

基于替代算法的案例推理灾害救助系统^①

Case-Based Reasoning Disaster Salvation System Based on Alternative Arithmetic

武民民^{1,2} 宋良图¹ 黄伟^{1,2} 胡宜敏^{1,2} (1.中国科学院合肥智能机械研究所 安徽 合肥 230031;
2.中国科学技术大学 自动化系 安徽 合肥 230026)

摘要: 为了能够有效的对灾后地区及时的给予物质和资金方面的救助,采用 CBR 推理技术,根据以往历史案例,合理的调拨物质和资金,提出了替代算法这一案例调整算法,基本实现系统最初设计的要求,使用结果表明此系统具有相当高的应用价值。

关键字: 灾害救助 专家系统 案例推理 案例调整 替代算法

1 引言

基于案例推理(Case Based Reasoning, 简称 CBR)是近年来人工智能领域中兴起的一项重要的问题求解和学习的推理技术。案例推理能利用过去经验中的特定知识即具体案例来解决新问题,它通过寻找与之相似的历史案例,把它重新应用到新问题的环境中来。也就是采用检索历史案例,获得当前工况相似特征参数的匹配案例,根据具体情况对匹配案例解决方案进行修整,然后应用于当前情况。

2 工作原理及相关技术

一个典型的 CBR 问题求解过程的基本步骤可以归纳为四个主要过程:案例检索(Retrieve)、案例重用(Reuse)、案例修正(Revise)和案例保存(Retain),因此 CBR 亦称为 4R,其工作过程如图 1^[1]。在 CBR 中,通常把待解决的问题或工况称为目标案例(Target Case),把历史案例称为源案例检索,得到与目标案例相类似的源案例,由此获得,源案例的集合称为案例库(Base Case)。从图 1 中可以了解 CBR 解决问题的基本过程为:一个待解决的新问题出现,这个就是目标案例;利用目标案例的描述信息查询过去相似的案例,即对案例库进行检索,得到与目标案例相类似的源案例,由此获得对新问题的一些解决方案;如果这个解答方案失败将对其进行调整,以获得一个能保存

的成功案例。这个过程结束后,可以获得目标案例的较完整的解决方案;若源案例未能给出正确合适的解,则通过案例修正并保存可以获得一个新的源案例。

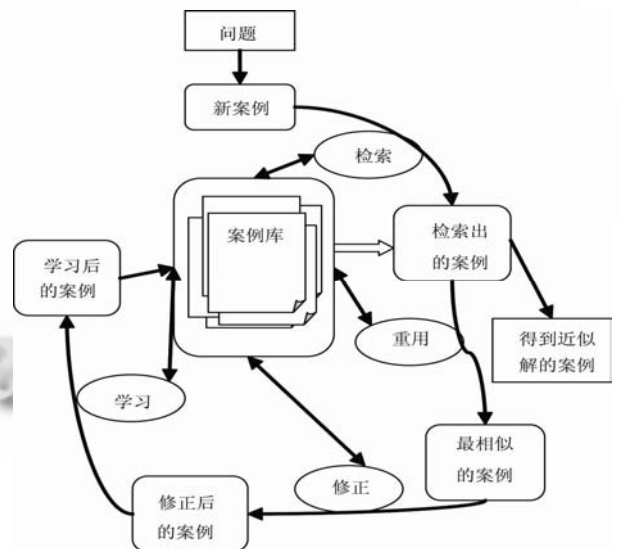


图 1 案例推理周期示意图

在案例推理过程中,案例表示、案例检索和案例调整是案例推理研究的核心问题。绝大多数现有的案例推理系统基本上都是案例检索和案例重用的系统,本系统我们将采用替代算法来实现对案例的调整。

定义 1^[2] 四元组 $S=(U,A,V,f)$ 是一个知识表达系统,其中, U 为一非空有限对象的集合,称为论域;

① 基金项目:国家 863 计划项目(2006AA10Z23702);国家科技支撑计划项目(2006BAD10A1410)

收稿时间:2008-09-12

$A=C \cup D, C \cap D = \Phi$, C 称为条件属性集, D 为决策属性集; $V= \cup V_a, a \in V_a, V_a$ 是属性 a 的值域; f 是 $U \times A \rightarrow V$ 的一个函数。

定义 2. 解元素 d 的非冗余属性空间 $\Omega_d = \overline{\Omega_n}$, 其中 $\Omega_d = \{a \in C \mid \text{Surpc}(a,d)\}$ 是解元素 d 的冗余空间。解元素 d 的参数 $\text{Para}(d)$ 的非冗余属性空间 $\Omega_{dp} = \overline{\Omega_{nsp}}$, 其中, $\Omega_{nsp} = \{a \in C \mid \text{Surpc}(a, \text{Para}(d))\}$ 是 d 的参数的冗余空间。

定义 3. 差异属性集 $\text{DiffAttr}(e,t) = \{a_e \in C_e \mid a_i \neq a_e, a_i \in C_i, e \in E, t \in T\}$, 其中 C_e 为案例属性集, C_i 为目标问题属性集, a_e 为案例 e 的属性, a_i 为目标问题 t 的属性。

定义 4. 设有解元素 $d \in D$, 解元素 d 的激活函数 $d = f(a(1), a(2) \dots a(k)) = d_u, u=0, 1, 2 \dots r$, 当且仅当 $\forall a^{(j)} \in A_u^{(j)}, j=1, 2 \dots k$ 。其中, k 是 Ω_d 中案例属性的个数, r 提该激活函数所能激活的解元素总数。 $A_u^{(j)}$ 为自变量 $a(j)$ 关于解元素 d 激活的无效范围, $A_u^{(j)}, u=1, 2 \dots r$, 为自变量 $a(j)$ 关于解元素 d_u 激活的有效范围。

定理. 设解元素 d 的某一参数 $\text{Para}(d)$ 满足 $\text{Para}(d) = f(a(1), a(2) \dots a(k))$; 其中 $a(j) \in \Omega_{dp}$ 由, $j=1, 2 \dots m$ 。若 d 被激活且 $\Omega_{dp} \cap \text{DiffAttr}(e,t) \neq \Phi$, 则需用替代调整实现解元素 d 的调整。^[3]

证明: 采用反证法。假设当解元素 d 被激活且 $\Omega_{dp} \cap \text{DiffAttr}(e,t) \neq \Phi$, 一定 $\exists a(i)$ 即在参数函数的自变量中, 又在差异属性集中, 而此时不采用替代调整也能实现关于解元素 d 的调整。即该属性对 $\text{Para}(d) = f(a(1), a(2) \dots a(k))$ 不起任何作用, 这与该属性是关于解元素 d 的参数非冗余相矛盾。

证毕。

3 系统结构

3.1 基于 CBR 的灾害救助系统包括:

案例库的管理: 将完成案例库的添加, 维护等管理功能。

规则库的管理: 将完成规则的输入, 添加, 修改等管理功能。

灾害救助: 在 GIS 的支持下, 将接收到的用户的输入条件分析处理后, 找到相关的案例, 完成所需救灾物质和救灾资金的科学推断。

3.2 系统实现中解决的关键问题

3.2.1 案例表示

案例表示要解决两个问题: 案例的属性定义即决

定用哪些特征信息描述案例; 案例的表示形式即确定一种适当的知识表示结构来描述案例内容。

案例的属性定义:

灾害的案例系统保存历史上已发生过的灾害案例, 将灾种、灾害的发生时间、发生地点、死亡人数等作为案例的属性。上述环境因素作为案例的属性描述, 考虑到环境状态因素的多样性及系统的通用性, 因此系统在初始化中也允许用户修改这些特征属性来更精确的描述案例。

案例的表示形式:

本系统采用属性值的方法来表示案例。设范例库中有 K 个范例: $CB = \{c_1, c_2, \dots, c_k\}$, 每个范例 c_i 有 N 个属性 $S = \{s_1, s_2, \dots, s_n\}$, 案例 $c_i = \{\text{CaseID}, (s_1, \text{value}_1), (s_2, \text{value}_2), \dots, (s_n, \text{value}_n)\}$, 其中 CaseID 代表案例号, 本系统用灾害发生日期作为案例号, s_i 代表属性名, value_i 代表属性取值, 取值类型分二类, 定性类型: 取值为具体的灾种, 如旱灾, 台风等; 数值类型: 取值为精确的数值, 如倒塌房屋数等; 权值的确定:

影响灾后物质和资金调拨的因素有几种, 每种所起的作用不同, 因此采用相对比较法确定各属性权重。将所有属性分别按行和列, 构成一个矩阵, 根据三级比例标度, 属性两两比较进行评分, 并记在矩阵相应位置, 再将各属性评分值按行求和, 得到各属性评分总和。最后, 进行归一化处理, 求得各属性的权重值^[4]。设有 n 个属性 $x_1, x_2, x_3, \dots, x_n$ 对前提条件进行描述, 按三级比例标度两两比较评分, 其分值设为 a_{ij} 。三级比例标度的含义是:

$$a_{ij} = \begin{cases} 1, & x_i \text{ 当 } x_i \text{ 比 } x_j \text{ 重要时;} \\ 0.5, & \text{当 } x_i \text{ 与 } x_j \text{ 同样重要时;} \\ 0, & \text{当 } x_i \text{ 比 } x_j \text{ 不重要时。} \end{cases}$$

评分值构成矩阵 $A = (a_{ij})_{m \times n}$ 。显然, $a_{ii} = 0.5, a_{ij} + a_{ji} = 1$, 属性 x_i 的权重值 w_i 为:

$$w_i = \frac{\sum_{j=1}^n a_{ij}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}}, (i = 1, 2, \dots, n) \quad (1)$$

3.2.2 灾后救助程序设计

救助程序分两部分: 案例检索和案例调整。案例检索找出与本次灾害属性最相近一个案例, 再根据此案例的调拨物质和资金给出此次灾害的相应处理方案。

最邻近算法^[5]是目前应用较多的案例检索算法, 应用前提是: 第一, 假定案例的属性之间是相互独立的, 全局的相似度是由属性的局部相似度加权平均计算求

出,第二,属性的重要性(权重)是通过知识工程师或机器学习技术从专家那里获得的。该算法的应用难题是权重系数的确定,以及案例库较大时检索效率会下降。

因为灾害案例的属性之间的独立性比较强。权重通过专家评分计算得到,由于不同地区的案例存在不同案例库中,因此避免了案例库过大的问题而产生的效率问题。本系统选用最邻近算法进行案例检索。具体预测程序流程如下:

step1: 数据预处理

通过输入新案例的属性值,形成待求目标案例:

Goalcase{(灾害种类, value),(受灾人数, value),(死亡人数, value),(转移人数, value),(倒塌房屋数, value),(受损房屋数, value)};

step2: 案例的局部相似性度量

将目标案例和案例库中的源案例相对应的属性分别进行相似度比较。

◆定性描述类型:直接匹配,按匹配与否计算相似度是 0 或 1;

◆数值描述类型:设 s, g 分别表示源案例和目标案例的对应属性值,属性相似度为:

$$sim(s, g) = \begin{cases} 0 < g \leq 0.5s \text{ 或 } g \geq 1.5s \\ -4(g - 0.5s)(g - 1.5s) / s^2, \text{ 其他} \end{cases} \quad (2)$$

step3:用最邻近算法对案例的所有属性计算综合相似度

$$Sim(C_i) = \sum_{j=1}^m \omega_j Sim(C_{ij}) \quad (3)$$

其中, $Sim(C_i)$ 表示案例库中第 i 个案例与目标案例的综合相似度, w_j 为第 j 个属性的权重, $Sim(C_{ij})$ 表示第 i 个事例的第 j 个属性与目标案例的第 j 个属性的相似度,即局部相似度,已由 step2 中得出。

step4: 计算相似案例的个数

通过以上检索,可以得到入选案例库集合中所有案例与待求问题案例的相似度,系统将根据专家给定的一个“阈值”与这些相似度进行比较,凡相似度超过“阈值”的案例将被认定为相似案例。统计检索到的相似案例个数 CaseCount。

step5: 救助系统将用检索算法得到相似案例的个数除以检索的年的个数,求得一个平均的案例发生个数 avercount。这个个数其实反应了在待求问题的条件下每年案例发生的可能性,个数越多案例发生的可能性越大。使用基于规则的推理给出灾害预警级别。规则集由专家给出,放在规则库中,规则库的管理由相

应的管理模块完成。通过规则推理,即可确定灾害的预警级别,从而做出相应的对策。

3.2.3 替代算法实现

step1: 根据 DiffAttr(e,t)查“解元素参数对照表”,确定影响待调整解元素 d 的参数的属性集;

step2: 根据定理判断是否满足替代调整的条件,若满足则转 step3,否则停止;

step3: 调用函数 $Para(d)=f(a(1),a(2)\cdots a(m))$,输入目标问题的属性值,计算目标问题的 $Para(d)$;

step4: 将新计算出来的 $Para(d)$ 替代原案例的 $Para(d)$ 作为目标问题中解元素 d 的参数值。

4 试验结果与总结

在试验中,我们以替代算法和基于相似度为权重的系统进行比较,以系统的运行时间和所拨资金与实际所拨资金的偏差作为比较点,我们得出如下表:

表 1 系统比较表

	运行时间	偏差
基于替代算法	156s	0.18
基于相似度	186s	0.53

通过上表我们可以看出,此系统较以前开发的系统从时间上略有提升,但在准确度上有大幅度改进。

本系统能够有效的实现对灾后所需物质和资金的预测,满足实际应用的需要。但在本系统还有某些方面值得继续探讨,例如,在计算相似度方面我们使用的是二次函数,是否可用正态分布函数替代,以更好查找相似的案例;在案例检索时,我们使用的是最近相邻法,此方法亦有其致命之处,当案例数量比较多时,算法复杂度将成指数增长,这时我们应该考虑归纳索引法来进行解决,这些都值得我们在以后的进行探讨。

参考文献

- 1 曲明,郝红卫.CBR 与 RBR 相结合的实时专家系统设计与实现.计算机工程,2004,30(18):144-146.
- 2 张文修.粗糙集理论与方法.北京:科学出版社,2001:198-199.
- 3 常春光,崔建江,汪定伟,胡琨元.案例推理中案例调整技术的研究.系统仿真学报,2004,16(6):1262.
- 4 Craw S, Wiratunga NRC, Rowe R C. Learning adaptation knowledge to improve case-based reasoning. Artificial Intelligence, 2006,170:1180-1182.
- 5 Lee M. A study of an automatic learning model of adaptation knowledge for case base reasoning. Information Sciences, 2003,155:65-69.