

一个基于 AHP 的 CBR 系统实例检索模型^{*}

屈喜龙, 杜娟, 孙林夫

(西南交通大学 CAD 工程中心, 四川成都 610031)

摘要: 分析了实例检索在 CBR 系统中的重要性, 指出了传统检索方法的不足。借助语意距离来刻画相似实例定性属性间的距离。通过一个具体例子来证明这种模型的有效性和可行性。

关键词: 基于实例的推理; 层次分析法; 实例检索

中图法分类号: TP18

文献标识码: A

文章编号: 1001-3695(2005)04-0033-02

A Model for Cases Indexing in CBR System Based on AHP

QU Xi-long, DU Juan, SUN Lin-fu

(CAD Engineering Center, Southwest Jiaotong University, Chengdu Sichuan 610031, China)

Abstract: Analyzes the importance of cases indexing in the CBR system, and give out the defaults of traditional indexing methods. Uses fuzzy distance to describe the distance between the linguistic attributes in the similar cases. Proves the effective and feasibility of this model by a concrete case.

Key words: CBR (Case-Based Reasoning); AHP (Analytic Hierarchy Process); Cases Indexing

基于实例的推理 (CBR) 是近年来人工智能界一致看好的一种方法, 在处理复杂问题和多属性决策时, CBR 往往是优选方法。基于实例的推理优点很多, 其中最重要的一条是它能模拟人的思维进行问题求解和决策^[1]。

基于实例的推理过程通常包括检索、重用、修正、校阅和系统更新。开发 CBR 系统应紧紧围绕这五步展开。在这五个步骤中, 实例库中有足够的实例是必不可少的, 实例检索是 CBR 系统的关键, 没有有效的实例检索, CBR 系统将一无所成。两个实例间的相似性度量在实例搜索中扮演了重要的角色, 所以 CBR 系统有时也称为相似性搜索系统。传统的 CBR 系统主要使用最近相邻检索法、归纳索引法和知识导引法这三种实例检索策略。这三种实例检索策略比较适合定性属性的实例检索, 而对数量型属性的实例检索, 特别是模糊数量型属性的实例检索, 则显得无能为力。在实际应用中, 还存在着大量的定量属性和混合属性的实例检索问题。本文则引入语义距离来刻画相似实例中两对应属性之间的相似程度, 采用 AHP 方法来描述新问题与各个实例之间的相似程度。该方法具有计算简单和结果准确的优点, 适用于定量属性、定性属性和混合属性的实例检索, 有较强的适应性。

1 实例属性间的距离

要计算实例之间的相似度, 必须考虑组成一个实例的各个属性相似度综合在一起的效应。实例之间的相似度也常常是通过距离来定义的。对于定量属性, 常用的距离定义有^[2]:

$$(1) \text{ 绝对值距离 } d_{ij} = \sum_{k=1}^N |i_k - j_k|$$

其中 i_k 与 j_k 分别表示实例 i 和实例 j 的第 k 个属性值。

$$(2) \text{ 欧氏距离 } d_{ij} = \sqrt{\sum_{k=1}^N (i_k - j_k)^2}$$

其实, 这就是空间几何距离。

$$(3) \text{ 麦克斯基距离 } d_{ij} = \left[\sum_{k=1}^N |i_k - j_k|^q \right]^{1/q}$$

(4) 对于定性属性, 不使用一个简单的定量数来描述, 我们采用语义距离来描述两个定性属性之间的距离^[1]。在语义上, 一个模糊区间数 $[a, b] / CF$ 表示该模糊数落在 $[a, b]$ 中的可能度为 CF , 当 $[a, b]$ 取最大区间时, 满足 $CF=1$ 。一个模糊中心数 $(c, r) / CF$ 表示该模糊数落在以 c 为中心, r 为半径的“超球”之中的可能度为 CF , 当 r 为最大偏差时, $CF=1$ 。论域为实数域时, 两模糊数 $A, B: [a_1, b_1] / CF_1$ 和 $[a_2, b_2] / CF_2$ 之间的语义距离定义为

$$SD(A, B) = (w_a |a_1 - a_2|^u + w_b |b_1 - b_2|^u + w_{CF} |CF_1 - CF_2|^u)^{1/u}$$

其中 $w_a \geq 0, w_b \geq 0, w_{CF} \geq 0$, 且 $w_a + w_b + w_{CF} = 1; u \geq 1$, 取整数值。

当 $CF_1 = CF_2 = 1$ 时, 若取 $u=1$, 有

$$SD(A, B) = w_a |a_1 - a_2| + w_b |b_1 - b_2| \quad (1)$$

实际上, u 的取值视具体问题而定。此处的模糊算子“ u ”可取普通的乘。

在下面的讨论中, 均假设 $CF=1, u=1$, 当 $CF \neq 1$ 及 $u \neq 1$ 时, 情况类似。

2 实例检索模型

2.1 属性之间相似程度描述

为了使讨论具有普遍性, 下面假设实例的前 m 个为定量属性; 后 $n-m$ 个属性为定性属性, 取值均为模糊区间数。

设需要求解的实例 C_0 为问题实例, 它的 n 个属性为

$$C_0 = (C_{01}, C_{02}, \dots, [d_{0k}, b_{0k}], \dots, [d_{0m}, b_{0m}])$$

不妨设 C_p, C_q 为实例库中的实例, 但 C_p, C_q, C_p, C_q 的 n 属性

有 $C_p = (C_{p1}, C_{p2}, \dots, [d_{pk}, b_{pk}], \dots, [d_{pm}, b_{pm}])$

$C_q = (C_{q1}, C_{q2}, \dots, [d_{qk}, b_{qk}], \dots, [d_{qn}, b_{qn}])$

C_p 与 C_0 的第 j_1 个定性属性之间的语义距离为

$$SD(a_{pj1}, a_{0j1}) = w_a \cdot |d_{pj1} - d_{0j1}| + w_b \cdot |b_{pj1} - b_{0j1}| \quad (2)$$

同样地, C_q 与 C_0 的第 j_1 个定性属性之间的语义距离为

$$SD(a_{qj1}, a_{0j1}) = w_a \cdot |d_{qj1} - d_{0j1}| + w_b \cdot |b_{qj1} - b_{0j1}| \quad (3)$$

2.2 模型的构建

美国著名的运筹学专家 T. L. Satty 于 1977 年提出的层次分析法 (Analytic Hierarchy Process, AHP) 是一种定性和定量相结合的决策方法^[3]。AHP 把复杂的问题分解为各个组成因素, 将这些因素按支配关系分组形成有序的递阶层次结构: 最上层的目标层, 中间的准则层和最底层的方案层, 通过两两比较的方式确定层次中诸因素相对重要性总的顺序。其求解过程如下:

(1) 对于目标层, 以特定的目标为准则, 将准则层的各个准则之间的相对重要性进行两两比较 (表 1), 得到判断矩阵, 记为 R_B 。

表 1 各个准则间两两比较的重要性

语言上的评价	评分值	语言上的评价	评分值
绝对的重要	9	略为重要	3
重要得多	7	一样重要	1
很重要	5	中间值	2, 4, 6, 8

(2) 以准则层的各个属性为目标, 将方案层中的各个方案两两比较, 得到各个准则所对应的判断矩阵:

$$R_C^k = \begin{bmatrix} k & k & k & \dots & k \\ 11 & 12 & 13 & \dots & 1M \\ k & k & k & \dots & k \\ 21 & 22 & 23 & \dots & 2M \\ \dots & \dots & \dots & \dots & \dots \\ k & k & k & \dots & k \\ M1 & M2 & M3 & \dots & MM \end{bmatrix} \quad k = 1, 2, 3, \dots, N \quad (4)$$

式中 N 是准则的个数, M 是方案的个数。

(3) 对每一个矩阵的每一列进行归一化处理。让矩阵中的每个元素除以它所在列的总和, 相当于每个元素在它所在列所占的比重:

$$ij = \frac{ij}{\sum_{i=1}^n ij} \quad (5)$$

式中 ij 为判断矩阵中的元素经过归一化后的值。

然后, 对矩阵中的每一行求和, 就得到该行所对应的准则或方案的相对权值, 求的权值同样进行归一化处理:

$$W_i = \frac{r}{\sum_{j=1}^n r} \quad (6)$$

(4) 根据第 (3) 步的结果, 可以求得准则层对应目标层的权重向量: W_B 与各方案对应每一个准则的权重向量, 即有向量组: W_C^i 。由 W_B 构成的矩阵 R_{WB} 乘以向量组 W_C^i 构成的矩阵 R_{WC} , 最后得到各个方案的综合排序。

本文把实例库中各个实例与需要求解问题实例对应的各个属性之间的距离作为方案层, 实例的各个属性为评价准则层, 从实例库中选择与问题实例最相似的实例为目标层。

3 例子

以企业采购决策为例说明上面方法的应用。设有采购管理面对一个新的采购决策, 把这个要解决的决策抽象成一个实例, 通过基于 AHP 的检索方法, 从实例库查出与问题实例相似的实例, 并根据相似度大小给出排序, 决策者可以提取最相似

的实例的求解策略, 作决策时参考。

由于篇幅所限, 为简单起见, 设由需解决的决策抽象而成的实例 C_0 包括如下属性: 商品价格 (B_1)、与公司的距离 (B_2)、数量 (B_3)、供货商声誉 (B_4)、商品质量 (B_5), 前三个为定量属性, 后两个为定性属性。设实例库中有 C_1 、 C_2 、 C_3 和 C_4 共四个实例。其具体数据如表 2 所示。

表 2 各个实例的具体数据表

	B_1	B_2	B_3	B_4	B_5
C_0	150	880	7000	(3, 5)	(7, 8)
C_1	120	200	5000	(8, 9)	(4, 6)
C_2	220	1100	1500	(5, 7)	(3, 6)
C_3	115	400	6000	(4, 6)	(5, 7)
C_4	300	450	10000	(7, 8)	(8, 9)

设 $SD(0, 1)$, $SD(0, 2)$, $SD(0, 3)$ 与 $SD(0, 4)$ 分别表示问题实例与实例库中的四个实例的距离, 为简单起见, 定量属性间距离采用差的绝对值; 定性属性之间的语义距离利用式 (1), 其中 w_a , w_b 分别取 0.4 与 0.6。实际应用中可以根据情况而定, 原理一样。它们之间的距离如表 3 所示。

表 3 C_0 与实例库中各个实例对应的属性间的距离

	B_1	B_2	B_3	B_4	B_5
$SD(0, 1)$	30	680	2000	4.4	2.4
$SD(0, 2)$	70	220	5500	2.0	2.8
$SD(0, 3)$	35	480	1000	1.0	1.4
$SD(0, 4)$	150	430	3000	3.2	1.0

设准则层各个准则两两之间比较得到判断矩阵为

$$R_B = \begin{bmatrix} 1 & 2 & 3 & 1/5 & 1/7 \\ 1/2 & 1 & 2 & 1/6 & 1/8 \\ 1/3 & 1/2 & 1 & 1/7 & 1/9 \\ 5 & 6 & 7 & 1 & 1/2 \\ 7 & 8 & 9 & 2 & 1 \end{bmatrix}$$

根据式 (5), 归一化处理后, 得到矩阵为

$$R_B = \begin{bmatrix} 0.072 & 0.114 & 0.130 & 0.057 & 0.076 \\ 0.036 & 0.057 & 0.087 & 0.047 & 0.061 \\ 0.024 & 0.290 & 0.043 & 0.041 & 0.059 \\ 0.361 & 0.343 & 0.304 & 0.285 & 0.266 \\ 0.506 & 0.457 & 0.391 & 0.570 & 0.532 \end{bmatrix}$$

根据式 (6), 得到 $W_B = (0.090, 0.059, 0.091, 0.304, 0.456)$

根据表 3, 针对准则 B_1 , 两两比较, 可以得到判断矩阵为

$$R_C^1 = \begin{bmatrix} 1 & 3/7 & 6/7 & 1/5 \\ 7/3 & 1 & 2 & 7/15 \\ 7/6 & 1/2 & 1 & 7/30 \\ 5 & 15/7 & 30/7 & 1 \end{bmatrix}$$

根据同样的计算方法, 得到四个实例对于准则 B_1 的权重向量: $W_C^1 = (0.105, 0.246, 0.123, 0.526)$

同理, 得到四个实例对于准则 B_2, B_3, B_4 和 B_5 的权重向量:

$$W_C^2 = (0.375, 0.122, 0.265, 0.238)$$

$$W_C^3 = (0.174, 0.478, 0.087, 0.261)$$

$$W_C^4 = (0.415, 0.189, 0.094, 0.302)$$

$$W_C^5 = (0.315, 0.368, 0.184, 0.132)$$

最后, 由 W_B 构成的矩阵 R_{WB} 乘以向量组 W_C^i 构成的矩阵 R_{WC} 。得到各个方案对于总的权重向量为 $(0.317, 0.332, 0.146, 0.243)$ 。

与 C_0 最相似的实例是 C_3 , 其次是 C_4, C_2, C_1 。决策者可以提取 C_3 的求解策略来参考, 如果其求解策略仍不能满足决策者的要求, 则可以对其求解策略进行修改, 直到满足要求为止。

交互,使用 JAXR 在 UDDI 注册中心发布服务,查找已注册的服务,并通过 JAX-RPC 来远程调用找到的服务。服务请求者通过 JAXM 组件提供的一种公用消息类型实现与服务提供者之间的交互^[3,5]。利用 Web 服务方式实现功能集成,使得网络化制造信息门户可以用一种公共的 Web 格式再现大量的服务,而企业则可以在这一公共平台上访问不同种类的资源,既满足了网络化制造需要资源集中而实际中企业资源分布性的难题,又解决了企业保持独立性与参与网络化制造的协同性的问题。

4 系统运行实例

以主题方式数据集成成为关键技术的 EAI 系统在浙江块状区域经济网络化制造系统中已经实现,系统运行稳定,为信息发布、知识共享、协同设计与制造、供应链管理等服务提供集成功能,充分体现了快速开发、成本降低、操作方便、开放性和通用性强等优势。系统采用目前较成熟的集成技术,以 XML 方式实现数据格式的转换,以 JMS 消息服务的方式实现系统之间的通信,以 Web 服务的方式实现企业之间的功能协同。

EAI 系统数据集成部分采用友好的图形界面,用户可以在服务器端方便地完成与公有主题相关的操作,以及主题库的管理;在客户端完成与私有主题相关的操作,并通过公有主题与私有主题之间的映射,如图 6 所示,实现主题信息的交互和数据共享。



图 6 EAI 系统数据集成客户端主题映射界面

EAI 系统功能集成部分以 Web 方式为用户提供服务,用户可以在 SOAP 客户端实现网上服务注册,服务查询以及获取服务等;在服务器端可以动态添加通用服务,如图 7 所示,并提供服务分类检索、服务注册管理等功能,具有良好的可扩展性。

5 总结与展望

区域经济和中小企业在经过高速发展后,已经具有了一定

的经济实力,但缺乏创新能力使经济发展进入了瓶颈阶段,而网络经济的到来,为这些经济实体带来了新的发展机遇。企业之间的协作日趋频繁,共享数据和信息的快速增长,使得网络化制造过程中的 EAI 技术得到了迅速发展。



图 7 EAI 系统功能集成服务器端服务发布界面

本文主要论述了网络化制造环境中 EAI 系统的一种解决方案。采用主题方式进行数据集成,解决异构数据源下数据共享和信息交互的问题,通过逻辑层与数据层的分离,减少了用户直接对数据源的操作,在一定程度上确保了数据的安全性;采用 Web 服务的方式实现功能集成,通过服务的封装、注册和获取,解决异构应用系统之间的协同工作问题。EAI 系统的整体开发采用了目前较成熟的技术和规范,通过实际应用验证了系统的可靠性和稳定性,是一个企业内部以及企业之间应用集成的较好地解决方案。然而,网络化制造是一个与时俱进的课题,网络化制造及其应用环境下的 EAI 技术也会随着应用需求的不断扩展,而被赋予新的使命。

参考文献:

[1] 范玉顺. 网络化制造的内涵与关键技术问题[J]. 计算机集成制造系统——CIMS, 2003, 9(7): 576-582.
 [2] 范玉顺, 刘飞, 祁国宁. 网络化制造系统及其应用实践[M]. 北京: 机械工业出版社, 2003.
 [3] Matjaz, B Juric, et al. J2EE EAI 编程指南[M]. 袁然, 汤代禄, 刘立君, 等. 北京: 电子工业出版社, 2002.
 [4] Willian A Ruh, Francis X Maginnis, et al. 企业应用集成[M]. 张博, 杨丽君, 等. 北京: 机械工业出版社, 2003.
 [5] 飞思科技产品研发中心. Java Web 服务应用开发详解[M]. 北京: 电子工业出版社, 2002.
 [6] 蔡铭, 林兰芬, 等. 网络化制造环境中制造资源的智能发现技术研究[J]. 计算机集成制造系统——CIMS, 2003, 9(7): 589-594.

作者简介:

马斌(1973-),男,新疆乌鲁木齐人,硕士研究生,主要研究方向为网络化制造、CIMS等;蔡铭(1973-),男,讲师,博士,主要研究方向为语义网、网络化制造、CIMS、AI等;林兰芬(1969-),女,副教授,主要研究方向为网络化制造、数字化建模、CIMS等;董金祥(1945-),男,教授,博士生导师,主要研究方向为计算机图形学、工程数据库、CIMS和 AI等。

(上接第 34 页)

4 结束语

基于 AHP 的 CBR 系统的实例检索模型是有效而且可行的。并且, AHP 方法本身是一种很成熟的方法,经过国内外许多专家学者的研究,取得了许多成果,如模糊层次分析法^[4]。因此,这种模型是可以借助建立在 AHP 基础上其他成果进一步完善。

参考文献:

[1] 张本生,于永利. CBR 系统案例搜索中的混合相似性度量方法

[J]. 系统工程理论与实践, 2002, (3): 131-136.
 [2] 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2002. 85-110.
 [3] 钟诗胜,王知行,等. 一个混合实例检索模型[J]. 软件学报, 1999, (10): 521-526.
 [4] 李培根,张洁. 敏捷化智能制造系统的重构与控制[M]. 北京: 机械工业出版社, 2003. 58-90.
 [5] 常大永,张丽丽. 经济管理数学模型[M]. 北京: 北京经济管理学院出版社, 1995. 146-154.

作者简介:

屈喜龙(1978-),男,湖南新邵人,博士研究生,研究方向为网络化制造系统研究;孙林夫(1963-),男, CAD 工程中心主任,教授,博士生导师,博士,研究方向为网络化制造系统研究。