基于本体的语义相似在维修案例推理中应用

闻敬谦 李 青

(北京航空航天大学 机械工程及自动化学院, 北京 100191)

摘 要: 案例推理(CBR, Case-Based Reasoning)方法基于历史经验对新问题进行求解,适于飞机维修保障的知识表达和应用.针对飞机维修保障过程中涉及故障记录、诊断分析、维修处理等诸多方面和专业的特点,对相关信息和专业知识的分类方式及相互关系进行了讨论,给出维修保障领域知识和案例知识定义.通过面向对象分析方法建立了产品、故障模式、故障原因和数据字典等领域知识的本体模型,实现知识的一致性描述;重点对案例本体构建和检索进行了研究,基于领域本体结构的深度和出度计算案例特征的语义相似性,给出维修保障案例匹配方法,在此基础上开发了飞机维修保障知识管理系统并给出应用实例.

关 键 词: 案例推理; 本体; 领域本体; 语义相似; 维修保障

中图分类号: TP 391

文献标识码: A **文 章 编 号**: 1001-5965(2009)02-0223-04

Ontology-based semantic similarity computation for maintenance CBR

Wen Jingqian Li Qing

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: CBR (case-based reasoning) finds solutions for new problems based on previous experience. It is suitable for such applications as aircraft maintenance knowledge acquisition, representation and reuse. With consideration of some special characteristics of aviation maintenance, such as spot recording, reliability analysis and fault disposal, firstly the classification and the relationship of the relevant maintenance knowledge and information were investigated, and then the definitions of domain knowledge and case knowledge of maintenance were provided. Ontology models for product structure, maintenance inventory, failure mode, failure causation, and data dictionary were developed and the consistency in the knowledge description was realized. The case model architecture and retrieval technique were studied in detail, and an ontology-based semantic similarity computation method was proposed and applied to maintenance case recall and match. The method combined the depth and out-degree as two main influences for the similarity of concept model. Finally, an aircraft maintenance knowledge management system was developed. An application example shows that the method can benefit the aircraft maintenance.

Key words: case-based reasoning; ontology; domain ontology; semantic similarity; maintenance

新型飞机技术含量高,训练任务重,相关承制单位的维修、带教和加改装工作频繁,迫切需要提高服务保障综合水平,而人员是最终决定维修保障能力的重要因素.由于维修保障工作涉及面广、综合性强,相应的维修保障知识具有很强的个体性和经验性,实施维修保障知识管理成为提高服务效能的重要途径.

维修保障知识应用研究可分为两类,一类为采用模糊集、神经网络等方法对大量检测数据进行分析,获得征兆与故障间推理规则[1-2];另一类针对离散型数据,主要采用计算知识的相似度为相关人员提供参考[3-4].针对航空维修保障事务性及偶发性特点,适于采用基于案例推理的启发式方法获取规则.

收稿日期: 2008-07-20

基金项目: 北京市共建项目资助

作者简介: 闻敬谦(1978-),男,陕西西安人,博士生, wenjq@me.buaa.edu.cn.

案例匹配主要采用最近相邻算法、归纳法、层次检索和基于知识检索等方法,基于语义相似案例推理取得良好的效果. 文献[5]基于本体给出识别及量化属性间映射关系的方法,引入语义相似对最近相邻算法进行了改进;文献[6]基于最近相邻法、归纳法和知识法给出面向对象的案例索引机制;文献[7-8]给出改进基于本体概念相似度计算方法.

1 航空维修保障相关知识

飞机维修保障需通过对多方面信息的综合分析,以分析故障关联并合理制定维修方案. 如根据天气、飞行科目、维护时机等信息模拟系统状态进行故障隔离;参考故障类型、组织特点、人员资质等因素进行维修保障计划. 结合故障树分析(FTA, Fault Tree Analysis)、故障模式和影响分析(FMEA, Failure Mode and Effect Analysis)中对故障原因、影响范围、改进措施等内容的分析要求,按照分析目标和方法建立维修保障领域知识结构,并以案例形式实现知识组织和应用. 给出领域知识和案例知识定义.

定义 1 领域知识是指不同专业或应用中涉及内容的知识抽象,知识具有相同或相似的结构,适于某一具体领域的分析.

定义 2 案例知识是对维修保障事例的规范化描述,是故障诊断、分析及处理等内容的综合.

美国航空航天局(NASA, National Aeronautics and Space Administration)通过对大量航空事故案例的分析,得出影响航空安全的十多种主要因素,分为训练与教学、天气、维修与维修人员、检查单与符合性等方面. 以NASA故障案例为参照,根据某航空制造企业在飞机使用过程中汇总的故障经验内容,维修保障知识可分为案例情景、诊断分析知识和故障处理知识等,关键要素及说明见表1.

表1 维修保障知识

应用	相关领域	要素
案例情景	产品结构	材料、工艺、加工路线、装配关系等
	故障环境	天气、维护人员、地理位置等
	发生时机	任务类型、故障时间等
诊断分析	故障模式	故障现象、检测值、影响程度等
	分析记录	试验及检验报告、会议记录等
	故障原因	设计缺陷、制造缺陷、维护问题等
维修处理	处理结论	产品改进、人员培训、环境设备改进
	实施方案	备件清单、维修组织、维修工艺等
	归零总结	安全性、任务性、经济性及满意度影响

其中产品结构、故障模式、故障原因等领域知识的数据量大、层次关系复杂,具有很强的逻辑规则;而故障环境、发生时机等领域知识,由于数据量有限而多采用枚举方式;分析记录、实施方案及归零总结等领域知识以文字描述为主. 维修保障案例由各领域知识组成,主要以案例情景和诊断分析知识为检索条件,以获得相似故障的历史案例,为维修保障人员提供故障处理参考. 因此,建立对各类知识的统一表述和组织不仅能够为信息收集及共享提供参考规范,也是实施知识应用的关键基础.

2 基于本体的知识建模

由W3C组织倡导的本体概念是一种明确的、形式化的规范描述,为共享概念模型提供了良好的知识表述支持及强大的语义方法. 采用本体对维修保障各类知识进行建模,能够为不同领域知识及规则提供描述框架及规范,构建易于扩展的术语词典,实现知识的统一描述和组织.

参照面向对象建模方法,本体模型定义可由概念 T、概念属性集 A_c、关系 R_t、关系属性集 A_r、概念层次 H 和约束断言 X 的六元组构成,即本体可表示为 $\{T, A_c, R_t, A_r, H, X\}$. 对维修保障知识管理,知识本体化的关键在于明确概念的定义、属性及层次关系,以下给出维修保障知识相关本体定义.

定义 3 产品本体 O_P 可表示为六元组:$O_P = \{I_P, D_P, F, R_P, P_C, K_V\}$,其中 I_P 为产品唯一标识,如型图号或件号;D_P 为产品名称等描述信息;F 为故障模式集;R_P 为可靠性信息集,如平均工作时间等;P_C 为工艺路线,用于批性问题分析;K_V 为关键技术指标集,如导管的长度、压力范围等.

定义 4 故障原因本体 O_R 可表示为五元组:$O_R = \{I_R, D_R, T_R, A_R, S_R\}$,其中 I_R 为故障原因唯一标识;D_R 为原因描述,包括摘要和详细信息;T_R 为原因类型,分为设计、制造、维护等;A_R 为分析记录集;S_R 为解决措施记录.

定义 5 维修保障案例本体 O_C 可表示为八元组:$O_C = \{I_C, E, T_C, O_P, O_F, O_R, D_C, R_C\}$,其中 I_C 为案例标识;E 为故障情景;T_C 为故障时机;O_P 为相关产品本体;O_F 为相关故障模式本体;O_R 为故障原因本体;D_C 为案例处理;R_C 为参考案例集.

定义 6 数据字典本体 O_D 可表示为四元组:$O_D = \{I_D, T_D, K_D, D_D\}$,用于枚举类信息的知识化,其中 I_D 为唯一标识;T_D 为分类;K_D 为关键词;D_D 为描述信息.

定义 7　故障模式本体 O_F 可表示为八元组: $O_F = \{ I_F, T_F, D_F, A_F, P_F \}$,其中 I_F 为故障模式唯一标识; T_F 为用于检索的分类信息; D_F 为模式描述,包括摘要和详细信息; A_F 为影响度集,分为可靠性、任务性、经济性等; P_F 为父级故障模式.

对产品、故障模式、故障原因的本体定义和扩展可建立起层次化结构的领域知识体系,形成知识积累的基础结构;对于环境、维护时机等则采用数据字典实现规范化;而对于制造工艺、组织机构等企业现有格式化信息则可直接引入. 通过案例本体建立起与各领域知识、数据字典及基础数据的相互关联,采用语义规则反映实际应用中逻辑关系,构成便于知识共享和应用的体系结构. 相应的知识层次及相互关系如图 1所示.

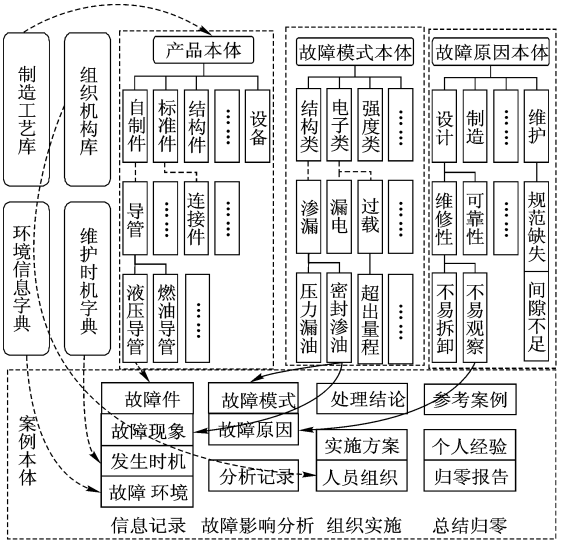

图 1　维修保障知识体系结构示意图

随着知识主体认知能力的不断提高,通过子类扩展、属性扩展等方法可使领域知识和案例知识的本体模型得以细化,逐步形成完整的知识框架,并可通过属性约束、基数约束、同义关系等丰富的语义规则使知识模型更符合实际. 为保证知识模型的合理性及系统进行知识处理的有效性,基于本体的知识定义及演化一般应遵循以下原则: ①替换原则. 由于本体类应用时附加有相应的语义规则,因此上位本体类规则应同样适用于下位本体类,即保证推理过程中可由下位实例替换上位实例. ②完整原则. 不同概念定义应具有明确的边界,各本体类应给出完整的概念特性及规则定义,不依赖于应用环境,以避免不同概念间语义冲突. ③最小原则. 本体类中应尽量减少特性,仅定义标识该本体类所必须的特性,以便于进行结构扩展及提高运行有效性.

案例知识和领域知识构成维修保障层次化知识体系,不但可满足不同领域知识的规范化、专业化要求,而且为案例知识提供良好的可扩展性. 其中,领域知识以描述知识的泛化关系,实现对知识的规范化及结构化描述为目标;案例知识则以建立案例与各领域知识关系为目标,通过建立案例本体与领域知识本体间的使用、聚集等关系,实现案例与领域知识的有机结合,为维修保障知识的应用和共享奠定了坚实基础.

3　基于语义相似的案例匹配

案例推理的关键在于如何快速、准确地从案例库中获得相似案例,针对以上建立的领域知识本体的知识细化功能和案例知识本体的知识综合功能,分别从结构层次的相似性及特征相似性给出相应的语义相似度计算方式. 对于领域知识本体,在由各概念结点构成的继承关系树结构中,本体概念按一定分类原则逐层进行细化,各概念特征与相似度无明确关系,相似度主要由结构特征及概念的层次深度决定. 假设两个概念 A, B 在层次网络中最近的祖先结点 P,则概念 A, B 自 P 点开始存在差异,自 P 点后概念 A, B 的细化方式并不影响其相似度. 因此,计算相似度主要考虑结点 c 的深度及出度密度两方面因素.

1) 深度因素,概念关系树中父结点比子结点更为抽象,因此层次越深的两个结点的相似度越大,可给出深度权重密度单调递增函数:

$$D_{ep}(x) = \frac{1}{2} + \frac{\arctan(x - \alpha_1)}{\pi} \qquad x \in 1, \cdots, N$$

其中 α_1 为中间层数,由 $\frac{\max(D_{ep})}{2}$ 取上限整数获得.

2) 密度因素,某一结点 c 出度大于 1时,密度越大说明该处概念细化程度越高,其细分的各结点间相似度越大,假设出度权重均相同,可得出密度权重模型:

$$D_{es}(c) = \begin{cases} 1 - \dfrac{1}{D_{eg}(c)} & D_{eg}(c) > 1 \\ 1 \end{cases}$$

其中 $D_{eg}(c)$ 为节点 c 的出度.

记根节点 R 到 P 结点的结点序列为 $\{ c_1, c_2, \cdots, c_n \}$,其中 $R = c_1$, $P = c_n$,最大层次深度为 m 层,则某领域知识中概念 A, B 的语义相似度计算式为

$$S(A, B) = \alpha_2 \times \sum_{i=1}^{n} (D_{es}(c_{i-1}) \times D_{ep}(c_i))$$

其中

$$\alpha_2 = \left(\sum_{i=1}^{m} D_{ep}(x) \right)^{-1}$$

案例是实现维修保障知识应用和共享的主体,案例本体中以领域知识或字典内容为特征,特征间逻辑关系为隐性的经验知识.因此,维修保障案例知识本体以向维修保障人员提供知识参考为目标,主要由特征值及附加规则决定其相似度,并且仅以共有特征进行计算相似性,忽略差异特征,最大限度扩大相似案例范围,以适应维修保障案例有限的特点.可得不同案例 C_a 和 C_b 间相似度计算式

$$S(C_a, C_b) = \sum_{i=1}^{n} \beta_i \times S(A_i, B_i)$$

其中,n 为共同特征数;β_i 为调节因子,且满足 $\sum_{i=1}^{n} \beta_i = 1$,用于设置特征所占权重.式中 $S(A_i, B_i)$ 为第 i 个特征相似度,当特征是领域知识时即为知识概念的语义相似度计算式;当特征为字典内容时则为关键字间相似度值,关键字相似度则可在相应的字典中给定;当特征为非格式描述文本则需进行分词处理后计算词汇序列的相似度.

4 应用实例

在基于 B/S结构开发的某航空制造企业飞机维修保障系统中,以维修保障中故障处理作为案例,针对故障较为频繁的导管类零备件进行分析.系统采用 Protégé建立知识的本体模型,基于 Jena实现基于本体的语义规则推理.

图 1中产品总层次数为 7,"液压导管"层次数为 5;故障模式总层次数为 7,"液压漏油"层次数为 4;故障原因总层次数为 3,"不易观察"层次数为 3.以故障判断为目标,设置相应调节因子为 0.38, 0.27, 0.16,相似度阈值为 0.5.典型界面见图 2.



图 2 基于语义维修保障案例检索结果界面

5 结束语

通过本体建立维修保障相关知识的一致性表述和组织,采用语义关联可描述不同领域知识间逻辑关系,并基于本体的语义相似实现了案例的检索和重用,为维修保障人员提供有效的经验获取和利用手段.但在领域知识本体的细化方面需结合模糊归类法、粗糙集等方法深入研究.

参考文献 (References)

[1] Naresh R, Sharma V, Vashisth M. An integrated neural fuzzy approach for fault diagnosis of transformers[J]. Power Delivery, 2008, 23(4): 2017 - 2024

[2] Zhao Feng, Koutsoukos X, Haussecker H, et al. Monitoring and fault diagnosis of hybrid systems[J]. Systems, Man, and Cybernetics, Part B, 2005, 35(6): 1225 - 1240

[3] 李青,史雅琴,周扬.基于案例推理方法在飞机故障诊断中的应用[J].北京航空航天大学学报, 2007, 33(5): 622 - 626
Li Qing, Shi Yaqin, Zhou Yang. CBR methodology application in fault diagnosis of aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(5): 622 - 626 (in Chinese)

[4] 魏传锋,李运泽,王浚,等.航天器热故障诊断专家系统推理机的设计[J].北京航空航天大学学报, 2005, 31(1): 60 - 62
Wei Chuanfeng, Li Yunze, Wang Jun, et al. Design on inference engine of the spacecraft thermal fault diagnosis expert system [J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(1): 60 - 62 (in Chinese)

[5] 李锋,魏莹.分布式环境下基于语义相似的案例检索[J].计算机工程, 2007, 33(9): 28 - 30
Li Feng, Wei Ying. Approach to case retrieval in distributed environment based on semantic similarity calculation[J]. Computer Engineering, 2007, 33(9): 28 - 30 (in Chinese)

[6] 杨奕飞,瞿元新.基于 CBR和 RBR的航天测控设备故障诊断专家系统[J].飞行器测控学报, 2005, 24(3): 23 - 31
Yang Yifei, Qu Yuanxin. A diagnosis expert system for aerospace TT&C equipment based on CBR and RBR[J]. Journal of Spacecraft TT&C Technology, 2005, 24(3): 23 - 31 (in Chinese)

[7] Rodriguez M Andrea, Egenhofer Max J. Determining semantic similarity among entity classes from different ontologies [J]. Knowledge and Data Engineering, 2003, 15(2): 442 - 456

[8] 黄果,周竹荣.基于领域本体的概念语义相似度计算研究[J].计算机工程与设计, 2007, 28(10): 2460 - 2463
Huang Guo, Zhou Zhurong. Research on domain ontology-based concept semantic similarity computation[J]. Computer Engineering and Design, 2007, 28(10): 2460 - 2463 (in Chinese)