

基于张量分解技术的知识图谱实体扩展与布局优化

李寿兴¹
LI Shouxing

摘要

当前知识图谱扩展优化过程中, 主要利用图嵌入方法完成实体扩展, 其可能忽略三元组中关系和实体的交互细节, 导致扩展优化后的知识图谱在链路预测任务上所得结果准确性较低。基于此, 文章提出了一种基于张量分解技术的知识图谱实体扩展与布局优化方法。将初始知识图谱表示为三维张量, 利用张量分解级数对其进行挖掘, 捕获潜在语义特征信息, 从而确定一系列新实体, 实现知识图谱实体扩展。针对扩展后的知识图谱进行表示学习, 获取实体之间的推理关系。通过 DSM (设计结构矩阵) 解耦, 对实体之间的复杂关系进行简化, 按照 DSM 建模结果调整实体位置和关系路径, 完成知识图谱布局优化。实验结果表明: 利用该方法扩展优化后的知识图谱执行链路预测任务, 最终预测结果的常规平均排名、平均倒序排名、Hits@3 指标值分别为 3.9、11.2 和 0.8, 有效提升了知识图谱在链路预测任务上的准确程度。

关键词

张量分解; 知识图谱; 实体扩展; 设计结构矩阵; 布局优化

doi: 10.3969/j.issn.1672-9528.2025.12.018

0 引言

数字化时代, 各领域的信息量出现爆发式增长, 如何实现有效的信息处理和决策分析成为一个难题。知识图谱的出现, 改变了多个领域的信息处理方式, 作为一种由实体和关系组成的结构化网络, 能够存储大量零散数据, 并捕获这些数据之间的潜在联系, 帮助用户理解现有知识的同时, 为信息搜索、智能推荐等功能的实现提供了支撑。然而, 随着用户需求的不断增长, 简单的知识图谱已难以满足应用需求, 针对知识图谱进行实体扩展与布局优化成为一个重要的研究问题。

宋宝燕等^[1]提出融合关系模式和类比迁移的方法, 考虑知识图谱实体关系的语义层次结构, 划分出多种关系模型, 并引入类比对象生成方法, 根据每个关系模式的特性生成对应的三元组, 将新生成三元组嵌入到原始图谱中, 即可实现知识图谱扩展补全。该方法的计算效率较高, 但类比迁移过程中会出现信息损失, 进而影响链路预测结果。陈强等^[2]提出融合任务知识的方法, 运用参数调节后的多模态编码器学习不同模态数据, 生成多模态实体向量表示信息。通过循环神经网络对表示向量进行学习, 实现多模态特征的融合与过

滤。借助同构图网络实现特征更新, 即可得到补全扩展后的知识图谱。该方法应用拓展性较好, 但处理多模态数据时对齐不准确, 会导致信息融合错误, 从而影响链路预测。杨强等^[3]提出基于改进图卷积神经网络的方法, 构造基于改进图卷积神经网络的编码器, 通过对原始知识网络进行循环学习, 捕获图谱中的邻居节点与关联路径, 利用注意力机制筛选出贡献程度较高的节点和路径, 将其输入 InteractE (知识图谱嵌入) 模型中进行解码训练, 最终生成补全优化后的知识图谱。该方法的泛化能力较强, 但网络参数取值不合理, 会直接造成链路预测不准确。周粤等^[4]提出基于双子图和注意力机制的方法, 以实体和关系为核心, 建立面向原始知识图谱的双子图, 借此呈现出潜在关联信息。将双子图依次输入编码器网络和注意力机制网络中, 更新实体关系结构和交互特征, 再借助解码器实现更新后特征向量的评分预测, 找出最佳特征补全到知识图谱中, 实现图谱扩展更新。该方法的推理过程具有可解释性, 但双子图划分不合理, 会导致链路预测准确性下降。

考虑到现有方法的应用效果不佳, 为满足现代化工作需求, 提出基于张量分解技术的实体扩展与布局优化方法。通过对知识图谱三维张量进行分解, 捕获原始图谱中的潜在语义特征, 即可生成新的扩展实体。借助学习表示方法和 DSM 解耦方法, 对实体扩展后的知识图谱完成推理关系捕获和布

1. 滇西应用技术大学 云南大理 671000

[基金项目] 云南省教育厅科学研究基金项目“知识图谱在教育突发事件智能决策中的应用研究”(2024J1122)

局优化, 最终得到结构简单、内容表达丰富的知识图谱, 以便实现高质量的数据管理和应用决策。

1 设计基于张量分解技术的知识图谱实体扩展与布局优化方法

1.1 建立基于张量分解的实体扩展方案

借助 VGG19 网络模型, 对原始知识图谱中包含的特征信息进行转换和融合, 生成涵盖多模态信息的三维张量, 再展开张量分解操作, 即可生成可扩展的新实体。

以任意选取的图像特征和文本特征为例^[5], 通过 VGG19 网络模型中的一个全连接层对其进行变换, 最终得到:

$$\begin{cases} V' = V \times W_1 + a_1 \\ D' = D \times W_2 + a_2 \end{cases} \quad (1)$$

式中: V 、 D 分别表示图像特征和文本属性特征; W_1 、 W_2 分别表示特征变换所需的映射矩阵; a_1 、 a_2 分别表示特征变换所需的参数矩阵。

将转换后的特征向量输入到第二个全连接层中, 通过映射处理可以得到多模态特征融合结果^[6], 用公式表示为:

$$\begin{aligned} Q_{(V,D,T)} &= P'_{(V,D)} + T \\ &= (P_{(V,D)} \times W_3 + a_3) + T \\ &= [(V' + D') \times W_3 + a_3] + T \end{aligned} \quad (2)$$

式中: T 表示嵌入向量; W_3 、 a_3 分别表示多模态特征融合所需的映射矩阵和参数矩阵; P 表示初步融合结果; P' 表示二次融合结果; Q 表示多模态最终融合结果。

运用张量分解技术处理多模态融合特征, 即可将一个高维特征向量表示为多个低维映射特征向量, 具体操作原理如图 1 所示。

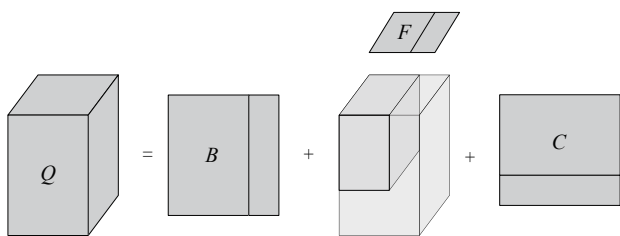


图 1 张量分解原理

按照图 1 对式 (2) 输出的多模态融合特征进行张量分解, 最终生成的分解结果为:

$$Q \approx \sum_{l=1}^L \beta_l B_l \otimes C_l \otimes F_l \quad (3)$$

式中: L 、 l 分别表示张量分解次数和秩; β 表示权重向量; B 、 C 、 F 表示张量分解后形成的特定因子矩阵。

通过上述张量分解操作, 可以捕获原始知识图谱中包含的潜在特征信息, 由此确定新的实体, 实现知识图谱实体扩展。

1.2 捕获知识图谱实体推理关系

面向实体扩展后的知识图谱, 捕获其包含的实体推理关系, 是实现知识图谱布局优化的前提^[7]。实际运算过程中, 先定义扩展后的知识图谱为:

$$G = \{\varphi, R, U\} \quad (4)$$

式中: G 表示知识图谱; φ 表示张量分解扩展后生成的实体集合; R 表示关系集合; U 表示三元组集合。

以知识图谱中任意一个实体对为例, 将其包含的两个实体视分别为头实体和尾实体, 并定义二者之间的关系路径为:

$$f = (b_1, e_1, \dots, e_{k-1}, b_k), b \in R, e \in \varphi \quad (5)$$

式中: f 表示任意两个实体对之间的关系路径; b 表示关系; e 表示中间实体; k 表示扩展后知识图谱内实体数量。

按照上述操作, 捕获知识图谱中所有实体对之间的关系路径, 完成知识图谱实体间推理关系的初步获取。为了去除其中冗余关系, 通过公式为关系路径计算出对应的综合能量函数。

$$E(a, f, c) = \|a' + f' - c'\| = \|f' - b'\|, a, c \in \varphi \quad (6)$$

式中: E 表示综合能量函数; a 、 c 分别表示头实体和尾实体; a' 、 b' 、 c' 表示实体向量; f' 表示关系路径向量。

将综合能量函数计算结果代入到公式中, 即可实现对实体推理关系的筛选^[8]。

$$\begin{cases} E(a, f, c) \leq \psi, & \text{保留} \\ E(a, f, c) > \psi, & \text{删除} \end{cases} \quad (7)$$

式中: ψ 表示判断阈值。

按照公式所示的判断规则, 重新捕获知识图谱实体推理关系, 作为后续布局优化的基础。

1.3 生成知识图谱 DSM 布局优化结果

基于知识图谱实体间的推理关系, 调整实体位置和关系路径^[9], 是实现布局优化的核心环节。将 DSM 解耦算法应用到布局优化过程中, 有利于生成质量更好的优化结果。

整理好所有实体之间的推理关系后, 通过布尔模型 DSM 初步分析描述实体依赖关系, 构造初始设计结构矩阵。针对该矩阵进行 DSM 解耦运算时, 需要将原始矩阵变换为下三角矩阵, 再筛选出所有的空行和空列, 将其从矩阵中移出, 即得到原始矩阵对应的可达矩阵。根据可达矩阵获取扩展知识图谱中的强连通子集和耦合活动集。需要注意的是, 若设计结构矩阵无法直接转化为下三角矩阵, 则需要执行撕裂操作^[10], 该操作的执行顺序依据信息依赖程度确定。

$$\begin{cases} \gamma_i = \sum_{j=1}^{i-1} g_{ij} \times w + \sum_{j=i+1}^n g_{ij} \times w_j \\ \rho_j = \sum_{i=1}^{j-1} g_{ji} \times w + \sum_{i=j+1}^n g_{ji} \times w_j \end{cases} \quad (8)$$

式中： γ 表示任务信息输入； ρ 表示任务信息输出； $i、j$ 表示任务编号； g 表示耦合任务集； n 表示集合内任务数量； w 表示反馈权重。

$$\eta_q = \min(\mu_i) = \min(\gamma_i / \rho_j)$$

(9)

式中： η 表示撕裂耦合块； q 表示撕裂耦合块编号； μ 表示信息关联程度综合度量值； $\min$ 表示最小化取值函数。

利用式(8)和式(9)，确定所有任务子集的撕裂操作顺序，并按顺序完成设计结构矩阵内所有元素的解耦运算，即可从另一方面求出可达矩阵。

利用可达矩阵内包含的信息，对知识图谱实体推理关系进行简化，并将关系密切的实体移动到相邻区域，实现实体布局的整体优化，生成具有更优数据管理性能的知识图谱。

2 实验分析

2.1 实验准备

基于张量分解技术设计知识图谱实体扩展与布局优化方法后，为了验证新方法的应用效果，基于地级市A市市长信箱诉求数据，进行实验测试分析。对近三年市长信箱诉求数据进行的统计，共获取2500条涉及所辖县区政府部门的数据（已全部脱敏），将其均分为5个子数据集（A1~A5）。为了便于处理各个数据集内的数据信息，构建了针对A市政府部门的初始知识图谱，如图2所示，后续实体扩展与布局优化实验即基于此图谱进行。

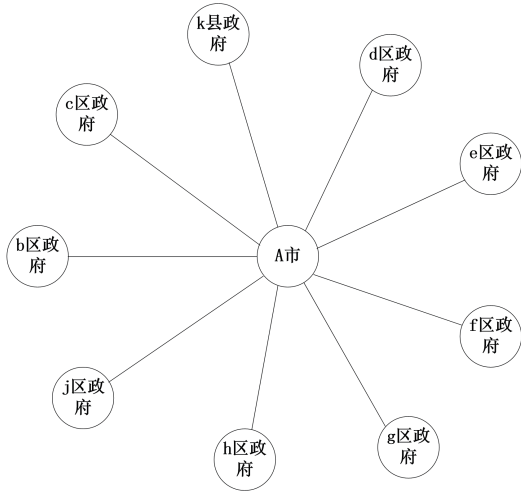


图2 初始知识图谱

从图2可以看出，现有的知识图谱涵盖的政府部门不够全面，在待处理数据量不断增长的背景下，难以满足市长信箱诉求数据管理需求，需要对其进行扩展优化处理。因此，分别运用新设计方法、融合任务知识的方法^[2]和基于改进图卷积神经网络的方法^[3]调整，并运用新生成知识图谱面向不同数据集执行链路预测任务，对比不同知识图谱的预测结果准确程度，体现新方法的应用优势。

2.2 实体扩展结果

应用新方法进行实体扩展前，需要运用VGG19网络模型对初始知识图谱进行多模态特征提取和融合，以便捕获潜在语义特征，生成新的扩展实体，在这一实验阶段，设置的主要实验参数如表1所示。

表1 参数设置

参数名称	数值
实体嵌入维度区间	[30,50,100,150,200]
关系嵌入维度区间	[30,50,100,150,200]
学习率区间	[0.0005,0.004,0.005,0.01]
学习率衰减率区间	[1,0.995,0.99,0.95]
批量标准化补偿区间	[50,100,500]
Dropout 区间	[0.3,0.4,0.5]

在表1所示的参数条件支撑下，实现知识图谱特征捕获和融合，并对融合后的多模态特征进行张量分解，最终得到图3所示的实体扩展结果。

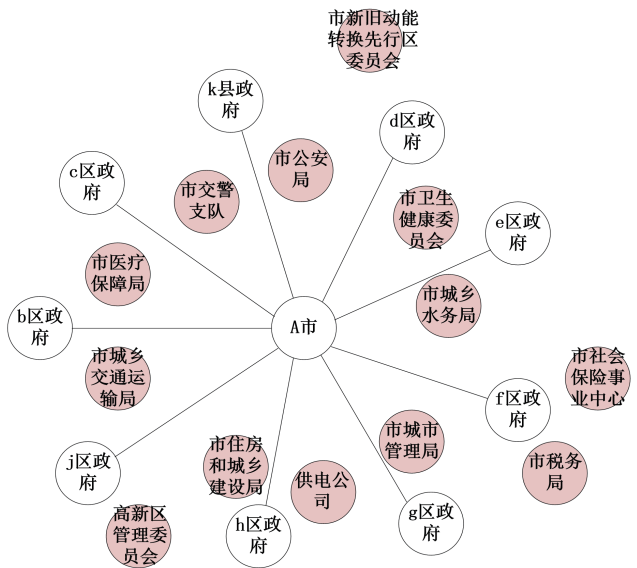


图3 知识图谱实体扩展结果

如图3所示，通过张量分解计算，在原始知识图谱内扩展了13个部门实体，这些部门实体在市长信箱诉求数据中频繁出现，将其添加到知识图谱中，有利于更好地处理市长信箱诉求数据。

2.3 布局优化结果

基于不同部门实体之间的关联性，运用DSM解耦算法完成知识图谱实体与关系的布局优化，最终得到图4所示的新型知识图谱。如图4所示，通过布局优化调整了图谱中不同实体的位置，并将存在密切关系的部门实体转移到一处，呈现简化后的关系路径。这一操作极大提升了知识模型的自身性能，为实现高质量数据处理打下基础。

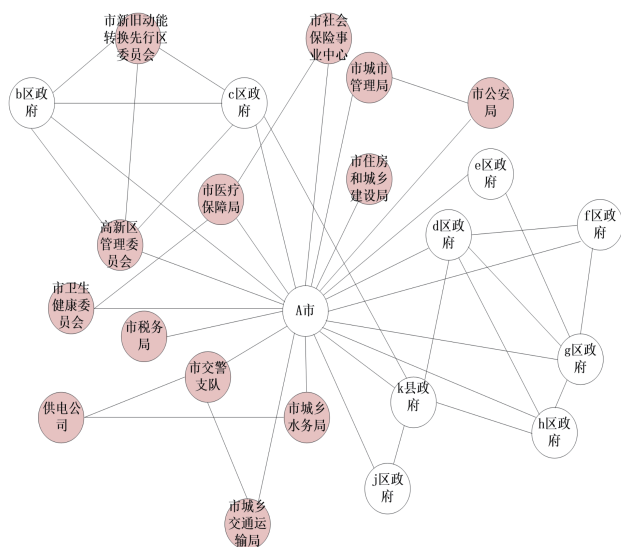


图4 布局优化结果

2.4 方法性能对比

运用新方法完成知识图谱实体扩展与布局优化,生成新的知识图谱后,以此应用其余两种文献提出方法构造新的知识图谱。应用3个新的知识图谱面向不同数据集执行链路预测任务,并利用常规平均排名、平均倒序排名、Hits@3三个评价指标,对知识图谱给出的链路预测结果准确程度进行评估。

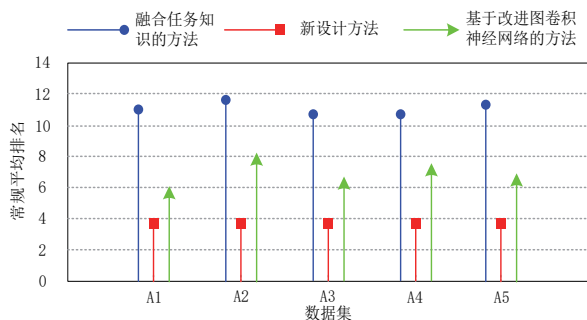
$$M = \frac{1}{|S|} \sum_{o=1}^{|S|} r_o \quad (10)$$

$$\hat{M} = \frac{1}{|S|} \sum_{o=1}^{|S|} \frac{1}{r_o} \quad (11)$$

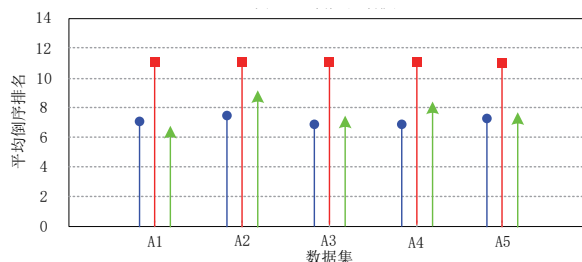
$$H = \frac{1}{|S|} \sum_{o=1}^{|S|} \prod (r_o \leq 3) \quad (12)$$

式中: M 表示常规平均排名; $\hat{M}$ 表示平均倒序排名; H 表示排名在前3的真实三元组的比例(Hits@3); S 表示测试集中真实三元组的总数(需要预测的链路数量); o 表示三元组编号; r 表示排名函数。

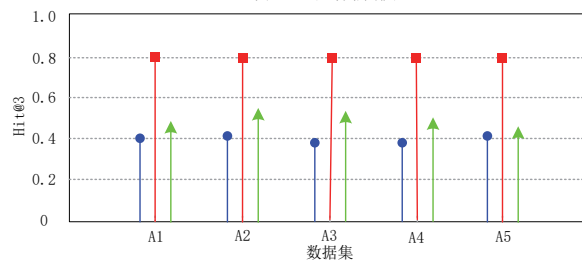
基于式(10~12),对3个新型知识图谱的实验应用效果进行深入分析,最终得出图5所示的对比结果。



图(a) 常规平均排名



图(b) 平均倒序排名



图(c) Hits@3

图5 常规平均排名、平均倒序排名、Hits@3对比结果

从图5可以看出,运用新方法构造的新型知识图谱应用后,所得链路预测结果的常规平均排名取值为3.9,相比其余方法大幅下降,平均倒序排名、Hits@3取值分别稳定在11.2和0.8,与另外两种方法相比大幅提升。综上所述,通过新研究方法构造出的知识图谱,具有更优的数据管理性能,能够实现准确的链路预测,表明新研究方法的应用价值。

3 结语

本文研究了知识图谱实体扩展与布局优化问题,提出了一种基于张量分解技术的新方法。利用张量分解技术强大的特征挖掘能力,有效捕获知识图谱中的潜在特征信息,生成准确的实体扩展结果。进而,通过表示学习模型捕获扩展图谱中实体间的推理关系,并借助DSM解耦优化实体布局、简化关系路径,获得布局优化的知识图谱。该方法显著提升了知识图谱的数据管理能力,有助于推动知识图谱在更广泛领域的应用。

参考文献:

- [1] 宋宝燕,刘杭生,单晓欢,等.融合关系模式和类比迁移的知识图谱补全方法[J].计算机科学,2025,52(3):287-294.
- [2] 陈强,张栋,李寿山,等.融合任务知识的多模态知识图谱补全[J].软件学报,2025,36(4):1590-1603.
- [3] 杨强,周欣,程军军,等.基于改进图卷积神经网络的知识图谱补全方法[J].计算机应用与软件,2025,42(1):82-89.
- [4] 周粤,范永胜,桑彬彬,等.基于双子图和注意力机制的知识图谱补全方法[J].计算机应用研究,2025,42(1):93-99.
- [5] 符玉飞,周莲英,丁腊春.融合知识图谱的KGCN推荐算法优化研究[J].计算机与数字工程,2024,52(12):3590-3594.

贝叶斯优化 SVM-RF 的下肢运动意图识别

吴仁愿^{1,2} 唐文超^{2*} 樊晓龙²
WU Renyuan TANG Wenchao FAN Xiaolong

摘要

人口老龄化加剧和下肢障碍患者增多, 推动了下肢康复机器人技术发展。精准识别人体运动意图是人机交互、智能康复及机器人控制的核心环节, 但仍面临诸多挑战。为此, 文章提出了一种基于贝叶斯优化的支持向量机与随机森林融合模型 (SVM-RF(BO))。该模型采用多模型融合策略, 结合贝叶斯优化调参技术, 在随机森林与支持向量机的基础上, 引入逻辑回归作为元学习器。经过实验表明, 该模型的识别准确率达到 97.97%。所提研究成果不仅将促进康复机器人的未来发展, 还能更好地辅助患者进行康复运动。

关键词

下肢康复机器人; 人体运动意图; 贝叶斯优化; 支持向量机; 随机森林

doi: 10.3969/j.issn.1672-9528.2025.12.019

0 引言

随着社会老龄化进程加快及下肢功能障碍患者数量增多, 人们对可穿戴式智能康复运动辅助机器人的关注度不断提升, 相关康复机器人技术也在持续迭代进步^[1]。其中, 随着人工智能与生物医学工程的深度融合, 人体运动意图识别作为人机交互 (human-machine interaction, HMI) 与

康复医疗领域的核心技术, 近年来受到广泛关注。该技术通过解析人体神经、肌肉或运动信号, 实时推断用户的动作意图, 为智能假肢控制、外骨骼机器人辅助、脑卒中康复训练等场景提供关键技术支撑。Abidi^[2] 提出一个自适应混合网络 (AHN) 来有效地预测人类运动意图, 将收集的 EEG/EMG 信号转换为频谱图图像并发送到 AH-CNN-LSTM, 为了增强预测, 使用改进的黄鞍山羊算法 (IYSGA) 对 AH-CNN-LSTM 和 AH-CNN-Res-LSTM 技术中的参数进行了优化。Ansah 等^[3] 提出了一种仅基于大腿安装惯性测量单元 (IMU) 的新型意图识别方法和一种基于事件的特征提取方法, 该方法在不同速度和对象上实现了良好的泛化性能。该系统在 Leave-One-Subject-Out 验证中对所有测试对象的准确率均高于 99%。邹锋等^[4] 结合 SEMG 信号与创新的非洲秃鹫优化算法提出了一种新的人体运动意图识别算法, 识别准确率能够达到 92.71%。赵伟明

1. 安徽理工大学煤炭安全精准开采国家地方联合工程研究中心 安徽淮南 232001
 2. 安徽理工大学人工智能学院 安徽淮南 232001
- [基金项目] 淮南市科技计划立项项目 (2023A3112); 煤炭安全精准开采国家地方联合工程研究中心开放基金 (EC2023015); 安徽理工大学青年教师科学研究基金资助项目 (XCZX2021-01); 安徽理工大学引进人才科研启动基金 (2022yjrc18); 机械工业矿山采选装备智能化重点实验室开放基金项目 (2022KLMI03)

- [6] 胡晶. 基于 SparkGraphX 的交通动态图谱分析与优化 [J]. 电脑与信息技术, 2025,33(2):69-73.
- [7] 叶志鸿, 吴运兵, 戴思翀, 等. 多级融合知识图谱补全模型 [J]. 计算机科学与探索, 2025,19(3):724-737.
- [8] 姬淼, 井娟丽, 张子琦, 等. 发电运行技术的知识图谱优化研究 [J]. 电子元器件与信息技术, 2025,9(3):74-76.
- [9] 杨荣泰, 邵玉斌, 杜庆治, 等. 基于子图结构语义增强的少样本知识图谱补全 [J]. 北京邮电大学学报, 2024, 47(4): 71-76.

- [10] 牛增贤, 刘海峰, 徐伟峰, 等. 基于扩展 Span 表示的电力变压器运维知识抽取与知识图谱构建 [J]. 吉林大学学报 (理学版), 2023, 61(5): 1112-1122.

【作者简介】

李寿兴 (1990—), 男, 云南大理人, 硕士, 讲师, 研究方向: 计算机应用、教育信息化。

(收稿日期: 2025-07-01 修回日期: 2025-12-08)