一种新的知识表示模型及其规则推导

花文健，刘作良，杨 凡
(空军工程大学 电讯工程学院，陕西 西安 710077)

摘 要:目前指挥决策支持系统中的知识系统大多只能表示清晰信息,当给定对象、属性及属性值,便能完全肯定对象是否具有这个属性(值)。但实际情况并不是这样,对象和属性(值)关系常常是不确定的。本文提出FPS模型解决这种不确定性,并在此基础上,为概念归纳推理定义了模糊粗糙集,为对象的排序定义了相似性测度、距离测度和熵测度,用模糊集方法定义了概念的上下近似集,并得出了决策规则。

关键词:决策支持;知识表示;模糊PS模型;上近似集;下近似集;决策规则
中图分类号:TP311.3　　**文献标识码**:A　　**文章编号**:1009-3516(2003)06-0044-04

指挥决策支持系统是指挥自动化系统的核心,指挥自动化系统效能直接取决于指挥决策支持系统效能的发挥,而构建合理的知识库是指挥决策支持系统的关键环节。当前,指挥决策支持系统的知识库中对对象的描述大多只限于与属性关系的精确描述和属性值的精确取值,而现实的指挥决策过程中,大量的描述是不确定、不精确甚至不完全的[1],这就使得基于精确描述的知识库系统不能很好满足指挥决策支持的要求。因此,需要在知识库中引入一种能描述对象与属性之间不确定性的机制。这种机制允许存在知识表示的不确定性,而决策支持系统应具备利用不确定性进行归纳推理的功能。

传统的知识表示模型使用固定的属性集描述对象,每个对象的描述由1个元组(若干属性——属性值对构成的集合)实现。知识表示采用表的形式,表是由要被表示的对象集(Θ)、属性集(AT)、属性值集合($VAL=\cup_{a\in AT}VAL_a$)及一个信息函数$f:\Theta\times AT\rightarrow VAL$,即一个四元组($\Theta,AT,VAL,f$)构成。这样的表示方式至少从两个方面限制了知识表达的能力。首先,所有的对象必须统一描述,而且,f对于每个属性最多只能取一个值,要求其定义在所有的$\Theta\times AT$元素上。因此,在这样的一个数据库中的元素就完全被限制在由统一的属性集合构成的描述上,所有的属性也必须对所有的元素有意义;所有的描述也必须完全确知,对一个给定的对象,任何属性都不能取多个值。基于这种思想的知识表示模型称为PS模型(Property Sets Model)。模型中,属性——属性值对称为特性(Property),每一个对象都可以用某些特性进行描述。但实际上,并不是所有对象都适用于相同的特性集合来描述。在所考察的论域当中,有些对象明确具有这些特性,另一些对象则不能确定是否完全具备。其次,这种方法描述一个对象时缺乏一种“裕度”——一个对象要么完全具备某种特性,要么根本不具备这种特性,这是不符合事实的。如,以敌机距我防卫区边界的距离r为判断属性,当$r\leq180$ km时,决定出动我机截击,而当$r=185$ km、187 km、189 km这些值时,已经可以完全据此决定出击了。这里,决定实施出击方案A(视为一个对象)的判断根据是与特性“距离=(r,180 km)”有一定的近似性,即已经在一定的程度上符合了“距离”特性。

为了改进上述缺陷,需要建立一种新的知识表示模型。基本思路是在原有的PS模型基础上,引入一个程度因子(在0和1之间取值)来表示对象具备属性——属性值对的程度,这样等于为每一个特性描述每一个对象时指定了一个程度值,这个“扩展”了的模型我们称之为FPS模型(Fuzzy Property Sets)。FPS模型继承采用PS模型中的基本描述单元——原子概念(Atomic Concepts),对象的描述表示为定义在原子概念集上

收稿日期:2003-05-26
基金项目:国家“863”计划基金资助项目(306-04-02-01)
作者简介:花文健(1976-),男,北京人,博士生,主要从事智能决策支持系统研究;
刘作良(1938-),男,四川成都人,教授,博士生导师,主要从事指挥自动化系统设计与分析研究.

的模糊隶属函数[2]。利用模糊对象描述模型,可以定义出任何子集(概念) $A\subseteq\Theta$ 的模糊 Rough 上、下近似集合。文献[3]中提出的关于一个集合 A 的上下 Rough 近似分别表示可能属于 A 的元素的最小集合和肯定属于 A 的元素集合。在此基础上定义了相似性测度,用以测量一个模糊描述的对象与某个给定的概念的上、下近似的接近程度。根据接近程度将对象排序,根据上近似排序的结果可以理解为保守排序,根据下近似排序的结果可理解为非保守排序,最后得到决策规则。

1　PS 模型(Property Sets Model)

根据文献[4]的定义,这里只简要的阐述 PS 模型的基本概念。令 $\Theta=\{\theta_1,\theta_2,\Lambda,\theta_n\}$ 是论域 U 中的一个子集。令 $\Pi=\{\pi_1,\pi_1,\Lambda,\pi_n\}$ 是可能与所有对象或可能与一部分对象有关的所有已知特性集合,其中,$\pi=(a,v)$,$a\in AT,v\in VAL$ 。AT 是属性集合,VAL 是 $|AT|$ 维的属性值空间。属性 a 在这里是特性的类型。如 π =(颜色,迷彩),那么特性 π 的类型 a 是“颜色”。P 为一个二元关系,表示“具备此特性”,$\theta\in\Theta$ 如果能被特性 π 表示,则 $(\theta,\pi)\epsilon P$,简写为 ${}_\theta P_\pi$。

对于任何 $\theta\in\Theta$,与 θ 相关的特性集表示为:$[\theta]=\{\pi\in\Pi\mid(\theta,\pi)\in P\}$ 。

给定一个特性集合Π,每个具备 $\pi\in\Pi$ 的对象集记为:$[\pi]=\{\theta\in U\mid(\theta,\pi)\in P\}$,

不具备此特性的对象集合记为 $[\hat{\pi}]=\{\theta\in U\mid(\theta,\pi)\notin P\}=U-[\pi]$。

由此,可以构造原子概念$[c_i]$和原子概念集合$\{[c_i]\}$:

$$[c_0]=[\hat{\pi}_1]\cap[\hat{\pi}_2]\cap\Lambda\cap[\hat{\pi}_n]$$

$$[c_1]=[\pi_1]\cap[\hat{\pi}_2]\cap\Lambda\cap[\hat{\pi}_n]$$

$$\cdots\qquad\cdots$$

$$[c_N]=[\pi_1]\cap[\pi_2]\cap\Lambda\cap[\pi_n]\ ,\ N=2^n-1。$$

所有原子概念的集合称为概念空间 C。显然,单个原子概念是等价类,在这个等价类中的元素不可辨识。每个对象 $\theta\in\Theta$ 精确的被一个原子概念确切地标示。至多存在着$|\Theta|$ 个非空的原子概念。任何在$[c_i]$中的对象可以用$[c_i]$确定。当给定具备特性$[\theta]$的对象 $\theta\in\Theta$ 时,θ 所属的原子概念 c_θ为

$$c_\theta=(\bigwedge_{\pi_i\in[\theta]}[\pi_i])\wedge(\bigwedge_{\pi_j\notin[\theta]}[\hat{\pi}_j])\ ,\ c_\theta\in C。$$

PS 模型的知识表示方法定义了一个表,表中的取值为 $\psi:\Theta\times\Pi\rightarrow\{0,1\}$ 。以军用机动装备为例说明。假设有:[车型1]={(颜色,迷彩),(速度,快),(气缸数,12)};[车型2]={(速度,快),(气缸数,4),(换档方式,手动)};[车型3]={(颜色,迷彩),(换档方式,自动)},[车型4]={(颜色,迷彩)}。概念为“功能强的车”其中包括{车型1,车型2}。表1所示为 PS 模型,表2所示为 FPS 模型。

表1　PS 模型

对象	π_1	π_2	π_3	π_4	π_5	π_6	决策类
车型1	1	1	1	0	0	0	+
车型2	0	1	0	1	1	0	+
车型3	1	0	0	0	0	1	-
车型4	1	0	0	0	0	0	-

表2　FPS 模型

对象	π_1	π_2	π_3	π_4	π_5	π_6	决策类
车型1	0.1	0.8	0.2	0	0	0	+
车型2	0	0.5	0	1	0.9	0	+
车型3	0.5	0	0	0	0	0.25	-
车型4	1	0	0	0	0	0	-

其中 $\pi_i(i=1,2,\cdots,6)$ 分别表示(颜色,迷彩)、(速度,快)、(气缸数,6)、(气缸数,4)、(换档方式,手动)和(换档方式,自动)。

2　FPS 模型

表1使用完全确定的对象—属性信息描述了对象。存在的问题是:如果对象的描述是模糊的怎么办?据此可以指定一个程度值 α_{ij}表示某对象 θ_i具有特性 π_j 的程度。此时,ψ 映射变为 $\psi':\Theta\times\Pi\rightarrow[0,1]$,$\alpha_{ij}\in[0,1]$ 。表2是一个可能的实例。但我们感兴趣的问题是如何用原子概念去表示每一个对象。

2.1　模糊对象描述

利用模糊集理论,模糊对象的描述视为是以原子概念集 C 为论域定义的模糊集。定义$\{(\beta,c)\}$ 为基本

元素集,$c\in C$,β是c的隶属度,$\mu_B(c)=\beta$定义了一个模糊集$\tilde{B}$。$(1,c)$表示原子概念肯定出现(即概念c发生),$(0,c)$表示原子概念肯定不出现(即概念c不发生)。令$\tilde{\pi}_j$表示π_j的模糊表示,则有:

$$\tilde{\pi}_j=(\bigcup_{c_k\Rightarrow\pi_j}\{(1,c_k)\})\cup(\bigcup_{c_k\Rightarrow\hat{\pi}_j}\{(0,c_k)\})=\bigcup_{k=0}^{N}\{(p_{jk},c_k)\} \tag{1}$$

其中,$c_k\Rightarrow\pi_j$意为所有含有π_j的$[c_k]$等价类,$c_k\Rightarrow\hat{\pi}_j$意为所有含有$\hat{\pi}_j$的$[c_k]$等价类,当$\Rightarrow$关系成立时,$p_{jk}$分别等于1和0。

PS模型中,每个$\theta_i\in\Theta$位于特性集合的交集中,并可视为特性的函数,利用这个思想及式(1)进行拓展,将θ_i的模糊描述表示为

$$\tilde{\theta}_i=\{(\alpha_{i1},\tilde{\pi}_1)\}\cap\{(\alpha_{i2},\tilde{\pi}_2)\}\cap\Lambda\cap\{(\alpha_{in},\tilde{\pi}_n)\}=\bigcap_{j=1}^{n}\{(\alpha_{ij},\bigcup_{k=0}^{N}\{(p_{jk},c_k)\})\} \tag{2}$$

其中α_{ij}为θ_i具有π_j的程度。

为了化简上式,我们在这里设定一种变换。设$\{\alpha,\{(\beta,c)\}\}=\{(\delta,c)\}$,其中$\delta=1-\alpha+\beta(2\alpha-1)$。$\alpha$是有序对集合中表示$c$的否定事件发生的程度。$\alpha=0$时,$\{(\delta,c)\}=\{(1-\beta,c)\}$,概念$c$不发生;$\alpha=1$时,$\{(\delta,c)\}=\{(\beta,c)\}$,概念$c$发生;$\alpha=0.5$时,$\{(\delta,c)\}=\{(0.5,c)\}$,完全不能确定概念$c$是否发生。基于这样的变换,在功能上保证了$\alpha$与$\beta$的一致性。故而有:

$$\{(\alpha,\bigcup_{k=0}^{N}\{(\beta_k,c_k)\})\}=\bigcup_{k=0}^{N}\{\alpha,\{(\beta_k,c_k)\}\}=\bigcup_{k=0}^{N}\{(\delta_k,c_k)\},\text{其中}\ \delta_k=1-\alpha+\beta_k(2\alpha-1) \tag{3}$$

将式(3)代入式(2)可以化简得到$\tilde{\theta}_i$的描述(其中,$\delta_{ijk}=1-\alpha_{ij}+p_{jk}(2\alpha_{ij}-1)$):

$$\tilde{\theta}_i=\bigcap_{j=1}^{n}(\bigcup_{k=0}^{N}\{(\delta_{ijk},c_k)\})=\bigcup_{k=0}^{N}(\bigcap_{j=1}^{n}\{(\delta_{ijk},c_k)\}=\bigcup_{k=0}^{N}\{(\min_j\{\delta_{ijk}\},c_k)\} \tag{4}$$

式(4)是在原子概念集合上对每一个对象的模糊表示。这种描述就是将对象在原子概念集合上表示为一个模糊集合,每个对象$\theta_i\in\Theta$都可用隶属函数$\mu_{\theta_i}(c_k)=\min_j\{\delta_{ijk}\}$描述。这种模糊表示方法是PS模型表示的扩展。当$c_i=c_\theta$时,$\mu_\theta(c_i)=1$,其他情况时,$\mu_\theta(c_i)=0$。

2.2 概念的上近似和下近似

根据文献[5]提供的方法,我们定义集合A的上近似为$\bar{A}=\bigcup_{\theta\in A}\tilde{\theta}$,根据式(4),有

$$\begin{aligned}\bar{A}&=\bigcup_{\theta_i\in A}(\bigcup_{k=0}^{N}\{(\min_j\{\delta_{ijk}\},c_k)\})=\\&\bigcup_{\theta_i\in A}\{(\min_j\{\delta_{ij0}\},c_0),(\min_j\{\delta_{ij1}\},c_1),\Lambda,(\min_j\{\delta_{ijN}\},c_N)\}=\\&\bigcup_{\theta_i\in A}\{(\zeta_{i0},c_0),(\zeta_{i1},c_1),\Lambda(\zeta_{iN},c_N)\}\qquad\text{式中},\ \zeta_{ik}=\min_j\{\delta_{ijk}\}\end{aligned}$$

根据上近似的定义[4],我们可得到:$\bar{A}=\bigcup_{k=0}^{N}\{(\max_{\theta_i\in A}\{\zeta_{ik}\},c_k)\}$。

由于下近似$\underline{A}$满足:$\underline{A}=\bar{A}-\overline{\Theta-A}=\bar{A}\cap-(\overline{\Theta-A})=(\bigcup_{\theta\in A}\tilde{\theta})\cap-(\bigcup_{\theta\in(\Theta-A)}\tilde{\theta})$,故:

$$\begin{aligned}\bar{A}&=(\bigcup_{k=0}^{N}\{(\max_{\theta_i\in A}\{\zeta_{ik}\},c_k)\})\cap(-\bigcup_{k=0}^{N}\{(\max_{\theta_i\in(\Theta-A)}\{\zeta_{ik}\},c_k)\})=\\&\bigcup_{k=0}^{N}\{(\min(\max_{\theta_i\in A}\{\zeta_{ik}\},1-\max_{\theta_i\in(\Theta-A)}\{\zeta_{ik}\}),c_k)\}。\end{aligned}$$

至此,已经得到了对于任何概念$A(A\subseteq\Theta)$上下近似集合的模糊表示方法。

2.3 对象排序

对象$\theta_i\in\Theta$是否属于A的范围内是运用知识表示系统解决现实问题的重要内容。基于经典集合论的分类识别方法,考察θ_i的外延(论域上的子集)是否是给定概念的外延(也是论域上的子集)的子集;模糊集理论的方法考察$\tilde{\theta}$外延的隶属函数与概念A外延的特征函数之间的关系,从而确定出在多大的程度上$\tilde{\theta}$包含于A或不包含于A。上述两种思路均可归结为考察对象与已知概念的相似性,利用这一启示,我们考虑应用上节表示方式中所蕴含的知识建立一个比较简单直观的方法解决问题。为此,定义如下3个测度用以测量$\tilde{\theta}$与A之间的相似性:

相关性测度:$$S_{BC}=\frac{\mu_B(x_1)\mu_C(x_1)+\Lambda+\mu_B(x_M)\mu_C(x_M)}{\sqrt{(\mu_B{}^2(x_1)+\Lambda\,\mu_B^2(x_M))(\mu_C{}^2(x_1)+\Lambda\,\mu_C^2(x_M))}}$$

距离测度:$D_{BC} = |\mu_B(x_1) - \mu_C(x_1)| + \Lambda + |\mu_B(x_M) - \mu_C(x_M)|$

熵测度[6]:$E_{BC} = 1 - \gamma$,其中:$\gamma = H(\frac{1}{2}B + \frac{1}{2}C) - \frac{1}{2}(H(B) + H(C))$

$$H(x) = -\sum_i \mu_X(x_i)\log \mu_X(x_i)$$

$$\mu_{\frac{1}{2}B}(x) = \frac{1}{2}\mu_B(x)\ ,\ \mu_{B+C}(x) = \mu_B(x) + \mu_C(x)$$

当给定一个概念 A 及其上近似集 $\overline{A}$和下近似集$\underline{A}$,运用这3种测度中的一种可得到未知对象集合与上、下近似集合的相似度排序,与上近似集的相似度排序可以认为是保守排序,与下近似的的相似度排序可以认为是非保守排序。

3 结束语

FPS模型完善了PS模型,使得对每一对象都可以描述为原子概念集上的模糊集,使得系统具备了描述特性与对象间的不确定关系。利用模糊集定义了上、下近似集合,结合定义在两个模糊集合上的相似性测度得出保守决策规则和非保守决策规则。从总体上来说,FPS较之PS模型,系统的表达能力有了很大的提高。

参考文献:

[1] 刘作良,许锦州,江光杰.指挥自动化系统[M].北京:解放军出版社,2001.

[2] Zadeh L A. Fuzzy Sets[J]. Information and Control,1965,8:338 - 353.

[3] Pawlak Z. Rough Sets - Theoretical Aspects of Reasoning About Data[M]. Dordrent: Kluwer Academic Publisher,1991.

[4] Wojciech P. Ziarko(Ed.) Rough Sets,Fuzzy Sets and Knowledge Discovery[M]. London:Springer - Verlag Publisher,1994.

[5] Dubois D Prade H. Rough Fuzzy Sets and Fuzzy Rough Sets[J]. International Journal of General Systems , 1990,17:191 - 209.

[6] Wong S K M ,Yao Y. A Statistical Similarity Measure[A]. in Proc:10th Annual International ACM SIGIR Conference on Reseach and Development in International Retrieval[C]. New Orleans:Springer - Verlag Publisher,1987.3 - 12.

[7] 张明智,娄寿春,何章明.指挥控制系统决策支持需求研究[J].空军工程大学学报(自然科学版),2001,2(3):22 - 25.

(编辑:门向生)

A Novel Knowledge Representation and Induction Decision Rules

HUA Wen - jian, LIU Zuo - liang, YANG Fan

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

Abstract:The FPS model is put forward to remedy the problem in traditional knowledge representation. Based on this, a fuzzy rough set is defined for concept inductive inference, three measures, i.e. similarity measure, distance measure and entropy measure are defined for the object rank, then upper and lower approximations of concept are defined by using this fuzzy set method , and decision rules are achieved.

Key words:decision support; knowledge representation; FPS model; upper approximation; lower approximation; decision rules