

文章编号：1005-8451（2014）02-0007-04

动车组故障知识库的研究与设计

王延翠，崔玉龙，吴冬华

（南车青岛四方机车车辆股份有限公司 技术中心，青岛 266000）

摘要：动车组的在途紧急故障处理受到地域、时间和人员配备的约束，目前在国内缺少相应的系统为之提供支持。本文介绍知识、知识表示、知识发现等技术，研究、设计动车组故障知识库，通过知识库强大的推理功能，为动车组故障处理提供技术支持。

关键词：动车组；故障知识库；知识管理；知识表示

中图分类号：U266.2 : TP39 **文献标识码：**A

Research and design of EMU's Fault Knowledge Base

WANG Yancui, CUI Yulong, WU Donghua

(Technology Center, CSR SIFANG CO.,LTD, Qingdao 266000, China)

Abstract: The emergency fault dealing of Electric Motor Train Unit (EMU) in transit was under restrictions of geographical, time, and personnel allocation, and now it was lack of corresponding system to provide support for EMU in the country. This paper introduced the current status of knowledge, knowledge representation, and knowledge discovery, designed and implemented the Fault Knowledge Base of EMU, provided support for the fault processing by the powerful reasoning capabilities of Knowledge Base.

Key words: EMU; Fault Knowledge Base; knowledge management; knowledge representation

随着人工智能技术的迅速发展，在国内，已经有一些研究所和高校在实际运用领域对知识库及其管理系统进行了理论研究，并在其设计和研发等方面取得了研究成果^[1]。刘晓冰等^[2]通过面向对象的建模技术和方法构建知识的模型，以对象模型映射到关系数据库的方式存储知识，并以数据库（DB）的组织策略组织知识库（KB），提出面向大规模客户化产品协同设计过程的知识库构建方法。李爱平^[3]通过设计知识的建模和基于因特网的知识共享的实现技术，对基于因特网的设计知识库系统框架进行了相关研究。刘一^[4]采用面向对象的思想对汽车制动专家系统及其知识库的构建进行了研究。钟佩思^[5]对面向并行设计智能决策支持的知识库系统进行了研究。黄平^[6]则在飞机设计领域对知识库系统进行了研究。

在已经开发出的知识库系统中，比较有代表性的就是由浙江大学研究和开发的ZKBE系统。它是一个大容量的知识库系统，能够支持实际知识库系统的各个方面，如知识的获取和编程、知识的组织存储、知识的查询与管理，知识操作和

收稿日期：2013-07-21

基金项目：国家“八六三”计划项目（SS2012AA040912）。

作者简介：王延翠，工程师；崔玉龙，工程师。

统一性维护等。虽然如此，我国的知识库系统至今仍然处于试验阶段，没有统一的规范，真正实现商品化还需更深层次的研究。

本文依托国家“八六三”计划先进制造技术领域面向制造业的核心软件开发项目中复杂装备运维服务专业化构件系统与开发课题（课题编号：SS2012AA040912），研究并设计了动车组故障知识库。

1 知识库基础简介

知识库由人工智能（AI）与数据库（DB）两项计算机技术有机结合而成，是合理组织关于某一特定领域的陈述性知识和过程性知识的集合，它与数据库的区别是：知识库并非是向用户提供检索到的信息，知识库中不仅包含了大量的简单事实，还包含了规则、过程性知识和策略性知识，知识库系统可以根据输入的数据信息向用户提供对知识进行分析判断的结果。

在实际应用过程中，由于知识库本身的特性，知识库的基本结构是层次结构。其层次由下至上分别是：事实知识、用来控制事实的知识、策略。

用来控制事实的知识通常用规则和过程等表示，策略以中间层知识为控制对象，它常常被认为是规则的规则。

1.1 知识与知识表示

1.1.1 知识

知识是人类在社会实践过程中认识和总结的经验的综合，是经过整理、加工、解释和转换的信息，是一个或多个信息的关联，是人类进行一切智能活动的基础。人工智能的求解是以知识为基础的，一个程序具备的知识越多，它的求解问题的能力也就越强^[7]。

在知识领域，知识一般都是通过数据、信息和知识之间的区别来进行定义，数据、信息、知识的层次关系如图1所示。



图1 数据、信息和知识的层次关系

1.1.2 知识表示

知识表示（Knowledge Representation）是研究以何种形式将有关问题的知识存入计算机，以便计算机进行处理，即将知识符号化并输入给计算机的过程和方法。它把从知识源处获得的关于事实、推理和判断的知识进行归纳概括并建立起关系，然后将知识表达成计算机易于利用和表达的符号。知识表示是人工智能研究最活跃的领域之一^[8]。

知识表示包含两层含义：(1) 用给定的结构，按一定的原则和组织方式表示知识；(2) 解释所表示知识的含义。知识表示的目的不仅仅是解决知识在计算机中的存储问题，更重要的是要使这种表示方法能够方便地运用知识和管理知识。

1.2 知识表示的方法

知识表示有较多方法，任何表示知识的方法都必须满足以下两个条件：(1) 能表达事实性的知识；(2) 表达这些事实间的联系。目前，比较常用和有效的方法有逻辑表示法、产生式表示法、

语义网络表示法、框架表示法和过程式知识表示法等。

1.3 知识发现与获取

拥有知识是智能系统区别于其它计算机系统的重要标志，而知识的质量和数量又是决定智能系统性能的关键因素，系统推导出的结论的准确性，主要取决于系统的知识是否完整、准确和组织合理等。因此，知识发现和获取成为系统的关键要素。

1.3.1 知识发现

数据库知识发现（KDD, knowledge discovery in databases）术语出现于1989年，Fayyad认为，KDD是从数据集中识别出有效的、新颖的、潜在有用的，以及最终可以理解的模式的非平凡过程^[9]。知识发现就是从数据集中抽取和精化出新的模式^[10]。目前比较具有代表性的知识发现系统有：SAS Enterprise Miner, Intelligent Miner, Clementine以及MSMiner^[11]。

知识发现的过程可粗略地理解为3部曲：数据准备、数据挖掘（DM, Data Mining）以及结果的解释评估。其中，数据准备包括数据选取、数据预处理、数据变换3个子步骤。知识发现主要是为了从数据集提取有用的知识，需要通过抽样方法从数据集中选取目标数据，并对其进行预处理。数据预处理一般可能包括推导计算缺值数据、消除重复记录、完成数据类型转换等。得到的预处理数据经过变换处理即可进入数据挖掘操作阶段。数据变换是从初始特征中找出真正有用的特点，以减少数据挖掘时需要考虑的特征或变量个数，而数据挖掘则是通过相应的挖掘算法对数据进行抽取。数据挖掘阶段所抽取的信息需要经过用户或者机器的评估与解释才能成为有用的知识。知识发现的过程如图2所示。

1.3.2 知识获取

知识获取是指将用于求解某专门领域问题的知识从拥有这些知识的知识源（人类专家、书籍、文献等）中抽取出来，并以适当的形式表示，然后存储到计算机（知识库）中的过程。

通常知识的来源有3种途径：

(1) 从领域专家那里得到有用的知识。通过对专家所提供的知识进行分析、综合、去粗存精，归纳出可建立知识库的知识。(2) 从系统的设

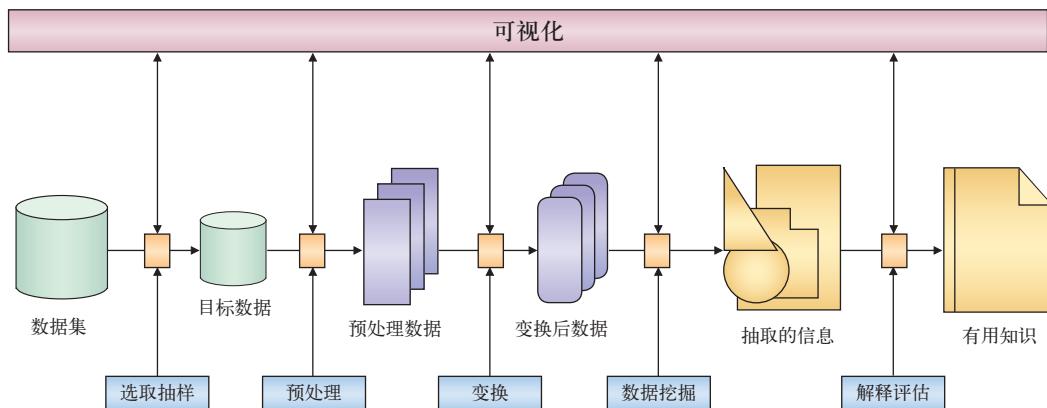


图2 知识发现的过程

计资料（包括设计原理、设计图样、各种技术报告、分析报告、试验数据等）中获取知识。（3）系统自身的运行实践。这需要在实践中学习，总结出新的知识。

2 动车组故障知识库的设计

由于知识推理过程中需要扫描检索整个知识库，若将所有的信息数据都存放于知识库中，势必会影响到整个系统的性能，因此本系统将所有不用于推理的事实性知识都单独存放于综合数据库（DB）中。

目前，知识库（KB）和DB的结合主要采用系统耦合的方式来实现，包括“强耦合”和“弱耦合”。强耦合是指数据库管理系统（DBMS）既管理规则库，又管理事实库，但是采用这种方法的知识库设计复杂程度较高，不易实现。弱耦合是将知识库管理系统（KBMS）和DBMS作为两个独立的子系统结合起来，令它们分别管理规则库和事实库。为了提高可靠性、可维护性、可理解性和可测性，动车组故障知识库的构建采用弱耦合方式，即对规则库和事实库分别进行管理。

因此，动车组故障知识库是全局性的，可被所有的规则访问；而规则之间不能相互调用，它们之间的联系只能通过综合数据库进行操作。这些特点使动车组故障知识库可以对综合数据库、产生式规则和控制系统进行相对独立的修改。

2.1 故障知识的获取

本知识库有专门的故障知识获取模块，负责建立、修改与扩充故障知识库，与专家知识审核提交系统相结合，对故障知识库的一致性、完整

性和正确性进行维护。

知识获取的方法按照推理能力来分，可以分为4种：

(1) 没有推理能力的人工获取方法；(2) 利用知识编辑工具的半自动知识获

取方法；(3) 具有推理能力的自动知识获取方法；(4) 具有高级学习功能的超水平的自主式获取方法。

根据动车组故障数据和故障维修的特殊性，本知识库的故障知识主要来源于以下4个方面：

(1) 动车组故障维修手册、动车组途中紧急故障处理手册；(2) 对动车组车载信息数据进行处理，挖掘出相应的规则；(3) 专家对一种从未出现且系统无法给出解决方案的故障提出较好的处理措施并正确维修后，将其通过知识管理系统录入到事实库和规则库。(4) 知识在推理过程中通过机器学习，生成新的知识。

2.2 故障知识库中知识的表示

动车组的故障知识主要包括动车组发生故障的原因、故障的解决方案、发生该故障的置信度等信息，因此，结合动车组故障信息领域的知识特点，采用产生式表示法来表示故障知识。即IF(条件)，THEN(动作，结论)。只要IF后面的条件或者事实成立，就可以得到THEN后面的动作或者结论。

2.3 故障知识库的建立

由于采用基于产生式规则的知识表示方法，因此，动车组故障知识库的建立包括对事实库和规则库的建立。在建立事实库时，故障知识需要按照专家提供的动车组故障字典的分类逐一添加入库，放入不同的故障类型表中。规则库是知识库的一个重要组成部分，进行推理的基础就是知识库中用规则表示的知识。

对规则库中的规则建立要求遵循两个原则：

(1) 规则库中定义的规则应当尽可能最简化，以避免不必要的冗余；(2) 故障规则之间不能够

有冲突存在。

根据以上步骤和要求,结合系统自身的实际运行需要,利用已经获取的故障知识,本文所建的动车组故障知识库采用的数据库系统为 Oracle 10g, 分别建立了规则表 CRH2_KNOWLEDGE_RULE 和事实表 CRH2_KNOWLEDGE_FACT。规则表主要记录规则编号、规则前提的个数、规则前提条件列表、规则置信度、规则结论等信息,其表格式如表 1 所示。

表1 规则表

字段命名	字段名称	数据类型	说明
ID	规则编号	NUMBER	主键, 唯一性约束
PREMISENUM	规则前提的个数	NUMBER	
Premise	规则前提	VARCHAR2(500)	数字 [& 数字]
CONFIDENCE	规则置信度	NUMBER	0~1 之间
CONCLUSION	规则结论	VARCHAR2(500)	
...

事实表主要记录故障编号、故障类别、故障名称、故障原因、故障发生概率、故障处理措施等信息。其表格式如表 2 所示。

表2 事实表

字段命名	字段名称	数据类型	说明
ID	故障编号	NUMBER	主键, 唯一性约束
FAILURECLASS	故障类别	VARCHAR2(50)	来源于故障字典
FAILURENAME	故障名称	VARCHAR2(50)	
CAUSE	故障原因	VARCHAR2(500)	
PROBABILITY	故障发生概率	NUMBER	0~1 之间
MEASURE	故障处理措施	VARCHAR2(500)	
...

每个产生式规则对应于表 CRH2_KNOWLEDGE_RULE 中的 1 个记录, 产生式规则条件之间的“与”(AND) 连接可以直接存储在规则表中, 而对于产生式规则的“或”(OR) 连接, 需要转换成 AND 连接方式来表达。例如: 规则 IF a OR b, THEN c AND d, 转化后变成两条规则存储在规则库中, 如下所示。

Rule 1: IF a, THEN c and d

Rule 2: IF b, THEN c and d

利用传统的方法提取动车组故障规则来建立知识库, 既耗时又需要复杂处理。而且, 大量冗余的规则知识会形成不必要的庞大的知识库系统, 这会使规则的维护变得低效, 从而阻碍一个高质量知识库的建立。因此, 动车组故障知识库从获得的大量数据中提取关联规则, 不但可以有效减

少规则的数量, 而且可以在不损失有效数据信息的情况下缩小知识库的规模。由于获取关联规则不是本文研究的重点, 在此不再过多叙述。

3 结束语

研究并设计了动车组的故障知识库, 介绍动车组故障知识的获取方法以及其知识来源, 通过产生式规则表示法来表示动车组故障知识, 建立故障知识库的事实表和规则表, 对其故障知识和规则的获取和建立做了详细设计和实现说明, 并列出其部分字段, 可为动车组故障处理提供技术支持。

参考文献:

- [1] 徐勇, 杨柯. 一种面向对象知识库的构造和维护 [J]. 计算机工程, 2000, 26 (9): 60-62.
- [2] 刘晓冰, 杨春立, 孙伟德, 等. 产品设计知识库建立方法研究 [J]. 计算机集成制造系统, 2003, 9 (8): 621-625.
- [3] 李爱平. 基于因特网的设计制品知识库系统框架的研究 [J]. 中国机械工程, 2002, 13 (14): 1212-1216.
- [4] 刘一. 面向对象的汽车制动系专家系统及其知识库的构建 [D]. 吉林: 吉林大学, 2003.
- [5] 钟佩思. 面向并行设计智能决策支持的知识库系统研究 [D]. 哈尔滨: 哈尔滨工业大学, 1999.
- [6] 黄平. 飞机构件 CAD 设计知识库系统研究 [D]. 西安: 西北工业大学, 2000.
- [7] OLES F J. An application of lattice theory to knowledge representation[J]. Theoretical Computer Science. 2000. (249): 163-196.
- [8] 史忠植. 知识发现 [M]. 北京: 清华大学出版社, 2011.
- [9] Chandrasekaran B, Johnson TR, Smith JW. Task structure Analysis for Knowledge Modeling[J]. Communication of the ACM. 1992, 35(9), 24-137.
- [10] 胡运发. 数据与知识工程导论 [M]. 北京: 清华大学出版社, 2003.
- [11] Xuefeng Kang, Hongyu Zhou. Object-oriented knowledge representation based on DB technique[J]. Journal of Harbin Institute of Science and Technology, 2001, 6(3): 1-3.

责任编辑 杨利明