

基于形式概念的汽轮机相似故障案例检索

方名菊¹, 章筠², 徐正国¹

(1. 浙江大学 工业控制技术研究所, 工业控制技术国家重点实验室, 浙江 杭州 310027;

2. 上海电气集团股份有限公司 中央研究院, 上海 200070)

摘要: 作为故障诊断的一种方法, 案例推理从以往的历史案例中检索相似案例作为新事件的参考解。在整个案例推理过程中, 案例检索是最为关键的一步。案例检索的内容是以一定形式表达的案例, 案例检索的方式是计算新事件与历史案例间的相似度。判断新事件与历史案例的相似度主要考虑两个问题, 一是属性间的相似度, 二是属性的权重。形式概念作为知识的一种表示方式, 不仅能表达案例与属性之间关系, 还能反映属性的权重。本文在基于形式概念的知识表达方式上, 提出了一种改进的相似度计算方法。在改进算法中, 引入了参数。在具体应用中, 通过选取合适的值, 能为每个属性合理地分配相应的权重, 使得检索结果更为精确。其效果在汽轮机故障诊断的实际应用中得到了检验。

关键词: 案例检索; 历史案例; 属性; 形式概念; 相似度计算

中图分类号: TP277

文献标志码: A

文章编号: 1672-3767(2017)04-0024-07

DOI: 10.16452/j.cnki.sdkjzk.2017.04.004

Case Retrieval for Faults of Steam Turbines Based on Formal Concept

FANG Mingju¹, ZHANG Yun², XU Zhengguo¹

(1. National Laboratory of Industrial Control Technology, Institute of Industrial Process Control,

Zhejiang University, Hangzhou, Zhejiang 310027, China;

2. Central Academe, Shanghai Electric Group Co. Ltd, Shanghai 200070, China)

Abstract: As a fault diagnosis method, case-based reasoning retrieves similar cases from history case base as reference solutions to new events. Case retrieval, the most critical step in the case-based reasoning process, is based on a certain expression of the case and is done by calculating the similarities between the new event and history cases. Two main factors should be taken into consideration when designing a similarity measure: similarity between attributes and weight assigned to each attribute. As a representation of knowledge, formal concept not only represents the relationship between cases and attributes, but it also reflects the weight of each attribute. In this paper, an improved similarity measure was proposed based on the knowledge representation of formal concept. Parameter K was introduced into this new measure. By choosing appropriate value for specific applications, the measure could assign corresponding weight to each attribute, thus obtaining more accurate case retrieval results. The proposed similarity measure was tested in the practical fault diagnosis of a steam turbine, which verified its effectiveness.

Key words: case retrieval; history cases; attributes; formal concept; similarity measure

收稿日期: 2017-03-11

基金项目: 国家自然科学基金项目(61473254); 工业控制技术国家重点实验室项目(ICT1703)

作者简介: 方名菊(1993—), 女, 四川荣县人, 硕士研究生, 主要从事故障诊断方面的研究。

徐正国(1979—), 男, 江苏无锡人, 副教授, 博士生导师, 主要从事控制系统可靠性分析、复杂系统性能退化分析故障诊断方面的研究, 本文通信作者。E-mail: xzgj@zju.edu.cn

1 引言

1.1 汽轮机故障诊断方法研究

汽轮机作为发电设备的一部分,其结构复杂,容易发生故障,一旦发生故障应及时诊断并维修,否则将对电力系统造成极大的压力,甚至可能导致大规模停电,对生产生活造成极大的不便。目前关于汽轮机故障诊断的方法有很多,如基于模型的故障诊断方法^[1]、基于故障树的故障诊断方法^[2]、基于神经网络的故障诊断^[3]、基于规则的故障诊断^[4]、基于案例的故障诊断^[5]等等。

基于模型的故障诊断^[1]直接利用对象原理构建诊断模型,可以避免专家知识获取困难的问题,但模型的建立是在诸多假设和简化条件下得到,在实际应用中结果并不一定准确。故障树^[2]将故障原因和结果通过多层次树节点连接形成。故障树的建立需要对系统的机理有深入的了解,并且该方法的灵活性和自适应能力差。人工神经网络^[3]将提取的参数特征向量作为网络输入,故障类型作为网络的输出,达到故障诊断的目的。该诊断方法具有并行性、容错性以及自学习能力等优点,但是学习样本获取困难,神经网络训练和结构的优化问题一直难以解决。规则推理^[4]根据专家的以往实践经验、专家分析问题和解决问题的思路,提取知识中存在的规则,进行故障诊断。规则推理存在知识获取瓶颈,并且其灵活性、自适应能力和学习能力差。基于案例的故障诊断方法^[5]将以往故障发生时的现象、原因以及解决方案以案例的形式存储到案例库中,以案例属性为依据,通过算法从案例库中检索相似案例,获取相应的问题解决方案。案例推理不需要了解系统机理,也不需人从实例中提取规则,降低了知识获取的负担,并且案例推理具有很强的扩展性、学习性和自适应能力。

由于工业现场的复杂性,设备运行机理比较复杂,无法利用机理进行建模或者构造故障树。同时,在样本不充足的情况下,构建的神经网络会不准确,而案例推理能根据自身的学习能力和自适应能力,弥补该缺点。因此,在机理不明确和样本不充足的情况下,本文选择案例推理的方法进行故障诊断。

1.2 案例知识的表达形式

案例推理是基于某种表达形式的案例知识。目前案例知识存在多种表达形式,比如框架表示法^[6]、语义表示法^[7]、面对对象的表示方法^[8]、形式概念表示法^[9]等等。

框架^[6]是描述对象属性的一种数据结构,将一个对象所包含的所有属性限定在一个框架中,并通过属性间的关系进行组织,框架之间通过属性间的相互关系建立框架网络。框架表示法无法表示过程性知识。语义网络^[7]是一种用实体及其语义关系来表达知识的有向图,结点代表实体,弧代表语义关系,存在关系的两个实体之间通过弧连接。语义网络主要描述了实体之间的语义关系,并且只能表达确定性知识。面向对象表示法^[8]将对象所包含的属性以及相应的方法封装起来,对象之间以某种方式进行联系。面向对象表示法知识创建过程工作量大、性能低。形式概念的表示方法^[9]从对象集中提取相同属性,根据对象之间的关联性建立层次结构。形式概念既表达了数据间的层次结构,也表达了对象与属性间、对象与对象间、属性与属性间的关联性。

在汽轮机案例中,无论是对象间或者是属性间,虽然具有关联性,但并没有明确的语义关系。因此,语义表示法并不能用于汽轮机的案例表示。而相对于形式概念而言,面向对象表示法工作量大、性能低。因此,针对具有关联性的过程性知识,我们选择形式概念进行汽轮机案例知识表示。

2 基于形式概念的相似度算法比较

2.1 形式概念

形式概念^[9]是一种数据分析和知识表示的方法,主要与三个概念相关:形式背景、形式概念和概念格。

形式背景为 (G, M, I) 的组合, G 表示对象集, M 表示属性集, I 表示对象集与属性集的关系,即 $(g, m) \in I$ 表示对象 g 含有属性 m 。

在形式背景的基础上创建形式概念,形式概念为 (A, B) 的一个组合,其中 A 表示对象集, B 表示属性集,并且满足 $h(A) = B, r(B) = A$,其中 $h(A) := \{m \in M | gIm \text{ for all } g \in A\}, r(B) := \{g \in G | gIm \text{ for all } m$

$\in B$)。 $h(A)=B$ 表示对象集 A 所包含的全部相同属性集合等于 B 集合, $r(B)=A$ 表示包含属性集 B 的全部对象集合等于 A 集合。

概念格为形式概念的层次数据结构, 将形式概念按一定顺序连接起来就构成了概念格。概念格中除包含形式概念这个元素外, 还包含偏序关系。偏序关系表示方法为: $(A_1, B_1) \leq (A_2, B_2)$, 即 $A_1 \subseteq A_2 (B_2 \subseteq B_1)$, 其中 (A_2, B_2) 处于上层结构, (A_1, B_1) 处于下层结构。

例如现有形式背景如表 1 所示, 在形式背景的基础上, 利用概念格创建算法^[10] 创建相应的概念格, 如图 1 所示。

在图 1 中, 圆圈内的 C 代表一个形式概念, 上方的 Obj 和下方的 Attr 分别表示该形式概念所包含的对象和属性, 并且下层概念包含上层概念的属性, 上层概念包含下层概念的对象, 括号中的数字表示属性或对象在概念格中出现的频次。

2.2 基于形式概念的相似度算法研究

本节先给出 Tadrat^[11] 和 Shi^[12] 的相似度计算方法, 并在此基础上, 提出改进算法。

2.2.1 Tadrat 提出的算法

Tadrat^[11] 基于形式概念的相似度计算公式为:

$$\begin{aligned} \text{sim}(C_P, C_N) = & \frac{1}{2} \left(\frac{\sum_{u \in I_N \cap I_P} (\log \frac{N}{F_{a_u}})^2}{\sqrt{\sum_{i \in I_N} (\log \frac{N}{F_{a_i}})^2} \sqrt{\sum_{j \in I_P} (\log \frac{N}{F_{a_j}})^2}} \right) \\ & + \frac{1}{2} \left(\frac{\sum_{v \in E_N \cap E_P} (\log \frac{N}{F_{c_v}})^2}{\sqrt{\sum_{h \in E_N} (\log \frac{N}{F_{c_h}})^2} \sqrt{\sum_{l \in E_P} (\log \frac{N}{F_{c_l}})^2}} \right) \end{aligned} \quad (1)$$

其中: N 代表概念格中形式概念的总数目; I_N, I_P 分别表示新事件和形式概念所包含的属性集; E_N, E_P 分别表示新事件和形式概念所包含的对象集; $F_{a_u}, F_{a_i}, F_{a_j}$ 表示属性 u, i, j 在概念格中出现的频次, u 表示新事件属性集与形式概念属性集所包含的相同属性, i 表示新事件所包含的属性, j 表示形式概念所包含的属性。 $F_{c_v}, F_{c_h}, F_{c_l}$ 表示对象 v, h, l 在概念格中出现的频次, k 表示新事件的初始对象集与形式概念对象集所包含的相同对象, h 表示新事件的初始对象集, l 表示形式概念的对象集。

新事件所包含的初始对象集通过计算新事件所包含的属性集与概念格中形式概念所包含属性集的相同属性个数, 直接将属性相同个数最多的形式概念的对象集赋给新事件, 作为新事件的对象集。

该算法的不足在于: 在相似度计算中引入了案例信息, 转移了属性权重比例。并且案例信息为不精确信息, 会降低计算结果的准确性。

2.2.2 Shi 提出的算法

Shi^[12] 的基于形式概念的相似度计算方法为:

$$\begin{aligned} & \text{sim}(C_P, C_N) \\ = & \frac{\sum_{o \in m \cap n} \log \frac{N}{F_o}}{\sum_{o \in m \cap n} \log \frac{N}{F_o} + \frac{|m-n|}{|m \cup n| - |m \cap n|} \sum_{r \in m-n} \log \frac{N}{F_r} + \frac{|n-m|}{|m \cup n| - |m \cap n|} \sum_{s \in n-m} \log \frac{N}{F_s}} \end{aligned} \quad (2)$$

其中, m, n 分别表示形式概念和新事件的属性集, F_o, F_r, F_s 表示属性 o, r, s 在概念格中出现的频次, o 表示新事件属性集与形式概念属性集所包含的相同属性, r 表示形式概念包含而新事件不包含的属性, s 表示新

表 1 形式背景
Tab. 1 Formal context

	Attr1	Attr2	Attr3
Obj1	1	1	
Obj2	1		1
Obj3	1	1	1

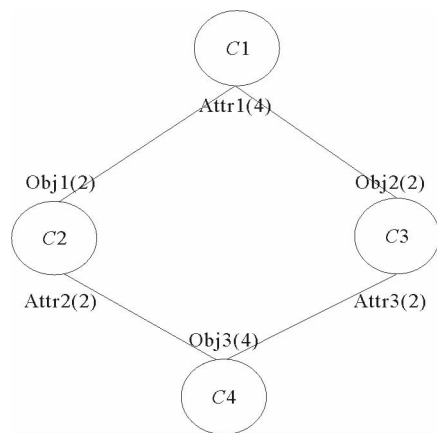


图 1 概念格
Fig. 1 Formal lattice

事件包含而形式概念不包含的属性。

2.2.3 改进算法

上述两种相似度计算方法都存在一个共同点:属性在形式概念中出现的频次越大,其权重越小。改进算法在保持上述特性的前提下,更改了权重函数,并引入了可调系数 k , k 可根据不同的应用进行调节。

$$\text{sim}(C_P, C_N) = \frac{\sum_{x \in m \cap n} e^{-\frac{F_x}{N}}}{\sqrt{\sum_{y \in m} e^{-\frac{F_y}{N}}} \sqrt{\sum_{z \in n} e^{-\frac{F_z}{N}}} \quad (3)$$

其中, x 表示新事件属性集与形式概念属性集所包含的相同属性, y 表示形式概念所包含的属性, z 表示新事件所包含的属性, F_x, F_y, F_z 表示属性 x, y, z 在概念格中出现的频次, k 为可变正数。

算法中包含了可调节参数 k , 通过调节 k 值可以改变不同频次参数间的相对权重, 参见图 2。这样针对不同的应用, 可以选取合适的值获得更精确的结果。而 Tadrat 和 Shi 的算法中, 相对权重则是固定的, 无法调节。图 2 假定形式概念总数为 50, 分别取 k 等于 1, 2, 3, 4, 5 求取属性的权重。

3 实例

本文研究的内容为汽轮机故障, 以其中的“#1 高加正常疏水调节阀阀位偏大”故障模式为具体仿真对象, 进行算法验证。文中的数据来自于某电厂, 但由于历史案例信息表达的不规范, 历史案例的故障现象描述中省略了一些进行故障原因判别的必要属性, 导致无法由它所描述的现象推出它所给定的故障原因。因此, 结合专业知识对历史案例进行了整理, 提取出该故障模式下的特征模式集。整理出的特征模式集可以被看做是部分历史案例集。

在进行相似度计算之前, 先要创建 #1 高加正常疏水调节阀阀位偏大故障特征模式集下的概念格。先将特征模式集(表 2)转化为形式背景(即对象与属性的 0~1 关系表), 如表 3 所示。然后在形式背景的基础上, 根据形式概念的定义以及相应的算法^[10] 创建形式概念和概念格, 如图 3 所示。表 4 表示每种特征模式所对应的故障原因。

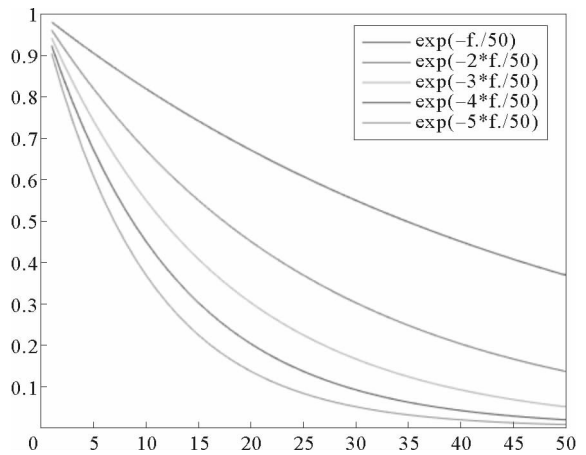


图 2 $k=1, 2, 3, 4, 5$ 的属性权重分配图

Fig. 2 The distribution of attribute weight

表 2 #1 高加正常疏水调节阀阀位偏大故障特征模式集

Tab. 2 Fault feature modes of #1 shell_side_normal_drain_valve_position larger than normal

	#1 高加正常疏水调节阀	#1 高加水位	#1 紧急事故疏水阀	疏水端差	#2 高加正常疏水调节阀
1	偏大	正常	关闭	正常	不变
2	偏大	正常	关闭	正常	增大
3	偏大	偏高	关闭	正常	不变
4	偏大	偏高	关闭	正常	增大
5	偏大		动作	正常	
6	偏大			增大	

表 3 “#1 高加正常疏水调节阀阀位偏大”形式背景

Tab. 3 Formal context of #1 shell_side_normal_drain_valve_position larger than normal

K	Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8	Attr9
1	1	1			1		1	1	
2	1	1			1		1		1
3	1		1		1		1	1	
4	1		1		1		1		1
5	1			1			1		
6	1					1			

注:Attr1: #1 高加正常疏水调节阀阀位偏大;Attr2: #1 高加水位正常;Attr3: #1 高加水位偏高;Attr4: #1 紧急事故疏水调节阀打开;Attr5: #1 紧急事故疏水调节阀阀位关闭;Attr6: 疏水端差增大;Attr7: 疏水端差正常;Attr8: #2 正常疏水调节阀阀位正常;Attr9: #2 正常疏水调节阀阀位增大。

针对汽轮机 #1 高加正常疏水调节阀阀位偏大该类故障模式,改进算法中选取等于 6 进行仿真。

新事件的表现特征为 { #1 高加正常疏水调节阀阀位偏大、#1 高加水位正常、疏水端差增大、#2 高加正常疏水调节阀阀位正常 }, 即新事件的属性集为 { Attr1, Attr2, Attr6, Attr8 }。在利用 Tadrat 相似度计算方法进行相似度计算时,求出新事件的初始对象集为 { Obj2 }。针对新事件,利用三种相似度计算方法进行仿真,仿真结果如图 4 所示,其中横坐标表示形式概念,纵坐标表示相似度。表 5 列出了三种相似度计算方法检索出的最相似案例以及相应的故障原因。

从图 4 中可以看出,Tadrat 算法和 Shi 的算法得出的结论都与 C₁₁ 相似,而本文的算法与 C₃ 最为相似。经过验证,该事件的事故原因为高加汽侧严重泄漏,与本文提出的相似度方法计算结果相同,而与 Tadrat 和 Shi 相似度方法的计算结果不同。并且结合专业知识进行分析,在相关的所有参数中,一旦疏水端差增大,其他参数对于结果的辅助分析作用弱到几乎可以忽略。

4 k 的选取

本文利用 UCI(University of California Irvine,加州大学尔湾分校)数据库中的 balance-scale、blogger 数据库对 k 的选取规则进行研究。balance-scale 数据集总共包含 625 个对象,以其中前 100 个对象为训练对象,建立形式概念,然后分别取不同数量的测试样本对不同取值的 k 所导致的计算精确度做比较,比较结果如图 5 所示,其中横坐标表示 k 的值,纵坐标表示准确率。blogger 数据集总共包含 100 个对象,以前 20 个对象为训练对象,建立形式概念。也是选取不同数量的测试样本对不同取值的 k 所导致的计算精确度做比较,比较结果如图 6 所示。结合图 5 和图 6,可以发现,对于某个数据集样本,存在一个确定的 k 使得计算结果的准确率最高,并且该值与样本数无关。在本文的实例中,

表 4 故障原因

Tab. 4 Fault reason

特征模式	故障原因
1	阀门轻微堵塞
2	高加汽侧轻微泄漏
3	阀门堵塞
4	高加汽侧泄漏
5	阀门堵塞较严重或高加汽侧泄漏较严重
6	严重泄漏

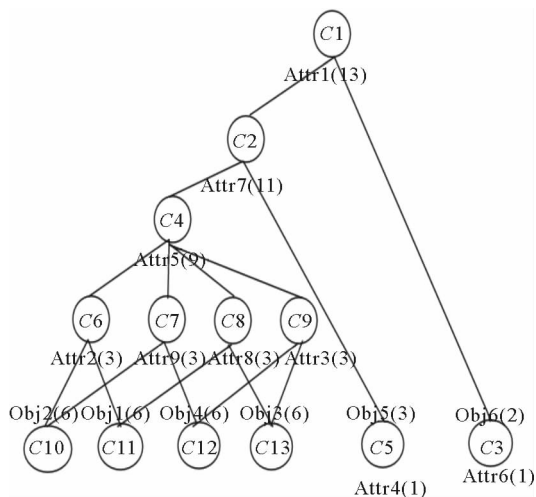


图 3 “#1 高加正常疏水调节阀阀位偏大”概念格

Fig. 3 Formal lattice of #1 shell_side_normal_drain_valve_position larger than normal

通过试验法发现 k 为 6 时, 计算出的结果更为准确。由于 k 的取值与样本数无关, 因此在本文中确定的最优的 k 值可以作为 #1 高加正常疏水调节阀阀位偏大故障模式类案例检索的 k 值。

5 结语

本文选取形式概念作为知识的表示方式, 形式概念既描述了过程性知识也表达了知识间的关联性。基于形式概念的知识表示方法, 在 *Tadrat* 和 *Shi* 的相似度计算方法的基础上提出了改进算法。相较于其他两种相似度算法, 本文提出的相似度算法更改了权重函数, 并引入了可变参数。针对不同的应用, 选取合适的系数, 可以使得权重分配更为合理, 计算结果更为准确。

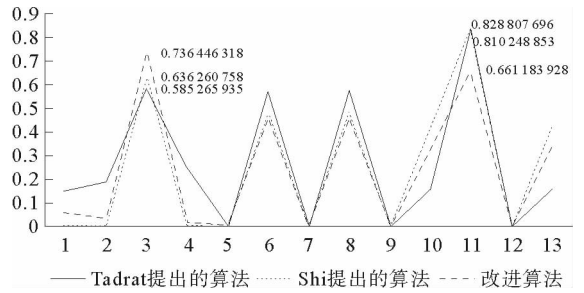


图 4 相似度方法比较

Fig. 4 Comparison of similarity measures

表 5 检索结果

Tab. 5 Retrieval results

形式概念	属性集	对象集	故障原因
C_3	#1 高加正常疏水调节阀阀位偏大、疏水端差增大	Obj6	高加汽侧严重泄漏
C_{11}	#1 高加正常疏水调节阀阀位偏大、#1 高加水位正常、#1 高加紧急事故疏水阀关闭、疏水端差正常、#2 高加正常疏水调节阀阀位正常	Obj1	阀门轻微堵塞

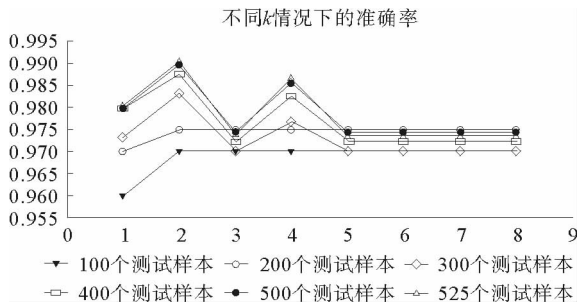


图 5 Balance-scale 数据集

Fig. 5 Balance-scale database

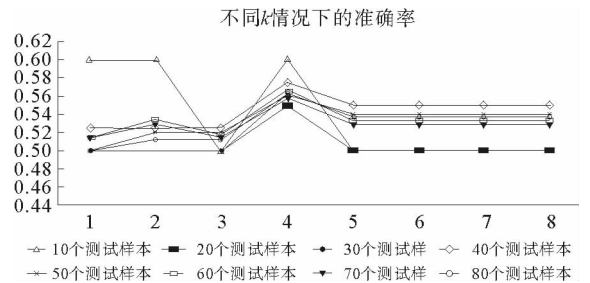


图 6 Balance-scale 数据集

Fig. 6 Balance-scale database

对汽轮机 #1 高加正常疏水调节阀阀位偏大故障模式下的历史案例进行了仿真, 分别利用三种相似度计算方法进行验证, 结果表明本文提出的方法计算结果更为准确。本文所涉及到的案例知识为明确知识, 但存在某些故障模式, 其案例内容包含模糊知识。接下来我们将针对模糊知识, 考虑如何利用形式概念进行模糊知识表达, 并且设计相应的相似度计算方法。

参考文献:

[1]高松, 刘志刚, 徐建芳, 等. 基于模型诊断和专家系统的牵引变压器故障诊断研究[J]. 铁道学报, 2013, 35(7): 42-49.

- GAO Song, LIU Zhigang, XU Jianfang, et al. Research on fault diagnosis for traction transformer on the basis of model-based diagnosis and expert system[J]. Journal of the China Railway Society, 2013, 35(7): 42-49.
- [2] 宗群, 李光宇, 郭萌. 基于故障树的电梯故障诊断专家系统设计[J]. 控制工程, 2013, 20(2): 305-308.
ZONG Qun, LI Guangyu, GUO Meng. Design of diagnostic expert system for elevator system based on FTA[J]. Control Engineering of China, 20(2): 305-308.
- [3] 郭建. 基于BP神经网络的汽轮机故障诊断系统研究[D]. 广州: 华南理工大学, 2014.
- [4] FRYE D, ZELAZO P D, PALFAI T. Theory of mind and rule-based reasoning[J]. Cognitive Development, 1995, 10(4): 483-527.
- [5] KOLODNER J L. An introduction to case-based reasoning[J]. Artificial Intelligence Review, 1992, 6(1): 3-34.
- [6] 冯明杰. 基于框架的语义题库推荐系统的研究与应用[D]. 成都: 电子科技大学, 2014.
- [7] CHANG T H, CHU Y H, LEE C C, et al. A semantic representation of energy-related information in future smart homes [J]. Energy & Buildings, 2012, 47(47): 169 - 179.
- [8] KHAN A A, CHAUDHRY I A. Object oriented case representation for CBR application in structural analysis[J]. Applied Artificial Intelligence, 2015, 29(4): 335-352.
- [9] SINGH P K, ASWANI K C, GANI A. A comprehensive survey on formal concept analysis, its research trends and applications[J]. International Journal of Applied Mathematics and Computer Science, 2016, 26(2): 495-516.
- [10] 纪彤坤. 概念格 Chein 算法的研究与改进[D]. 广州: 华南理工大学, 2012.
- [11] TADRAT J, BOONJING V, PATTARINTAKORN P. A new similarity measure in formal concept analysis for case-based reasoning[J]. Expert Systems with Applications, 2012, 39(1): 967-972.
- [12] SHI C, LAI L, FAN J, et al. Similarity model based on CBR and FCA[C]// IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. IEEE Computer Society, 2016: 597-603.

(责任编辑: 傅 游)