

基于 RDF 的语义知识超图存储研究

宋雪雁, 张伟民, 张祥青

(吉林大学商学与管理学院, 长春 130012)

摘要 针对资源描述框架 (resource description framework, RDF) 存储效率低、难以存储复杂语义关系等问题, 本研究引入超图理论, 探索一种融合超图理论的语义知识图谱存储模型, 以期实现 RDF 对超图数据的存储, 为其他学者利用 RDF 构建知识超图提供参考。本研究构建了适用于超图结构的语义知识超图 (semantic knowledge hypergraph, SKH) 模型, 与语义知识图谱 (semantic knowledge graph, SKG) 进行对比, 分析其存储效率与复杂语义关系存储能力, 并论述其在知识检索、知识推理、数据转换与可视化等方面的应用。研究结果发现, SKH 模型具有比 SKG 更优的存储效率与复杂语义关系存储能力, SKG 的知识检索和知识推理方式也适用于 SKH 模型, SKH 模型数据在一定程度上能与 SKG 数据相互转换, SKH 模型具有更加多元且表意丰富的可视化方式, 对于信息资源管理领域复杂语义存储具有重要意义。

关键词 RDF; 语义知识超图; 知识存储; 超图理论

Storage of Semantic Knowledge Hypergraph Based on a Resource Description Framework

Song Xueyan, Zhang Weimin and Zhang Xiangqing

(School of Business and Management, Jilin University, Changchun 130012)

Abstract: To address the low storage efficiency and difficulty in storing complex semantic relationships in a resource description framework (RDF), the hypergraph theory is introduced to explore a semantic knowledge graph storage model that integrates the hypergraph theory to realize the storage of hypergraph data based on the RDF and provide a reference for other scholars to use the RDF to build knowledge hypergraphs. We construct a semantic knowledge hypergraph (SKH) model suitable for hypergraphs, analyze its storage efficiency and storage capacity of complex semantic relations by comparing it with a semantic knowledge graph (SKG), and discuss its applications in knowledge retrieval, knowledge reasoning, data conversion, and visualization. We established that the SKH model has better storage efficiency and complex semantic relationship storage capacity than the SKG. The methods of knowledge retrieval and knowledge reasoning of SKG are also applicable to SKH. The SKH model data can be transformed into SKG data considerably, and the SKH model provides a diversified and expressive visualization method. It is crucial for complex semantic storage in the field of information resource management.

Keywords: RDF; semantic knowledge hypergraph; knowledge storage; hypergraph theory

收稿日期: 2022-09-30; 修回日期: 2023-01-11

基金项目: 国家社会科学基金一般项目“数据驱动的档案文献资源知识构建与知识服务研究”(21BTQ109)。

作者简介: 宋雪雁, 女, 1980 年生, 博士, 教授, 博士生导师, 主要研究领域为知识组织与知识服务; 张伟民, 通信作者, 男, 1999 年生, 博士研究生, 主要研究领域为知识组织与知识服务, E-mail: zhangweiminjl@163.com; 张祥青, 女, 1993 年生, 博士研究生, 主要研究领域为知识组织与知识服务。

0 引言

资源描述框架 (resource description framework, RDF) 是万维网联盟推荐的一种基于图的数据模型^[1], 其使用统一资源标识符 (uniform resource identifier, URI) 命名事物及其关系, 构成三元组, 该模型允许在不同的应用程序之间混合、开放和共享结构化与半结构化数据, 具有促进数据集成的功能, 在视觉上具有易于理解的优势, 是在 Web 上进行数据交换的标准模型^[2]。语义知识图谱 (semantic knowledge graph, SKG) 是基于 RDF 模型的一种应用, 用于实现知识的存储与呈现、关联与发布, 对于信息资源管理领域数字资源的知识化描述与开发具有重要价值。目前, 语义知识图谱数据层的构建方式通常是基于关系数据库或二维表, 通过 RDB2RDF 映射软件或根据其原理将关系数据库中的元组映射为 RDF 三元组形式。然而, 语义知识图谱间的关联关系大多只能连接两个实体, 难以完全满足传统关系数据库向 RDF 的转换。因此, 需要在符合既有 RDF 标准的基础上, 探索新的知识存储方式, 以实现语义知识图谱对复杂关系的存储与可视化。

超图是图的延伸, 打破了普通图谱只能实现“点-边-点”描述的局限, 使得一条边连接多个顶点成为可能, 也使得以图为拓扑结构的语义知识图谱存储复杂关系成为可能。2019年, 中国电子技术标准化研究院发布的《知识图谱标准化白皮书》中提到, 超图拥有简单图无可比拟的复杂关系表示方式, 将是未来知识存储的研究热点, 超图技术的探索将是知识图谱下一个研究方向, 然而超图还没有较为理想的可视化方案^[3]。据此, 本研究以提升语义知识图谱普适性及其性能为目的, 探索超图理论在语义知识图谱存储中的应用。

1 文献回顾

1.1 知识存储相关研究

知识存储是保存知识化开发成果并提供后续利用的重要环节。目前, 国内外对知识存储进行深入的研究相对较少, 主要停留在存储工具的利用方面, 且对于关系数据库、非关系数据库等知识图谱知识存储方式的观点不一。关系数据方面, 高劲松等^[4]认为, OWL (web ontology language) 格式难以进行后期管理, 建议采用关系型数据库 MySQL 对关联数据进行存储。非关系数据库知识存储主要有

图数据库、RDF 三元组数据库等方式。Zhang 等^[5]认为, 基于图结构的存储是知识图谱研究的主流。张琪等^[6]、Wei 等^[7]、Zhao 等^[8]均使用 Neo4j 存储知识图谱; 还有学者利用图数据库 AllegroGraph 进行知识存储^[9]。杭婷婷等^[10]对比 RDF 存储与图存储方式, 认为 RDF 存储适用于小规模、操作简单的图谱, 表达能力优于图存储。陈涛等则根据不同存储方式对知识图谱进行划分, 认为基于 RDF 模型是语义知识图谱, 也被称为关联数据, 而基于图数据库存储的为广义知识图谱^[11], 并对比关系数据库与三元组数据库, 认为三元组数据具有易扩展、易维护、易访问、易迁移、高并发等优势^[12]。除此之外, 有学者利用非关系数据库 MongoDB 进行知识存储^[13]。

尽管图数据库在图计算、图挖掘等方面具有显著优势, 但对数据缺乏严格的约束, 容易造成数据质量参差不齐; 而基于 RDF 模型并使用 URI 标识实体资源的三元组数据库具有严谨的数据格式和规范的国际标准, 有利于实现语义互联互通。因此, 本研究探讨的是基于 RDF 存储的语义知识图谱。

1.2 超图相关研究

图结构被广泛应用于成对关系的知识存储中, 而某些场景中存在难以用图结构存储的数据, 超图比图更加泛化, 具有强大的表达能力^[14]。在超图有关理论研究方面, 学者们对超图的张量谱性质^[15]、一致超图边连通性与最大边连通性^[16]、稀疏超图^[17]、超网络关键节点识别^[18]、加权超图最优推理近似算法^[19]等问题展开深入探讨, 为基于超图的实践探索提供理论支撑。在超图实践研究方面, 陈文杰^[20]、刘高等^[21]、于亚新等^[22]基于超图分别提出推荐算法 SCRH (scientific collaboration recommendation based on hypergraph)、产品设计知识智能推荐方法、EBSN (event-based social networks) 个性化推荐及优化算法, 并在知识推荐领域取得了一系列成果。此外, 超图还被广泛应用于节点分类^[23]、文本分类^[24]、异质网络社区发现算法^[25]、异构数据可视化分析^[26]等诸多领域。

现有研究通常是基于超图强大的表示能力进行探索, 对超图存储方式的研究相对较少。超图作为一种有限集的组合学, 用于论证组合问题结果有时会更简洁^[27], 而知识图谱知识存储的本质即存储知识及其关联, 可以视作以语义为纽带的复杂知识组合。因此, 引入超图将为语义知识图谱领域带来新

的研究方向。

1.3 超图与知识存储交叉研究

在超图与知识存储交叉领域, 学者们对知识超图架构与实践、知识存储方式等进行了探讨。高峰等^[28]提出了针对金融领域知识的时序超图表示模型; 李豪等^[29]针对电力智库, 将时间、层次、事实三个维度进行组合, 提出三维知识超图架构, 然而并未将其与 RDF 结合, 详细论述其知识存储模式; 田玲等^[30]提出涵盖事理、概念、实例的知识超图三层架构, 该架构将模式层进一步划分为事理层与概念层, 然而并未针对 RDF 的缺陷为实例层知识存储提出具体解决方案。

学者们对知识超图存储方式也进行了一系列探索。Munshi 等^[31]为了加强知识图谱对复杂情景的表示能力, 介绍了一种“超图-图”(hypergraph-graph, HG(2))的数据存储格式, 并尝试将 RDF 与 RDFs 元素映射为 HG(2)^[32]; Chernenkiy 等^[33]认为, RDF 模型缺乏简单的方式来描述多元关系, 并提出应用元图(metagraph)解决 RDF 知识表示局限; Terekhov 等^[34]也提出应用元图进行复杂的知识表示, 元图的知识表示架构改变了 SKG 平面化的结构, 使知识超图不仅可以存储知识, 也可存储多维、多层次知识结构; Krótkiewicz^[35]针对语义知识库设计面向关联元模型的超图数据模型, 提出将边视为角色集的思想。尽管学者们未详细论述如何基于 RDF 数据格式存储元图, 但这种知识结构为本研究探索适用于 RDF 存储格式的知识超图模型提供了借鉴。

总体而言, 对知识图谱知识存储方式的研究相对较少, 通常是对现有存储工具的应用, 少有学者深入分析现有知识存储模式尤其是 RDF 三元组的不足, 进而改进现有存储方式, 而超图相关理论、方法已经相对成熟, 且已有学者从理论框架与存储方式等方面对应用超图理论实现复杂知识存储进行了探索。尽管大多数学者并未详细阐述基于 RDF 的知识超图存储方案, 但其超图架构、HG(2)、元图等均对本研究具有较高的借鉴意义。基于此, 本研究探索将超图理论融合 RDF 三元组的存储方式, 以期实现复杂知识的存储, 并分析如何基于该存储方式进行知识检索、推理、数据转换与可视化。

2 语义知识超图存储模型构建

语义知识图谱通常分为模式层与数据层。模式层是对数据层数据的描述与约束, 内容简洁, 不存

在复杂语义存储问题, 因此, 本研究主要对数据层进行讨论, 数据层可以表示为 $G=\{E,R,F\}$, 其中 E、R、F 分别表示实体集合、关系集合与事实集合, 每条事实以三元组 (h,r,t) 表示^[36]。作为语义知识图谱的拓扑结构, 图(graph)可以表示为 $G=\{V^s,E^s\}$, 其中 V^s 表示节点, E^s 表示边。与图类似, 超图也可表示为 $H=\{V^h,E^h\}$, 其中 V^h 表示节点, E^h 表示超边。当秩为 2 时, 超图就是普通的图^[27], 因此超图可以视为图的一般形式, 图则是特殊的超图。

Munshi 等^[31]提出的 HG(2) 模型将问题空间(problem space)划分为图(problem space graph, PSG)与超图(problem space hypergraph, PSH)两层, 并用连接器(connectors)连接两层数据。然而, 在现实中难以将图与超图完全分离。Chernenkiy 等^[33]、Terekhov 等^[34]均推荐的元图, 是通过元顶点将扁平的图结构立体化, 实现复杂关系的分层与嵌套表示。

本研究综合 HG(2) 与元图模型的优点, 并结合 RDF 语义标准及其三元组存储特点, 构建了简化且适用多场景的语义知识超图(semantic knowledge hypergraph, SKH)存储模型, 该模型可以表示为

$$SKH=\{V,HEV,E,HE\}$$

其中, V 表示超图中的普通节点; HEV 表示特殊的超边节点, 用节点替代超边以实现 RDF 存储超边; E 表示边, 既可以连接普通节点, 也可以连接超边节点, 可以表示为 $E=(v_1,v_2)$, 其中 $v_1,v_2\in V\cup HEV$; HE 表示超边, 用于连接 HEV 及其对应超边上的节点。SKH 模型中, V 与 HEV、E 与 HE 通过本体构建进行语义划分。

SKH 模型如图 1c 所示, 其中普通节点 $v_1,v_2,v_3,v_4,v_5\in V$, 超边节点 $hev_1,hev_2\in HEV$, 边 $e_1,e_2,e_3\in E$, 超边 $he_1,he_2,he_3,he_4,he_5,he_6\in HE$ 。HG(2) 模型实现了图与超图的分层表示, 元图则通过嵌套的方式, 提供多层级的立体知识结构。SKH 模型则有效将 HG(2)、元图模型以及 RDF 数据模式融合, 既保证了超图节点关系的有序性, 又兼顾了元图的灵活性, 同时符合现有 RDF 语义标准, 可以通过三元组的形式进行知识存储。相比于现有 RDF 模型, SKH 可以实现部分节点的聚合, 优化存储空间, 适用于多场景的语义知识存储。

SKH 模型为复杂数据属性的存储提供聚合方式, 如图 2 所示, 当实体具有多组属性相同但属性值不同的三元组时, 将该实体作为所有三元组的主语会出现语义混乱, 难以区分属性值间的组合关

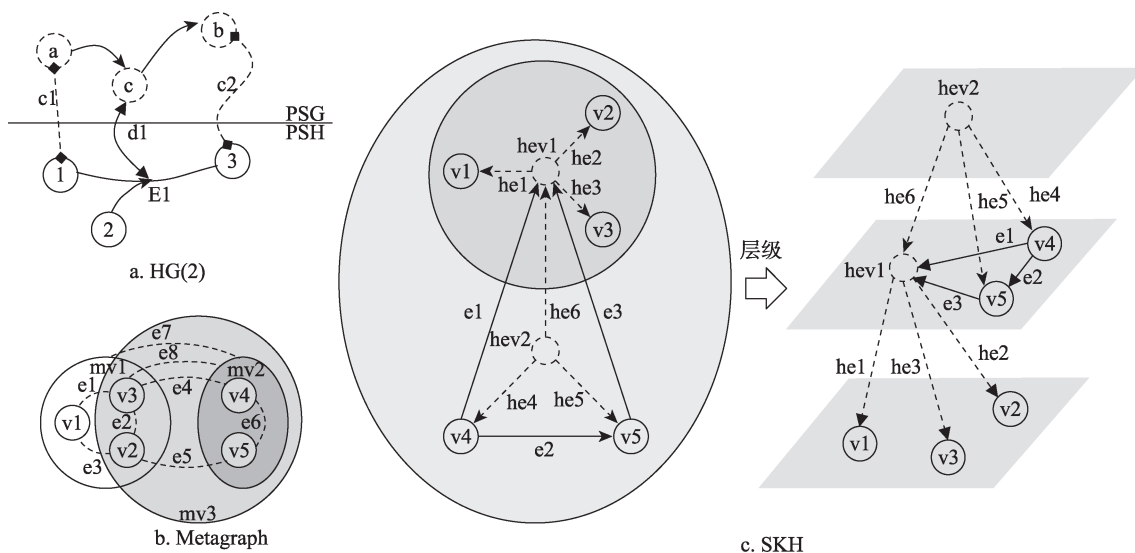
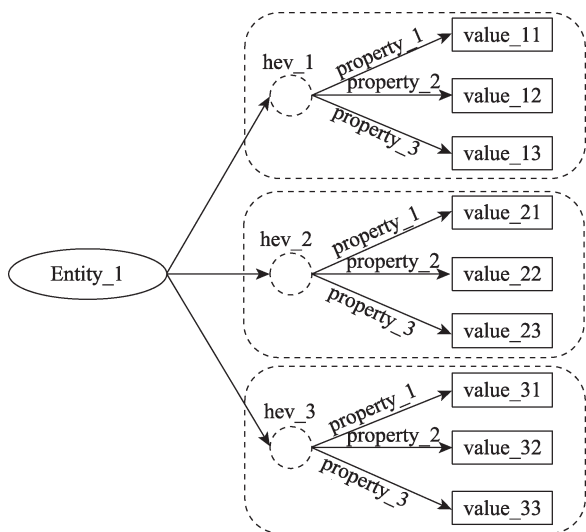
图1 HG(2)^[31]、元图^[33]与SKH模型

图2 SKH模型属性存储

系,而SKH模型中超边节点为各组属性三元组提供属性空间,此时超边节点作为属性的集合,共同构成该实体。因此,在对复杂属性存储时,SKH模型中的超边节点可以视作比实体低一阶的特殊实体。

SKH模型不仅可以优化复杂数据属性的存储方式,还提升了复杂语义关系的存储效率与效果。语义知识图谱的构建通常是基于传统关系数据库,基于SKH模型的知识超图也可以基于传统关系数据库构建。因此,探究SKH模型对复杂语义关系存储的优化,可以围绕关系数据库主要关系类型及其向RDF的映射展开讨论。根据关系涉及的实体类型及其在关系中的语义,本研究将关系数据库实体间联系分为二元实体联系与多元实体联系。其中,多元实体联系又根据同一实体类型中参与构成关系的实

体数量不同分别进行讨论。

2.1 二元实体联系存储解析

二元实体联系指仅涉及两个实体的联系,二元实体联系根据集合中实体的对应关系可划分为一对一联系(1:1)、一对多联系(1:n)以及多对多联系(m:n)三类(图3)。其中,两个实体既可以分属于不同实体型,如书籍与作者之间的联系;也可以同属于一个实体型,如一篇文献与另一篇文献之间具有引用关系,而两篇文献均为同类实体。在关系数据库中,一对一联系与一对多联系的存储模式相似,均可以采用图3d和图3e所示的两种关系模式,而多对多联系只能采用图3e所示的独立关系模式。将二元实体关系由关系数据库映射为RDF较为简洁,而且较难出现语义上的混乱,只需要根据RDB2RDF规则映射为<实体,关系,实体>即可。因此,仅存在二元实体关系时,SKH模型采用与SKG一致的存储方式。

二元实体联系是目前语义知识图谱常涉及的联系类型,然而并非真正反映客观世界的实体联系。时间实体、地点实体与其他多数实体具有联系,但对特定问题进行研究时,会忽略研究意义较小的时间实体、地点实体,或将其设为属性,以简化实体联系;当研究问题需要建立多个实体联系时,大多SKG将其简化为二元实体联系分别建立实体间的关联。例如,书籍、出版时间与出版社之间的联系,往往构建书籍与出版时间、书籍与出版社间的关联,即建立<书籍实体,出版时间,时间实体>和<书籍实体,出版社,出版社实体>两个三元组,而忽视出

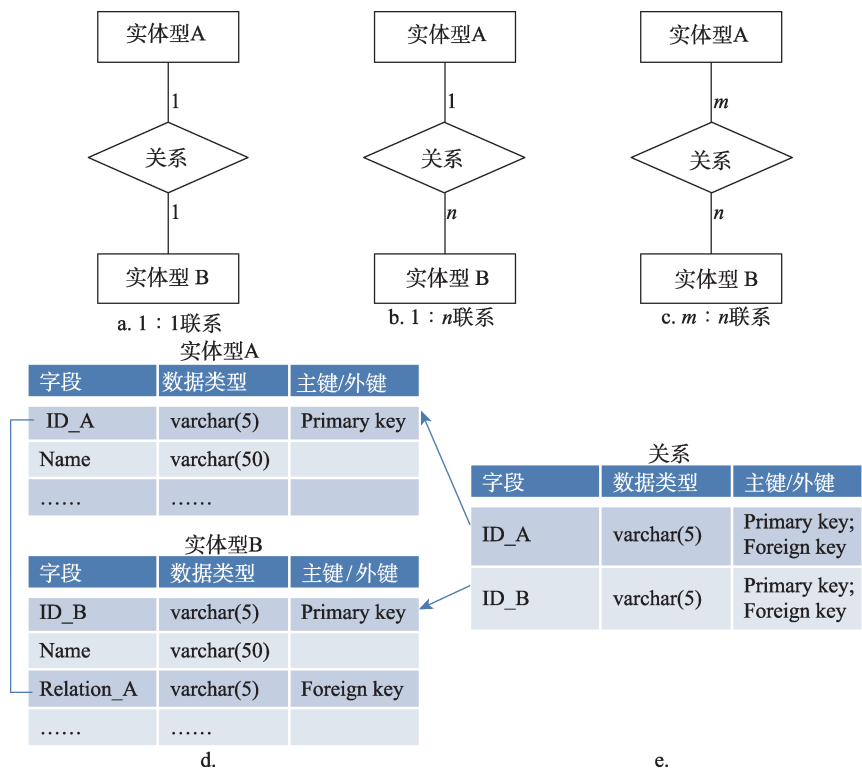


图3 二元实体关系及其数据库关系模式

出版社与出版时间的关联，即便建立二者间的联系，也不具有实际意义，且该方式难以区分二元关系所属的多元关系或作为独立二元关系存在^[35]。因此，二元实体联系可以视作 SKG 构建的主要方式，可以极大简化问题分析流程，迅速得到有效结论，但该方式所建立的多元实体联系较为松散。

2.2 多元实体联系存储解析

现实世界中的联系往往是多元的，多元实体联系涉及 3 个及以上、具有不同语义的实体型。假设有 k 元实体联系，其对应关系为 $n_1 : n_2 : \cdots : n_k$ 。其中， $k > 2$ ， $n_i = 1, 2, 3, \cdots$ ， $i = 1, 2, 3, \cdots, k$ 。若 n_i 全部为 1，则在关系数据库存储模式与二元实体一对一联系相似；若 n_i 中仅有一个不为 1，则在关系数据库存储模式与二元实体一对多联系相似；若 n_i 中存在两个及以上不为 1，则在关系数据库存储模式与二元实体多对多联系相似，即需要建立独立的关系模式。多元实体联系适用于多种领域，以非物质文化遗产领域“宜兴紫砂陶制作技艺”师承知识存储为例，假定师承关系为三元联系“师承(师傅,徒弟,技艺)”，据此探讨 SKH 模型对多元实体联系存储的优势。

(1) 1 : 1 : 1 联系存储优势分析

假定在某一时期，师傅、徒弟、技艺的联系是

1 : 1 : 1，即一位徒弟仅能选择一位师傅、学习一门技艺，师傅同样也只能教授一名徒弟、传授一门技艺，一门技艺也只能由一位徒弟、师傅掌握，且该时期内一个人不能同时作为徒弟和师傅出现。将其存储为 RDF 存储模式，在不考虑互反关系时，需在其中选择一个实体作为三元组主语，其他实体作为宾语，如图 4a 和图 4b 所示，假设对于砂陶制作技艺中“陶刻”，由“Master_1”传授给“Apprentice_1”，以徒弟作为主语，以师傅和技艺作为宾语，建立图中关系①②，为加强语义表达，三元联系的 3 个实体应两两间均建立关系，即共计 6 个三元组。随着联系所涉及实体不断增加，其所需存储的三元组数量也逐渐增加，存储三元组数量与实体数量呈现二次曲线关系，关系每增加一元，增加后达到 k 元关系，关系存储所需的三元组增加 $2 \times (k-1)$ (图 4c)，导致多元联系的存储具有极高的冗余性。

SKH 模型可以有效减少三元组存储的复杂度，如图 4d 和图 4e 所示，通过超边节点，建立实体与实体间的并列关系，在关系涉及实体较多时，SKH 可以极大地减少存储所需的三元组 (图 4f)。然而，在现实世界中， n_i 均为 1 的多元联系极其稀少，大多数实体均随时间变化而与多个实体产生联系，甚至在同一时间，一个实体也会承担一个关系的不同

@prefix RelationST: <http://127.0.0.1:8000/RelationST#>.
 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

RelationST:Skill_1 rdf:type RelationST:Skill; foaf:name "陶刻".
 RelationST:Apprentice_1 rdf:type RelationST:Apprentice.
 RelationST:Master_1 rdf:type RelationST:Master.

RelationST:Skill_1 RelationST:beLearnedBy RelationST:Apprentice_1.
 RelationST:Skill_1 RelationST:beTaughtBy RelationST:Master_1.

RelationST:Apprentice_1 RelationST:learn RelationST:Skill_1.
 RelationST:Apprentice_1 RelationST:learnFrom RelationST:Master_1.

RelationST:Master_1 RelationST:teach RelationST:Apprentice_1.
 RelationST:Master_1 RelationST:teachSkill RelationST:Skill_1.

a. SKG存储 (turtle格式)

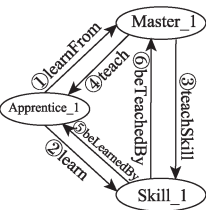
@prefix RelationST: <http://127.0.0.1:8000/RelationST#>.
 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

RelationST:Skill_1 rdf:type RelationST:Skill; foaf:name "陶刻".
 RelationST:Apprentice_1 rdf:type RelationST:Apprentice.
 RelationST:Master_1 rdf:type RelationST:Master.

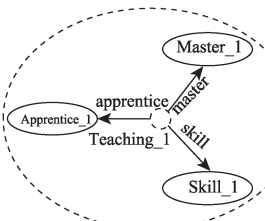
RelationST:Teaching_1 rdf:type RelationST:Teaching.

RelationST:Teaching_1 RelationST:apprentice RelationST:Apprentice_1.
 RelationST:Teaching_1 RelationST:master RelationST:Master_1.
 RelationST:Teaching_1 RelationST:skill RelationST:Skill_1.

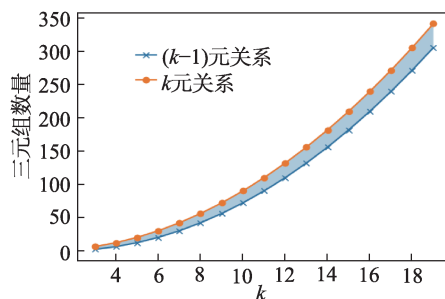
d. SKH模型存储 (turtle格式)



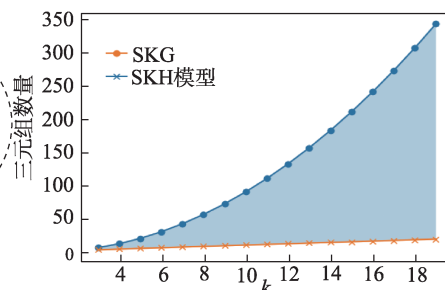
b. SKG存储图解



e. SKH模型存储图解



c. SKG存储 k 元与 $(k-1)$ 元关系对比



f. SKG与SKH模型知识存储对比

图4 1:1:1三元师承关系以及 k 元与 $(k-1)$ 元关系三元组数量对比

角色,如三元师承关系中,一个师傅可能会教授多个徒弟多门技艺,徒弟可能既向师傅学习技艺,同时又招收门徒传承技艺,因此,需要进一步探讨更为复杂的多元关系。

(2) 1:1:n联系存储优势分析

假设师傅、徒弟、技艺的联系为1:1:n,即徒弟可以学习,师傅可以传授多门技艺,其余假定条件不变。基于该假定条件,其SKG存储与SKH模型存储方式、图解及其对比如图5所示。其中,图5a、图5d分别表示一名徒弟、一名师傅以及“陶刻”“镶嵌”两项技艺在SKG与SKH模型存储的三元组;图5b、图5e表示对应的图解;图5c表示 k 元1:1:n关系中 k 与 n 逐渐增加时,SKG与SKH模型存储所需三元组数量;图5f表示相较于SKG,SKH模型对存储效率的优化比例。在SKG存储中,除技艺与技艺不存在语义关联外,均需在实体间构建语义关联。当师傅传授 n 门技艺时,共计 $2 \times (1+2n)$ 个三元组;当拓展至 k 元关系时,则需 $(k-1)(k-2+2n)$ 个三元组。而SKH模型仅需构建超边节点与普通节点间的关联,当师傅传授 n 门技艺时,共需 $(2+n)$ 个三元组,存储空间优化比例达60%~75%;当拓展至 k 元关系时,则仅需 $(k-1+n)$ 个三元组,联系所涉及实体越多,即 k 越大,SKH

模型存储效果越优。当 $k>10$ 时,SKH模型对SKG存储的优化比例可以达到90%以上。由此可见,对于多元实体联系存储,SKH模型存储效率相比于SKG具有显著优势,且使得语义关联更加紧密。

(3) 1:n₁:n₂联系存储优势分析

相比于1:1:n联系,1:n₁:n₂联系更为常见且更为复杂。用集合 A 、 B 分别表示一个关系中 n_1 、 n_2 所在位置实体集合,若集合 A 、 B 中每一个实体均与另一集合中全部实体关联,则本研究称该关系为完全对应关系;反之,则称为非完全对应关系。

假设师傅、徒弟、技艺的联系为1:n₁:n₂,即一位师傅可以传授多名徒弟多种技艺,其关系图解如图6所示。完全对应的师承关系在SKG中通常以图6a的方式存储,而在SKH模型中不仅优化了存储空间,还具有多种存储方式,可以在普通节点数量固定、其他语义信息缺失或模糊的情况下表示不同含义。若师承关系中传承时间、传承地点等其他信息缺失,在图6a中,SKG仅能表示师傅传授了两名徒弟两个技能,无法表示两名徒弟是否共同学习两项技能、是否两项技能在同一时间学习;而在SKH模型中则可以清晰表示,图6c表示两名徒弟共同学习两项技能,图6d表示师傅在不同时间单独传授两名徒弟技能,徒弟之间可能互不相识。除

@prefix RelationST: <http://127.0.0.1:8000/RelationST#>.
 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

RelationST:Skill_1 rdf:type RelationST:Skill; foaf:name "陶刻".
 RelationST:Skill_2 rdf:type RelationST:Skill; foaf:name "镶嵌".
 RelationST:Apprentice_1 rdf:type RelationST:Apprentice.
 RelationST:Master_1 rdf:type RelationST:Master.

RelationST:Skill_1 RelationST:beLearnedBy RelationST:Apprentice_1.
 RelationST:Skill_2 RelationST:beLearnedBy RelationST:Apprentice_1.
 RelationST:Skill_2 RelationST:beTaughtBy RelationST:Master_1.

RelationST:Apprentice_1 RelationST:learn RelationST:Skill_1.
 RelationST:Apprentice_1 RelationST:learn RelationST:Skill_2.
 RelationST:Apprentice_1 RelationST:learnFrom RelationST:Master_1.

RelationST:Master_1 RelationST:teach RelationST:Apprentice_1.
 RelationST:Master_1 RelationST:teachSkill RelationST:Skill_1.
 RelationST:Master_1 RelationST:teachSkill RelationST:Skill_2.

a. SKG存储 (turtle格式)

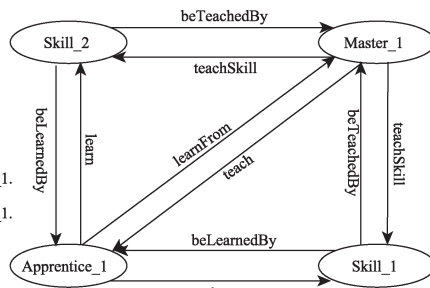
@prefix RelationST: <http://127.0.0.1:8000/RelationST#>.
 @prefix foaf: <http://xmlns.com/foaf/0.1/>.
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.

RelationST:Skill_1 rdf:type RelationST:Skill; foaf:name "陶刻".
 RelationST:Skill_2 rdf:type RelationST:Skill; foaf:name "镶嵌".
 RelationST:Apprentice_1 rdf:type RelationST:Apprentice.
 RelationST:Master_1 rdf:type RelationST:Master.

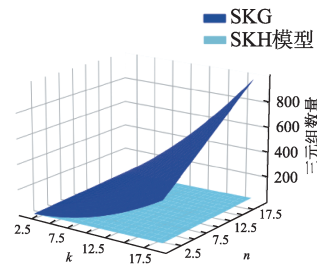
RelationST:Teaching_1 rdf:type RelationST:Teaching.

RelationST:Teaching_1 RelationST:apprentice RelationST:Apprentice_1.
 RelationST:Teaching_1 RelationST:master RelationST:Master_1.
 RelationST:Teaching_1 RelationST:skill RelationST:Skill_1.
 RelationST:Teaching_1 RelationST:skill RelationST:Skill_2.

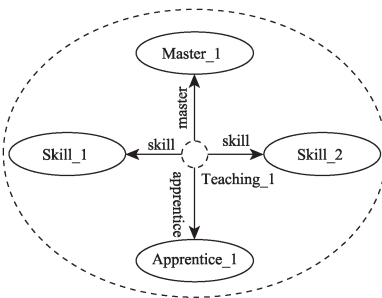
d. SKH模型存储 (turtle格式)



b. SKG存储图解



c. SKG与SKH模型知识存储对比

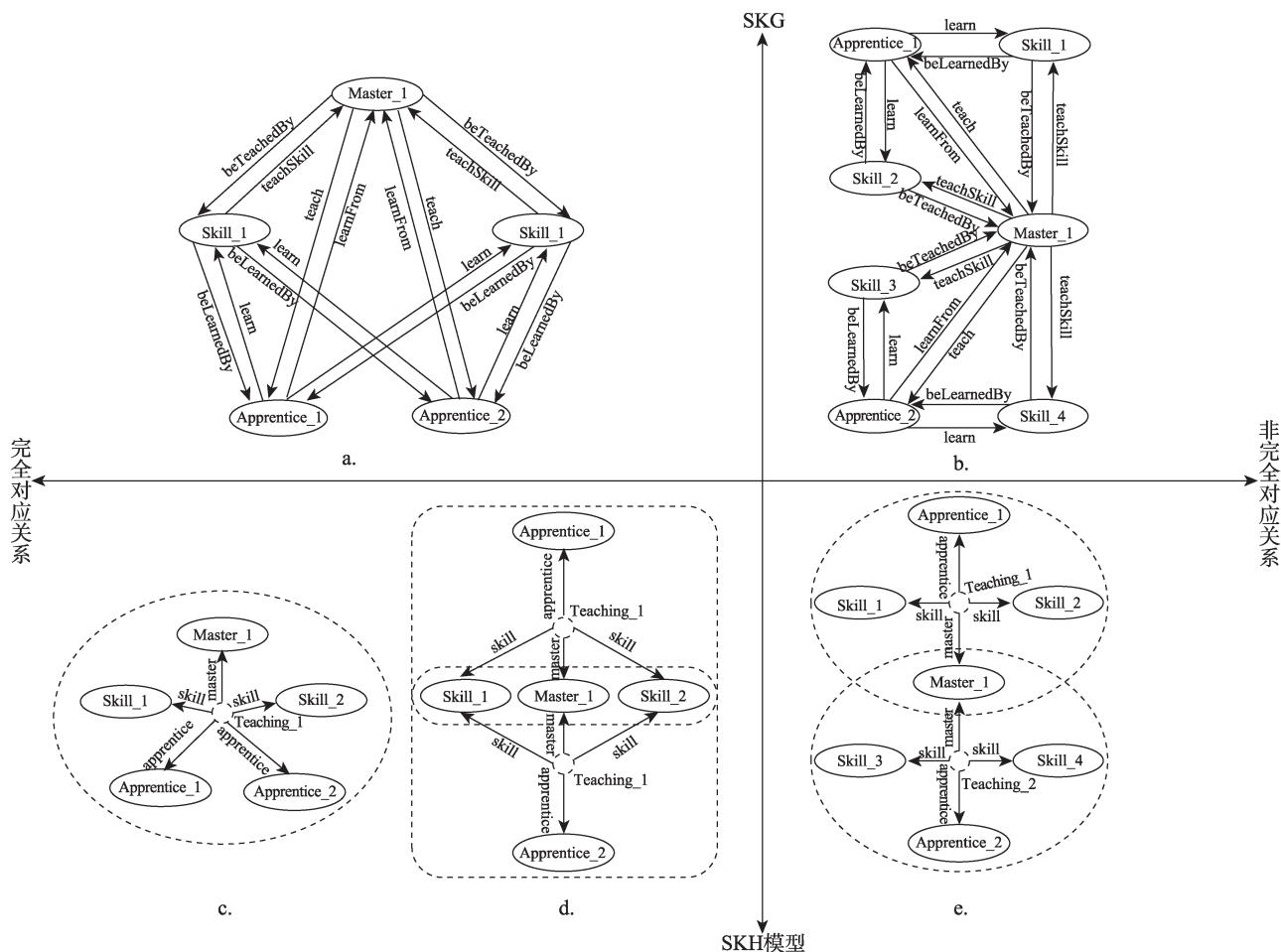


e. SKH模型存储图解

		n				n	
		2	20			2	20
k	3	60.00%	73.17%	k	12	91.56%	94.36%
	4	72.22%	81.75%		13	92.22%	94.77%
	5	78.57%	86.05%		14	92.79%	95.12%
	6	82.50%	88.64%		15	93.28%	95.42%
	7	85.19%	90.37%		16	93.70%	95.68%
	8	87.14%	91.61%		17	94.08%	95.91%
	9	88.64%	92.55%		18	94.41%	96.11%
	10	89.81%	93.29%		19	94.71%	96.30%
	11	90.77%	93.88%		20	94.94%	96.46%

f. SKH模型存储优化比例(SKG-SKH)/SKG

图5 1:1:n三元师承关系的SKG存储与SKH模型存储

图6 1:n₁:n₂三元师承关系图解

此外, SKH模型还可以对其他可能的语义关系进行有效存储。例如, 两项技能在不同的时间传授、两项技能在不同时间且单独传授等。图6b和图6c表示非完全对应的师承关系, 即师傅因材施教, 对不同的徒弟传授不同技能, 非完全对应的关系是在完全对应关系的基础上, 去除部分实体联系简化而成, SKH模型对非完全对应的语义关系存储具有同样的优势。相比之下, SKH模型存储更为简洁、高效, 且对部分信息缺失或模糊的关系具有兼容性, 对复杂知识具有更优的语义表达能力。

(4) $n_1 : n_2 : n_3$ 联系存储优势分析

相比于上述三类联系, $n_1 : n_2 : n_3$ 联系更为普遍, 师傅、徒弟、技能之间是多对多的联系, 师承关系中实体的关系更加自由, 更接近现实世界的实体关系。SKG与SKH模型对该联系的存储如图7所示。由于SKG对多元关系的存储主要是分解为多个二元关系, 联系松散, 存在语义缺失。图7a中两位师傅、两项技能、两名徒弟之间是完全对应关系, 对该存储具有多种可能的解读, 如两名师傅均将两项技能传授给两名徒弟, 或“Master_1”将“Skill_1”传授给“Apprentice_1”, 将“Skill_2”传授给“Apprentice_2”, “Master_2”将“Skill_2”传授给“Apprentice_1”, 将“Skill_1”传授给“Apprentice_2”等。因此, 对于存在三元及以上多对多的联系,

SKG因存储结构松散导致语义缺失、关系混乱, 必须采用更高一阶的实体进行知识聚拢, 即SKH模型中的超边实体, 图7b和图7c分别存储图7a可能代表的两种语义关系, 前者表示两位师傅共同传授、两位徒弟共同学习两项技能, 后者表示两位师傅均传授两项技能, 但每位师傅将每项技能传授给一位徒弟。除此之外, 还存储其他可能存在的语义关系, 加之信息缺失或模糊可能产生关系的多样性, 以及联系涉及实体增多导致关系的复杂性, 使得SKH模型在复杂语义关系存储方面更具优势。

综合上述分析, 对于 k 元实体联系 $n_1 : n_2 : \dots : n_k$, 其中 $k > 2$, $n_i = 1, 2, 3, \dots$, $i = 1, 2, 3, \dots, k$, 存在当 $n_i = 1$ 时, 均会以联系比例为1的实体为中心, 形成孤立的知识集合, 需通过其他联系打破知识孤立的局面; 仅存在不超过两个 n_i 大于1时, SKG可以明确存储该联系的语义关系, SKH模型则优化了SKG存储, 同时在无法明确某些实体时可以表示更加丰富的语义, 如已知时间不等但无法确定具体时间时, 可以通过超边节点划分实体所属的时间空间; 当存在超过两个 n_i 大于1时, SKG无法明确存储其语义关系, 而SKH模型则具有良好的兼容性, 可以存储更为复杂的语义关系, 又因为将关系实例化为超边节点, 故可以为其附加属性, 增强语义存储能力。

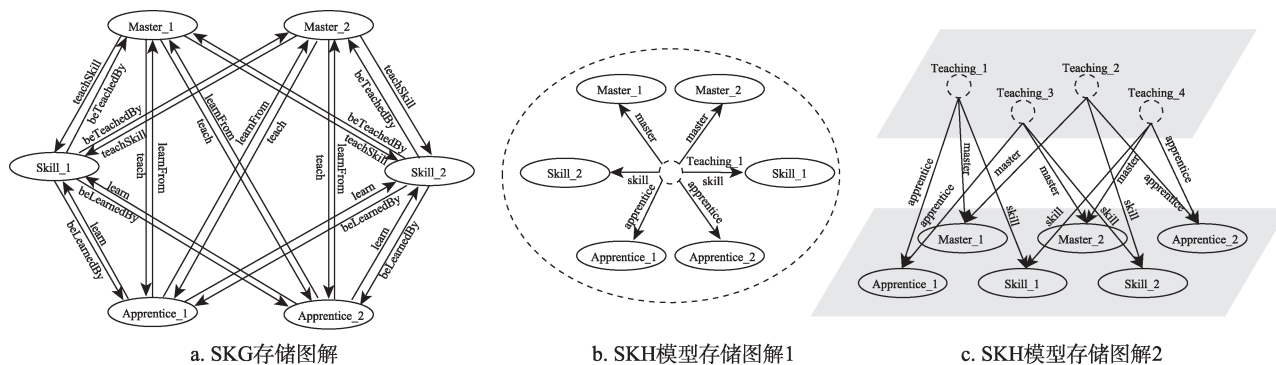


图7 $n_1 : n_2 : n_3$ 三元师承关系图解

3 语义知识超图模型应用

SKH模型是基于RDF, 结合HG(2)、元图等提出的知识存储模型, 以超图适用于RDF为目的, 因此, SKH模型在检索、推理等方面的应用与SKG相似, 且可以通过一定映射规则将关系数据库的数据映射为SKH模型数据, SKG与SKH模型数据在一定程度上也可以相互转换, 具有优越的兼容性;

在可视化方面, 由于其层级性, SKH模型具有更加多元且含义丰富的呈现方式。

3.1 SKH模型检索与推理

与SKG检索相似, SKH模型也使用SparQL进行检索, 但存在一定差异。SKG建立实体与实体间直接的联系, 因而检索时可以直接检索与已知实体相关的实体或属性; 而SKH模型借助超边实体存储

复杂知识，实体与实体间缺少直接关联，因而检索时需以超边实体为中介，通过间接的方式获取所需知识。同时，超边实体的使用，使得数据结构层级更加复杂，将超边实体视为普通实体上一层级，而超边实体又能作为普通实体参与构成另一超边，层层嵌套，结构复杂。

如图 8 所示，语义超图共包含 4 个层级：将用于描述“Entity₁”的一系列属性视为层级 0，将用于对这些属性分组、作为容器的“hevP₁”等超边实体视为层级 1，将“Entity₁”等实体与描述超边实体“hevE₁”的属性均视为层级 2，将超边实体“hevE₁”视为层级 3。SKH 模型层级的复杂性，导致检索时需通过自下而上和（或）自上而下的方式

层层检索。例如，若要通过“Entity₂”检索“Entity₁”，则必须先寻找上一层楼级实体，即超边实体“hevE₁”，再进行检索，无法直接通过实体检索超边上另一实体。图 8a 中检索具有关系“he₁”“he₂”，以及属性“propertyE₁”“propertyE₂”的超边，仅涉及层级 2 与层级 3 中实体与属性；图 8b 中检索实体以及实体每个状态所对应的属性，通过自上而下的检索方式，检索从层级 0 到层级 2 共 3 个层级的实体与属性。SKH 模型的层级性反映了由微观粒子到客观世界的组成方式，相应的层级检索反映了人类从浅入深、由宏观到微观的思考问题、认识世界的过程，有助于借助数字技术实现语义层面的数字孪生。

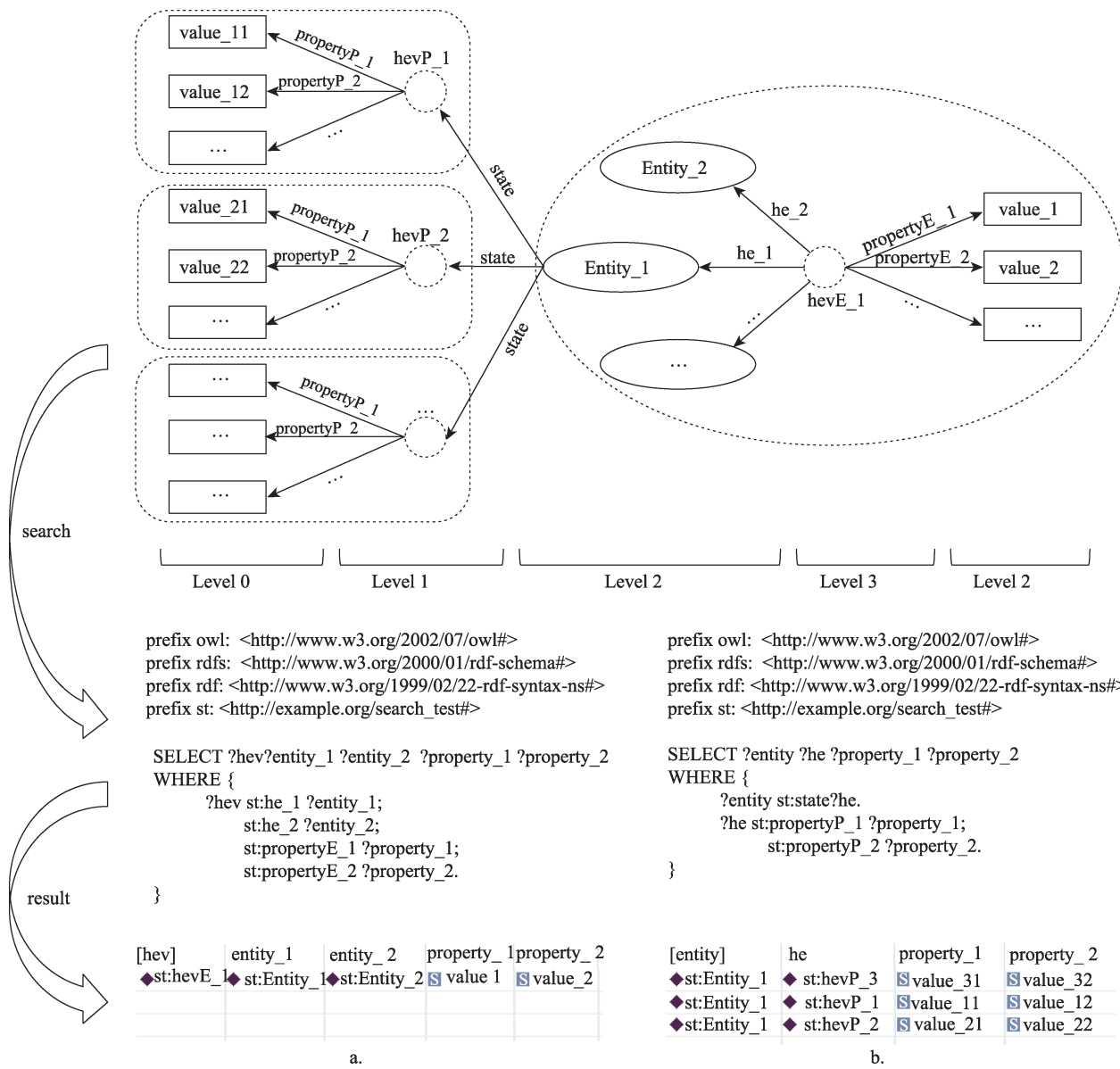


图 8 SKH 模型检索

知识推理是通过利用已有规则或编写自定义的推理规则,发现隐含知识关联的过程。对于已有规则,SKH模型同样适用,但降低了部分推理规则的使用频率,如互反关系、对称关系等;而部分推理规则因SKH的层级性产生更为重要的作用,如传递关系在分析事物多级构成要素时具有重要意义。对于自定义推理规则,SKH模型与SKG均基于RDF三元组存储,均可使用SWRL等编写推理规则,检索符合推理条件的若干三元组,构建新的语义关联,实现知识推理。通过知识推理,SKH模型不仅可以挖掘隐含知识,还可以将超边分解,转换为普通边。对于挖掘隐含知识,SKH模型由于其精准的语义存储能力,推理结果比SKG更为精确,又因其复杂的层级结构,推理规则的编写更为复杂;SKH模型对超边上节点关系的推理,则可以将超边重构为若干普通边,实现超边向普通边的转换或两者共存于同一知识图谱,同时满足多场景的不同需求。

3.2 SKH模型数据转换

与RDB2RDF (relational database to RDF) 映射相似,关系数据库也可以根据一定规则映射为符合SKH模型的数据。复杂语义关系向SKH模型映射如图9所示,同一实体具有多组属性(图9a)时,将关系数据库中每一个元组的主键与超边节点命名空间结合,创建超边实体,作为外键的、属性真正描述的实体映射为普通实体,并作为三元组的首实体与超边实体建立联系,属性值描述超边实体;具有复杂的多元实体关系或关系具有属性描述(图9b)时,与实体多组属性映射相似,主键映射为超

边实体,外键映射为普通实体,属性描述超边实体。若主键包含多个字段,则可映射为空节点或参照万维网联盟推荐的Direct Mapping创建超边实体。

尽管两者均建立超边实体与普通实体间联系,并以属性描述超边实体,但其含义截然不同。前者以普通实体作为首实体,超边实体作为尾实体,在语义上比普通实体更低一阶,且两者之间的边为普通边,属性通过超边实体提供的属性空间,间接描述普通实体;后者则以超边实体为首实体,普通实体为尾实体,在语义上超边实体是比普通实体更高一阶的实体,且两者之间的边为超边,属性旨在描述超边实体,为普通实体丰富语义环境。

SKH模型与SKG具有一定兼容性,且在一定程度上也可以相互转换,公众可以自由选择两种知识存储方式。如图10所示,SKG知识通过SPARQL查询将知识提取为二维表,按超边实体进行分组,将其映射为SKH数据,实现SKG向SKH模型知识的转换,SKH模型知识通过知识推理可以转换成SKG知识。对于目前已存在的大量语义知识图谱,SKH模型以其优越的兼容性,既可以将已有的语义知识图谱转换为基于SKH模型存储的知识,以优化知识存储,并在其基础上进行复杂语义存储,又可以将SKH模型中简单知识转换成基于SKG存储的知识,适用于多类人群,为广泛推广奠定基础。

3.3 SKH模型可视化

与SKG存储方式相比,基于SKH模型存储的知识可视化方式更为多元。SKG存储的知识通常采用扁平化的二维可视化工具,如Echarts、Neo4j、

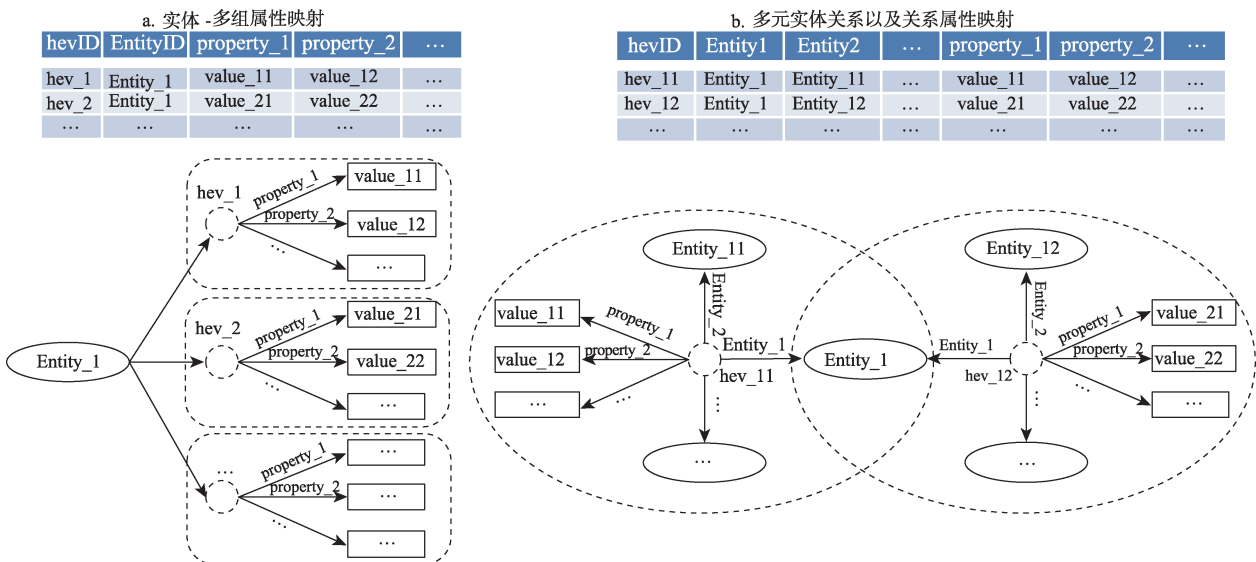


图9 关系数据库向SKH模型映射

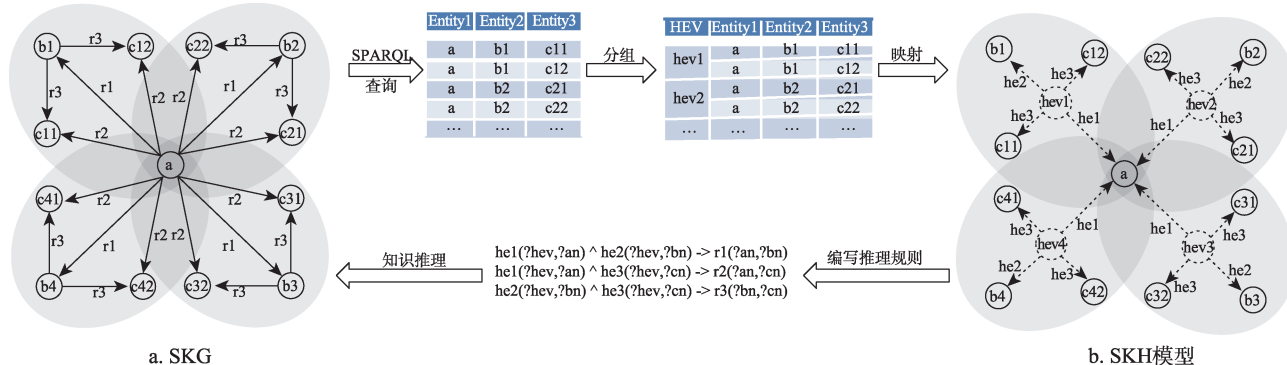


图 10 SKG与SKH模型数据转换

D3等,也可以采用空间化的三维布局。由于SKG数据结构单一,除所属类别外,缺乏其他明确统一的分层依据,因此,一般不进行分层表示。而SKH模型引入超边关系与超边实体,增添了统一的聚类依据,可以明确划分节点与节点间的上下级关系,因此,SKH模型具有更加多元的可视化方案。

SKH模型可视化方案如图11所示,SKG的非层级化平面布局与空间布局同样适用于SKH模型,

且在空间布局中具有与SKG可视化不同的内涵。以基于力引导算法的非层级化空间布局为例,SKG节点以关系为纽带,互相牵引,没有关系的节点间相互排斥,最终低中心度节点向高中心度节点聚拢,可能呈现区域性分布特点,而每一个区域可能表示具有共同特征的某类节点。如以时间为中心,周围聚集了同一时间出生的人、创造的事物等;亦可能表示一系列关系紧密的节点,如在用户信息透明的

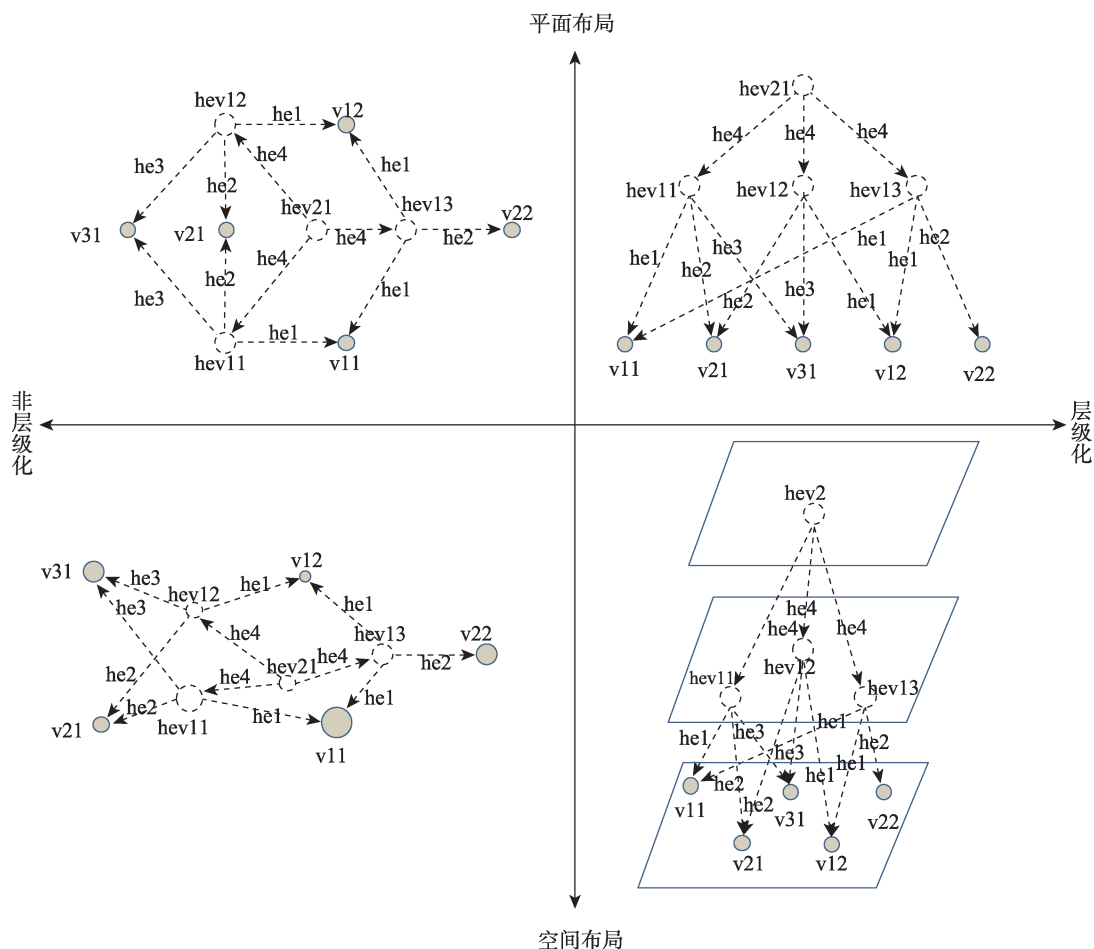


图 11 SKH模型可视化

假设条件下,将不同在线社区的用户及其交互关系绘制社会网络,每个社区极易形成自己的“小世界”,内部用户联系紧密。除此之外,还存在信息缺失导致关系缺失,最终形成无规则聚集等多种可能。

在仅存在超边的数据结构中,SKH模型中以超边实体为中心,层层聚集且结构分明;当数据同时含有超边与普通边时,可以在力引导算法中为SKH模型超边设置较高的“劲度系数”与较短的“自然长度”,使超边连接的实体相互聚集,普通边连接实体的聚集程度次之,无关联的节点尽量离散,形成局部聚集、层次更加清晰的网状结构,SKH模型中超边实体为超边所连实体提供了环境信息,因此,围绕超边实体聚集形成的区域性特点具有相对明确的含义。由于SKH模型的层级性,其可视化布局不仅可以实现实体间聚集,不同区域也向其上级实体聚拢,多层嵌套,表意更加丰富,并且反映了由微观到宏观的世界组成方式。

SKH模型非层级化的可视化方式反映了客观世界的组成方式,而层级化的可视化方式则清晰地由粗粒度到细粒度、由浅入深地分层呈现存储知识的架构。SKH模型中,超边连接实体,均视为首实体位于尾实体上一级,据此SKH模型可视化将实体(含超边实体)划分为若干层,普通边既可以存在于每层内部,也可以进行跨层实体连接,而超边仅能进行跨层连接,若该层含有超边实体,则将超边实体视为普通实体参与关系构建。非层级化可视化方式呈现以超边实体为中心的区域性特点,而层级化可视化方式则为每个实体赋予层级,每个层级为实体提供相对独立的知识空间,清晰呈现各层实体间关系。层级化的可视化方式有利于表示超图与普通图混合存在的SKH模型知识,然而为了进行清晰的层级划分,需要对基于SKH模型存储的知识结构进行精细设计,增加了SKH模型的使用难度,同时也使得其存储的知识更加规范。

4 总结与展望

本研究基于语义知识图谱知识存储局限,借鉴HG(2)与元图模型,提出简易、通用的SKH模型,以非物质文化遗产领域师承关系为例,围绕关系数据库主要关系类型及其向RDF的映射,分析得出SKH模型比SKG知识存储效率更加优化,且可以存储内涵更加丰富、结构更加复杂的语义知识,对复杂实体属性存储具有优势。SKH模型可以沿用SKG的检索与推理方式,且其优越的语义存储能力

使得检索与推理结果更为精确,其复杂的层级结构使得检索与推理更为复杂。借助检索、推理、RDB2RDF映射等方法,SKH模型在一定程度上可以与关系数据库、SKG等知识存储方式互相转换,具有优越的兼容性,适用于多场景知识存储,同时基于SKH模型的知识存储结构具有更加多元、清晰且表意丰富的可视化方式。

SKH模型尽管在语义存储方面具有广阔前景,但仍存在一定局限性。首先,本研究以非连续的实例阐释SKH模型对知识及其语义关系的存储,但未在已知语义知识较为完整的领域体系进行具体应用;其次,尽管SKH模型实现了对复杂语义的存储,但随着语义不断复杂,其检索与推理规则更加复杂,在一定程度上影响了检索与推理效率;最后,目前利用语义知识图谱进行知识存储最为普遍,尽管对于简单语义关系,二者之间可以相互转换,但转化过程相对较为烦琐,对使用者要求较高,并且由于SKG知识存储的局限,基于SKH模型的复杂语义关系难以向语义知识图谱转换。

针对上述研究局限性,未来研究可在以下方面继续深化:①以具体领域的完整语义知识作为研究对象,应用SKH模型进行语义存储,探索其在实践应用与理论研究中的差异,持续完善SKH模型,提升其普适性和实用性,提高信息资源管理等领域语义知识存储效能。②简化SKH模型检索与推理方法,通过预定义规则等方式提高其检索与推理效率。③优化数据转换流程,通过将规则集成化等方式简化数据转换过程,提升SKH模型的可用性。

参 考 文 献

- [1] W3C RDF Working Group. RDF 1.1 concepts and abstract syntax [EB/OL]. (2014-02-25) [2023-01-07]. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- [2] W3C RDF Working Group. Resource description framework (RDF) [EB/OL]. (2014-02-25) [2023-01-07]. <https://www.w3.org/2001/sw/wiki/RDF>.
- [3] 中国电子技术标准化研究院. 知识图谱标准化白皮书(2019版) [R/OL]. (2019-08-01) [2022-04-23]. <http://www.cesi.cn/images/editor/20190911/20190911095208624.pdf>.
- [4] 高劲松, 张强, 李帅珂. 可移动文物的知识图谱构建及关联数据存储——以湖北省博物馆为例[J]. 现代情报, 2022, 42(4): 88-98.
- [5] Zhang F, Wu J Z, Nie Y L, et al. Research of knowledge graph technology and its applications in agricultural information consultation field[C]// Proceedings of the 39th IEEE International Per-

- formance Computing and Communications Conference. Piscataway: IEEE, 2021: 1-4.
- [6] 张琪, 王东波, 黄水清, 等. 史书多维知识重组与可视化研究——以《史记》为对象[J]. 情报学报, 2022, 41(2): 130-141.
- [7] Wei J Z, Liu R. An approach of constructing knowledge graph of the hundred schools of thought in ancient China[C]// Proceedings of the 2019 ACM/IEEE Joint Conference on Digital Libraries. Piscataway: IEEE, 2019: 335-336.
- [8] Zhao H X, Pan Y L, Yang F. Research on information extraction of technical documents and construction of domain knowledge graph[J]. IEEE Access, 2020, 8: 168087-168098.
- [9] 高劲松, 马倩倩, 周习曼, 等. 文献知识元语义链接的图式存储研究[J]. 情报科学, 2015, 33(1): 126-131.
- [10] 杭婷婷, 冯钧, 陆佳民. 知识图谱构建技术: 分类、调查和未来方向[J]. 计算机科学, 2021, 48(2): 175-189.
- [11] 陈涛, 刘炜, 单蓉蓉, 等. 知识图谱在数字人文中的应用研究[J]. 中国图书馆学报, 2019, 45(6): 34-49.
- [12] 陈涛, 李惠, 张永娟, 等. LIBRA技术理论及其在史料图像资源中的应用[J]. 大学图书馆学报, 2022, 40(4): 64-74.
- [13] 吴鹏, 刘恒旺, 丁慧君. 基于本体和NoSQL的机械产品方案设计的知识表示与存储研究[J]. 情报学报, 2017, 36(3): 285-296.
- [14] 胡秉德, 王新根, 王新宇, 等. 超图学习综述: 算法分类与应用分析[J]. 软件学报, 2022, 33(2): 498-523.
- [15] 王蝶, 康丽英. 一般超图的张量谱性质[J]. 运筹学学报, 2023, 27(1): 138-148.
- [16] 张磊, 牛倩楠, 任海珍. 一致超图的边连通性和最大边连通性[J]. 山西大学学报(自然科学版), 2021, 44(6): 1079-1085.
- [17] 上官冲, 葛根年. 稀疏超图: 从理论到应用[J]. 中国科学: 数学, 2023, 53(2): 187-216.
- [18] 周丽娜, 常笑, 胡枫. 利用邻接结构熵确定超网络关键节点[J]. 计算机工程与应用, 2022, 58(8): 76-82.
- [19] Li W, Wang R J, Jia X F. The optimal inference approximate algorithm in weighted hypergraph based on granular computing[C]// Proceedings of the 5th International Conference on Computer Science and Network Technology. Piscataway: IEEE, 2017: 273-276.
- [20] 陈文杰. 基于超图的科研合作推荐研究[J]. 数据分析与知识发现, 2023, 7(4): 68-76.
- [21] 刘高, 黄沈权, 龙安, 等. 基于超图网络的产品设计知识智能推荐方法研究[J]. 计算机应用研究, 2022, 39(10): 2962-2967.
- [22] 于亚新, 张文超, 李振国, 等. 基于超图的EBSN个性化推荐及优化算法[J]. 计算机研究与发展, 2020, 57(12): 2556-2570.
- [23] 吴越, 王英, 王鑫, 等. 基于超图卷积的异质网络半监督节点分类[J]. 计算机学报, 2021, 44(11): 2248-2260.
- [24] 白思萌, 牛振东, 何慧, 等. 基于超图注意力网络的生物医学文本分类方法[J]. 数据分析与知识发现, 2022, 6(11): 13-24.
- [25] 赵宇红, 张晓楠. 基于超图和 k -means改进的异质网络社区发现算法[J]. 计算机应用与软件, 2021, 38(10): 290-296.
- [26] Xie C, Zhong W, Xu W, et al. Visual analytics of heterogeneous data using hypergraph learning[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(1): Article No.4.
- [27] C. 贝尔热. 超图——有限集的组合学[M]. 卜月华, 张克民, 译. 南京: 东南大学出版社, 2002.
- [28] 高峰, 郑丽丽, 顾进广. 面向多元时序关系的金融知识图谱表示与构建[J]. 山西大学学报(自然科学版), 2022, 45(4): 873-883.
- [29] 李豪, 周爽. 基于三维知识超图的电力智库知识服务平台建设[J]. 智库理论与实践, 2022, 7(3): 84-92, 99.
- [30] 田玲, 张谨川, 张晋豪, 等. 知识图谱综述——表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
- [31] Munshi S, Chakraborty A, Mukhopadhyay D. Theories of hypergraph-graph (HG(2)) data structure[C]// Proceedings of the 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies. Piscataway: IEEE, 2014: 204-207.
- [32] Munshi S, Chakraborty A, Mukhopadhyay D. Integrating RDF into hypergraph-graph (HG(2)) data structure[C]// Proceedings of the 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies. Piscataway: IEEE, 2014: 208-212.
- [33] Chernenkiy V, Gapanyuk Y, Nardid A, et al. Using the metagraph approach for addressing RDF knowledge representation limitations[C]// Proceedings of the Conference of 2017 Internet Technologies and Applications. Piscataway: IEEE, 2017: 47-52.
- [34] Terekhov V, Gapanyuk Y, Kanev A. Metagraph representation for overcoming limitations of existing knowledge bases[C]// Proceedings of the 2021 28th Conference of Open Innovations Association. Piscataway: IEEE, 2021: 458-464.
- [35] Krótkiewicz M. Hypergraph approach towards ontology design in association-oriented metamodel[J]. Cybernetics and Systems, 2019, 50(2): 132-153.
- [36] Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514.

(责任编辑 王克平)