

# 应急辅助决策中案例表示与检索方法研究

仲秋雁\*, 郭素, 叶鑫, 张英菊

(大连理工大学 管理与经济学部, 辽宁 大连 116024)

**摘要:** 针对突发事件发生后如何进行应急决策的问题, 基于应急案例的层次结构提出了应急案例的 XML 表示方法, 解决了异构应急案例的统一表示问题. 在案例检索算法中, 设计了一种基于结构相似度和属性相似度的双重案例检索算法, 有效克服了传统最近邻算法中由于案例属性值缺失而无法进行案例相似性计算的缺陷. 台风领域案例检索算法应用过程的实例说明了案例推理在应急辅助决策中的实用性, 为应急指挥者提供了一种科学而有效的辅助决策方法.

**关键词:** 应急辅助决策; 案例推理; XML; 案例检索; 相似度

**中图分类号:** TP39      **文献标志码:** A

## 0 引言

近年来, 突发事件频繁发生, 引起了各国政府的高度重视. 如何在突发事件发生后快速高效地进行应急指挥决策以将灾害的损失降到最小, 是目前人们普遍关注的问题.

案例推理(case-based reasoning, 简称 CBR) 是人工智能领域的一种类比推理方法, 通过检索案例库中与当前问题相类似的案例, 从中选择相似性高的一个或多个进行适当地修改或调整, 从而得到当前问题的求解结果. 案例推理的思想符合人们的思维习惯, 当人们遇到问题时, 总是根据以前的经验和方法, 从成功或失败的经历中总结出当前问题的解决方案. 一个典型的 CBR 过程主要包括 4 个部分<sup>[1]</sup>: 案例检索、案例重用、案例修正、案例保存. 由于案例检索的前提是将案例进行合理表示, 目前有一些学者将案例表示也加到 CBR 过程中作为 CBR 的首要步骤.

近年来随着学者对 CBR 的研究逐步深入, CBR 的应用范围也在不断扩大, 在医疗诊断、法律咨询、企业经营以及故障诊断等领域 CBR 都得到了广泛的应用, 证明了 CBR 的实用性和有效性.

本文将案例推理方法应用于应急辅助决策中,

主要解决案例推理中的两个最关键的问题: 应急案例表示与检索算法设计.

## 1 应急案例的表示方法

应急案例的表示是指把过去发生的各类突发事件保存为应急案例的形式. 合理的案例表示不仅能够使问题表述更加清晰, 也可以使案例推理更加高效、可靠. 目前, 案例有多种表示方式, 如语义网络、框架、XML、决策树、神经网络和面向对象技术等<sup>[2]</sup>.

应急领域案例结构复杂, 不易于描述, 具有显著的非结构化特点. 本文利用 XML 在描述复杂非结构化问题方面所具有的优势, 设计了一种基于 XML 的应急案例表示方法. 一次突发事件被视为一个案例. 每一个案例可以包含若干个元事件(根据突发事件的发生发展过程、性质和机理, 具有共同特征的灾害事件划归为一类元事件), 例如台风元事件、暴雨元事件等. 根据元事件在一定时期内的相对稳定情况, 每一个元事件又可以分为若干个状态, 例如台风元事件可以分为形成、登陆、减弱等状态. 每一状态又有多个特征属性, 需要分别记录特征属性的值. 基于这种层次结构的案例表示方式如图 1 所示.

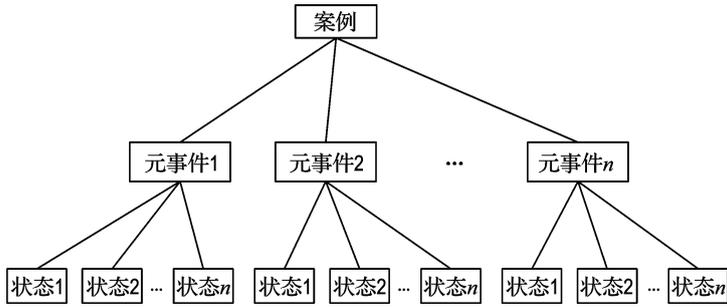


图 1 基于层次结构的案例表示方式

Fig. 1 The method of case description based on the hierarchical structure

根据图 1 所示的层次结构,本文设计出案例的 XML 表示形式,结构示例如下:

```

<? xml version="1.0" encoding="GBK" ?>
<案例 案例编号="001" 案例名称="格美" 死亡人数="
" 受伤人数=" " 失踪人数=" " 直接经济损失=" " 间接经济
损失=" " 其他已造成的后果=" " 评估结果=" " 案例来源="
" 案例编制时间=" " 备注=" " >
  <元事件 元事件编号="01" 元事件名称="台风" 备注="
" >
    <状态 状态 ID="01" 状态名称="状态 1" 发生地点="
福建省平和县境内" 经度="118.2°E" 纬度="
25°N" 发生日期="2006-07-26" 台风等级="热带风
暴(8-9 级)" 中心气压="985 百帕" 移速="15 公里/
小时" 移向="偏西" 风速="13 米/秒" 近中心风力="
9 级" 降雨量="100 毫米" 行动方案="进入四级
应急响应状态" 损失=" " />
    <状态 状态 ID=" " ..... />
    .....
  </元事件>
  <元事件 ..... >
    <状态 状态 ID=" " ..... />
    <状态 状态 ID=" " ..... />
    .....
  </元事件>
  .....
</案例>
  
```

采用 XML 语言来描述案例具有如下优势:

(1) 当一个案例涉及不同的领域并且结构非常复杂时,使用 XML 可以将案例用层次结构清晰地表达出来,便于人们理解案例的复杂内容,为案例的进一步研究工作奠定了基础。

(2) XML 可以充当不同数据源之间的数据交换媒介,实现不同数据源之间的资源共享,使得案

例可以被不同的应用程序处理,完成不同应用实体之间的信息传递<sup>[3]</sup>。

## 2 案例检索算法

一个好的案例检索算法应该能够快速高效地从案例库中检索出与当前问题最相似的案例。在历史案例库与当前问题中,信息可能不够全面,最初关于问题的记录也可能不很准确,这些问题可能会造成某些属性值的缺失。对于传统检索算法——最近邻算法来说,它有个公认缺陷就是无法计算具有缺失属性值的案例之间的相似性。另一方面,只用属性值的相似度来衡量历史案例与当前问题的相似性可能不够全面和准确。为了解决这些问题,本文提出一种基于结构和属性的双重案例检索算法,根据历史案例与当前问题的结构相似度和属性相似度,计算两者的综合相似性,避免了属性值缺失无法进行相似性计算的问题,更全面地反映了当前问题与历史案例之间的相似性,提高了案例检索的效率和准确度,适应应急领域案例形式复杂多变、结构化程度低的特点,取得了较好的应用效果。

应急案例的每一类元事件中,都有一些特征属性来描述它的状态,本文把描述某类元事件状态的特征属性全集称为该类元事件的问题变量。每一个元事件在案例中均用状态来描述其从产生、发展到结束的不同时期的变化情况,而一个状态可以表述突发事件在某一特定时期内的相对稳定的状况,因此,在案例检索中,只要计算一类元事件当前问题状态与历史案例状态的相似度,根据相似度从高到低排列,就可以根据与当前问题状态最相似的历史案例状态的应急处置方案来指导当前突发事件的应急行动,从而达到辅助决策

的目的。

### 2.1 结构相似度的计算

在应急领域各个元事件中，不同的问题变量在决策过程中的重要程度不同，所以要对不同的问题变量分配不同的权重，同一个元事件的问题变量权重之和为1。

由于当前问题状态和历史案例状态的非空特征属性可能不尽相同，首先要计算结构相似度。在某类元事件中，本文将所有问题变量的集合设为  $P$ ，当前问题状态的非空属性集合设为  $Q$ ，历史案例状态的非空属性集合设为  $C$ ，根据决策者认为的当前问题状态和历史案例状态重要程度的不同，结构相似度有两种计算方法：一种方法是将当前问题状态与历史案例状态视为同等重要，一种是以当前问题状态为基准，即认为当前问题状态更重要。两种结构相似度的计算方法如下：

(1)  $sim_s(Q, C) = Q$  和  $C$  交集属性的权重之和 /  $Q$  和  $C$  并集属性的权重之和<sup>[4,5]</sup>

即

$$sim_s(Q, C) = \frac{W_{Q \cap C}}{W_{Q \cup C}} = \frac{\sum_{i=1}^m \tau_{w_i}}{\sum_{j=1}^n \tau_{w_j}} \quad (1)$$

其中  $sim_s(Q, C)$  为  $Q$  与  $C$  的结构相似度， $\tau_{w_i}$  为  $Q \cap C$  中第  $i$  个属性的权重， $\tau_{w_j}$  为  $Q \cup C$  中第  $j$  个属性的权重， $m$  为  $Q \cap C$  中属性的个数， $n$  为  $Q \cup C$  中属性的个数。

(2)  $sim_s(Q, C) = Q$  和  $C$  交集属性的权重之和 /  $Q$  中所有属性的权重之和<sup>[6]</sup>

即

$$sim_s(Q, C) = \frac{W_{Q \cap C}}{W_Q} = \frac{\sum_{i=1}^m \tau_{w_i}}{\sum_{k=1}^l \tau_{w_k}} \quad (2)$$

其中  $\tau_{w_k}$  为  $Q$  中第  $k$  个属性的权重， $l$  为  $Q$  中属性的个数。

在这里，引入缺失度的概念，用来表示当前问题状态中属性值的缺失程度，其计算方法如下：

$$\theta = \frac{W_{P-Q}}{W_P} = 1 - W_Q \quad (3)$$

式中： $\theta$ 为缺失度； $W_{P-Q}$ 为  $P$  与  $Q$  差集属性的权重之和； $W_P$  为  $P$  中属性的权重之和，值为1； $W_Q$  为  $Q$  中属性的权重之和。

### 2.2 属性相似度的计算

应急领域包含了各种各样的突发事件，每种突发事件的特征属性又多种多样，不尽相同，由此本文根据应急领域中特征属性所包含信息内容的不同，将属性划分为4种类型<sup>[7]</sup>，分别为

(1)确定数属性：这种属性值通常是精确的数值。

(2)确定符号属性：这种属性值通常用明确的术语表示，是一种确定性的符号描述。

(3)模糊概念属性：这种属性值可以认为是一个概念变量，所有这样的属性值可构成一个集合。集合中每一属性值对应一个模糊概念，如图2所示。

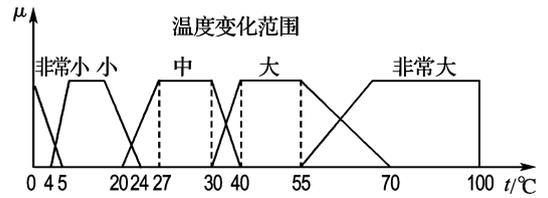


图2 模糊概念属性实例

Fig. 2 An example of fuzzy linguistic attribute

(4)模糊数或模糊区间属性：指一个不确定的数或一个没有确定边界的区间，如图3所示。

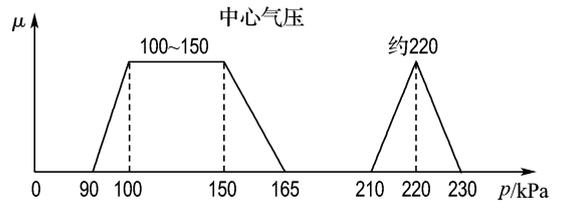


图3 模糊数或模糊区间属性实例

Fig. 3 An example of fuzzy numeric or fuzzy interval attribute

针对上面4种属性类型，本文分别采用了不同的属性相似度计算算法<sup>[7]</sup>。

(1)确定数属性

确定数属性相似度的计算方法有很多，本文采用加权的海明距离反函数的方法计算确定数属性的相似度，计算公式为

$$sim_a(q_i, c_i) = 1 - dist(q_i, c_i) = 1 - |q_i - c_i| / |A_{maxi} - A_{mini}| \quad (4)$$

式中： $sim_a(q_i, c_i)$ 为  $Q$  和  $C$  第  $i$  个确定数属性的相

似度;  $q_i$  和  $c_i$  分别为  $Q$  和  $C$  第  $i$  个确定数属性的值;  $A_{\max i}$  和  $A_{\min i}$  分别为第  $i$  个确定数属性的最大值和最小值.

## (2) 确定符号属性

确定符号属性是用术语表示属性值, 需要比较  $q_i, c_i$  的含义, 相似度计算公式为

$$\text{sim}_a(q_i, c_i) = \begin{cases} 1; & q_i = c_i \text{ 或 } q_i \subset c_i \\ 0; & q_i \neq c_i \end{cases} \quad (5)$$

## (3) 模糊属性

对于模糊概念、模糊数、模糊区间属性, 本文采用文献[7、8]所设计的方法, 应用模糊属性相似度算法(FSM)来计算属性间的相似度. 为计算简便, 本文采用基于梯形的模糊集合来模拟模糊属性, 其形状函数如下:

$$L(x) = R(x) = \max\{0, 1 - x\} \quad (6)$$

所以模糊集  $M$  的隶属函数表示如下:

$$\mu_M(x) = \begin{cases} L\left[\frac{m-x}{p}\right]; & x \leq \underline{m} \\ 1; & \underline{m} \leq x \leq \bar{m} \\ R\left[\frac{x-\bar{m}}{q}\right]; & x \geq \bar{m} \end{cases} \quad (7)$$

式中:  $\underline{m}, \bar{m}, p, q$  为参数. 在三角形模糊集合中,  $\underline{m} = \bar{m}$ .  $p$  和  $q$  的取值随属性的不同而不同. 在模糊概念属性中,  $p$  和  $q$  的值一般由领域专家确定; 对于模糊数或模糊区间属性,  $p$  和  $q$  一般取为  $c \underline{m}$  和  $c \bar{m}$ ,  $c$  的默认值为 0.1.

本文中用相对面积法来计算两个模糊属性间的相似度, 该方法将计算所得两个隶属函数对应面积的重叠率作为模糊集间的相似度, 具有既准确又简单的优点. 计算公式如下:

$$\text{sim}_a(q_i, c_i) = \frac{A(q_i \cap c_i)}{A(q_i) + A(c_i) - A(q_i \cap c_i)} \quad (8)$$

式中:  $A$  代表相应隶属函数与横坐标所围的面积;  $q_i \cap c_i$  为模糊集  $q_i$  和  $c_i$  的交集;  $q_i \cup c_i$  为模糊集  $q_i$  和  $c_i$  的并集.

## 2.3 综合相似度的计算

正如本文前面所述, 一个元事件的各问题变量被分配了不同的权重, 各权重之和为 1. 由于  $Q$  和  $C$  中特征属性可能不同, 在计算相似度时, 只能依赖于它们共有的特征属性, 由此需根据共有特

征属性的权重比例重新分配权重. 根据重新分配的权重及各属性之间的相似度, 可以计算所有属性的加权相似度, 方法如下:

$$\text{sim}_{aw}(Q, C) = \sum_{i=1}^m \left[ \frac{\omega_i}{W_{Q \cap C}} \text{sim}_a(q_i, c_i) \right] \quad (9)$$

式中:  $\text{sim}_{aw}(Q, C)$  为  $Q$  和  $C$  的属性加权相似度,  $m$  为  $Q \cap C$  中属性的个数,  $\omega_i$  为  $Q \cap C$  中第  $i$  个属性的权重,  $W_{Q \cap C}$  为  $Q \cap C$  中属性的权重之和,  $\text{sim}_a(q_i, c_i)$  为  $Q \cap C$  中当前问题状态与历史案例状态第  $i$  个属性的相似度.

本文用综合相似度  $\text{sim}_c(Q, C)$  来表示  $Q$  和  $C$  的相似性, 即当前问题状态与历史案例状态的相似性, 计算公式如下:

$$\text{sim}_c(Q, C) = \text{sim}_s(Q, C) \times \text{sim}_{aw}(Q, C)$$

## 3 实例分析

本文选取两个案例作为示例, 来说明案例检索算法的具体应用方法并检验其有效性.

设历史案例为  $X$ , 当前问题案例为  $Y$ .  $X$  与  $Y$  的 XML 表示形式为

```
<? xml version="1.0" encoding="GBK" ?>
<案例 案例编号="01" 案例名称="X" ..... >
  <元事件 元事件编号="011" 元事件名称="台风" 备注="..." >
    <状态 状态 ID="01" 状态名称="案例 X-台风-状态 1" 发生地点="台湾台东偏东方向的海面" 经度="128.7°E" 纬度="21.7°N" 发生日期="2005-07-19" 台风等级="强热带风暴" 中心气压="984 百帕" 移速="25 公里/小时" 移向="偏西" 风速="28 米/秒" 近中心风力="10 级" 降雨量="50 毫米" 行动方案="进入四级应急响应状态"/>
    .....
  </元事件>
  .....
</案例>
<? xml version="1.0" encoding="GBK" ?>
<案例 案例编号="02" 案例名称="Y" ..... >
  <元事件 元事件编号="021" 元事件名称="台风" 备注="..." >
    <状态 状态 ID="01" 状态名称="案例 Y-台风-状态 1" 发生地点="浙江省杭州以南方向 459 公里" 经度="119.9°E" 纬度="26.3°N" 发生日期="2007-08-06" 台风等级="台风" 中心气压="970 百帕" 移速
```

=“15 公里/小时” 移向=“偏西” 风速=“40 米/秒” 近  
中心风力=“12 级” 降雨量=“” 行动方案=“”/>

.....

</元事件>

.....

</案例>

本文将台风的特征属性分成几种类型(见表 1),分别为确定数(CN)、确定符号(CS)、模糊数(FN)。由于状态名称、发生日期不参加相似度计算,未给定类别和权重。由于应用经度和纬度来描述地点并进行相似度计算,发生地点不再参加计算,未给定类别和权重。所有属性的权重之和为 1。

根据国家标准《热带气旋等级》(GB/T 19201—2006),热带气旋分为热带低压、热带风暴、强热带风暴、台风、强台风和超强台风 6 个等级,本文用 1 至 6 个数字分别代表,以便于相似度计算。

首先计算案例 X 中台风元事件的状态 1 与案例 Y 中台风元事件的状态 1 之间的结构相似度,此示例进行相似度计算时选择结构相似度的第一种方法即将当前问题案例与历史案例视为同

等重要。案例 X-台风-状态 1 与案例 Y-台风-状态 1 中,交集属性的权重之和为 0.9,并集属性的权重之和为 1,因此案例 X-台风-状态 1 与案例 Y-台风-状态 1 之间的结构相似度为  $0.9/1=0.9$ 。然后计算属性加权相似度,用每一个属性的相似度乘以重新分配后的权重再求和,利用式(9)计算:

$$sim_{aw}(Q,C) = \sum_{i=1}^m \left[ \frac{w_i}{W_{Q \cap C}} sim_a(q_i, c_i) \right] = 0.951 \times$$

$0.05/0.9 + 0.949 \times 0.05/0.9 + 0.8 \times 0.1/0.9 + 0.939 \times 0.15/0.9 + 0.667 \times 0.15/0.9 + 1 \times 0.1/0.9 + 0.762 \times 0.15/0.9 + 0.818 \times 0.15/0.9 = 0.837$ , 这样案例 X-台风-状态 1 与案例 Y-台风-状态 1 之间的属性加权相似度为 0.837。最后计算综合相似度为  $0.9 \times 0.837 = 0.753$ 。这样就得出结论:案例 X-台风-状态 1 与案例 Y-台风-状态 1 的综合相似度为 0.753。属性信息及相关计算结果如表 1 所示。决策者如果可以接受案例 X-台风-状态 1 与案例 Y-台风-状态 1 之间的相似度,就可以根据历史案例 X 中台风元事件的状态 1 的应急处置来指导当前问题案例 Y 中台风元事件的状态 1 的应急指挥,必要时对解决方案进行修正,辅助决策者进行应急决策。

表 1 属性信息及相关计算结果

Tab. 1 The information of attributes and the calculation results

属性名称	类别	权重	案例 X 台风某状态的属性值	案例 Y 台风某状态的属性值	属性相似度
状态名称		0	案例 X-台风-状态 1	案例 Y-台风-状态 1	
发生地点		0	台湾台东偏东方向的海面	浙江省杭州以南方向 459 公里	
经度/°E	CN	0.05	128.7	119.9	0.951
纬度/°N	CN	0.05	21.7	26.3	0.949
发生日期		0	2005-07-19	2007-08-06	
台风等级	CN	0.10	3	4	0.800
中心气压/hPa	CN	0.15	984	970	0.939
移速/(km·h <sup>-1</sup> )	CN	0.15	25	15	0.667
移向	CS	0.10	偏西	偏西	1
风速/(m·s <sup>-1</sup> )	CN	0.15	28	40	0.762
近中心风力	CN	0.15	10 级	12 级	0.818
降雨量/mm	FN	0.10	50	—	

### 4 结 论

本文针对应急案例结构复杂的特点,提出了一种基于 XML 的应急案例表示方法,这种方法

具有层次分明、结构清晰、适用于多种类型应急案例表示等优点;设计了基于结构相似度和属性相似度的双重案例检索算法,从而有效避免了传统检索算法的缺陷;最后通过台风的实例说明了案

例检索算法的实现过程. 实例证明本文基于案例推理方法设计的应急辅助决策中的案例表示与检索方法能够准确地从历史案例库中检索出与当前问题相似的案例, 从而有效地辅助决策者根据历史案例的应急指挥调度方案进行当前问题的应急决策.

## 参考文献:

- [1] SHIU S C K, PAL S K. Case-based reasoning: concepts, features and soft computing [J]. *Applied Intelligence*, 2004, **21**(3):233-238
- [2] 周凯波, 魏莹, 冯珊. 基于案例推理的金融危机预警支持系统[J]. *计算机工程与应用*, 2001, **14**(9):18-21
- [3] 李玲娟, 汤文字, 王汝传. 基于 XML 的案例表示和案例库构造方法[J]. *计算机应用研究*, 2007, **24**(11):70-73
- [4] 闫利国, 贺飞. XML 文档结构相似测度研究[J]. *计算机应用研究*, 2006, **23**(3):44-46
- [5] LIAN W, CHEUNG D W L, MAMOULIS N, *et al.* An efficient and scalable algorithm for clustering XML documents by structure [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2004, **16**(1):82-96
- [6] 鲁远, 万常选. XML 信息检索中结构相似度计算[J]. *情报杂志*, 2007, **26**(11):109-111
- [7] 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
- [8] 路云, 吴应宇, 达庆利. 基于案例推理技术的企业经营决策支持模型设计[J]. *中国管理科学*, 2005, **13**(2):81-87

## Research on method of case representation and retrieval in emergency aid decision-making

ZHONG Qiu-yan\*, GUO Su, YE Xin, ZHANG Ying-ju

( Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China )

**Abstract:** Focusing on a discussion about how to make emergency decisions after the occurrence of emergent events, a representation method based on XML is proposed based on the hierarchical structure of emergency cases, which can solve the unified representation problems of the heterogeneous emergency cases. A two-layer case retrieval algorithm combining structure similarity and attribute similarity is designed, which can effectively overcome the shortcomings of traditional nearest neighbor algorithm that fails to calculate the similarities between the cases with the missing values. Finally, the application of case retrieval algorithm to typhoon analysis is used to illustrate the practicality of case-based reasoning (CBR) approach in emergency aid decision-making, which is testified a scientific and effective aid decision-making method for the emergency officials.

**Key words:** emergency aid decision-making; case-based reasoning; XML; case retrieval; similarity degree