

文章编号:1671-251X(2018)05-0026-05

DOI:10.13272/j.issn.1671-251x.2017120026

基于形式概念分析的煤矿事故本体构建

朱佳¹, 王向前^{1,2}, 张宝隆¹, 刘敏¹

(1.安徽理工大学 经济与管理学院,安徽 淮南 232001;

2.复旦大学 管理科学与工程博士后流动站,上海 200433)

摘要:为解决煤矿事故知识管理混乱、联动不足等问题,构建了基于形式概念分析的煤矿事故本体。首先分别构建以煤矿领域的主题词表和文本集为数据源的形式背景,并通过并叠置运算得到异构资源的形式背景,然后利用概念格构造工具将形式背景转换为概念格,最后从概念格中抽取概念、属性和实例组织成煤矿事故本体。利用煤矿事故本体对煤矿事故知识进行组织和形式化表示,可实现煤矿事故知识的共享和重用,为煤矿安全预警提供帮助。

关键词:煤矿安全预警;煤矿事故;煤矿事故知识;本体;形式概念分析

中图分类号:TD67 文献标志码:A

网络出版地址:<http://kns.cnki.net/kcms/detail/32.1627.tp.20180418.1643.004.html>

Construction of coal mine accident ontology based on formal concept analysis

ZHU Jia¹, WANG Xiangqian^{1,2}, ZHANG Baolong¹, LIU Min¹

(1.School of Economics and Management, Anhui University of Science and Technology,

Huainan 232001, China; 2.Post-doctoral Station of Management Science and

Engineering, Fudan University, Shanghai 200433, China)

Abstract:In order to solve problems of management confusion and linkage lack of coal mine accident knowledge, a coal mine accident ontology based on formal concept analysis was constructed. Firstly, formal contexts are separately constructed based on thesaurus and text set of coal mine, and formal context of heterogeneous resources is obtained through apposition-overlap operation. Then the formal context is converted to concept lattice by use of concept lattice construction tool. Finally, concept, attribute and instance are extracted from the concept lattice to form coal mine accident ontology. The coal mine accident ontology can organize and formally express coal mine accident knowledge, and realize sharing and reuse of coal mine accident knowledge, so as to provide help for coal mine safety warning.

Key words:coal mine safety warning; coal mine accident; coal mine accident knowledge; ontology; formal concept analysis

0 引言

中国大多数煤矿企业已经对瓦斯、水灾、火灾等煤矿事故进行自动化监控,有效保障了煤矿安全生

产。但煤矿企业没有统一的标准来规范监控系统中的煤矿事故知识,导致监控系统得到的海量数据共享不足、利用率低^[1],煤矿事故知识管理混乱、联动不足。本体是共享概念模型明确的形式化规范说

收稿日期:2017-12-11;修回日期:2018-04-15;责任编辑:盛男。

基金项目:国家自然科学基金资助项目(51474007)。

作者简介:朱佳(1994—),女,安徽六安人,硕士研究生,主要研究方向为领域本体构建、数据挖掘,E-mail:191746203@qq.com。

引用格式:朱佳,王向前,张宝隆,等.基于形式概念分析的煤矿事故本体构建[J].工矿自动化,2018,44(5):26-30.

ZHU Jia, WANG Xiangqian, ZHANG Baolong, et al. Construction of coal mine accident ontology based on formal concept analysis [J]. Industry and Mine Automation, 2018, 44(5): 26-30.

明^[2],可对知识进行规范化、系统化的组织和表示,实现知识的共享和重用。孟现飞^[3]利用本体构建了煤矿事故预警知识库,可提高安全预警的准确率;马莉^[4]构建了煤矿事故应急预案本体,可辅助应急救援人员快速、准确地做出应急决策;支向阳等^[5]构建了瓦斯灾害预警本体,可利用本体的语义推理实现瓦斯预警,提高了瓦斯预警的准确率;药慧婷等^[6]通过研究掘进工作面的生产方式,构建了掘进工作本体,可为智能决策系统提供帮助。但上述本体大多是利用人工构建,工作量大且易受开发者主观影响。鉴此,本文采用形式概念分析(Formal Concept Analysis, FCA)方法构建煤矿事故本体,可半自动挖掘出隐含的概念及概念之间隐含的层次关系,大大降低工作量和开发者的主观影响,为煤矿安全生产信息化发展提供支持和帮助。

1 FCA

FCA 由 Wille^[7]在 1982 年提出,其作用是运用数学模式进行概念的挖掘、排序和表示。在 FCA 中,所有属于某个概念的对象集合叫做概念的外延,所有属于这个概念的对象 of 公共属性集合叫做概念的内涵。根据内涵和外延的确定,可利用二元关系表达出所有的概念及概念间关系,并从中提取出概念层次关系——概念格^[8]。关于形式背景的主要定义如下:

定义 1 形式背景 K 是由对象集合 G 、属性集合 M 及对象集合 G 与属性集合 M 之间的二元关系 I 组成的, $K=(G,M,I)$ 。 $(g,m) \in I$ 或 gIm 表示对象 g 具有属性 m 。

定义 2 假设集合 X 是对象集合 G 的一个子集,则 X 中所有对象公共属性的集合为

$$f(X) = \{m \in M \mid \forall g \in X, gIm\}$$

相应地,设集合 Y 是属性集合 M 的一个子集,则具有 Y 中所有属性的对象的集合为

$$g(Y) = \{g \in G \mid \forall m \in Y, gIm\}$$

定义 3 若 $(X_1,Y_1),(X_2,Y_2)$ 是任意形式背景的 2 个概念,且 $X_1 \subseteq X_2$,则可认为 (X_1,Y_1) 是 (X_2,Y_2) 的子概念,并记为 $(X_1,Y_1) \leq (X_2,Y_2)$,关系 “ $\leq$ ”表示概念的序层次,简称为序。形式背景中所有有序层次组成的概念集合称为概念格。

2 基于 FCA 的本体构建流程

本文提取煤矿领域两大主要数据资源——主题词表和文本集中的基本知识构建本体,相比于利用

单一数据源构建本体,这样可获取更多的概念和关系。由于主题词表和文本集存在根本上的差异,在煤矿事故本体构建中需要融合这 2 种资源。基于 FCA 的煤矿事故本体构建流程:

(1) 搜集煤矿领域经典且完备的主题词表,参照煤矿主题词表中独特的属(S)、参(C)、分(F)、代(D)、族(Z)等,从中全面地提取有关煤矿事故的概念与概念间关系等,然后对其进行形式化处理,生成基于主题词表的形式背景。

(2) 选取一些关于煤矿领域的文本集,然后对文本集中的文献进行自然语言处理,再根据主题词表中的词性结构,从中提取可作为对象和属性的基本概念,最后形式化构建基于文本集的形式背景。

(3) 将构建的 2 种煤矿事故知识的形式背景进行并叠置处理,再运用概念格工具构造基于异构资源的概念格。

(4) 将基于异构资源的概念格与本体结构进行映射,生成基于异构资源的煤矿事故本体^[9-12]。

3 煤矿事故本体构建

3.1 异构资源的形式背景处理

本文选择《煤炭科技文献检索词典》和《中国分类主题词表》作为基础数据来源,选取的部分煤矿事故主题词表见表 1。

表 1 部分煤矿事故主题词表
Table 1 Part thesaurus of coal mine accident

参照项	煤矿事故[0701]
D	矿井事故
	矿山事故
F	爆破事故
	机电事故
	冒顶事故
	煤尘爆炸事故
	煤岩突出事故
	瓦斯爆炸事故
	煤与瓦斯突出事故

从主题词表中选取有关煤矿事故的核心概念及主题词表特有的词间关系进行形式化处理,详细操作为以具体煤矿事故为形式对象,以该煤矿事故所属类别为形式属性,即可得到基于主题词表的形式背景,见表 2。

根据《煤炭科技文献检索词典》中的相关概念,选取一些有关煤矿事故的文献,从文献中抽取摘要

表 2 基于主题词表的形式背景

Table 2 Formal context based on thesaurus

对象	属性		
	矿山水灾	矿山火灾	煤矿事故
瓦斯爆炸事故			✓
煤尘爆炸事故			✓
透水	✓		
突水	✓		
内因火灾		✓	

和题目形成文本集。由于纯文本属于非结构化数

据,在进行形式背景提取时,首先需要对文本集数据进行自然语言处理。本文使用汉语词法分析系统 ICTCLAS 对煤矿事故文本集进行切词处理,得到一些带有词性的数据集,经过切词处理后的数据集还需要过滤停用词,如“的”、“了”等^[10]。然后,根据汉语的语法特性从数据集中提取核心名词、动词和宾语等。最后,将提取的核心名词作为形式对象,将所有语句中与该核心名词对应的宾语作为形容该对象的形式属性,这样就可得到基于文本集的形式背景,见表 3。

表 3 基于文本集的形式背景

Table 3 Formal context based on text set

对象	属性					
	煤矿事故	破坏生产环境	产生有害气体	突发性强	火源明显	事故多发
瓦斯爆炸事故	✓	✓	✓	✓	✓	✓
煤尘爆炸事故	✓	✓	✓	✓	✓	
岩尘危害	✓					
煤与瓦斯突出事故	✓	✓	✓	✓		✓
透水	✓	✓				
突水	✓	✓		✓		
内因火灾	✓	✓	✓			

煤矿事故复杂多样,涉及内容较多,因此需要集合多种数据源的信息。煤矿事故主题词表与文本集在结构与属性等方面区别极大,利用这 2 种不同结构的数据资源构建的本体各有优缺点:主题词表容易确定概念之间结构关系,但主题词表中语义与属性都不够丰富;文本集中具有大量的概念和概念关系,但不容易提取。若单独以其中一种结构的数据资源为基础去构建本体,得到的本体都是不完善的,因此,需要在煤矿事故本体构建过程中融合异构资源^[13]。利用并叠置运算融合异构资源的主要步骤:

(1) 依照《煤炭科技文献检索词典》和《中国分类主题词表》,得到基于主题词表的形式背景 $K_1 = (G_1, M_1, I_1)$,其中 G_1, M_1, I_1 分别为基于主题词表的对象集合、属性集合及 G_1 与 M_1 之间的二元关系。

(2) 对文本集中的内容进行自然语言处理,针对核心词汇提取对象和属性,得到基于文本集的形式背景 $K_2 = (G_2, M_2, I_2)$,其中 G_2, M_2, I_2 分别为基于文本集的对象集合、属性集合及 G_2 与 M_2 之间的二元关系。

(3) 因为 $K_1 = (G_1, M_1, I_1)$ 和 $K_2 = (G_2, M_2,$

$I_2)$ 满足 $G_1 \subseteq G, G_2 \subseteq G, M_1 \subseteq M, M_2 \subseteq M$,即 K_1 和 K_2 属于同域形式背景,可对其进行并叠置运算,获得合并后的形式背景 $(G_1 \cup G_2, M_1 \cup M_2, I_1 \cup I_2)$ 。

利用并叠置运算获得异构资源的形式背景,见表 4。

3.2 概念格构造

形式背景过多会使概念格的结构过于复杂,可对其进行适当约简,从而实现有效信息的提取和完整概念格的简单构建。根据文献[14]的方法进行约简,利用概念格构造工具 Concept Explorer 1.3,输入约简后的形式背景,即可得到基于形式背景的概念格。Hasse 图是一种概念格描述工具,可实现概念及其层次关系的可视化。根据表 4 得到基于 Hasse 图的概念格,如图 1 所示,其中每个节点代表一个概念,且每个概念节点的属性为与其连接的所有上级概念节点属性的集合,每个概念节点的对象为与其连接的所有下级概念节点对象的集合。利用形式背景构建概念格的过程就是将各种概念聚类的过程。

3.3 本体生成

先对概念格进行处理:命名顶端概念节点、标注

表 4 并叠置后的形式背景
Table 4 Formal context after apposition-overlap operation

对象	属性							
	矿山水灾	矿山火灾	煤矿事故	破坏生产环境	产生有害气体	突发性强	火源明显	事故多发
瓦斯爆炸事故			✓	✓	✓	✓	✓	✓
煤尘爆炸事故			✓	✓	✓	✓	✓	
岩尘危害			✓					
煤与瓦斯突出事故			✓	✓	✓	✓		✓
透水	✓		✓	✓				
突水	✓		✓	✓		✓		
内因火灾		✓	✓	✓	✓			

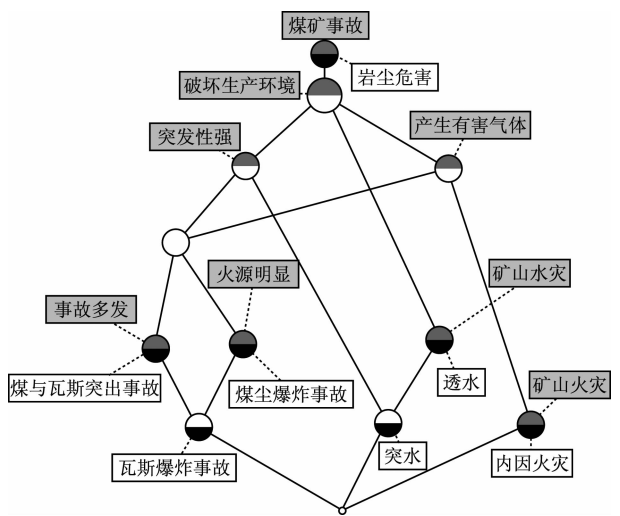


图 1 基于 Hasse 图的概念格

Fig. 1 Concept lattice based on Hasse diagram

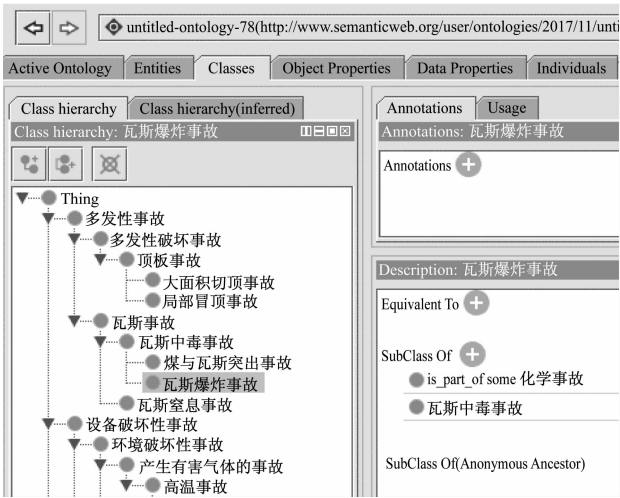


图 2 煤矿事故本体

Fig. 2 Coal mine accident ontology

中间概念节点、删除底端概念节点。再对处理过的概念格进行转换:将概念节点关系转换为概念与概念间的关系;概念节点的属性转换成本体的数据属性;对象转换成本体概念的实例。

利用概念格构建本体的具体步骤:首先,将概念格中顶端概念节点抽取出来作为煤矿事故本体的一级概念;然后点击顶端概念节点,从概念格中找到与其连接的所有概念节点,根据这些概念节点的位置及与顶端节点的连线情况,挖掘出概念节点与顶端节点之间的结构关系,再根据对应的结构关系将这些概念节点作为顶端节点的子类或相关类,这样从上往下依次挖掘完最后一个概念节点后即可得到所有概念的基本结构关系^[8];最后利用本体编辑工具 Protégé 4.3 及 OWL(Web Ontology Language,网络本体语言)对由 FCA 得到的概念、属性和实例等进行构建^[15],得到煤矿事故本体,如图 2 所示。

4 结语

首先,从主题词表和文本集中抽取了煤矿事故的一些基本对象和属性,通过并叠置运算构成异构资源的形式背景;然后,利用概念格构造工具 Concept Explorer 将形式背景构建为概念格;最后,利用本体编辑工具 Protégé 对概念、属性和实例进行形式化的表达,得到基于 FCA 的煤矿事故本体。该本体对煤矿事故知识进行了规范化、系统化的组织和表示,方便煤矿安全管理人员集成、共享和重用煤矿事故知识,且可为煤矿安全预警提供帮助。然而,在煤矿事故本体构建过程中仍存在一些不足:由于收集到的煤矿事故概念的数量相对较少,隐藏的结构关系挖掘得不够全面,使得构建的本体具有一定的局限性,影响本体的实用性;FCA 虽然可降低本体构建的工作量,但这种半自动化本体构建方法仍比较费时费力。因此,搜集更多的文本资料、挖掘

更多的有效信息,以完善本体的概念、属性和实例等,以及本体的自动化构建将是下一步研究的重点。

参考文献(References):

- [1] 郭华.煤矿瓦斯监控系统的本体模型研究[D].太原:太原科技大学,2014.
- [2] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering: principles and methods[J]. Data and Knowledge Engineering, 1998, 25(1/2): 161-197.
- [3] 孟现飞.基于本体的煤矿事故预警知识库模型及其应用[D].徐州:中国矿业大学,2014.
- [4] 马莉.本体的煤矿数字化应急预案系统研究[J].西安科技大学学报,2014,34(2):216-223.
- MA Li. Digitalized system of coalmine emergency plan based on ontology[J]. Journal of Xi'an University of Science and Technology, 2014, 34(2): 216-223.
- [5] 支向阳,吕岳东,陈立潮,等.基于本体的瓦斯灾害预警模型研究[J].太原理工大学学报,2010,41(4): 388-391.
- ZHI Xiangyang, LYU Yuedong, CHEN Lichao, et al. Gas safety warning model based on ontology[J]. Journal of Taiyuan University of Technology, 2010, 41(4): 388-391.
- [6] 药慧婷,陈立潮,潘理虎.掘进工作面本体模型研究[J].工矿自动化,2015,41(9):93-96.
- YAO Huiting, CHEN Lichao, PAN Lihu. Research of ontology model of heading face[J]. Industry and Mine Automation, 2015, 41(9): 93-96.
- [7] GANTER B, WILLE R. Formal concept analysis: mathematical foundations [M]. Berlin: Springer-Verlag, 1999.
- [8] 刘萍,胡月红.基于FCA和关联规则的情报学本体构建[J].现代图书情报技术,2012(2):34-40.
- LIU Ping, HU Yuehong. Development of domain ontology in information science based on FCA and association rules[J]. New Technology of Library and Information Service, 2012(2): 34-40.
- [9] 张云中.基于形式概念分析的领域本体构建方法研究[D].长春:吉林大学,2009.
- [10] 滕广青,毕强.基于概念格的异构资源领域本体构建研究[J].现代图书情报技术,2011(5):7-12.
- TENG Guangqing, BI Qiang. A study on domain ontology construction from heterogeneous resources based on concept lattice [J]. New Technology of Library and Information Service, 2011(5): 7-12.
- [11] BENDAOU R, NAPOLI A, TOUSSAINT Y. Formal concept analysis: a unified framework for building and refining ontologies[M]. Berlin: Springer-Verlag, 2008.
- [12] 孙利.基于主题词表和FCA的海事本体构建研究[D].大连:大连海事大学,2010.
- [13] 韩道军,甘甜,叶曼曼,等.基于形式概念分析的本体构建方法研究[J].计算机工程,2016,42(2): 300-306.
- HAN Daojun, GAN Tian, YE Manman, et al. Research of ontology construction method based on formal concept analysis[J]. Computer Engineering, 2016, 42(2): 300-306.
- [14] 杨丽,徐扬.基于形式背景的概念格约简及其修复[J].计算机工程,2008,34(9):22-24.
- YANG Li, XU Yang. Concept lattice reduction and reparation based on formal context [J]. Computer Engineering, 2008, 34(9): 22-24.
- [15] 邱璇,李端明,张智慧.基于FCA和异构资源融合的本体构建研究[J].图书情报工作,2015,59(2): 112-117.
- QIU Xuan, LI Duanming, ZHANG Zhihui. Research on ontology construction based on FCA and heterogeneous resources integration[J]. Library and Information Service, 2015, 59(2): 112-117.