

# 程序在线测评系统的知识图谱构建与应用

宋玮情, 崔焕庆, 杨峻铸

(山东科技大学计算机科学与工程学院, 山东 青岛 266590)

**摘要:**近年来,各高校为提高学生编程能力,上线了大量程序在线测评系统用于计算机类专业课程教学。为清晰了解此类系统中数据对教学的影响,首先依据程序在线测评系统分析数据包含的本体,经过数据清洗,运用Neo4j构建知识图谱。通过对学习路径、学情分析和教学路径的可视化分析得出结论:程序在线测评系统知识图谱可将教学信息以知识图谱的形式呈现,使学习者能清楚了解自身水平,提升学习兴趣,实现教师高效化教学和学生自主化学习。

**关键词:**知识图谱;程序在线测评系统;Neo4j图数据库;可视化

**DOI:** 10.11907/rjdk.212040

开放科学(资源服务)标识码(OSID):



中图分类号:G434

文献标识码:A

文章编号:1672-7800(2022)001-0075-04

## Construction and Application of Knowledge Graph for Program Online Evaluation System

SONG Wei-qing, CUI Huan-qing, YANG Jun-zhu

(School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

**Abstract:** In recent years, a large number of program online evaluation systems for the teaching of computer courses are launched in universities to improve students' programming ability. In order to better clearly understand the impact of data in such systems on teaching, analyze the ontology contained according to the data of the program online evaluation system, and then use neo4j to construct the knowledge map after data cleaning. Through the visual analysis of learning path, learning situation analysis and teaching path, it is concluded that the knowledge map of program online evaluation system can express the teaching information in the form of knowledge map, so that learners can clearly understand their own level, improve their learning interest, and realize the high efficiency of teachers' teaching and students' autonomy of learning.

**Key Word:** knowledge graph; program online evaluation system; Neo4j graph database; visualization

## 0 引言

随着大数据时代来临,数据价值日益增加。面对多样化的数据形式,传统关系型数据库只能处理简单的数据关系,而知识图谱能高效地存储和应用复杂的关系知识,受到了学者们的广泛关注和研究,目前已成为知识应用和资源管理的重要技术之一<sup>[1]</sup>。知识图谱由节点和边组成,其中节点表示概念和实体,概念指抽象出的事物。实体指具体事物,基本特征通过属性表示;边表示实体与概念之间的关系。知识图谱本质是一个具有属性的实体,通过关系相互连接所构建的网状知识库,并以结构化的形式描述客

观世界的概念、实体和两者之间的相互关系<sup>[2]</sup>。因此,知识图谱不仅表示语义关系,还能联系各类信息构建关系网络,从“关系”的角度分析问题。

近年来,传统课堂教学模式存在教师教学强度大、学生学习效率低、教学资源利用率低等问题。为了解决上述问题,提出了程序在线评测(Online Judge, OJ)系统,该系统在程序设计等课程的教学、竞赛中受到广泛应用,获取了海量的数据资源,这些资源是计算机教育的重要组成部分<sup>[3]</sup>。如何从数据资源中分析出有价值的信息辅助教学是目前的研究热点。基于此,本文结合知识图谱对OJ系统的数据进行处理,构建OJ知识图谱,并通过挖掘、分析数据,为个性化教学和学习提供数据支撑,最后以个例分析的形式

收稿日期:2021-08-12

基金项目:全国高等院校计算机基础教育研究会计算机基础教育教学研究项目(2021-AFCEC-064);山东科技大学优秀教学团队建设计划项目(JXTD20180503);山东科技大学青年教师教学拔尖人才培养项目(BJ20200505)

作者简介:宋玮情(1996-),女,山东科技大学计算机科学与工程学院硕士研究生,研究方向为知识图谱;崔焕庆(1979-),男,博士,山东科技大学计算机科学与工程学院教授,研究方向为大数据分析、图计算和位置服务;杨峻铸(1998-),男,山东科技大学计算机科学与工程学院硕士研究生,研究方向为知识图谱。本文通讯作者:宋玮情。

式讨论 OJ 知识图谱对学情分析与路径规划的实际应用价值。

## 1 相关工作

知识图谱由三元组表示,表达式为 $\langle head, relation, tail \rangle$ ,简记为 $\langle h, r, t \rangle$ 。其中, $h$ 和 $t$ 分别表示三元组的头实体和尾实体, $r$ 表示头实体和尾实体之间的关系,每个节点和关系都通过键值对(Key-value)进行表示<sup>[4]</sup>。知识图谱可通过关系型数据库、RDF三元组库和图数据库进行存储。其中,图数据库是基于图形式存储的新型非关系型数据库,具有存储灵活、查询高效等优势,适合表示如(实体,属性名,属性值)、(实体1,关系,实体2)等形式数据,并且关联查询无需复杂操作<sup>[5]</sup>。

马腾等<sup>[6]</sup>建立了个性化教学知识图谱,将知识图谱融入教师备课、课堂教学、课下作业布置、课外辅导等一系列教学环节当中,自动为学生提供感兴趣的学习内容,实现了精准化和个性化教学。黄梅根等<sup>[7]</sup>创建了初、高中知识图谱,基于知识图谱使用改进的认知诊断模型诊断学生对特定知识点的认知程度。Yan<sup>[8]</sup>以“国际贸易实务”课程为研究实例,通过融合教学理论与教学资源构建知识图谱,帮助学生提高学习效率;通过培养逻辑思维能力和改进教学方法,提高教师教学水平。Zheng<sup>[9]</sup>从MOOC网站上获取课程的相关信息,对实体和关系进行提取,构建高等教育知识图谱。

本文基于Neo4j构建OJ数据的知识图谱,给出了OJ数据的预处理过程。然后,Web端通过可视化方式展现OJ数据的网络关系,进一步分析OJ知识图谱在学情分析、个性化教学和学习方面的应用。

## 2 OJ知识图谱的本体设计

为了构建实用性强、结构合理的OJ知识图谱,本文设计主要从2个方面考虑:①OJ系统中实体和关系的组织;②现实应用对OJ知识图谱的功能需求。OJ系统中包含的节点种类较少,但节点之间的关系数量很多,需要设计明确的知识架构作为创建知识图谱的基础<sup>[10]</sup>,即定义OJ知识图谱中包含各实体的概念以及它们之间的关系。如表1所示,OJ知识图谱中主要包含测试、题、学生、老师4类实体,为了准确描述各实体信息,定义了每类实体的相关属性。实体属性分为对象属性和数据属性,其中对象属性描述实体之间的关系,值为相关的另一实体;数据属性表示实体的属性,值为XSD(XML Schema Definition)的数据类型。基于上述设计,将各类实体和关系相互组合构成了一个纵横交错复杂的知识图谱,如图1所示(彩图扫OSID码可见,下同)。

## 3 构建OJ知识图谱

### 3.1 数据采集

首先需要采集、分析数据。OJ系统使用关系型数据库

Table 1 Entity attribute table

表1 实体属性表

属性	属性类型	Domain	Range
测试名	数据属性	测试	xs:string
错误情况	数据属性	测试	xs:string
完成数	数据属性	题	xs:integer
正确数	数据属性	题	xs:integer
问题描述	数据属性	题	xs:string
用户名	数据属性	学生	xs:string
提交数	数据属性	学生	xs:integer
通过数	数据属性	学生	xs:integer
开设	对象属性	测试	老师
包含	对象属性	测试	题
成绩	对象属性	题	学生
顺承	对象属性	题	题

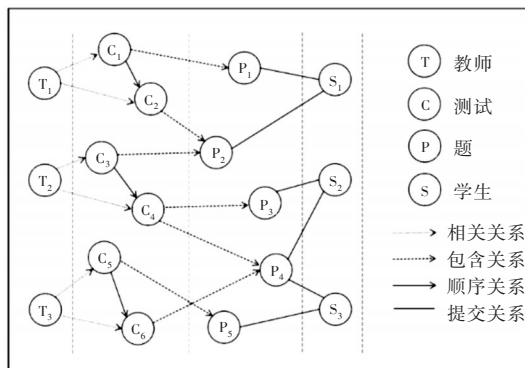


Fig. 1 Ontology structure of OJ knowledge graph

图1 OJ知识图谱的本体结构

存储数据,先将数据以逗号分隔的CSV文件导出,为了防止乱码,所有CSV文件格式都使用UTF-8进行编码。如表2所示,文件存储了OJ系统的实体、关系及它们之间的属性,每条记录都拥有唯一的ID属性,然后利用Neo4j中Cypher的LOAD CSV命令将数据导入至Neo4j。

Table 2 CSV file table

表2 CSV文件表

CSV文件	说明
contest.csv	实体集,包含测试的基本信息
student.csv	实体集,使用OJ系统的学生
problem.csv	实体集,包含题目
solution.csv	关系集,用于创建学生实体与题目实体之间的关系
contest_property.csv	属性集,包含测试实体的类型属性
contest_problem.csv	关系集,用于创建测试、问题两类实体之间的关系

### 3.2 知识图谱的构建方式

知识图谱构建的主要方式有:自底向上、自顶向下和自动构建3种。自底向上的构建方式根据人们已知知识为主导,对相关领域概念进行层次划分,该方式适合构建通用型知识图谱;自顶向下的构建方式先从比较权威的公开数据集或专业的网站中获取数据信息,再通过技术手段获取所需的信息实体、关系及属性,适合在知识明确、关系清晰的领域构建知识图谱;自动构建则通过机器学习、数据自动获取、统计等技术构建知识图谱。由于OJ系统应用于特定领域,本文采用自顶向下的构建方式,将实体、关系及属性信息输入图形数据库Neo4j中实现知识图谱的可视化

展示及存储。

基于山东科技大学 OJ 系统中 2019 年程序设计基础(C 语言)实验课程数据构建知识图谱,其中数据包括 924 个实体(6 名老师、298 道题目、164 次测试和 456 名学生)和 153 612 条关系。图 2 展示了 OJ 知识图谱的部分内容,其中学

生、题目之间的提交关系占据了非常重要的地位,它们不仅展示了学生对某个题目所涵盖知识点的学习情况,还体现了学生的努力程度、学习能力和学习效果。此外,对这些关系进行分析有助于教师提供个性化辅导和教学,有助于实现学生学习效果的形成性评价。

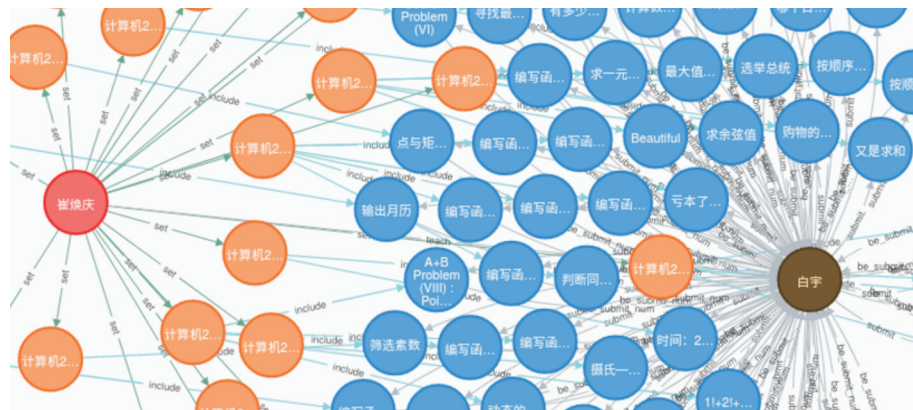


Fig. 2 Partial knowledge graph

图 2 部分知识图谱

### 4 OJ 知识图谱的应用

目前,在线教育领域存在老师获取反馈不及时、学生定位<sup>[11]</sup>和学习路径<sup>[12]</sup>不清晰、教师学生数据管理繁琐等问题,借助 OJ 知识图谱可有效解决上述问题。本文基于 Neo4j 的查询语言 Cypher 对知识图谱信息进行检索,分别从学生、教师和教学管理人员角度,以个例分析方式对 OJ 知识图谱的潜在应用进行说明。

从学生角度而言,由于学生存在学习努力程度、学习能力、兴趣爱好等个体差异,基于知识图谱分析学生题目提交和通过情况及各种信息,对学生进行画像<sup>[13-14]</sup>,能够更好地掌握每个学生的特点。进一步,根据 K-means 聚类算法分析学生的学习情况,提取学生之间的相似度,如图 3 所示。通过聚类分析后,选择相似用户有助于后续工作的研究。例如:如图 4 所示,通过分析学习者的相似度,结合知识图谱可个性化定制学生的学习路径<sup>[15]</sup>。

从教师角度而言,OJ 知识图谱帮助教师分析学生对不同问题的掌握程度,判断学生的学习水平和能力,实现学

情分析,从而有针对性地调整教学计划和提供个性化辅导。如图 5 所示,76.8% 的学习者在答题过程中出现答案错误,表明学习者对相关知识点的学习程度不够,还需加强练习。

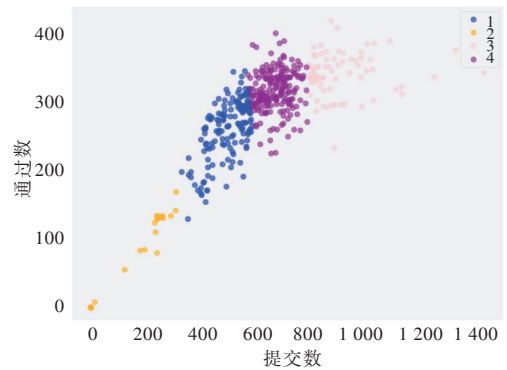


Fig. 3 Student clustering

图 3 学生聚类

从教学管理者角度而言,借助知识图谱能有效掌握教学质量。如图 6 所示,以两位老师(实体为红色节点)为例,分析代表测试集的橙色节点顺序,可生成每位教师的教学路径。如果当测试集由题目组成时,则分析题目的先后顺序可生成教学路径,进一步分析可为教师进行教学画像。



Fig. 4 Student learning path planning

图 4 学生学习路径规划

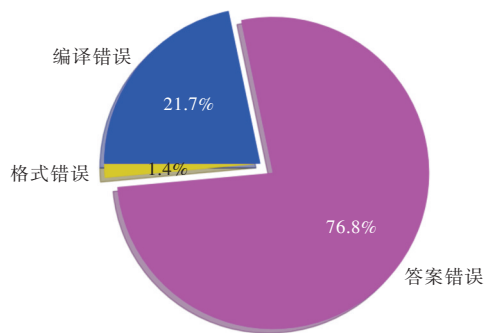


Fig. 5 Students' errors in one test

图5 一次测试中学生的错误情况

### 5 结语

本文对OJ知识图谱进行本体设计,通过图数据库Neo4j和程序在线测评系统数据构建OJ知识图谱,并利用可视化方法分析OJ知识图谱的潜在应用。首先,OJ知识图谱能显著提高学生学习兴趣和学习效果,其次能协助教师进行学情分析,协助教学管理人员进行评教和评学,OJ知识图谱能在一定程度上满足现实课程教学需求。未来将进一步完善OJ知识图谱,为智能问答、成绩预测分析、个性化课程推荐、关系挖掘等应用提供更强大的推动力。

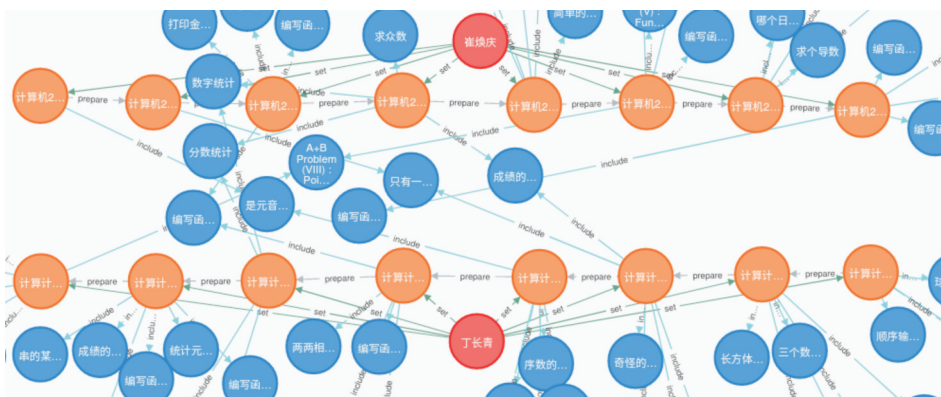


Fig. 6 Analysis of teaching paths of different teachers

图6 不同教师的教学路径分析

### 参考文献:

[1] RAVAT F, SONG J F, TESTE O, et al. Efficient querying of multidimensional RDF data with aggregates: comparing NoSQL, RDF and relational data stores [J]. International Journal of Information Management, 2020, 54: 102089.

[2] ZHOU H J, SHEN T T, LIU X L, et al. Survey of knowledge graph approaches and applications [J]. Journal on Artificial Intelligence, 2020, 2(2): 89-101.

[3] WANG M, HAN W T, CHEN W G. MetaOJ: a massive distributed online judge system [J]. Tsinghua Science and Technology, 2021, 26(4): 548-557.

[4] LI M D, SUN Z Y, ZHANG S H, et al. Enhancing knowledge graph embedding with relational constraints [J]. Neurocomputing, 2021, 429: 77-88.

[5] FALCAO T A, FURTADO P M, QUEIROZ J S, et al. Comparative analysis of graph databases for git data [J]. Journal of Physics: Conference Series, 2021, 1944: 012004.

[6] MA T, NI R K, LI Y R, et al. Research on the application of knowledge atlas in personalized teaching [J]. China Arab Technology Forum (Chinese and English), 2021(2): 177-180.

马腾,倪睿康,李艳茹,等.知识图谱在个性化教学中的应用研究 [J].中阿科技论坛(中英文),2021(2):177-180.

[7] HUANG M G, LIU C, DU H, et al. Cognitive diagnosis model based on knowledge map and its application in teaching aids [J]. Computer Science, 2021(S1): 644-648.

黄梅根,刘川,杜欢,等.基于知识图谱的认知诊断模型及其在教辅中的应用研究 [J].计算机科学,2021(S1):644-648.

[8] YAN L. The application of the blended teaching model based on the

knowledge graph in the course of "international trade practice" [J]. Advances in Educational Technology and Psychology, 2021, 5(5): 115-121.

[9] ZHENG Y, LIU R F, HOU J M. The construction of high educational knowledge graph based on MOOC [C]//Chengdu: Proceedings of 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2017), 2017.

[10] LIN Q, ZHU Y F, LU H, et al. Improving university faculty evaluations via multi-view knowledge graph [J]. Future Generation Computer Systems, 2021, 117: 181-192.

[11] ZHU J, ZHANG L J, LIANG W Y. A review of research on personalized adaptive learning driven by data [J]. Journal of South China Normal University (Natural Science Edition), 2020(4): 17-25.

朱佳,张丽君,梁婉莹.数据驱动下的个性化自适应学习研究综述 [J].华南师范大学学报(自然科学版),2020(4):17-25.

[12] ZHU H P, TIAN F, WU K, et al. A multi-constraint learning path recommendation algorithm based on knowledge map [J]. Knowledge-Based Systems, 2018, 143: 102-114.

[13] LANG Y K, WANG G Z. Personalized knowledge point recommendation system based on course knowledge graph [J]. Journal of Physics Conference Series, 2020, 1634: 012073.

[14] WU Y W, SUN C H, LI B. Knowledge graph helps discipline teaching: taking the basic course of college analog electronic technology as an example [J]. Software Guide, 2020, 19(12): 195-198.

吴彦文,孙晨辉,李斌.知识图谱助力学科教学:以大学模拟电子技术基础课程为例 [J].软件导刊,2020,19(12):195-198.

[15] NABIZADEH A H, JOSE, LEAL P, et al. Learning path personalization and recommendation methods: a survey of the state-of-the-art [J]. Expert Systems With Applications, 2020, 159: 113596.

(责任编辑:刘嘉文)