

装备保障数据仓库实施中的多维数据模型构建研究

谢峰, 孙江生, 代冬升, 梁伟杰

(机械技术研究所, 石家庄 050000)

摘要:目的 解决装备保障数据仓库实施中数据建模问题。方法 总结近年来多维数据模型的构建方法,对基于本体的多维数据模型构建方法进行优化,提出适于装备保障数据多维模型构建的“本体-层次”建模方法。结果 该方法的运用,能够在庞大的装备保障体系中,较为便捷地提取数据模型的维度与指标,简化了多维数据模型的构建过程。结论 为装备保障数据仓库的实施提供了有力支持。

关键词: 本体; 多维数据模型; 数据仓库; 装备保障

DOI:10.7643/issn.1672-9242.2014.05.026

中图分类号: TP311.1 **文献标识码:** A

文章编号: 1672-9242(2014)05-0137-05

Study on Multidimensional Data Modeling in Building Equipment Support Data Warehouse

XIE Feng, SUN Jiang-sheng, DAI Dong-sheng, LIANG Wei-jie

(Mechanical Technical Research Institute, Shijiazhuang 050000, China)

ABSTRACT: **Objective** To solve the issue of the construction of multidimensional data model in the implementation of equipment support data warehouse. **Methods** This paper summarized multidimensional data modeling methods in recent years, optimized the ontology-based multidimensional data modeling method and put forward a new method of establishing multidimensional data model-"ontology-AHP" construction method. **Results** This method conveniently extracted the dimension and indexes for the data model from the large equipment support system and simplified the construction process of multidimensional data modeling. **Conclusion** It provided strong support for the implementation of equipment support data warehouse.

KEY WORDS: ontology; multidimensional data model; data warehouse; equipment support

近年来,数据仓库技术在装备保障领域得以应用,解决了因装备保障数据量急速增长、数据分析需求不断提高带来的现实难题。现代信息技术的使

用,有效提高了装备保障的效率,为装备战斗力的快速生成提供了有力支撑^[1-2]。不同于普通企业数据仓库的构建,在装备保障数据仓库实施过程中,因其

收稿日期: 2014-06-05; 修订日期: 2014-07-30

Received: 2014-06-05; Revised: 2014-07-30

作者简介: 谢峰(1986—),男,山东泰安人,硕士研究生,主要研究方向为装备保障信息化。

Biography: XIE Feng(1986—), Male, from Tai'an, Shandong, Master graduate student, Research focus: equipment support informatization.

体系庞大,面临许多新的技术问题,这些问题的存在对于项目实施提出了新的挑战^[3]。多维数据模型构建是装备保障数据仓库实施中的核心问题,通过对传统建模方法的总结分析,文中提出基于“本体-层次”的多维数据模型构建方式。在装备保障数据仓库实施实践中,该方法对多维数据模型的构建过程提供了有力支持,具有良好的实践应用价值。

1 多维数据模型构建概述

传统的多维数据模型构建方法主要有基于关系表^[4]、E-R 模型^[5]和 UML^[6]等构建方式。这些方式大都是立足于传统数据库中数据的建模方式,在构建多维数据模型的过程中都有一定的局限性。2007 年,Oscar Romero 和 Alberto Abello 提出基于本体的半自动维度建模方法^[7],在该方法的指导下,陆群进行了海洋环境数据仓库多维模型的构建实践,取得良好效果^[8]。该方法能较为便捷地生成维度候选集,却并未完成最后维度的确定。徐俊良采用本体与业务混合驱动的方式实现了维度的筛选与确定,为维度的确定提供了一种新方法^[9]。在混合驱动的多维数据模型构建中,业务的理解较为复杂,无形中增加了开发人员的工作负担和数据仓库开发周期。文中使用一种体系化的多维数据模型构建方法,从本体生成,到候选维度生成,再到维度确定,实现了装备保障数据多维数据模型的完整构建。多维数据模型维度构建实施步骤如图 1 所示。

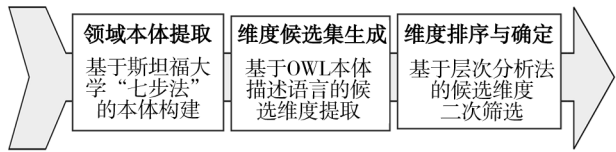


图 1 “本体-层次”方法多维数据模型构建实施步骤
Fig.1 "Ontology-AHP" based process for building multidimensional data models

2 装备保障领域本体生成

装备保障领域本体构建是装备保障信息化发展的必然。根据装备保障领域知识的研究和应用,进行科学化地提取和分类,形成形式化的概念,即本体。本体现阶段的主要构建方法主要有手工构建、

复用已有本体(半自动构建)以及自动构建本体等 3 种方法。胡金强对装备保障领域本体进行了研究,采用了半自动的方式进行构建,利用已有的装备保障资源目录,结合领域专家的参与,对装备保障领域本体构建进行了分析^[10]。借鉴上述文献的本体构建思想,并对常用的本体构建方法进行总结,文中使用较为常用的斯坦福大学的“七步法”,参考已有的知识库资源,对装备保障本体进行构建,实施过程如图 2 所示。

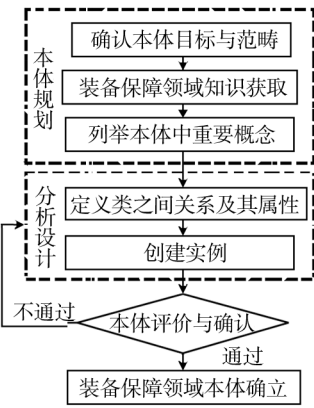


图 2 装备保障领域本体构建流程
Fig.2 Process of building equipment support business ontology

3 基于装备保障领域本体的维度候选集生成

3.1 本体的三元组表示

装备保障领域内的本体结构可以表示为一个三元组:

$$ESOntology := \langle C, R, Attr \rangle.$$

1) C 表示概念集合,主要指装备保障领域中的对象、类,如设备工具类、备件器材类、故障维修类等,形式化描述为:

$$C := \langle CID, Name, Syn, Abb, ParentC, ChildC, Domain, Description \rangle$$

分别表示概念标识符、概念名称、同义词、缩略词、父概念、子概念、领域以及概念的描述。

2) R 表示概念间的关系,及装备保障领域中定义的关系包括常用的关系(如 is-a, part-of, kind-of)等以装备保障领域中定义的关系(如基地、仓库与部队),形式化描述为:

$R := \langle RID, Name, FirstCID, SecondCID, Description \rangle$

分别表示关系标识符、关系名称、第一个概念标识符、第二个概念标识符、关系描述。需要说明的是,一些基本关系具有传递性,例如:如果存在 $c_1 \in C, c_2 \in C, c_3 \in C$ 且 c_1 is-a c, c_2 is-a c_3 , 则 c_1 is-a c_3 。

3) Attr 表示概念的属性集合,如维修器材类的属性有:器材名称、器材编码、生产厂家、单价、器材批次等表示为:

$Attr := \langle AttrID, Name, CID, RID, Type, Constraint, Default, Description \rangle$

其分别表示属性标识符、属性名称、概念标识符、关系标识符、属性取值类型、属性缺省值、属性描述。另外,属性可以附属于某一概念,也可以附属于某一关系。

3.2 OWL 描述本体中维度的生成策略

在数据仓库中,数据模型以星型结构或雪花型结构进行存储,这种存储结构利于数据的分析、挖掘,同时形成了维度与事实之间的特殊关系,即一对多的关系。其关系总结如下:对于每个事实都有且仅有 1 个维与之对应;对于每个维度来说也就有多个事实与之对应。基于维度与事实的这一特殊关系,结合本体概念的特殊描述形式,得到如下推论:在本体集中如果一个概念与另一个概念是一对多的关系,那么这个概念就有可能成为一个潜在的维。

OWL 是常用的本体描述语言,该语言中,将本体描述为概念、属性、实例等 3 部分,其中属性又分为对象属性和数据属性,对象属性能够完整描述类之间的相互关系。通过上文的分析,维度与事实的关系符合 OWL 描述语言中的如下描述:

本体中描述的概念 F 和 D,如果 F 的一个属性 R 具有约束 ‘allValuesFrom(D)’ 和 ‘minCardinality(1)’,那么就可以认为 D 和 F 之间的关系式是维度与事实。

其中,属性 R 的约束 allValuesFrom(D) 表示 F 中的实例全部来源于概念 D 的;属性 R 的约束 minCardinality(1) 表示概念 F 的实例数量至少为 1。

通过以上分析,从本体描述的 OWL 中遴选装备保障数据仓库维度集,其实现的伪代码如下:

```
for (classObj : ontology) { //循环本体中的概念
```

```
class = ontology.getNextClass(); //获取本体中的概念
```

```
Fact fact = new Fact(classObj); //为概念建立事实
```

```
for (property : ontology) { //循环本体中的属性
```

```
property = ontology.getNextProperty(); //得到本体中的属性
```

```
if (property.allValueFrom(class) and minCardinality(1)) { //存在 allValueFrom 和 minCardinality(1) 关系
```

```
dimension = new Dimension(property); //建立维度
```

```
fact.addDim(dimension); //为事实表添加维度}}}
```

4 基于层次分析法的维度确定

美国运筹学家 A. L. Saaty 于 20 世纪 70 年代提出的层次分析法 (Analytic Hierarchy Process, AHP), 是一种定性、定量分析相结合的多目标决策分析方法^[11-13]。应用该方法,对上文通过本体筛选出的维度候选集进行评价,得到符合装备保障需求的最优选维度集。应用层次分析法,通过以下步骤实现维度确定^[14-15]:建立装备保障层次结构模型、构造判断矩阵、层次单排序、判断矩阵的一致性检验、层次总排序权重计算。

1) 建立层次结构模型,确定各个评价指标。如图 3 所示。

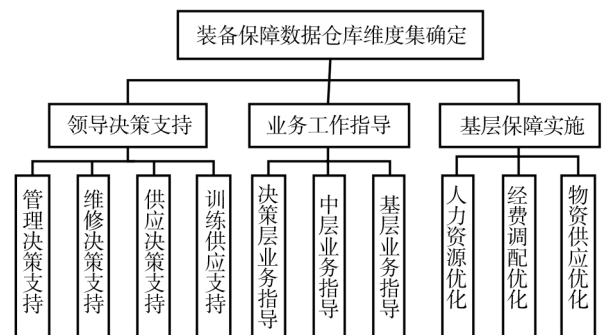


图 3 维度集评价指标体系

Fig. 3 Index system for dimensionality evaluation

2) 根据指标体系的层次结构构建判断矩阵,并

求出权重向量。

矩阵 $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ 为指标体系的比较判

断矩阵。矩阵中 a_{ij} 表示第 i 个指标对第 j 个指标的影响程度,反映一个指标对于另一指标的影响程度。矩阵中的 a_{ij} 须由领域内最相关实践人员评定给出,矩阵中取值见表 1。

表 1 比例标度含义
Table 1 Implications of the proportional scale

判断矩阵中 各元素取值	含义
1	两个目标相比,同样重要
2	两个目标相比,一个比另一个稍微重要
5	两个目标相比,一个比另一个明显重要
7	两个目标相比,一个比另一个重要很多
9	两个目标相比,一个比另一个显得极其重要
2,4,6,8	介于上述两相邻情况之间
以上各数的倒数	两目标反过来比较

表 2 为判断矩阵之一,是准则 B 层对于 A 层的权重判断矩阵。层次分析法是一种多层级的分析方法,这样,每一层对于上层的影响都有一定的权重分配。表 2 确定了 B 层的领导决策支持、业务工作指导、基层保障实施相对于 A 层装备保障数据仓库维度集确定相对重要性权重。

表 2 A-B 判断矩阵
Table 2 A-B judgment matrix

维度集合 确定 A	领导决策 支持	业务工作 指导	基层保障 实施
领导决策支持	1	5	3
业务工作指导	1/5	1	1/2
基层保障实施	1/3	2	1

3) 层次单排序权重确定并检验一致性,通常利用方根法近似值确定各评价指标的权重 W ,计算步骤如下:① 计算判断矩阵 A 每一行的积 $M_i =$

$\sum_{j=1}^n a_{ij} \quad (i = 1, 2, \cdots, n)$; ② 计算 M_i 的 n 次方根 $\overline{w_i} =$

$\sqrt[n]{M_i}$; ③ 计算矩阵的特征向量的元素值 $w_i = \frac{\overline{w_i}}{\sum_{j=1}^n \overline{w_j}}$, 则向量 $w = [w_1, w_2, \cdots, w_n]^T$ 即为所求特

征向量;④ 计算矩阵的最大特征值 $\lambda_{\max} = \sum_{j=1}^n \frac{(Aw)_i}{nw_i}$, 其中 $(Aw)_i$ 为向量 Aw 的第 i 个元素。

4) 实现一致性检验。为确保专家思维的一致性,针对专家所填写判断矩阵进行一致性检验。计算一致性指标 $CI = \frac{\lambda_{\max} - n}{n - 1}$, 然后根据 n 查表取得平均随机性指标 RI 的值。计算随机一致性比率 $CR = \frac{CI}{RI}$, 当 $CR > 0.10$ 时,认为层次分析排序结果不满足一致性,即权重系数的分配不合理,需要重新分配权重系数的值;当 $CR < 0.10$ 时,则不需要调整判断矩阵的元素取值,通过一致性检测。

5) 进行层次总排序获得维度排序。通过以上几个步骤,得到了一组元素对其上一层中某元素的权重向量。通过自上而下将单一准则下的权重进行合成,得到了各元素对于评价指标体系顶层目标的最终排序。该排序反映出各维度在装备保障工作中的重要程度。

通过以上 5 个步骤,在维度候选集中的对象以定量的方式给出了排序。用户根据排序,选取最具有价值的维度对象,生成最终的维度集,完成了装备保障数据仓库维度的生成。

5 结 语

面对装备保障地域广、层次复杂、对象多样的现实情况,数据仓库维度的选择亟待一种定性定量相结合的筛选方法。文中在汲取近年来多维数据模型构建方法的基础上,提出的“本体-层次”多维数据维度选取方法,改变了简单由客户需求定性生成维度集的单一方式,减少了多维模型构建中维度选择的盲目性。通过科学化、体系化的方法甄选维度集,使得模型维度的构建更符合装备保障的实际需求,为装备保障数据仓库的实施提供了有力支持。

参考文献:

[1] 阮拥军,董立宁,刘栋. 信息化战场环境下装备精确保

- 障探析[J]. 装备环境工程, 2012, 9(2): 54—56.
- RUAN Yong-jun, DONG Li-ning, LIU Dong. Research of Equipment Support in Informatization Battlefield Environment Analysis[J]. Equipment Environmental Engineering, 2012, 9(2): 54—56.
- [2] 单志伟. 装备综合保障工程[M]. 北京: 国防工业出版社, 2007.
- SHAN Zhi-wei. Integrated Equipment Support Project[M]. Beijing: Defense Industry Press, 2007.
- [3] 郑岩. 数据仓库与数据挖掘原理及应用[M]. 北京: 清华大学出版社, 2011.
- ZHENG Yan. Data Warehouse and Data Mining Principles and Application[M]. Beijing: Tsinghua University Press, 2011.
- [4] 陆昌辉. 基于关系数据模型的多维数据建模方法的研究[D]. 长沙: 国防科学技术大学, 2002.
- LU Chang-hui. Multidimensional Data Modeling Method Based on The Relational Data Model Research[D]. Changsha: National University of Defense Technology, 2002.
- [5] 严金贵. 基于 ER 模型的多维数据建模方法研究[D]. 重庆: 重庆大学, 2006.
- YAN Jin-gui. The Research of Multidimensional Data Modeling Based on ER[D]. Chongqing: Chongqing University, 2006.
- [6] 段薇, 戴莉萍. 基于 UML 的社会保险多维数据模型建模的研究[J]. 计算机与现代化, 2008, 10: 40—42.
- DUAN Wei, DAI Li-ping. On Multidimensional Data Model of Social Insurance Based on UML[J]. Computer and Modernization, 2008, 10: 40—42.
- [7] OSCAR Romero, ALBERTO Abello. A Survey of Multidimensional Modeling Methodologies[J]. International Journal of Data Warehousing & Mining, 2009, 5(2): 1—23.
- [8] 陆群. 基于领域本体的海洋环境数据仓库设计[D]. 沈阳: 东北大学, 2009.
- LU Qun. Domain Ontology-based Marine Environmental Data Warehouse Design[D]. Shenyang: Northeastern University, 2009.
- [9] 徐俊良. 混合驱动多维数据模型建模方法研究与应用[D]. 上海: 上海交通大学, 2012.
- XU Jun-liang. Research and Application of Hybrid-driven Data Warehouse Modeling Method[D]. Shanghai: Shanghai Jiaotong University.
- [10] 胡金强. 装备保障知识管理平台构建与应用研究[D]. 石家庄: 军械工程学院, 2010.
- HU Jin-qiang. Research on Construction and Application for Knowledge Management Platform of Equipment Support[D]. Shijiazhuang: Ordnance Engineering College, 2010.
- [11] 李千, 纪小柠, 牟牧. 基于 AHP 和 ANP 作战飞机系统效能评估[J]. 火力与指挥控制, 2009(S1): 92—94.
- LI Qian, JI Xiao-ning, MOU Mu. AHP and ANP based combat aircraft system performance evaluation[J]. Fire Control and Command Control, 2009(S1): 92—94.
- [12] 张志强, 徐斌, 何勇灵, 等. 基于 AHP 评价方法的发动机性能评价[J]. 兵工学报, 2008(5): 625—628.
- ZHANG Zhi-qiang, XU Bin, HE Yong-ling, et al. AHP Based Engine Performance Evaluation[J]. Journal of China Ordnance, 2008(5): 625—628.
- [13] 赵国存, 刘占岭. 基于 AHP-TOPSIS 的装备保障信息定量评价研究[J]. 装备环境工程, 2012, 9(4): 118—122.
- ZHAO Guo-cun, LIU Zhan-ling. Based on AHP and TOPSIS Equipment Support Information Quantitative Evaluation Research[J]. Equipment Environmental Engineering, 2012, 9(4): 118—122.
- [14] 梁波, 易建政, 蔡军锋, 等. 基于 AHP 的野战弹药防护效果评价[J]. 装备环境工程, 2007, 4(2): 73—76.
- LIANG Bo, YI Jian-zheng, CAI Jun-feng, et al. Field Ammunition Protection Effect Evaluation Based on AHP[J]. Equipment Environmental Engineering, 2007, 4(2): 73—76.
- [15] 蓝金辉, 马宝华, 李有文, 等. 层次分析法用于多传感器数据融合中属性权的求取[J]. 兵工学报, 2001(1): 74—77.
- LAN Jin-hui, MA Bao-hua, LI You-wen, et al. AHP Used in Multi-sensor Data Fusion in the Calculating of Property Rights[J]. Journal of China Ordnance, 2001(1): 74—77.