

文章编号:1671-251X(2019)06-0042-05

DOI:10.13272/j.issn.1671-251x.2019030013

煤矿巷道支护领域知识图谱构建

吴雪峰¹, 赵志凯^{2,3}, 王莉⁴, 翟勃⁵, 王卫龙⁵

(1. 中国矿业大学 徐海学院, 江苏 徐州 221008; 2. 中国矿业大学 物联网(感知矿山)研究中心, 江苏 徐州 221008; 3. 中国矿业大学 矿山互联网应用技术国家地方联合工程实验室, 江苏 徐州 221008; 4. 中国矿业大学 力学与土木工程学院, 江苏 徐州 221116; 5. 山东能源淄博矿业集团有限责任公司 信息中心, 山东 淄博 255100)



扫码移动阅读

摘要:针对煤矿巷道支护专家系统知识内容有限、知识难以融合共享、无法从非结构化数据中挖掘相关知识等问题,构建了煤矿巷道支护领域知识图谱。首先通过设计领域概念、关系及属性对煤矿巷道支护领域知识建模;然后从煤矿巷道支护领域结构化、半结构化、非结构化数据源获取知识,并基于深度学习模型 BI-LSTM-CRF 进行实体识别;最后利用图数据库 Neo4j 存储煤矿巷道支护领域知识,形成煤矿巷道支护领域知识图谱。煤矿巷道支护领域知识图谱可进一步提升煤矿巷道支护设计和管理效率,为煤矿巷道支护智能化管理提供知识支持。

关键词:巷道支护;知识图谱;支护专家系统;语义;知识库

中图分类号:TD79 文献标志码:A

Construction of knowledge graph of coal mine roadway support field

WU Xuefeng¹, ZHAO Zhikai^{2,3}, WANG Li⁴, ZHAI Bo⁵, WANG Weilong⁵

(1. Xuhai College, China University of Mining and Technology, Xuzhou 221008, China; 2. Internet of Things (Perception Mine) Research Center, China University of Mining and Technology, Xuzhou 221008, China; 3. The National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, China University of Mining and Technology, Xuzhou 221008, China; 4. School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China; 5. Information Center, Shandong Energy Zibo Mining Group Co., Ltd., Zibo 255100, China)

Abstract:In view of problems of limited knowledge content, difficult knowledge fusion and sharing, and inability to mining related knowledge from unstructured data existed in coal mine roadway support expert system, a knowledge graph of coal mine roadway support field was constructed. Firstly, knowledge of coal mine roadway support field is modeled by designing field concept, relationship and attribute. Then, knowledge is acquired from structured, semi-structured and unstructured data sources of coal mine roadway support field, and entity recognition is carried out based on deep learning model BI-LSTM-CRF. Finally, knowledge graph of coal mine roadway support field is formed by using graph database Neo4j to store knowledge of coal mine roadway support field. The knowledge graph of coal mine roadway support field can further improve efficiency of roadway support design and management, and provide knowledge support for intelligent management of coal mine roadway support.

Key words:roadway support; knowledge graph; support expert system; semantics; knowledge base

收稿日期:2019-03-05;修回日期:2019-05-24;责任编辑:盛男。

基金项目:国家重点研发计划资助项目(2017YFC0804401)。

作者简介:吴雪峰(1980—),男,江苏盐城人,讲师,硕士,主要研究方向为机器学习、矿山物联网技术,E-mail:cumtaust@126.com。

引用格式:吴雪峰,赵志凯,王莉,等.煤矿巷道支护领域知识图谱构建[J].工矿自动化,2019,45(6):42-46.

WU Xuefeng,ZHAO Zhikai,WANG Li,et al. Construction of knowledge graph of coal mine roadway support field[J]. Industry and Mine Automation,2019,45(6):42-46.

0 引言

随着煤矿巷道支护研究^[1-3]不断深入,以专家系统为代表的巷道支护信息化和智能化得到了广泛应用。杨仁树等^[4]采用基于规则和典型案例2条途径推理出巷道支护初始方案,运用FLAC3D二次开发技术实现了巷道支护方案专家系统;王连国等^[5]提出了软岩巷道锚注支护设计专家系统结构模型,并构建了锚注支护设计专家系统的核心推理机;张新蛮^[6]采用VC++6.0开发了霍州煤电集团巷道支护专家系统;肖福坤等^[7]开发了井巷掘进支护设计的智能专家系统;张金山等^[8]以知识、信息、经验和事实为基础,利用Xi-plus建立了矿山巷道支护设计专家咨询系统。大量专家系统的普及应用在一定程度上提升了巷道支护设计和管理效率。

煤矿巷道支护设计和管理需要大量的知识支持。巷道支护专家系统针对该难点建立了知识库,但由于特定范围内的专家知识内容有限,不同专家系统知识又难以融合共享,此外现有专家系统的知识库无法和开放领域的大规模知识库(如Freebase, Wikidata, DBpedia, YAGO等)建立链接,无法从非结构化数据中挖掘相关知识。

知识图谱是大数据时代知识表示的重要方式之一^[9],是实现认知智能的关键技术。目前,随着智能信息服务应用的不断发展,知识图谱已被广泛应用于智能搜索、智能问答、个性化推荐等领域。通过知识图谱能将信息、数据及链接关系聚集为知识,使信息资源更易于计算、理解,并形成一整套语义知识库。因此,本文构建了煤矿巷道支护领域知识图谱。

1 知识图谱构建流程

知识图谱是一种结构化的语义知识库,用于描述现实世界中的概念、实体、事件及其之间的关系,使得机器理解与解释成为可能。本体能够为知识图谱提供基本的数据模式,即概念及概念之间的关系。知识图谱在本体基础上,具有更加丰富的语义关系,尤其在实体信息和实体关系表示方面更加灵活和多样。

在知识图谱中,实体对应客观世界中的事物,属性代表事物的特征,概念是对具有相同属性事物的概括和抽象,事件可表示为不同时空状态下一系列实体及其联系。通过概念与概念、实体与实体、概念与实体、实体与属性之间的关系,形成网状的知识结构。这些关系是构成知识图谱的最基本知识单元,如果采用有向标记图来表示知识图谱,则图中每一条边对应1个三元组(Subject——主语,

Predicate——谓语, Object——宾语)。构建领域知识图谱就是要将相关领域中涉及的知识单元通过共同的节点(概念或实体)联系起来,形成该领域适用的大规模语义网络。

知识图谱构建主要有自顶向下和自底向上2种方式^[10]:自顶向下构建是指先为知识图谱定义好概念及关系模式,再将实体按照预定义好的模式加入知识库;自底向上构建是指从一些开放链接数据中提取出实体,选择其中置信度较高的加入知识库,将具有相似属性的实体进行抽象归纳,形成相应的概念,再逐层向上汇集,最终构建顶层概念本体模式。对于通用领域的知识,知识范围相对广泛,可采用自底向上的方式构建知识图谱;而对于特定领域,由于知识范围相对明确,一般采用自顶向下的方式构建知识图谱。本文将2种方式相结合,先通过设计领域概念、关系及属性进行知识建模,再从不同数据源中识别实体、关系、属性来获取知识,然后将获取的知识进行融合,存入知识库,形成煤矿巷道支护领域知识图谱,流程如图1所示。

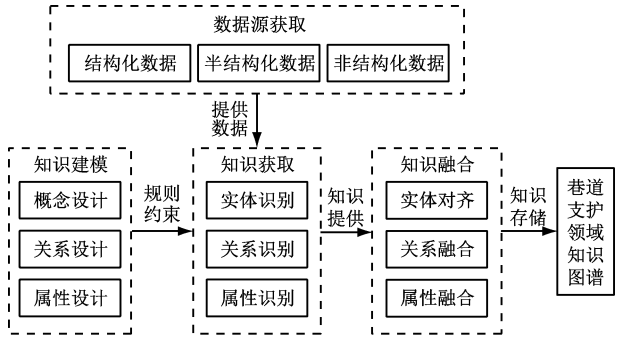


图1 煤矿巷道支护领域知识图谱构建流程
Fig.1 Construction process of knowledge graph of coal mine roadway support field

2 基于本体的知识建模

知识建模是构建知识图谱的首要步骤,是对知识结构进行设计的过程,也称为模式设计。基于本体的知识建模方法能够有效描述事物的基本特征、内在联系和演化规律。本体的目标是捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并从不同层次的形式化模式上给出这些词汇间相互关系的明确定义^[11]。本体作为共享的形式化概念模型,清晰定义概念之间的关系,使得被刻画的知识富含语义,具有良好的概念层次结构和对逻辑推理的支持。

为消除人与机器的语义鸿沟,在对本体进行描述和建模时,需要采用统一的建模语言,并遵循相应的本体建模准则。典型的本体建模准则是T. R. Gruber提出的5条准则^[12]:明确性(Clarity)、一致

性(Coherence)、可扩展性(Extendibility)、最小编码偏差 (Minimal Encoding Bias)、最小本体承诺 (Minimal Ontological Commitment)。

基于本体的知识建模主要从概念、属性、关系等方面进行设计。根据巷道支护设计和施工有关标准规范,将巷道支护方案相关概念归纳为 6 大类:工程类、地质类、支护类、施工类、标准规范类、案例类。每类概念又可细分为多个子概念,形成具有层级关系的概念体系。每个概念类具有的特征不同,通过属性进行标志。现实世界中的实体对象对应特定的概念,并且具有不同特征值。煤矿巷道支护领域部分概念和属性见表 1。

表 1 煤矿巷道支护领域部分概念和属性

Table 1 Some concepts and attributes of coal mine roadway support field

类别	概念	属性
工程类	矿区	矿区名称、编号、类型、位置等
	巷道	巷道名称、用途、服务年限、断面形式等
地质类	围岩	岩石类型、强度、稳定性、应力、结构面等
支护类	支护形式	支护类型、适用条件等
	支护锚杆	锚杆类型、适用条件等
	支护材料	材料名称、标准要求等
施工类	支护施工	施工工艺、作业名称、要求等
	支护监测	监测类型、内容、要求等
标准规范类	标准	文号、标题、实施日期等
	规范	文号、标题、实施日期等
案例类	支护方案实例	工程名称、支护方式、施工情况等

利用 Protégé 工具构建的煤矿巷道支护领域部分概念本体如图 2 所示,其中 Thing 为 Protégé 自带的顶层本体。

知识图谱中概念之间具有上下位关系,实体与概念之间存在类属关系,实体和实体之间存在各种语义关系。2 个实体之间可能存在的关系类型取决于这 2 个实体所属类型(即概念)。因此,基于上述概念设计结果,利用工程经验可确定主要的关系类型,如支护施工和支护材料实体之间具有“使用”关系。

3 煤矿巷道支护领域数据源和知识获取

煤矿巷道支护领域知识来源包括各类结构化数据、半结构化数据及非结构化数据。

(1) 结构化数据为存储在关系数据库或面向对象数据库中的数据。在行业信息化发展过程中,各类巷道支护专家系统的数据绝大多数存储于关系数

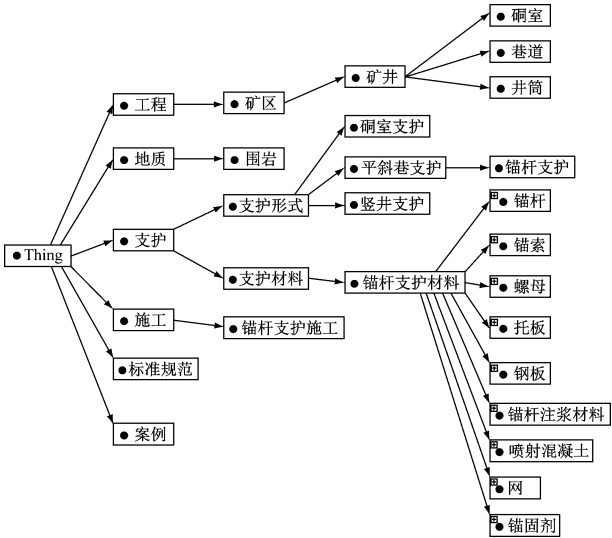


图 2 煤矿巷道支护领域部分概念本体

Fig. 2 Some concept ontology of coal mine roadway support field

据库中。关系数据库采用经典的关系模型,通常以二维表形式进行存储,直观且易于理解。本文在构建煤矿巷道支护领域知识图谱时,以结构化数据为基础,进而扩充其他数据。

(2) 各类行业网站的数据属于半结构化数据,其描述的内容属于特定领域,范围较窄,但具有相应的专业性,数据在一致性和完整性方面表现较好。由于网站内容自身的半结构化特性,通过特定的标签抽取其中的知识时,通常仅需按照其结构解析即可,如通过网络爬虫获取行业数据时,即可提取相应的语义标签。

(3) 文本作为非结构化数据是最丰富的数据资源,在巷道支护领域相关的手册、标准、规范等,以及互联网网页、开放链接数据、开放知识库、在线百科,甚至数据库中结构化数据的某些字段中,均存在大量文本。文本最大的特点是没有结构,因此从中学习知识的难度非常高。采用自然语言处理相关技术,可实现领域知识的自动识别和抽取^[13-14]。

实体识别是知识结构化和数字化的基础,在定义好的概念模式基础上进行实体识别,称为命名实体识别。命名实体识别不仅识别文本中的实体,还将其划分到指定类别(即预先定义的具有明确语义的概念),从而为实体的语义表示提供基础。如果该实体已经存在于知识库中,则只需将其链接到具有明确语义的实体,这个过程称为实体链接。作为知识图谱的基本单元,命名实体识别与链接是知识图谱构建的核心技术。由于人类知识具有复杂性、开放性、多样性和巨大规模等特征,知识通常动态递增。实体识别技术可检测文本中新实体,并将其加入现有知识库中,而实体链接技术可针对性地发现

特定实体的新知识。

以非结构化的文本数据为例,从文本中识别实体是获取知识的首要环节,常见的实体识别技术包括基于词典和规则的方法、基于统计机器学习的方法、基于深度学习的方法等。基于词典和规则的方法需要建立命名词典和获取实体名称的规则模板,适用于自然语言表述特征较明确的文本知识提取。基于统计机器学习的方法需要对获取知识的文本语料进行统计分析,选择能有效反映实体特性的特征集合,如单词特征、上下文特征、词典及词性特征、停用词特征、核心词特征及语义特征等,通过训练获得实体识别模型,该方法对语料库的依赖性很强。基于深度学习的方法具备强大的序列建模能力,能很好地捕捉上下文信息,同时具备神经网络非线性拟合能力,由于神经网络通常以向量作为输入,所以该方法需要进行词向量转换。本文采用基于深度学习的方法,运用典型的深度学习模型 BI-LSTM-CRF^[15]进行实体识别,流程如图 3 所示。

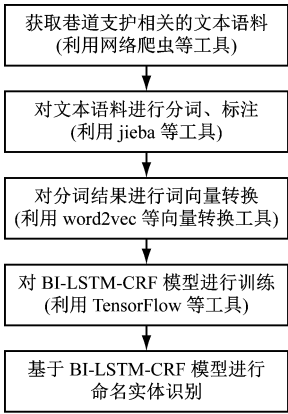


图 3 基于 BI-LSTM-CRF 模型的实体识别流程

Fig. 3 Entity recognition process based on BI-LSTM-CRF model

基于命名实体识别的结果可进一步识别实体之间的关系和实体属性,从自然语言处理的角度来看,关系识别、属性识别和实体识别的原理相似,都属于标注问题,不再赘述。

4 煤矿巷道支护领域知识存储

目前,知识图谱的知识存储主要通过关系型数据库管理系统、资源描述框架(Resource Description Framework, RDF)三元组、图数据库系统 3 种方式实现。① 借助关系型数据库管理系统存储知识,能充分发挥关系型数据库管理系统的稳定性和成熟性等优势,但没有直接表示多种实体关系的有效方式,通常需要借助大量表连接操作实现,因此在获取知识时运算开销大,难以应对超大规模复杂语义知识获取需求。② 采用 RDF 三元组存储

知识,能利用 RDF 较成熟的规范化表示机制,但使用 RDF 三元组描述语义关系要求必须事先定义好符合 RDF 规则的特定领域所需词汇表,在查询和管理方面还需借助特定的 RDF 工具。③ 利用图数据库系统存储知识,能充分发挥图数据库对图结构数据的表示能力,利用图数据库提供的管理工具,能快速、方便地处理各类复杂知识获取需求。目前,图数据库系统中 Neo4j 使用最普及。

本文利用 Neo4j 存储煤矿巷道支护领域知识,其中概念、实体、关系分别对应 Neo4j 中标签、节点、边/关系。Neo4j 提供 Cypher 命令对图数据进行操作,常用的语句包括 LOAD(批量导入 CSV 格式的数据)、CREATE(创建节点或关系)、MATCH(查找匹配节点或关系)、WHERE(设置匹配条件)、SET(设置节点或关系的属性值)、RETURN(返回需要的节点、关系或属性)。由于从多种数据源中提取的煤矿巷道支护知识已根据实体和关系类型分别存放在不同的 CSV 文件中,所以利用 LOAD 语句批量导入相应的知识。

批量导入概念/实体:

```
LOAD CSV FROM "file:///supportNode.csv" AS line
MERGE (n:Term {id: toInteger(line[0])})
SET n.name = line[1], n.caption=line[2]
```

批量导入关系:

```
LOAD CSV FROM "file:///supportRelation.csv" AS line
MATCH (from:Term{id: toInteger(line[0])},
(to:Term{id: toInteger(line[2])})
MERGE (from)-[r:{rcaption: toString(line[1])}]->(to)
```

将煤矿巷道支护知识批量导入 Neo4j 后形成煤矿巷道支护领域知识图谱。知识图谱是一种开放式的知识结构,如果不考虑存储工具的限制,理论上基于现有知识图谱能无限拓展领域相关各种类型知识,后续可利用 Cypher 命令实现新知识的动态集成和管理。

5 结语

从概念、关系、属性等方面对煤矿巷道支护领域知识建模,通过煤矿巷道支护领域结构化、半结构化、非结构化数据源获取知识,并将获取的知识融合后存储至图数据库 Neo4j 中,形成煤矿巷道支护领域知识图谱,可进一步提升煤矿巷道支护设计和管理效率,为煤矿巷道支护智能化管理提供知识支持。

参考文献(References):

[1] 左建平,文金浩,胡顺银,等.深部煤矿巷道等强梁支

护理论模型及模拟研究[J]. 煤炭学报, 2018, 43(增刊 1):1-11.

ZUO Jianping, WEN Jinhao, HU Shunying, et al. Theoretical model and simulation study of uniform strength beam support in deep coal mine roadway[J]. Journal of China Coal Society, 2018, 43(S1):1-11.

[2] 康红普, 王金华, 林健. 煤矿巷道支护技术的研究与应用[J]. 煤炭学报, 2010, 35(11):1809-1814.

KANG Hongpu, WANG Jinhua, LIN Jian. Study and applications of roadway support techniques for coal mines [J]. Journal of China Coal Society, 2010, 35(11):1809-1814.

[3] 王帆, 赵继云. 巷道支护支架液压缸位置闭环控制系统设计[J]. 工矿自动化, 2017, 43(10):74-77.

WANG Fan, ZHAO Jiyun. Design of closed-loop control system for hydraulic cylinder position of roadway support[J]. Industry and Mine Automation, 2017, 43(10):74-77.

[4] 杨仁树, 马鑫民, 李清, 等. 煤矿巷道支护方案专家系统及应用研究 [J]. 采矿与安全工程学报, 2013, 30(5):648-652.

YANG Renshu, MA Xinmin, LI Qing, et al. Research on expert system of roadway supporting schemes for coal mine and its application[J]. Journal of Mining & Safety Engineering, 2013, 30(5):648-652.

[5] 王连国, 陆银龙, 孙小康. 软岩巷道锚注支护智能设计专家系统及应用[J]. 采矿与安全工程学报, 2016, 33(1):1-6.

WANG Lianguo, LU Yinlong, SUN Xiaokang. Intelligent expert system for designing bolt-grouting support of soft-rock roadways and its applications [J]. Journal of Mining & Safety Engineering, 2016, 33(1):1-6.

[6] 张新奎. 巷道支护专家系统软件的开发与实现[J]. 中国矿业, 2012, 21(3):100-102.

ZHANG Xinman. Development and implementation of roadway support expert system [J]. China Mining Magazine, 2012, 21(3):100-102.

[7] 肖福坤, 孙豁然, 刘晓军, 等. 煤矿巷道支护智能决策系统[J]. 辽宁工程技术大学学报(自然科学版), 2004, 23(3):293-295.

XIAO Fukun, SUN Huoran, LIU Xiaojun, et al. Intelligence decision system for tunnel supporting in coal mine [J]. Journal of Liaoning Technical University (Natural Science Edition), 2004, 23(3):293-295.

[8] 张金山, 侯殿昆. 矿山巷道支护设计专家咨询系统 [J]. 金属矿山, 2003(5):7-9.

ZHANG Jinshan, HOU Diankun. Expert advisory system for the tunnel support design of underground mine[J]. Metal Mine, 2003(5):7-9.

[9] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报(自然科学版), 2017, 40(3):454-459.

LI Juanzi, HOU Lei. Reviews on knowledge graph research [J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(3):454-459.

[10] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3):582-600.

LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques [J]. Journal of Computer Research and Development, 2016, 53(3):582-600.

[11] 邓志鸿, 唐世渭, 张铭, 等. Ontology 研究综述[J]. 北京大学学报(自然科学版), 2002, 38(5):730-738.

DENG Zhihong, TANG Shiwei, ZHANG Ming, et al. Overview of Ontology [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2002, 38(5):730-738.

[12] GRUBER T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2):199-220.

[13] 任飞亮, 沈继坤, 孙宾宾, 等. 从文本中构建领域本体技术综述[J]. 计算机学报, 2019, 42(3):654-676.

REN Feiliang, SHEN Jikun, SUN Binbin, et al. A review for domain ontology construction from text [J]. Chinese Journal of Computers, 2019, 42(3):654-676.

[14] 马良荔, 孙煜飞, 柳青. 语义 Web 中的本体匹配研究 [J]. 计算机应用研究, 2017, 34(5):1287-1292.

MA Liangli, SUN Yufei, LIU Qing. Research on ontology matching on semantic Web[J]. Application Research of Computers, 2017, 34(5):1287-1292.

[15] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies, San Diego, 2016:260-270.