

模型的事例表示<sup>①</sup>

张洁, 潘启澍, 胡运权

(哈尔滨工业大学管理学院, 哈尔滨 150001)

**摘要:**基于事例推理的模型管理系统中,有效地表达事例是实现系统的基础,模型事例不是属性值事例,而是结构事例。本文针对原有表达事例方式的不足提出了一种模型事例的表达方法,该方法适合于各种数值类模型的表示,并能够脱离模型的语义,同时给出了此种事例库组织的初步方案,为以后的研究奠定了基础。

**关键词:**基于事例推理; 模型管理; 智能决策支持系统

**中图分类号:** TP399; F27 **文献标识码:** A **文章编号:** 1007-9807(2000)02-0090-05

## 0 引言

基于事例推理(case-based reasoning, CBR)或者基于事例学习(case-based learning, CBL)技术是针对传统 AI 知识获取的瓶颈问题提出的。因此,以 CBR 技术为基础的智能决策支持系统的研究成为解决 IDSS 问题的一条新途径。本文旨在解决 IDSS 中用 CBR 进行模型管理中存在的一些问题。

CBR 的思想是通过比较目前状态(事例)与过去曾经遇到的状态(事例),从而提出在新形势下的行动模型<sup>[1]</sup>。CBR 技术中包括两个最重要的部分,即:(1)索引,查找,修改事例的算法;(2)事例的表示,通过解决这两个问题的不同方式而形成了组成 CBR 系统的两大部件:事例库和求解器,而求解器又分为检查部件和推理部件两部分。事例表示问题是 CBR 的基础,所以解决事例的表示是引入 CBR 技术的先决条件。

## 1 研究进展及问题分析

## 1.1 研究进展

CBR 技术曾经成功地应用于分析,设计,计划,诊断等许多领域。CBR 技术用于模型方面,笔

者认为可以归纳为如下几类,但分类中不包含简单的属性值事例和与 DSS 无关的应用。

## (1) CBR 技术用于模型管理

这方面最有代表性的是 Liang<sup>[2]</sup>将 CBR 技术用于模型管理系统(model management system),其具体的细节本文后面作详细分析,再有 Robert<sup>[3]</sup>等也是作这方面研究,他的方法与 Liang 不同,采用框架表示事例,事例包括五个槽,索引集范围,输入参数,处理单元,输出参数,同类模型框架和模型框架(scope, C-inputs, processing-units, C-outputs, Abstraction-frame 和 module-frame)。该方法不能从构模角度反映模型事例的本质,并且其检索也以语义为基础,不能做为这方面研究工作的代表。

## (2) CBR 技术用于决策分析

这方面的研究以属性值事例为基础的研究占有很大比例。有事例组织结构方面的(象事例的索引结构、事例库的组织),也有许多研究算法方面的(比如事例分类及检索算法)。结构型事例用于这方面研究的比较少,值得一提的是 Govern<sup>[4]</sup>将 CBR 技术用于决策理论模型,借助影响图来表示事例,通过对影响图的结点和连结弧的检索和修改,适应于新问题。

## (3) CBR 技术用于模型求解

这方面的研究主要集中在一些 NP 问题上,

① 收稿日期:1999-03-29;修订日期:1999-05-24。

作者简介:张洁(1972-),男(汉族),黑龙江省牡丹江人,博士生。

包括用 CBR 技术求解 TSP<sup>[5]</sup>问题,用 CBR 技术求解 job-shop scheduling<sup>[6]</sup>等。

## 1.2 问题分析

在模型管理领域中要引入 CBR 技术,首先遇到的问题就是事例的表示,这是应用 CBR 技术的基础,事例的表示有许多方法,根据其最根本的特征可以把事例分成两类,即属性值事例和结构事例,结构事例又可以分为抽象事例、部分事例、子事例、复合事例、层次事例,在模型管理的应用中要表示一个模型不可避免地要用到结构事例,原因很明显,模型是现实世界的反映与抽象,而现实世界中的问题仅用简单的属性值来表达是不可能的,所以在这方面研究中遇到多为结构事例,Liang把模型事例表达为一个三元组,即:

$\langle E, S, F \rangle$

E 代表问题表述(problem representation)

S 代表一般模型图(generic model graph)

F 代表数学形式(mathematical form of the model)

E 部分又分为实体图(Entity graph)和属性层次图(attribute hierarchy),实体图反映问题的非数字特征,属性层次反映问题的数学特征并且按数量的单位分组,分层,一般模型图 S,脱胎于 Geoffrion<sup>[7]</sup>的结构化构模的一般图,反映模型的基本部分, F 数学形式部分是以数学公式的形式反映模型,分为四级,即模型实例,模型,模型类和理论模型,一级比一级抽象,Liang 的方法是 CBR 用于 MMS 的代表,但是 Liang 的方法有一些不足之处。

(1)反映问题的模式与联系不充分,实体和属性相分离,构模过程主要是一个抽象过程,从而得出问题的内在联系,如果两个模型内在联系一致,则是一类模型,Liang 是用实体图和一般模型图来反映联系的,而一般图中是反映出了数学联系,实体图只反映实体的关系,从而实体与属性的联系相脱离,不能完整地把握问题实质;

(2)目标不独立,因为对于同一个问题,目标不同模型构建的结果也不一样,虽然两个问题数学形式完全一样,但目标不同模型最后构建也是不一样的,例如同样一个车间,追求产量极大和只计算产量则模型也是不同的,但不同的目标却可以使用相同的数学公式,可以说目标是构模的起

点,因此应与问题和模型相独立;

(3)模型表达过于琐细,影响匹配效率和构模效果,Liang 的模型事例中用到单位层次涉及到单位的匹配,非常琐细,这就影响了 CBR 检索的效率,又因为构模过程是一个创造过程,如果两种事例越相似,其推理过程中的创新性越小,很多情况下问题环境差别很大或相似性极小则影响事例推理的效果。

## 2 模型的事例表示法

根据以上分析笔者提出一种模型事例表示法,一个事例仍然表示为一个三元组,但内容完全不同,即: $\langle P, F, G \rangle$ 三个元组,其中 P 代表问题,由适合于模型表示的改进 ER 图来表示,反映事实的抽象, F 代表约束条件和函数关系等数学形式,反映问题中的数学关系, G 目标是模型的最终目标,下面分别讨论。

### 2.1 问题表示

问题表示的关键是要反映问题的本质联系,而且考虑到 CBR 技术中的检索和匹配不能采用过于复杂的形式,问题表示采用改进 ER 图好处有两点:1. ER 适合于描述关系,ER 模型能反映问题内部的复杂关系,Liang 将属性与实例分开则不能反应其实体和属性的联系,所以对问题的表示这里直接采用 ER 模型,2. 便于数据与模型的结合,构建模型时数据与模型是相分离的,而模型求解时要寻求其结合,ER 是数据库设计工具,因此,便于将两者结合起来,这里的 ER 模型与数据库设计的 ER 模型稍有不同,本文将 ER 模型中属性分已知属性 K(known)和未知属性 U(unknown),其中,已知属性表示属性的值是已知的,即已知属性值属性,简称为已知属性,未知属性是指属性的取值是未知的,简称为未知属性,而且已知属性的数字型属性又分为限制型(用 L 代表)和常数值型(用 C 代表)两种,属性类分别写在属性后面的括号中,比如运输模型的这种 ER 图可以见图1所示。

其中实体用长方形、属性用椭圆、联系用菱形表示,这种 ER 图的基本概念与数据库设计的 ER 图一致,实际绘制时为了简化 ER 图,增加一些要求,首先,属性上是只反映与变量约束条件等数学

形式相关的,无关属性不反映在其中,而且对于那种只有一对多联系的实体如果该实体不含相关属性则可以省略,并可以把联系中的属性归为多方,其中相关属性是指属性值对构模的数学形式发生

影响的属性,通过这种变换方法可以简化 ER 图,便于问题匹配,可以证明这种省略不影响问题的表示及数据库的设计.

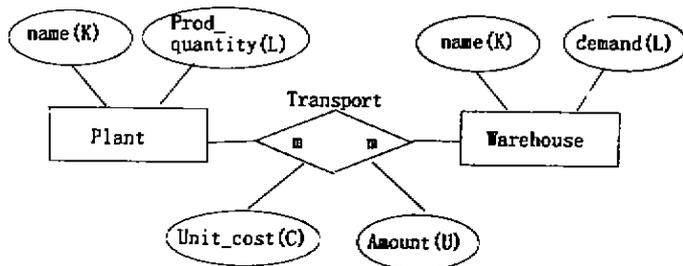


图1 运输问题的改进 ER 图

**命题1** ER 图中一对多或一对一的二元联系中,‘一方’如果没有相关属性,则‘一方’实体和关系可以省略.

**证明**

不失一般性,假设存在  $1:X$  关系. ‘一方’实体为  $ES$ , 多方为  $EM_1, EM_2, \dots, EM_n$ ,  $EM_i$  与  $ES$  之间存在二元联系为  $R_i$ ,  $R_i$  的属性集合为  $A_i$ ,  $EM_i$  的属性集合为  $(KEM_i, AEM_i)$ , 其中  $KEM_i$  代表码,  $AEM_i$  代表其它属性,  $ES$  中没有相关属性.

1) 因为  $ES$  中没有相关属性,所以可以去掉  $ES$  并把  $R_i$  中的属性  $A_i$  给相应多方实体  $EM_i$ , 然后,去掉关系  $R_i$ , 最后只留下  $EM_i$  实体.

2) 化简之前,  $R_i$  的关系数据模型为  $(KEM_i, A_i)$ ,  $EM_i$  的关系数据模型为  $(KEM_i, AEM_i)$ , 则实体与关系相合并得到关系模型  $(KEM_i, AEM_i, A_i)$ . 化简后关系模型直接表示为  $(KEM_i, AEM_i, A_i)$ , 因此两者的数据库关系模型是一致.

3) 改造的 ER 图中相关属性集为其表现出属性的子集,而函数关系都是由相关属性相互推导而来,在 1 的变换中,  $EM_i$  的属性不变化,  $R_i$  的属性归为  $EM_i$ , 所以相关的属性不会减少,因此不影响函数关系,因而对模型的表达不受影响.

**命题2** ER 图中存在三个或多个实体相联系时,某‘一方’如果没有相关属性,则此一方实体可以省略.(证明与上命题类似,注意此时联系不可省,因为此联系还涉及到其它实体)

这种 ER 图可以反映出问题的内在联系与本质内容. 比如运输问题的这种 ER 图有两个约束

性属性,所以必然存在两组约束条件. 决策变量存在于实体与实体之间的联系中即  $U$  类属性  $Amount$ . 除非为了表达问题的方便,决策变量只有一组最基本的,其他变量可以由一组基本决策变量推导得出. 如果图中有多组  $U$  属性则其间必然存在数学推导关系. 另外,属性的改变影响模型,如运输联系中运价由  $K$  变为  $U$  则模型发生变化成为非线性规划问题.

这种问题表示方法如果去掉实体联系和属性中的语义对实体和联系编号,只保留属性类型则可以反映问题的实质,如果两个问题的这种 ER 图相似则模型基本一致. 而且问题表示中可以很容易地设计出模型的数据库,即

(Plant\_name, Quantity)

(Warehouse\_name, Demand)

{Plant\_name, Warehouse\_name, Unit\_shipping\_cost, Shipping\_amount}

## 2.2 函数及约束的表示

这部分内容反映模型中的数学关系. 因为 ER 图中每个实体代表一个集合,所以首先赋予每一实体一个索引,然后给属性分配变量名,从而表达出整个模型的数学关系如下:

INDICES

Index Index set

i plant

j warehouse

VARIABLES AND ATTRIBUTE

Name Attribute Index

Q Plant\_Quantity i  
 D  
 Demand j  
 S  
 Unit\_shipping\_cost i,j  
 T  
 Shipping\_amount i,j  
 CONSTRAINTS  
 $\sum_i T_{ij} = D_j$   
 $\sum_j T_{ij} = Q_i$

其中 INDICES 代表索引集, VARIABLES AND ATTRIBUTE 以下是属性, 而且只列举了对数学形式有贡献的属性, CONSTRAINTS 是约束条件.

2.3 目标部分

对目标的分析可以采用如图2所示的属性关系图.

Name	Attribute	Goal
Z	Total_shipping_cost	Min

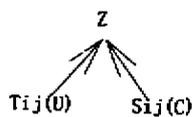


图2 目标的函数关系

根据对目标的分析得到数学关系的计算公式:

$$Z = \sum_i \sum_j T_{ij} S_{ij}$$

目标部分反映事例的最终求解目标, 运输问题目标为 minZ. 因为此种函数关系只是依赖于目标, 不同的目标有可能其目标的函数关系也不同.

将目标与其它部分独立出来的好处是, 目标不同则决策的对象也不同, 同样是上面的问题, 如果有一运输方案现在目标为求总运费 Z, 则目标是计算函数, 即

Calculation Z

但其基本的函数关系仍然一样, 只因为目标不同, 前者求解要用线性规划方法如单纯形法或表上作业法, 后者只是利用公式计算, 目标独立的另一个好处是便于多角度检索.

2.4 事例表示中三部分的关系

<P,F,G> 表示中包括 P 问题、F 函数及约束和 G 目标三个方面, 其目的是用于解决非属性一值事例, 即模型事例的表示问题. 三个部分分别对应事例的三个方面的内容, P 部分采用改造的 ER 图可以反映问题的实质, P 部分的属性是 F、G 部分函数关系表达的基础, F、G 都是 P 部分属性的不同函数表达形成的, 所以三者有密切的联系. 如果模型的 P 部分非常标准, 则完全可以由 P 部分直接得到 F、G 部分, 这种情况是在模型极为标准、典型的情况下才可能. 在实际的建模中, 更多的情况是遇到不能由 P 直接得到的问题. 原因有二点, 首先是实际问题非常复杂多变, 即使 P 部分完全一样, 模型也有可能不同. 其二是决策变量之间可能还有非常复杂的关系, 不可能自然由 P 得到 F 和 G, 可参考文献[8]中的实例. 所以 <P,F,G> 这三方面缺一不可, 实现了事例的表达.

2.5 事例库的组织

事例组织的结构模型如图3所示. 为 ER 图、数学关系和目标分别建立多角度索引, 可以从不同的角度来检索事例库, 其中以 ER 图索引为主要索引机制, 其它为辅助索引机制. 因为 ER 图的不同将对模型起很大的影响. 如在前面的 ER 中令运输的价格作为 U 属性, 因为单位运价可能随运量而递减, 那么这时的问题就转化为一个非线性规划问题. ER 索引可以包含图形的特征, 比如实体的数量, 联系的数量, 变量的位置等. 检索时可以快速找到相似的事例然后可以进一步匹配.

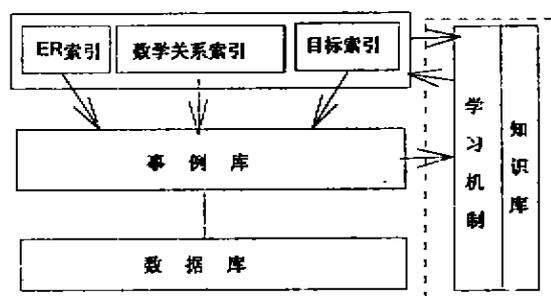


图3 事例库的组织

数学关系索引可以从约束的数量, 公式的线性与否等方面对事例分类. 目标索引可以从不同的目标对事例进行分类. 通过这种多角度索引可以加快检索的过程. 事例的索引中如果出现太多的相

似事例,则还可以通过其他方法帮助索引,如增加知识库,或者增加其他方面的索引。

事例的存储原则是保存典型事例,如果出现的事例结构完全相同,则不保留新事例,否则,对事例评价后将新事例存储起来以备后用。

### 3 结论

事例表示问题是一个基础问题,为了克服原

有系统的不足提出了一种模型事例的表示方法。此种方法可以在使用 CBR 时完全脱离语义。如果新事例的结构和原有事例完全一样,则新问题和原有问题是一致的,即使问题领域有很大差别,此模型实现了数据模型的分离,同时因采用改进的 ER 图,使得模型求解时与数据的结合问题也得到了解决,以后的研究是针对事例的检索匹配和适应的问题。

#### 参 考 文 献:

- [1] Harmon P. Case-based reasoning II [J]. *Intelligent Software Strategies*, 1991, 7(12): 1~8
- [2] Liang T P. Analogical reasoning and case-based learning in model management system [J]. *Decision Support Systems*, 1993, 10(2): 137~160
- [3] Robert T. chi et al. Case based reasoning to model building [C]. *Proceedings of the 26th Hawaii International Conference on System Science*, 324~332
- [4] Jim McGovern. Using case based reasoning for basis development in intelligent decision systems [J]. *European Journal of Operational Research*, 1994, 74: 40~59
- [5] Cunningham, et al. On the use of CBR in optimization problems such as the TSP [C]. *Proceeding of the Case Base Reasoning Workshop*, 1996, 401~410
- [6] Kazuo Miyashita et al. Modeling ill-structured optimization tasks through cases [J]. *Decision Support Systems*, 1996, (17): 345~364
- [7] Geoffrion A M. An introduction to structured modeling [J]. *Management Science*, 1987, 33(5): 574~588
- [8] Zhang Jie, Pan Qishu, Hu Yunquan. DVA, a new method of modeling [J]. *Journal of Harbin Institute of Technology*, (New Series), 1998, 5 (4): 43~46

## Model case representation

ZHANG Jie, PAN Qi-shu, HU Yun-quan

School of Management, Harbin Institute of Technology, Harbin 150001, China

**Abstract:** In case-based model management system the effective presentation of case is the prime thing. Model case is not attribute value case but structural case. In this paper, we put forward a kind of model case presentation method according to the deficiencies of preexisting method. The method can adapt to the numeric value model and independent of the semantics of the model. Then, we plan the preliminary scheme of the construction of the case base. The study found the basis for the succeeding research.

**Key words:** case-based reasoning; model management; intelligent decision support systems