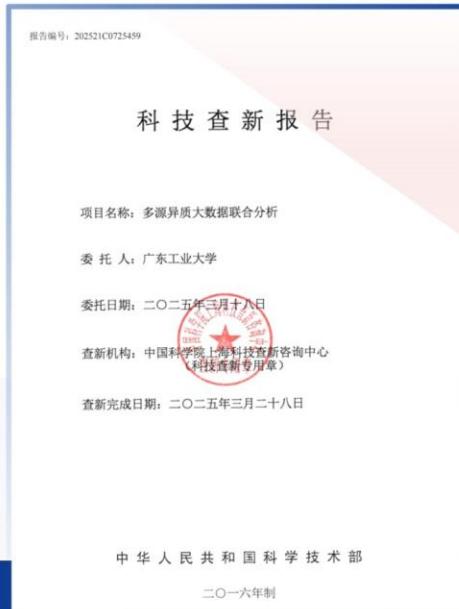
查新项目名称

多源异质大数据联合分析



查新结论

国内外文献检索中未见相同文献报道
本查新项目具有新颖性



论文引用成果

均发表于2024-2025年
累计被引次数100+

获知名学者多次正面引用与评价

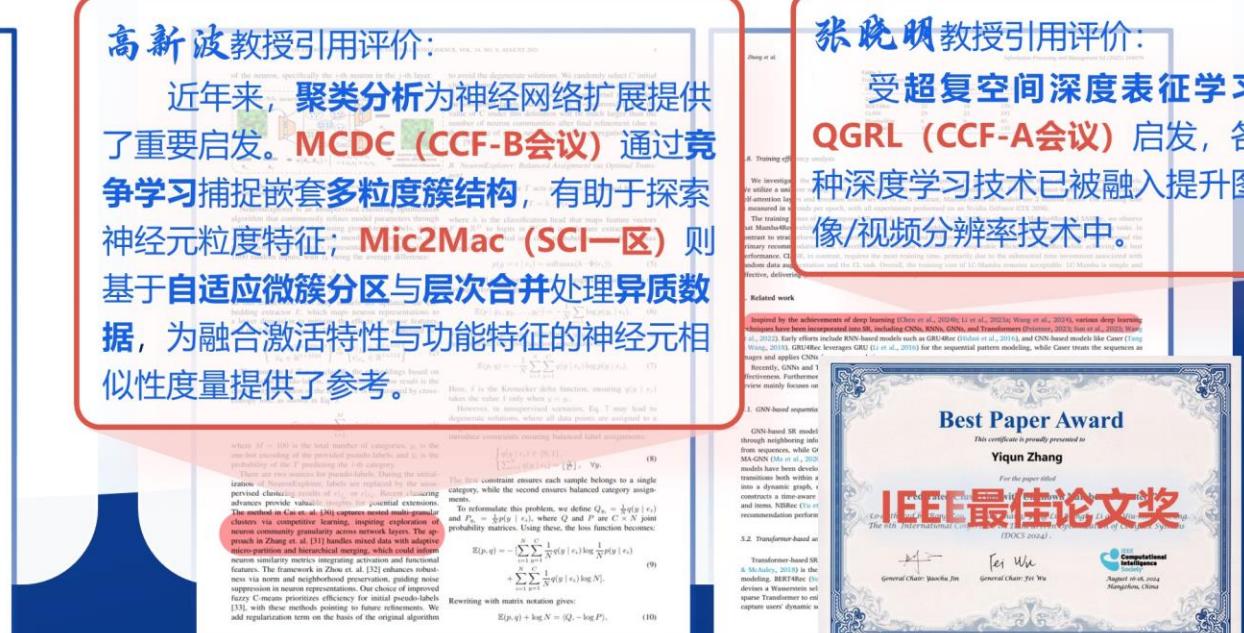
- 教育部长江学者 香港浸会大学讲席教授 张晓明
- 国家杰青 西安电子科技大学校长 高新波

高新波教授引用评价：

近年来，聚类分析为神经网络扩展提供了重要启发。**MCDC (CCF-B会议)**通过竞争学习捕捉嵌套多粒度簇结构，有助于探索神经元粒度特征；**Mic2Mac (SCI一区)**则基于自适应微簇分区与层次合并处理异质数据，为融合激活特性与功能特征的神经元相似性度量提供了参考。

张晓明教授引用评价：

受超复空间深度表征学习
QGRL (CCF-A会议)启发，各种深度学习技术已被融入提升图像/视频分辨率技术中。



已录用论文成果列表

项目简介 技术创新 成果展示 团队介绍 拓展延续



No.	名称		No.	名称	
1	Yiqun Zhang, ..., Zexi Tan, Xiaopeng Luo, Yuzhu Ji*, "Learning SelfGrowth Maps for Fast and Accurate Imbalanced Streaming Data Clustering", <i>TNNLS</i> , 2025. 探索动态数据分布自适应的高精度高效率非平衡数据聚类算法	一区SCI IF: 8.9	9	Yunfan Zhang, Yiqun Zhang*, Yang Lu, Mengke Li, Yiu-ming Cheung, "Asynchronous Federated Clustering with Unknown Number of Clusters", <i>AAAI'25</i> . 本地节点异步传输且合理簇数未知情形下的联邦聚类范式	CCF-A
2	Yuzhu Ji, Wei Hu, Guoji Gan, Zhiquan Long, Yiqun Zhang*, An Zeng*, "SMACNet: A Unified Framework for One-shot Talking Head Synthesis via Subtle Motion and Appearance Compensation", <i>TCE</i> , 2025. 提出能够捕捉动态数据中细微变化的深度表征与生成模型	一区SCI IF: 10.9	10	Junyang Chen, Yuzhu Ji, Rong Zou, Yiqun Zhang, Yiu-ming Cheung, "QGRL: Quaternion Graph Representation Learning for Heterogeneous Feature Data Clustering", <i>SIGKDD'24</i> . 引入超复空间深度表征学习机制突破异质特征数据聚类精度瓶颈	CCF-A
3	Junyang Chen, ..., and Zhixin Huang, "Quaternion Cross-Modality Spatial Learning for Multi-Modal Medical Image Segmentation", <i>JBHI</i> , 2024. 面向医学影像分割应用的距离度量空间学习深度模型	一区SCI IF: 6.8	11	Shenghong Cai, Yiqun Zhang*, ..., Hong Jia, Peng Liu, "Robust Categorical Data Clustering Guided by Multi-Granular Competitive Learning", <i>ICDCS'24</i> . 拓展多粒度竞争学习理论提升聚类分析鲁棒性	CCF-B
4	Rong Zou, Yunfan Zhang, ..., Zexi Tan, Yiqun Zhang* and Yiu-ming Cheung*, "SDENK: Unbiased Subspace Density-k-Clustering", <i>Neurocomputing</i> , 2025. 密度分布学习与子空间学习融合互补的无偏聚类算法	一区SCI IF: 6.5	12	Yunfan Zhang, ..., Yiqun Zhang*, Yiu-ming Cheung, "Towards Unbiased Minimal Cluster Analysis of Categorical-and-Numerical Attribute Data", <i>ICPR'24</i> . 提出异质特征数据最小簇理论实现低归纳偏执聚类分析	CCF-C
5	Yunfan Zhang, Rong Zou, Yiqun Zhang*, ..., "Adaptive Micro Partition and Hierarchical Merging for Accurate Mixed Data Clustering", <i>CAIS</i> , 2024. 提出数据微簇划分机制，引导高精度异质数据聚类	二区SCI IF: 4.6	13	Pengkai Wang†, Yunfan Zhang†, Yiqun Zhang*, ..., Yiu-ming Cheung, "Clustering by Learning the Ordinal Relationships of Qualitative Attribute Values", <i>IJCNN'24</i> . 进行异质特征数据取值顺序关系学习以增强聚类精度	CCF-C
6	Chuyao Zhang, Xinxi Chen, Zexi Tan, ..., Yuzhu Ji, Yiqun Zhang*, "Towards Clustering of Incomplete Mixed-Attribute Data", <i>Expert Systems</i> , 2025. 对异质特征数据进行缺失值补全以增强聚类精度	二区SCI IF: 2.3	14	Haoyi Xiao, Xinxi Chen, Xiaopeng Luo*, ..., Wei Ai, "MACL: Metric and Attribute Space Co-Learning for Qualitative Data Clustering", <i>ICIC'25</i> . 提出度量空间和属性子空间协同学习机制增强聚类精度	CCF-C
...
8	Rong Zou, Yunfan Zhang, Yiqun Zhang, ..., Yiu-ming Cheung*, "Federated Clustering with Unknown Number of Clusters", <i>IEEE DOCS'24</i> . 针对普遍的真实簇数未知情形提出簇分布自探索联邦聚类算法	EI IEEE 最佳论文奖	17	Zexi Tan, Tao Xie, Binbin Sun, Xiang Zhang, Yiqun Zhang* and Yiu-ming Cheung, "MEET-Sepsis: Multi-Endogenous-View Enhanced Time-Series Representation Learning for Early Sepsis Prediction", <i>PRICAI'25</i> . 面向脓毒症早期预测的多源内生视图增强时序表征学习	CCF-C

▶ 专家推荐

项目简介 技术创新 成果展示 团队介绍 拓展延续



沈卫明

加拿大工程院院士
华中科技大学教授
福耀科技大学讲席教授

推荐理由
摘选

“对无监督联邦学习和异质数据分析的研究具有前沿性和前瞻性，项目所提出的方案为解决该领域的关键基础性技术难题提供了创新思路。”



蔡宏民

国家杰出青年科学基金获得者
华南理工大学教授

推荐理由
摘选

“成果构成了联邦异质数据分析方法体系，理论性和科学性较强，研究具有前瞻性，拓展了数据科学和机器学习研究领域。在许多重要行业有良好应用前景，有望提升行业应用中的数据分析效能。”

▶ 团队成员

项目简介 技术创新 成果展示 **团队介绍** 拓展延续



张云帆
(负责人)

- 录用CCF-A会议一篇 (一作)
- 录用一区SCI期刊一篇 (一作)
- 录用一区SCI期刊一篇 (第一本科生作者)
- 录用四区SCI期刊一篇 (三作)
- 录用CCF-C会议两篇 (一作)
- 录用IEEE会议一篇 (获最佳论文奖)
- 获得本科生国家奖学金
- 已被香港高校直博全奖录取



谭泽熙

- 录用一区SCITop期刊一篇 (第一本科生作者)
- 录用一区SCI期刊一篇 (三作)
- 录用二区SCI/CCF-C期刊一篇 (三作)
- 录用四区SCI期刊一篇 (第一本科生作者)
- 录用CCF-C会议论文一篇 (一作)
- 录用EI会议论文一篇 (一作)
- 国家级大创项目第一负责人
- 计算机设计大赛国赛一等奖第一负责人
- 广东省级科技创新项目第一负责人



陈俊仰

- 录用一区SCI期刊一篇 (一作)
- 录用CCF-A会议一篇 (一作)
- 在投CCF-A会议一篇 (一作)
- 在投一区SCI期刊一篇 (二作)
- 保研至清华大学



陈欣禧

- 录用二区SCI/CCF-C期刊一篇 (第一本科生作者)
- 录用CCF-C会议一篇 (二作)
- 二审一区SCITop期刊一篇 (二作, 指导老师一作)
- 在投一区SCI期刊一篇 (一作)
- 国家级大创项目第一负责人



蔡升宏

- 录用CCF-B会议一篇 (一作)
- 在投CCF-A会议一篇 (一作)
- 在投一区SCITop期刊一篇 (二作)
- 国家级大创项目第一负责人
- 多所港澳高校直博全奖条件录取



甘国基

- 录用一区SCI期刊一篇 (第一本科生作者)
- 在投CCF-B会议一篇
- 实审&授权国家发明专利各一项
- 保研至广东工业大学



龙志全

- 录用一区SCI期刊一篇 (四作)
- 在投CCF-B会议一篇
- 实审&授权国家发明专利各一项

▶ 应用价值

项目简介 技术创新 成果展示 团队介绍 拓展延续



现实应用价值

✓ 与各三甲医院展开广泛合作与应用部署

(如广州医科大学附属第三医院、深圳市妇幼保健院)

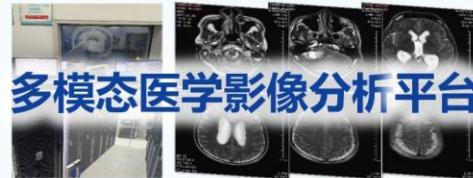
✓ 成果价值广受合作单位认可

预处理 + 预分析

医院本地数据

提取 + 融合

复杂临床信息



联合分析

甲方医院-乙方企业

国家发明专利

实审

已授权

通用分析工具

数据约简

簇数自动判定

异常检测

数据可视化

医学领域应用

辅助诊断

疾控与健康数据分析

国防军工领域应用

网络攻击检测

情报分析与整合

金融领域应用

信用风险评估

实时欺诈检测

显著实际意义 & 广泛应用价值

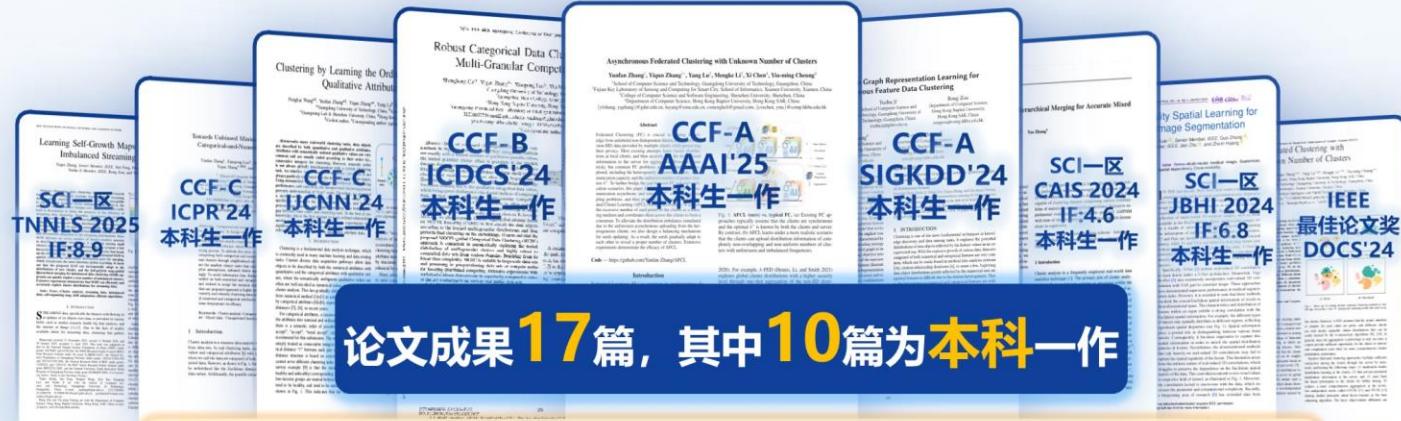
应用拓展



多源异质大数据 联合分析

智联异质数据孤岛·赋能精准分析决策

敬请各位
评委老师指导



团队已有成员保送至清华大学、全奖直博至香港高校

科学性

面向应用需求和技术现状解决系列**关键科学问题**:

- 率先统一**异质特征数据距离度量**
- 创新**自组织联邦聚类策略**
- 提出**原创持续联邦学习模型**

先进性

被知名学者**正面评价引用**, 查新报告证实**新颖性**:

- 异质数据基础测度: 信息融合损失降低高达**37.49%**
- 多源数据联邦聚类: 知识嗅探精度高达**90.05%**
- 动态环境持续学习: 知识更新时效提升高达**74.27%**

拓展性

- 为疾控健康数据分析、信用风险评估等**诸多领域**提供**通用方法和技术支持**