

基于可拓集的决策模型研究^{*}

王洪伟 吴家春 蒋 馥

(上海交通大学管理学院智能化管理与计算机仿真应用研究中心 上海200052)

Study on Decision Model Based on Extension Set

WANG Hong-Wei WU Jia-Chun JIANG Fu

(Research Center of Intelligent Management & Computer Simulation Application, Shanghai Jiao Tong University, Shanghai, 200052)

Abstract Decision model for incompatible problems is a difficult issue in knowledge-based systems, which is currently being focused by the artificial intelligence and management communities. Extension set is an emerging set theory for solving incompatible problem systems. Using extension set theory, this paper presents the formal conceptions of strategy and key strategy, establishes their matter-element models, and gives the strategy development method.

Keywords Decision model, Matter-element, Extension set

1 前言

决策过程的形式化是基于知识系统中策略自动生成首要解决的问题,对提高决策系统的科学性、可操作性起到重要的作用,尤其是不相容系统的决策模型研究,近年来更是深受人工智能、管理学界的关注,但目前仍是一个难点问题^[1]。

可拓集是近年来提出的一种解决不相容问题的集合理论,可拓集以物元为基本元素,用关联函数分析基本元素(决策对象各子系统间)的矛盾性,用物元变换化矛盾问题为相容问题^[2]。本文利用物元可拓集理论对决策模型中策略集、关键策略集以及策略的生成过程提供了形式化的描述与解决方法,这些工作对基于知识系统的问题求解的建模具有一定的参考价值。

2 物元与可拓集理论

定义2.1(物元) 以有序的三元组 $R=(N, c, v)$ 为描述事物的基本元,简称为物元。其中, N 表示事物, c 表示特征, v 表示 N 关于 c 所取的量值,这三者称为物元的三要素。

定义2.2(物元可拓性) 物元具有发散性、共轭性、相关性、蕴含性和可扩性,物元这5种基本特性被统称为物元的可拓性。物元可拓性从事物向外、向内、平行、变通和组合分解的角度提供事物拓展的多种可能性,从而为创造性思维和解决矛盾问题的解决提供依据。对物元可拓性的详细论述请见文^[2,5]。

定义2.3(物元变换) 物元 $R_0=(N_0, c_0, v_0)$ 改变为物元 $R=(N, c, v)$ 或若干物元 $R_1=(N_1, c_1, v_1), R_2=(N_2, c_2, v_2), \dots, R_n=(N_n, c_n, v_n)$, 称为物元 R_0 的变换。物元变换具有置换、分解、增删、扩缩四种基本类型和积、逆、或、与四种基本运算。

定义2.4(可拓集) 设 U 为论域(物元集), k 是 U 到实域 I 的一个映射, $T=(T_U, T_k, T_v)$ 为给定的物元变换, 称

$$\tilde{A}(T)=\{(u, y, y') \mid u \in T_U U, y=k(u) \in I, y'=T_k k(T_v u) \in I\}$$

为论域 $T_U U$ 上的一个可拓集, $y=k(u)$ 为 $\tilde{A}(T)$ 的关联函数,

$y'=T_k k(T_v u)$ 为 $\tilde{A}(T)$ 的可拓函数, 其中 $T_U U$ 表示对论域 U 的变换, T_k 为对关联函数 k 的变换, T_v 为对关联函数 u 的变换。

(1) 当 $T_U=e, T_k=e, T_v=e$ 时, $\tilde{A}(T)=\tilde{A}(e)$, 记 $\tilde{A}(e)=\tilde{A}=\{(u, y) \mid u \in U, y=k(u) \in I\}$, e 为幺变换, 并称

$A=\{(u, y) \mid u \in U, y=k(u) \geq 0\}$ 为 $\tilde{A}$ 的正域;

$\bar{A}=\{(u, y) \mid u \in U, y=k(u) \leq 0\}$ 为 $\tilde{A}$ 的负域;

$J_0=\{(u, y) \mid u \in U, y=k(u)=0\}$ 为 $\tilde{A}$ 的零界。

(2) 当 $T_U=e, T_k=e$ 时, $T_U U=U, T_k k=k$, 这时 $\tilde{A}(T)=\tilde{A}(T_v)$, 此可拓集为关于元素 u 变换的可拓集, 并称

$A_+(T)=\{(u, y, y') \mid u \in U, y=k(u) \leq 0, y'=k(T_v u) \geq 0\}$ 为 $\tilde{A}(T)$ 的正可拓域;

$A_-(T)=\{(u, y, y') \mid u \in U, y=k(u) \geq 0, y'=k(T_v u) \leq 0\}$ 为 $\tilde{A}(T)$ 的负可拓域;

$A_+(T)=\{(u, y, y') \mid u \in U, y=k(u) \geq 0, y'=k(T_v u) \geq 0\}$ 为 $\tilde{A}(T)$ 的正稳定域;

$A_-(T)=\{(u, y, y') \mid u \in U, y=k(u) \leq 0, y'=k(T_v u) \leq 0\}$ 为 $\tilde{A}(T)$ 的负稳定域;

$J_0(T)=\{(u, y, y') \mid u \in U, y=k(T_v u)=0\}$ 为 $\tilde{A}(T)$ 的拓界。

(3) 当 $T_U=e, T_v=e$ 时, $T_U U=U, T_v u=u, \tilde{A}(T)=\tilde{A}(T_k)=\{(u, y, y') \mid u \in U, y=k(u) \in I, y'=T_k k(u) \in I\}$, 此可拓集为关于关联函数变换的可拓集, 它同样有可拓域、稳定域和拓界。

(4) 当 $T_v=e$ 且 $T_U U-U \neq \emptyset$ 时, $T_v u=u$,

$$T_k k(u)=k'(u)=\begin{cases} k(u), & u \in U \cap T_U U \\ k_1(u), & u \in T_U U-U \end{cases}$$

$\tilde{A}(T)=\tilde{A}(T_U)=\{(u, y, y') \mid u \in T_U U, y=k(u) \in I, y'=k'(u) \in I\}$ 此可拓集为关于论域变换的可拓集。

可拓集描述了事物“是”与“非”的相互转化, 它既可用于描述量变的过程(稳定域), 又可以描述质变的过程(可拓域)。零界或拓界描述了质变点, 超过它们, 事物就产生质变。

^{*} 国家自然科学基金资助(70271038), 王洪伟 博士生, 研究方向: 信息系统集成化管理, 智能决策, 吴家春 副教授, 工学博士, 研究方向: 信息系统集成化管理、系统仿真, 蒋 馥 教授, 博导, 研究方向: 信息系统集成化管理、系统仿真。

3 基于物元可拓集的策略模型

3.1 决策系统的基本元素

一般决策模型中的基本概念有状态物元集、决策物元集、输入物元集、输出物元集和干扰物元集^[4]。

状态物元集是可以用来表征决策问题所涉及的系统及其环境特征状况的最小一组物元集。依据表述的对象不同,状态物元集分为系统状态物元集和环境状态物元集,它们可分别表示为

$$R(t) = \{r_1(t), \dots, r_n(t)\}$$

和

$$R^*(t) = \{r_1^*(t), \dots, r_n^*(t)\}$$

其中, $r_i(t) = (N_{r_i}(t), c_{r_i}(t), v_{r_i}(t))$, $r_i^*(t) = (N_{r_i^*}(t), c_{r_i^*}(t), v_{r_i^*}(t))$

决策物元集用

$$W(t) = \{w_1(t), \dots, w_m(t)\}$$

表示,其中 $w_i(t) = (N_{w_i}(t), C_{w_i}(t), v_{w_i}(t))$, 决策者可以采用控制手段对其中每一元素进行物元变换。

输入物元集是从环境到系统的流通或影响,可能是环境对系统的主动作用,也可能是系统的主动吸收,后者是决策物元集的一部分,称为可控输入,输入物元集中的不可控和可控部分可分别表示为

$$I(t) = \{i_1(t), \dots, i_l(t)\}$$

和

$$I^C(t) = \{i_1^C(t), \dots, i_l^C(t)\}$$

其中, $i_j(t) = (N_{i_j}(t), c_{i_j}(t), v_{i_j}(t))$, $i_j^C(t) = (N_{i_j^C}(t), c_{i_j^C}(t), v_{i_j^C}(t))$

输出物元集是系统到环境的流通或作用,表示为

$$O(t) = \{o_1(t), \dots, o_s(t)\}$$

其中, $o_i(t) = (N_{o_i}(t), c_{o_i}(t), v_{o_i}(t))$ 。

干扰物元集通常是指影响系统状态变化,决策者不可控制,甚至难以预测的系统内部或来自环境的随机干扰,分别表示为

$$V(t) = \{v_1(t), \dots, v_r(t)\}$$

和

$$V^*(t) = \{v_1^*(t), \dots, v_r^*(t)\}$$

其中 $v_j(t) = (N_{v_j}(t), c_{v_j}(t), v_{v_j}(t))$, $v_j^*(t) = (N_{v_j^*}(t), c_{v_j^*}(t), v_{v_j^*}(t))$ 。

除上述变量外,一般决策模型中还有目标、策略、衡量条件等基本概念。

目标是决策行为的预期后果,通常以期望的计划期末系统状态和决策实施期间 $[t_0, t_T]$ 系统的输出为主要内容,目标可用物元集表示,记为

$$Q = \{Q_1, Q_2, \dots, Q_l\}$$

若 $l=1$, 该决策问题为单目标决策问题。若 $l>1$ 为多目标规划问题。目标可以细化,建立目标的物元蕴含系统。

策略可以看作是为了实现一定目标所预设的选择性方案或手段。从可拓集的观点看,策略是决策者预定在计划期内对诸决策物元变量在不同时间进行物元变换或事元变换的一种组合,可记为

$$P(W) = \{T_{w_1}[t_0, t_T], \dots, T_{w_m}[t_0, t_T]\}$$

显然,一个策略为 m 维物元变换。一切可能的策略组成 m 维物元变换空间,称为策略集 $P = \{p_1(W), \dots, p_2(W), \dots, p_p(W)\}$ 。物元变换中引入时间参数可用来描述序贯决策或动

态决策。 T^* 为广义的可实现的物元变换集。

衡量条件,是决策者为了对策略实施可能出现的结果进行评判拟定的,用来衡量既定目标实现程度的标准。衡量条件可用特征元集来表示,记为

$$M = \{m_1, m_2, \dots, m_q\}$$

其中, $m_i = (c_i, V_i)$ 是特征元, V_i 都是数量化了的数值域。

3.2 基于物元可拓集的策略模型

(1) 决策系统状态物元模型

$$f: R \times P \times I \times O \times V \times V^* \rightarrow R$$

$$F: R^* \times P \times I \times O \times V \times S \rightarrow R^*$$

(2) 决策方案拟定物元模型

$$H: T^* \rightarrow P$$

其中 H 为基于物元可拓性物元变换规则,即具体的策略生成技术。

(3) 决策模型的输出物元模型

$$G: R \times P \times I \times V \rightarrow O$$

(4) 决策评价物元模型

$$\Phi_i: Q \times O \rightarrow Y$$

上述五种关系,组成了完整的决策的一般物元模型。当然,为了适应不同的需要,可以做不同的精细程度和不同形式的描述。

3.3 可行策略与最优策略的物元模型

3.3.1 可行策略的物元模型 设 $M = \{m_1, m_2, \dots, m_q\}$

为衡量条件,其中 $M' = \{m_1, m_2, \dots, m_s, s \leq q\}$ 为非满足不可条件, $C = (c_1, c_2, \dots, c_s, c_{s+1}, \dots, c_q)$ 为指标集权重向量,其

中 $\sum_{i=1}^q c_i = 1, c_i \geq 0$ 。建立关于 $V_1, V_2, \dots, V_q$ 的关联函数

$$k_i(x) \quad (i=1, 2, \dots, q)$$

把策略 P_j 关于各衡量条件 m_i 的关联函数值简记为

$$k_i(P_j) \quad (i=1, 2, \dots, q; j=1, 2, \dots, p)$$

并进行规范化处理。

定义3.1(可行策略) 设 $L = \{j | \min(k_{ij}) \geq 0, j \in \{1, \dots, p\}, i \in \{1, \dots, s\}\}$, 若 $L \neq \emptyset$, 令 $B = C \times K$, 若存在 $j_0 \in L$ 且满足 $b_{j_0} \geq 0 \quad (i=1, 2, \dots, q)$, 则称 P_{j_0} 为实现目标 Q 的一个可行策略或满意策略,记为

$$P_{j_0} \Rightarrow (Q @, R),$$

$Q @$ 表示目标 Q 实现。否则,若 $L \neq \emptyset$ 或 $\max_{i \in L} (b_i) < 0$, 则称此决策问题已有策略均不可行或不令人满意。

定义3.2(最优策略) 设 $L = \{j | \min(k_{ij}) \geq 0, j \in \{1, \dots, p\}, i \in \{1, \dots, s\}\}$, 若 $L \neq \emptyset$, 令 $B = C \times K$, 若存在 b_{j_0} 满足

$$b_{j_0} = \max_{i \in L} (b_i) \text{ 且 } b_{j_0} \geq 0$$

则称 P_{j_0} 为实现目标的最优策略。

性质3.1 若 $P = \{p_1, p_2, \dots, p_n\}$ 为决策系统的可行策略集,其中 p_{j_0} 为最优策略,则有

$$b_{j_0} \geq b_{j_i}, j=1, 2, \dots, n$$

3.3.2 可行策略的生成过程 策略的生成过程即策略集的开拓的过程,是可拓决策的核心技术,当已有的策略不能令人满意时,或者说策略集中不存在可行策略,就有必要对已有的策略进行开拓。

(1) 初始化阶段主要是在充分了解决策系统所处的初始状态(内部与外部)基础上,确定系统的目标物元、可控的决策物元及可能存在的内外干扰物元。

(2) 进行相容性分析,即在现有状态和策略作用下,决策

系统的输出是否达到预期的目标。若可以实现既定目标,则策略可行。否则进行下一步,生成可行策略。

(3)对系统目标进行可拓性分析^[2]。

(4)对系统条件进行可拓性分析^[2]。

(5)进行物元变换,得到若干个策略。

(6)优劣评价。进行经济评价(财务评价、国民经济评价)和社会评价,得到最优策略或可行策略。

3.4 关键策略的物元模型及生成过程

3.4.1 关键策略的物元模型 在复杂的决策系统中,有的决策者善于“抓住要害,理出头绪”制定出“四两拨千斤”的关键策略,从而打开局面。相反,有的决策者“棋差一着”,没能抓住问题的主要矛盾,从而“满盘皆输”。尽管人们对关键策略早有认识,但对其生成问题却很少有人研究,通常所说的关键策略可归纳为2种类型:

(1)策略的策略^[3]。为了实现某一目标,必须具备很多条件,而要使这些条件实现,又要有相应的策略。如果有一个特殊的策略,它实现时,这些策略就能实现,这个特殊的策略称为关键策略。

(2)对实现目标影响最大的策略^[6]。实现一个目标,有多个策略,它们构成了实现该目标的策略集,在这些策略中,对目标的实现有举足轻重作用的策略视为关键策略。

基于以上2种认识,本文从2个角度研究关键策略的物元模型。

(1)第1类物元模型

(a)关键特征:设目标物元 $R = (N_0, c_0, v_0)$, 且

$$R \leftarrow \begin{cases} r_1 \\ r_2 \\ \vdots \\ r_n \end{cases}$$

其中条件物元 $r_i = (N_i, c_i, v_i)$, $(i = 1, 2, \dots, n)$, $c_0 \sim c_i$ 。记 c_i 的相关特征集为 $c_r(c_i)$ ($i = 1, 2, \dots, n$), 令

$\{c\} = \{c_r(c_1)\} \cap \{c_r(c_2)\} \cap \dots \cap \{c_r(c_n)\} \cap \{c_r(c_0)\}$, 则

可取 $\{c\}$ 中对 $R@$ 影响最大者为关键特征 c_k 。

关键特征是待选特征集中对实现目标起关键作用的特征。关键特征可从待选特征关于目标特征的重要程度来确定。其大体分析步骤如下:

(1)确定目标物元或目标特征。

(2)确定待选特征,以目标特征的相关特征作为待选特征。

(3)分析待选特征关于目标特征的重要程度来确定。

(4)根据一定的阈值标准确定关键特征。

(b)关键事物:设 c_k 为关键特征, $\{N_i \sim (c_i)\}$ 为 N_i ($i = 1, 2, \dots, n$) 的相关事物集,记 $h r N_0 = \{m\}$, $s f N_0 = \{l\}$, 令

$$\{N\} = \bigcup_{i=1}^n \{N_i \sim (c_k)\} \cap (\{m\} \cup \{l\}),$$

则 $\{N\}$ 中对 $R@$ 影响最大者为关键事物 N_k 。

关键事物是可选事物集中对目标实现起着关键作用的事物。关键事物可从待选事物关于制定关键特征(我们把它们称为参照特征)的重要程度来确定,其分析步骤如下:

(1)选定作为参照特征的关键特征。

(2)确定待选事物。

(3)分析待选事物关于参照特征的量值。

(4)根据一定的阈值标准确定关键事物。

(c)关键量值:若 $c_k(N_k) \in V_k$ 时, $R@$, $c_k(N_k) \notin V_k$ 时, $R@$, 则称 V_k 为关键量值域,为 $c_k(N_k)$ 关键量值。

关键量值是待选量值集中对实现目标起关键作用的量值。关键量值可从待选量值对于实现目标的重要程度来确定。其大体分析步骤如下:

(1)确定目标物元。

(2)确定关键特征的待选量值。

(3)分析待选量值对目标物元的贡献程度。

(4)根据一定的阈值标准确定关键量值。

(d)关键物元:若 N_k, c_k, v_k 分别为关键事物、关键特征、关键量值,则称 $r_k = (N_k, c_k, v_k)$, $v_k \in V_k$ 为关键物元。

定义3.3(关键策略) 设 $R@$ 的关键物元是 $h = (m, c, v)$, $k(v)$ 为以 V 为正域的可拓集合的关联函数, $h(t_0) = (m(t_0), c, c(m(t_0))), k(c(m(t_0))) \leq 0$ 。若存在变换 T , 使

$$Th_0 = h(t) = (m(t), c, c(m(t)))$$

$$k(c(m(t))) \geq 0, t \in I = [t_0, t_1]$$

则称 T 为 $R@$ 的关键策略。

(2)第2类物元模型

关键策略是指策略集中对实现期望目标起决定作用的策略。为了对“关键策略”从定量方面进行刻画,特引入“影响函数”的概念。

定义3.4(影响函数) 设策略 p 为 $Q@$ 可行策略,即 $p \Rightarrow (Q@, R)$, $\{p_1, \dots, p_n\} \Rightarrow p$, 给从 $U = \{p, p_1, \dots, p_n\}$ 到实数域上的映射 $g(p, Q@)$, $p \in U$, 满足以下式子

$$g(p, Q@) \leq g(p, Q@)$$

$$\sum_{i=1}^n g(p_i, Q@) \geq g(p, Q@)$$

则称 $g(p, Q@)$ 为 p 关于目标实现的影响函数。我们用影响函数表示策略 p 对 $Q@$ 的贡献值。

定义3.5(关键策略) 设策略 $p \Rightarrow (Q@, R)$, $\{p_1, \dots, p_n\} \Rightarrow p$, $g(p, Q@)$ 为 p 关于目标的实现的影响函数。若 $g(p_{i_0}, Q@) = \max_{i \in \{1, 2, \dots, n\}} g(p_i, Q@)$, 则称 p_{i_0} 为 $Q@$ 的关键策略。

3.4.2 关键策略的物元模型的生成

(a)基于第1类物元模型的关键策略生成法

(1)建立目标物元的蕴含系。

(2)确立考虑的条件集 $\{r\} = \{r_1, \dots, r_n\}$, $r_i = (m_i, c_i, v_i)$, $v_i \in V_i$ ($i = 1, 2, \dots, n$)。

(3)确立 c_i 的相关特征集 $\{c_i \sim\}$ 。

(4)确定关键特征集,即求 $\bigcap_{i=1, 2, \dots, n} \{c_i \sim\} = \{c\}$ 。

(5)确定关键事物集: $\{M\} = \{m | c(m) \in V(c)\}$ 。

(6)确定关键量值集:对 $c \in \{c\}$, 取 $\{V_1(c)\} \cap \{V_2(c)\} \cap \dots \cap \{V_n(c)\} = V(c)$ 。其中 $V_i(c) = \{v | v \in V_i(c), f_i(v) = c, (m_i) \in V_i\} \neq \emptyset$ 。

(7)确定关键物元 $h = (m, c, c(m))$, $m \in \{M\}$, $c \in \{c\}$ 。

(8)求关键策略:若 $c(m) \notin V(c)$, 作变换 T 使 $Th = (m', c, c(m')), c(m') \in V(c)$ 则 T 为关键策略,若 T 为多个,则记为关键策略集 $\{T\}$ 。

(9)评价 $\{T\}$:使目标物元实现的可能性最大的关键策略 $T \in \{T\}$, 即为要生成的关键策略。

(b)基于第2类物元模型的关键策略生成法

(1)建立目标物元的蕴含系;

(2)建立在限制 L 下使 $R@$ 的策略集 $W = \{W_i, i = 1, 2, \dots, n\}$;

(3)建立策略集 W 上的影响函数 $g(W_i, R@)$;

(4)计算影响函数,取 $g(W_{i_0}, R@) = \max_{i \in \{1, \dots, n\}} g(W_i, R@)$

②), 则 W_{i_0} 为 R 的关键策略。

3.5 关键策略的生成过程

针对最优策略或满意策略, 寻找关键策略, 对系统进行重点管理。

(1) 建立系统矛盾蕴含系统^[2,5,6]。

(2) 根据共轭性分析 N 的结构。

$$N = hrN \otimes sfN = reN \otimes inN = apN \otimes ltN = ps(c)N \otimes ng(c)N$$

(3) 相关分析: 根据 N 的结构, 利用相关网, 列出与目标特征 c 相关的关键特征 $c_1, c_2, \dots, c_m$ 。

(4) 蕴含分析: 利用蕴含系方法, 寻找关键特征 $c_1, c_2, \dots, c_m$ 的最上位特征 $c_{i_1}, c_{i_2}, \dots, c_{i_p}$ 。

(5) 根据问题现状, 确定研究对象 N 关于最上位特征的值, 写出相应的物元 $R_1, R_2, \dots, R_p$ 。

(6) 进行物元变换, 得到若干个关键策略。

$$T_j R_j = R'_{i_j}, (j=1, 2, \dots, p; i_j=i_1, i_2, \dots, i_p)$$

(7) 针对扰动物元, 制定防范策略(前馈控制)与补偿策略(反馈控制)进行协调, 请参见文[5]。

结束语 策略模型与生成的形式化研究对提高基于知识的系统的性能起到重要的作用, 对知识管理系统的研究也有着指导意义。可拓集是近年来提出的一种解决不相容问题的

(上接第129页)

定义3 $F_{i_low} = b_{low} = \max\{F_i : i \in I_0 \cup I_3 \cup I_4\}$

$$F_{i_up} = b_{up} = \min\{F_i : i \in I_0 \cup I_1 \cup I_2\}$$

这样, 检查样本 i 是否满足优化条件就很容易了。例如, 假设 $i \in I_1 \cup I_2$, 我们只需要检查条件: $F_i < F_{i_low}$ 。假如这个条件满足, 那么就存在一对违反 KKT 条件的样本对 (i, i_{low}) 。

这样, 在我们循环遍历所有样本时, 对于一个给定的样本 i , 首先计算 F_i , 然后应用当前的 (b_{low}, i_{low}) , (b_{up}, i_{up}) 来检查样本 i 是否满足优化条件。如果满足, F_i 被用来更新 (b_{low}, i_{low}) 或 (b_{up}, i_{up}) 。如果不满足, 我们就选用 (i, i_{low}) 或 (i, i_{up}) 作为本次优化的样本对。例如, 如果 $i \in I_1 \cup I_2$ 并且 $F_i < b_{low}$, 即样本 i 违反了优化条件, 这时我们选择样本对 (i, i_{low}) 作为本次优化的两个样本。如果, 样本 i 满足优化条件的话, 那么我们就用 F_i 来更新 (i_{up}, b_{up}) 。即, 如果 $F_i < b_{up}$, 那么 $i_{up} = i$ 并且 $b_{up} = F_i$ 。

这样在 SMO 算法中每次外层循环选择 i_2 时我们不再用 (6) 式作为判优条件, 而是用 (7) 式作为判优条件, 这样就避免了由于不正确的值 β 而引起的不必要的开销。

6. 实验结果

本部分我们比较传统的 SMO 算法和我们改进的 SMO 算法之间的性能。我们实验的硬件环境是 128 兆内存, Pentium III 450MHz 的机器。软件环境是 Win2000 Professional + Matlab 6.1。我们选用一阶多项式函数作为核函数, 误差控制参数 $C=0.10$ 。

我们选用的数据集是 UCI 的 votes 和我们自己的 IGA 肾病诊断数据。

表1 数据集属性

数据集	样本维数	样本数
Votes	16	435
IGA 肾病	14	153

集合理论, 本文利用可拓集理论在策略的模型、形式化生成方法等方面作了探索, 初步的研究表明可拓集和物元理论可为策略的生成方法及策略协调等的研究提供了新的思路。在“CRM 中客户本体的建立方法及共享机制”课题中, 我们利用基于可拓集的策略模型分析了领域客户本体模型的建立方法, 探讨了本体中术语语义不一致问题, 取得了一定的效果。目前这些决策模型的可操作性还不太强, 需要大量领域专家的参与, 这需要今后进一步的研究。

参考文献

- 1 Cai Wen, He Bin, Zhang Yingli, Yang Chunyan. Key Strategy and Coordination Problem [J]. Shanghai: In: The 3rd Intl. Conf. On Management, (ICM'98)
- 2 蔡文, 杨春燕, 林伟初. 可拓工程方法[M]. 北京: 科学出版社, 1997
- 3 王洪伟. 关键策略的物元模型[J]. 系统工程理论与实践, 1998, 18(2): 106~109
- 4 李习彬. 一般决策模型与概念性决策模型[J]. 论文集: 科学决策与系统工程. 中国科学技术出版社, 1990
- 5 王洪伟. 策略协调问题的研究[D]. [广东工业大学硕士学位论文]. 1999
- 6 杨益民, 魏仲俊. 关键矛盾法及其在 CSE 中的应用[J]. 系统工程理论与实践, 1998, 18(1): 99~97

表2 votes 数据集运行效果

算法	运行时间(s)
SMO 算法	23.7
改进的 SMO 算法	15.6

表3 IGA 肾病数据集运行效果

算法	运行时间(s)
SMO 算法	16.9
改进的 SMO 算法	9.8

考虑到我们的 SMO 程序是用 Matlab 实现的, Matlab 是一种解释型的语言, 执行效率比较低, 如果改用 C 语言编程, 速度会更快一点。

结束语 本文通过对 SVM 分类算法的分析, 以及对其最常用的 SMO 算法的深入研究, 给出一种改进的 SMO 算法。用标准的 UCI 数据以及我们的肾病诊断数据测试的结果表明, 改进的 SMO 算法在性能上优于传统的 SMO 算法。

参考文献

- 1 Vapnik V. Nature of Statistical Learning Theory. John Wiley and Sons, Inc., New York, in preparation
- 2 Burges J C. A Tutorial on Support Vector Machines for Pattern Recognition. Bell Laboratories, Lucent Technologies. 1997
- 3 Keerthi S S, et al. Improvements to Platt's SMO Algorithm for SVM Classifier Design
- 4 Keerthi S S, et al. A Fast iterative Nearest Point Algorithm for Support Vector Machine Classifier Design
- 5 Platt J. Sequential minimal optimization: A fast algorithm for training support vector machines. Advances in Kernel Methods-Support Vector learning. Cambridge, MA: MIT Press, 1999. 185~208
- 6 边肇祺, 等. 模式识别. 清华大学出版社, 1988