

基于图嵌入的个性化心力衰竭管理运动处方推荐系统构建^{*}

张珂 鲍婷 吴蓉蓉 吾尔满 沈百荣

(四川大学华西医院/疾病分子网络前沿科学中心疾病系统遗传研究院 成都 610212)

〔摘要〕 目的/意义 促进心力衰竭 (heart failure, HF) 管理运动训练的临床实践。方法/过程 系统分析并整合相关文献、临床指南及专家共识, 构建 HF 与运动训练知识图谱 (knowledge graph, KG)。基于该 KG, 采用快速随机投影算法和 K-邻近算法, 构建图嵌入的推荐模型和个性化 HF 管理运动处方推荐系统。结果/结论 HF 与运动训练的 KG 共包括 2 703 个实例和 25 161 条关系。基于该 KG, 图嵌入的个性化推荐系统可提供安全、有效、多样的运动处方推荐。

〔关键词〕 心力衰竭; 运动处方; 医学知识图谱; 图嵌入; 推荐系统; 决策支持

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.06.012

Construction of a Graph Embedding Recommendation System for Personalized Exercise Prescription in Heart Failure Management

ZHANG Ke, BAO Ting, WU Rongrong, WU Erman, SHEN Bairong

Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network/West China Hospital, Sichuan University, Chengdu 610212, China

〔Abstract〕 **Purpose/Significance** To promote the clinical practice of exercise training intervention in the management of heart failure (HF). **Method/Process** The relevant literatures, clinical guidelines, and expert consensus are systematically analyzed and integrated, and a knowledge graph (KG) of HF and exercise training is constructed. Based on the KG, a graph embedding recommendation model and a personalized exercise prescription recommendation system for the management of HF are constructed by using the fast random projection algorithm and the K-nearest neighbors algorithm. **Result/Conclusion** In total, the KG of HF and exercise training includes 2 703 instances and 25 161 relations. Based on the KG, the graph-embedded personalized recommendation system provides safe, effective, and diverse exercise prescription recommendations.

〔Keywords〕 heart failure (HF); exercise prescription; medical knowledge graph (KG); graph embedding; recommendation system; decision support

1 引言

1.1 研究背景

心力衰竭 (heart failure, HF) 是各种病因导致的心室功能障碍引起的一组综合征, 是大多数心血管疾病 (cardiovascular disease, CVD) 的终末期。

〔修回日期〕 2022-12-21

〔作者简介〕 张珂, 硕士研究生; 通信作者: 沈百荣, 博士, 教授。

〔基金项目〕 国家自然科学基金项目 (项目编号: 32270690)。

近年来，随着人口老龄化，HF 死亡率和患病率都逐年升高^[1]。全球疾病负担研究显示，近 20 年来，全球 HF 患者绝对数量增加了 91.9%，且其中 29.9% 来自中国。无论对中国还是全球而言，HF 已造成严重的医疗和社会负担，如何有效防治 HF 已成为目前亟须解决的公共卫生问题之一^[2]。

既往研究表明，定期、系统的体力活动可以有效降低 HF 风险和死亡率，且随着研究的深入，其安全性和有效性得到进一步证实^[3]。越来越多的 CVD 指南将运动训练作为 HF 的重要预防干预措施，鼓励大部分 HF 患者长期运动训练^[4]。然而，针对 CVD 的运动训练临床实践情况并不理想，仅有 30%~40% 的临床医生或健康管理人士推荐其患者进行锻炼。造成这一问题的原因之一是现有相关指南主要面向普通人群，忽视了疾病严重程度、药物摄入和不同风险因素等个体特异性^[4]。因此，在临床实践中，医生难以仅仅依靠指南为具有复杂生理情况的患者开具安全有效的运动处方^[3]。

随着计算机技术的发展，数字医疗系统为这一困境提供了可能的解决方案^[5-7]。由欧洲预防心脏病协会与比利时哈塞尔特大学合作开发的首个交互式运动处方临床决策支持系统——运动处方日常实践和康复训练工具，通过综合考虑 CVD 的不同病种、不同风险因素和不同用药等多种临床情况，结合患者具体生理状况，自动生成运动训练推荐，为临床工作者提供有效和安全的运动处方决策支持^[8]。此外，由 Pescatello L S 等^[9]以美国运动医学学会和美国心脏协会的指南为基础开发的运动处方临床决策支持系统，优先个性化处方运动，通过综合考虑糖尿病、血脂异常、高血压和肥胖等多种 CVD 危险因素、官方指南和专家共识，为患者制定个性化运动处方，为医生和相关从业人员提供有效指导。

虽然以上两款解决方案已经初步考虑了患者的异质性，但其算法逻辑和数据支持仍然是基于临床指南和路径，与现有指南面临同样问题，即推荐结果过于广泛单一，难以为具有复杂生理情况的个体提供安全有效的个性化运动指导。

1.2 研究内容

为解决以上问题，本研究设计并构建针对 HF

管理的运动处方个性化推荐系统，基于标准化的体力活动和心力衰竭关系（physical activity – heartfailure, PAHF）数据集，设计并构建 PAHF 知识图谱，结合相似性算法，开发基于图嵌入的个性化运动处方推荐系统，为临床医生、科研工作者、公众提供可解释的个性化运动处方决策支持，从而促进运动训练防治 HF 的临床实践。

2 方法

2.1 系统构建流程

基于图嵌入的个性化 HF 管理运动处方推荐系统的构建流程主要包括知识图谱构建、图嵌入的推荐模型构建和用户交互界面构建 3 部分。首先从前期工作开源体力活动和心力衰竭关系知识库（physical activity – heart failureknowledge graph, PAHFKB）获取标准化 PAHF 数据集^[10]，基于该数据集构建 HF 与运动训练的知识图谱。然后采用投影算法对知识图谱进行表征，并以低维向量形式嵌入基于协同过滤的个性化推荐模型。最后基于该模型集成浏览器/服务器模式的用户交互界面，用于采集用户信息，反馈并可视化推荐结果，见图 1。

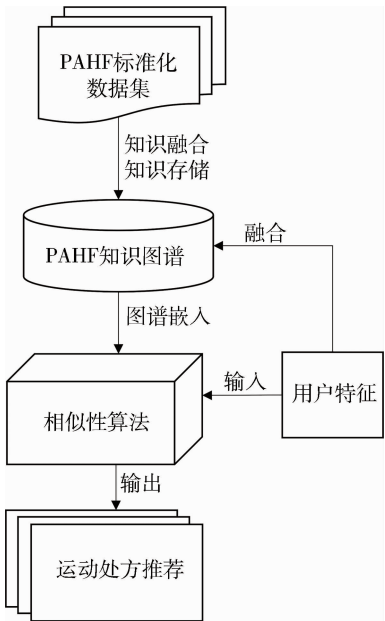


图 1 基于图嵌入的个性化心衰管理运动处方推荐系统构建流程

2.2 数据来源

本研究中知识图谱数据来源于 PAHFKB 知识库。依据现有文献、指南和共识，制定 PAHF 数据描述标准，并基于该标准，从相关文献中手动提取整合结构化 PAHF 数据集。

PAHF 数据集共纳入 1989—2020 年发表的 357 篇文献（原始研究 258 篇，系统综述/Meta 分析 99 篇），涉及来自 42 个国家的 900 000 余名研究对象（年龄 18 ~ 88 岁，女性 47%，男性 53%，无 HF 5%，纽约心脏病学会心功能分级 I 12%，分级 II 39%，分级 III 39%，分级 IV 5%）。该数据集共涵盖 1 010 条结构化 PAHF 关系数据，357 种 HF 防治的运动方案，333 种 HF 与运动训练相关的结局指标和 40 种风险因素组合。

2.3 知识图谱设计

知识图谱的构建包括本体（数据标准）构建、知识获取（数据采集）、知识融合和知识存储 4

步^[11]。其中数据标准构建和数据采集部分，已于前期工作 PAHFKB 中完成^[10]。基于 PAHFKB 中的标准化数据集，知识图谱构建包括知识融合和知识存储两部分，见图 2。

其中知识融合主要包括重复知识合并和实体关系重构，根据先验知识和临床决策需要，对 PAHF 中实体和关系采用三元组结构进行标准化描述。着重关注如何在对 PAHF 进行标准化的同时充分考虑个性化特征信息，因此在 PAHF 知识图谱中除了基本的 HF 和运动训练相关概念（HF 类型、HF 分级、HF 病因、运动方案、运动强度、运动频率、运动类型、运动方式和运动时长）外，还包括与运动处方决策相关的个体特异性概念：共病、用药史、手术、身体质量指数（body mass index，BMI）、年龄、性别、种族和地区。在知识存储部分，采用 Neo4j 图数据库管理系统实现基于图模型的存储。相比传统的关系模型存储，图模型能更直观地表达和处理节点和关系数据，且支持多种数据挖掘算法，是目前主流的知识图谱存储方式^[11]。

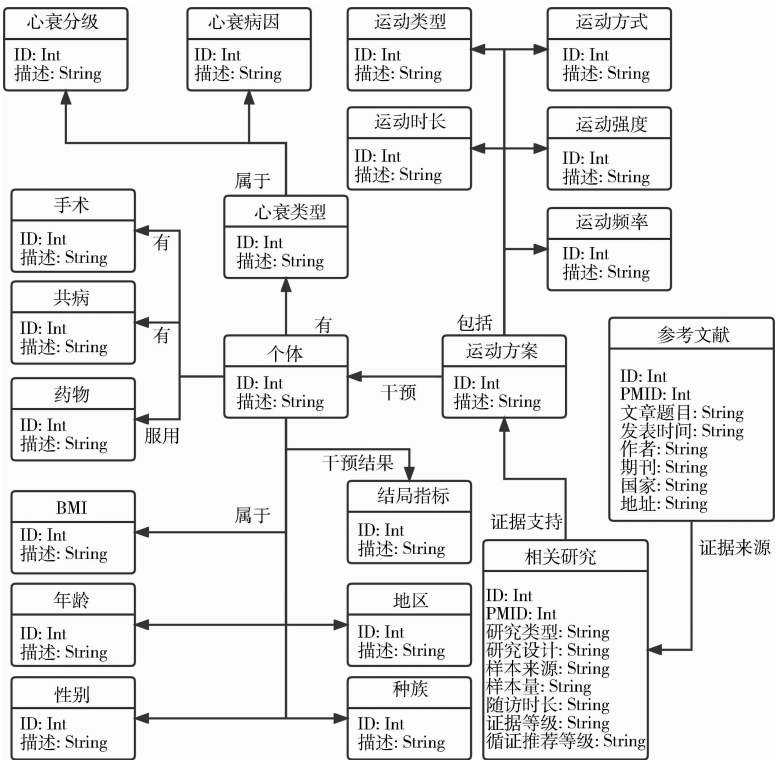


图 2 知识图谱逻辑模型

2.4 推荐模型设计

推荐系统算法主要包括：基于内容、基于协同过滤和混合模型 3 类^[12]。本研究采用基于特征协同过滤的推荐算法：通过计算知识图谱中已有“个体”节点与输入用户特征之间相似度寻找与用户的相似“个体”，再利用相似“个体”关联的运动处方为用户推荐。

该推荐算法主要包括图嵌入和推荐两个核心模块。首先，在图嵌入模块判断知识图谱是否已存在具有完全相同用户特征的“个体”节点，若存在则直接返回关联运动处方。否则将用户特征作为新节点嵌入知识图谱，并采用投影算法，将知识图谱的节点和关系信息映射在低维空间，表征其中的关键信息，并以向量形式存储。之后，在推荐模块中，将知识图谱的特征向量嵌入相似性算法，计算各节点之间特征相似性。最后，根据各节点的相似性评分给出适合的运动处方推荐，具体算法如下。

```
输入：用户个性化特征 UserInfor
输出：个性化运动处方推荐 ExRset
1 ifUserInfor exists in the KG then
2 ExRset = KG [UserInfor]. ExRset
3 else
4     merge UserInfor as a new node into KG
5 FastRP graph embedding
6     KNN similarity calculation
7 foreach KG [Individual] as n
8 n. KNN_ SIMILARITY > 0. 95
9 KG [UserInfor]. ExRset = n. ExRset
10 ExRset = KG [UserInfor]. ExRset
11 return ExRset
```

在图嵌入模块中，采用快速随机投影算法（fast random projection，FastRP）表征知识图谱。FastRP 是基于线性代数约翰逊－林登斯特劳斯引理（Johnson－Lindenstrauss Lemma）的随机投影算法，通过非常稀疏随机投影，在准确保留节点关系之间特征相似性的同时，快速将知识图谱投影到低维向量^[13]。相比于基于神经网络的传统图嵌入算法（如 Node2Vec 和 GraphSAG），FastRP 大幅缩短运算时间，并且同样保证投影结果质量。在推荐模块

中，采用无监督机器学习中的 K－邻近算法（K－nearest neighbors algorithm，KNN）计算相似性，通过比较“用户”节点和知识图谱已有“个体”节点特征间的特征计算相似度并推荐^[12]。

2.5 系统架构设计

为了确保用户能随时随地访问和使用个性化运动处方推荐系统，采用浏览器/服务器模式构建网页客户端。主要分为 4 层：底层操作系统采用 Windows 作为服务器、知识存储层采用 Neo4j 实现知识存储和检索、业务处理层采用基于 MVT 模式的 Flask 架构实现算法逻辑和数据处理、顶层用户层由 Bootstrap 和 Cytoscape 框架实现浏览器端的用户交互，见图 3。

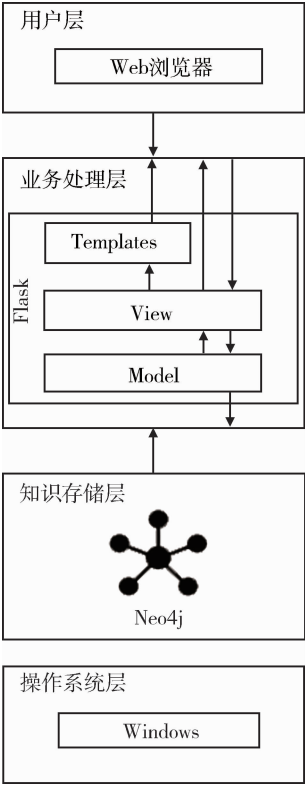


图 3 系统架构

3 结果

3.1 知识图谱数据统计

PAHF 知识图谱从结构化数据集 PAHF 中共抽

取 2 703 个实例和 25 161 条关系。其中个体相关实例 12 类，相关关系 4 类；运动干预相关实例 7 类，相关关系 3 类；与证据支持相关实例 2 类，相关关系 2 类，见表 1、表 2。

表 1 PAHF 知识图谱实例抽取结果

名称	数据量（个）	分类
个体	1 012	个体信息
心衰类型	5	个体信息
心衰分级	5	个体信息
心衰病因	10	个体信息
共病	24	个体信息
药物	8	个体信息
手术	7	个体信息
BMI	5	个体信息
年龄	5	个体信息
性别	3	个体信息
地区	31	个体信息
种族	13	个体信息
运动方案	358	运动信息
运动频率	51	运动信息
运动类型	8	运动信息
运动强度	5	运动信息
运动时长	83	运动信息
运动方式	23	运动信息
结局指标	333	运动信息
相关研究	357	证据信息
参考文献	357	证据信息
合计	2 703	

表 2 PAHF 知识图谱关联抽取结果

名称	定义域	值域	数据量（个）
属于	个体	心衰分级	11 632
		心衰病因	
		年龄	
		性别	
		种族	
		地区	
		BMI	
有	个体	心衰类型	3 830
		共病	
		手术	

续表 2

名称	定义域	值域	数据量（个）
服用	个体	药物	3 573
干预	运动方案	个体	533
包括	运动方案	运动频率	3 487
		运动强度	
		运动时长	
		运动类型	
		运动方式	
干预结果	个体	结局指标	1 216
证据支持	相关研究	运动方案	533
来源	参考文献	相关研究	357
合计			25 161

3.2 推荐系统应用

推荐系统用户交互模块主要包括用户信息采集、运动处方推荐和相关知识图谱可视化。

首先，用户在信息采集模块输入个体特征数据：心衰类型、心衰分级、心衰病因、共病、用药史、手术史、BMI、年龄、性别、地区和种族。采集完成之后，推荐模块根据用户特征数据，对知识图谱中已有数据进行匹配，并计算用户特征与知识图谱中“个体”节点间的相似性系数，选取其中完全匹配的节点所关联运动处方作为主要推荐结果，并将与用户特征相似性系数大于 0.95 的“个体”节点关联的运动处方作为扩展推荐结果。最后，推荐模块将运动处方相关节点和关系数据返回可视化模块进行展示。该模块还提供结果视图切换、数据筛选比较和推荐结果下载功能，以此帮助用户灵活对比和分析推荐结果。

以输入用户特征（Individual. id = 837）：“HF（心衰类型）”“NYHA II 级（心衰分级）”“糖尿病（共病）”“冠状动脉旁路移植（手术史）”“ACE - I（用药史）”“老年人（年龄）”为例。推荐系统共推荐运动处方 2 例，包括 2 种运动频率，2 种运动强度，3 种运动时长和 4 种运动类型供用户选择，见图 4。

上述推荐结果来自 3 个具有证据支持的“个体”节点，它们与用户节点特征相似性评分依次为

0.994 1 (Individual. id = 2076)、0.988 8 (Individual. id = 1942) 和 0.988 0 (Individual. id = 1944)。其中“个体”节点“1944”和“1942”及其相关运动处方, 来源于同一研究 (Reference. pmid =

21094549), “个体”节点“2076”来源于另一项研究 (Reference. pmid = 23690816)。根据循证医学评分标准, 这两项研究的证据等级和推荐等级分别为: b, Ib 和 c, IIb。

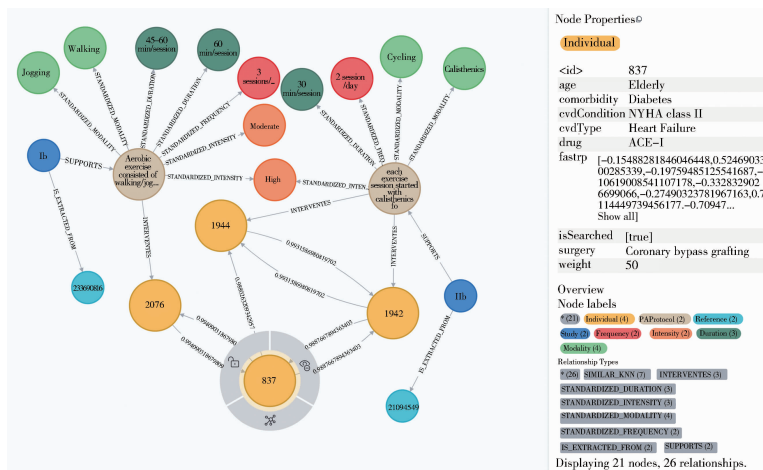


图4 个性化运动处方推荐系统推荐结果可视化

3.3 应用效果评价

传统推荐系统主要应用于“用户 - 商品”推荐模式,通过分析用户历史行为和偏好,从众多信息中为其推荐可能感兴趣的 商品或内容。用户对推荐结果的满意度是衡量传统推荐系统表现的主要指标。而个性化运动处方推荐系统旨在根据用户个人生理特征,为其推荐安全有效的个性化运动处方。因此,用户满意度或喜好并非首要评价指标。基于 HF 运动管理的应用场景,对个性化运动处方推荐系统的主要考察指标是其推荐结果的安全性、有效性和多样性^[14]。从前文图 2、图 4 可知,该系统推荐结果均提取于有证据支持的已发表文献,保证了推荐结果的安全性和有效性。此外,该系统在给出完全符合用户特征推荐的同时,基于相似性算法为用户推荐可能适合的运动处方,并给出运动处方的证据支持和相关参考文献信息,在保证推荐结果科学性的同时为用户提供更多选择。

和高风险人群的安全性和益处，临床指南也鼓励大部分稳定期的 HF 患者进行运动训练。但现有运动相关指南主要是面向人群的，无法为具有复杂情况的个体提供安全有效的个性化运动指导，因而运动训练的临床实践情况并不理想。

为促进 HF 运动管理的临床实践,本研究基于结构化 PAHF 数据集,构建了 HF 与运动训练的知识图谱,并采用 FastRP 和 KNN 算法实现基于图嵌入的个体特征协同过滤运动处方推荐系统,为临床工作者提供了安全有效的个性化运动处方推荐。相比于指南,本研究充分考虑个体异质性,可以提供更加详细多样的个性化运动处方推荐。此外,在保证推荐结果安全有效的前提下,考虑个体偏好,为其提供多种运动类型,从而提高个体对运动处方的接受度和依从性,实现运动干预的效用最大化。

参考文献

- 1 CLELAND J G F, VAN VELDHUISEN D J, PONIKOWSKI P. The year in cardiology 2018: heart failure [J]. European heart journal, 2019, 40 (8): 651 – 661.
- 2 BRAGAZZI N L, ZHONG W, SHU J, et al. Burden of heart failure and underlying causes in 195 countries and territories from 1990 to 2017 [J]. European journal of preven-

4 结语

尽管越来越多的研究证实运动训练对 HF 患者

- tive cardiology, 2021, 28 (15): 1682 - 1690.
- 3 TAYLOR R S, LONG L, MORDI I R, et al. Exercise - based rehabilitation for heart failure: Cochrane systematic review, meta - analysis, and trial sequential analysis [J]. JACC heart failure, 2019, 7 (8): 691 - 705.
 - 4 PELLICCIA A, SHARMA S, GATI S, et al. 2020 ESC guidelines on sports cardiology and exercise in patients with cardiovascular disease [J]. European heart journal, 2021, 42 (1): 17 - 96.
 - 5 SUTTON R T, PINCOCK D, BAUMGART D C, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success [J]. NPJ digital medicine, 2020, 3 (1): 17.
 - 6 SHEN B. Translational informatics: sports and exercise medicine [M]. Singapore: Springer, 2022.
 - 7 SHEN L, BAI J, WANG J, et al. The fourth scientific discovery paradigm for precision medicine and healthcare: challenges ahead [J]. Precision clinical medicine, 2021, 4 (2): 80 - 84.
 - 8 HANSEN D, DENDALE P, CONINX K, et al. The European Association of Preventive Cardiology exercise prescription in everyday practice and rehabilitative training (EXPERT) tool: a digital training and decision support system for optimized exercise prescription in cardiovascular disease. Concept, definitions and construction methodology [J]. European journal of preventive cardiology, 2020, 24 (10): 1017 - 1031.
 - 9 PESCATELLO L S, WU Y, PANZA G A, et al. Development of a novel clinical decision support system for exercise prescription among patients with multiple cardiovascular disease risk factors [J]. Mayo clinic proceedings innovations, quality & outcomes, 2020, 5 (1): 193 - 203.
 - 10 ZHANG K, BAO T, WU R, et al. PAHFKB: a knowledge base and an online service for personalized physical activity in prevention and intervention of heart failure [EB/OL]. [2022 - 07 - 10]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4147835.
 - 11 谭玲, 鄂海红, 匡泽民, 等. 医学知识图谱构建关键技术及研究进展 [J]. 大数据, 2021, 7 (4): 80 - 104.
 - 12 秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述 [J]. 中国科学: 信息科学, 2020, 50 (7): 937 - 956.
 - 13 CHEN H, SULTAN S, TIAN Y, et al. Fast and accurate network embeddings via very sparse random projection [EB/OL]. [2022 - 08 - 21]. <https://dl.acm.org/doi/10.1145/3357384.3357879>.
 - 14 芦威. 推荐结果多样性的评价方法及优化算法研究 [D]. 北京: 北京交通大学, 2019.

(上接第 71 页)

- 29 CAO J, KARRAS P, KALNIS P, et al. SABRE: a sensitive attribute bucketization and redistribution framework for t - closeness [J]. The VLDB journal, 2011, 20 (1): 59 - 81.
- 30 SORIA - COMAS J, DOMINGO - FERRER J, SANCHEZ D, et al. T - closeness through microaggregation: strict privacy with enhanced utility preservation [J]. IEEE transactions on knowledge and data engineering, 2015, 27 (11): 3098 - 3110.
- 31 FANG Y, ASHRAFI M Z, NG S K. Privacy beyond single sensitive attribute [C]. Berlin: International Conference on Database and Expert Systems Applications, 2011.
- 32 WANG R, ZHU Y, CHEN T S, et al. Privacy - preserving algorithms for multiple sensitive attributes satisfying t - closeness [J]. Journal of computer science and technology, 2018, 33 (6): 1231 - 1242.
- 33 王平水, 王建东. 一种基于聚类的个性化 (1, c) - 匿名算法 [J]. 计算机工程与应用, 2012, 48 (23): 5.
- 34 黄玉蕾, 林青, 戴慧珏. 基于多敏感值的个性化隐私保护算法 [J]. 计算机与数字工程, 2016, 44 (9): 1761 - 1765, 1800.
- 35 朱理奥, 曹天杰. 量化敏感度的个性化数据发布模型 [J]. 计算机应用与软件, 2022 (8): 39.
- 36 SOLANAS A, MARTINEZ - BALLESTE A, DOMINGO - FERRER J. V - MDAV: a multivariate microaggregation with variable group size [C]. Rome: 17th COMPSTAT Symposium of the IASC, 2006.
- 37 徐瑞. 边缘计算中的数据隐私保护技术研究 [D]. 北京: 北京交通大学, 2020.
- 38 唐明坤, 钱庆, 张丽鑫, 等. 生物医学数据匿名化工具 ARX 研究及启示 [J]. 中华医学图书情报杂志, 2022, 31 (2): 19 - 29.