

模糊信息系统及其相应的关系数据库

洪 国 彬

[计算机科学(电脑)系]

摘要 从模糊信息系统的角度出发,对于各种类型的模糊关系数据库的现状,系统地进行归类、分析,同时也通过一些例子说明利用目前的 RDB 来实现模糊数据库所存在的难点和问题。

关键词 模糊信息系统,模糊关系数据库,属性值模糊信息系统,属性模糊信息系统

0 引言

传统的数据库只能用来处理精确而又良构的数据。现实的世界,特别是复杂系统中存在着大部分模糊数据,甚至是不完整的数据,传统的数学模型及方法已经远远不能适应这方面的要求。因此,近年来致力于研究模糊关系数据库的大有人在,相应地从理论上或实现方法上提出互相独立的模糊关系数据库的数学模型,有实现的,有的只进行了代数运算的公式化,但是这些模型在对模糊信息的表达中令人感到杂乱、茫然,无法有一种全面认识的感觉。为此,本文从模糊信息系统的观点来看待模糊关系数据库,以便归类和分析,同时也通过一些例子说明,利用目前的 RDB 来实现模糊关系数据库所存在的难点和问题。

1 模糊信息系统

信息系统是一种以信息为对象的数据模型,相应地,模糊信息系统是以模糊信息为对象的信息系统模型。为了更清楚地说明问题,明确地区分信息系统,以及不同性质的模糊信息系统,下面根据文献[1]给出定义。

定义1 一个信息系统 S 是一个四元组: $S = \langle X, A, V, P \rangle$, 其中, X 为对象的集合, A 为描述对象的属性集合, V 为每个属性的值域的并集, 即若 $a \in A$, a 的值域为 V_a , 则有 $V = \bigcup_{a \in A} V_a$; P 为一函数, $P: X \times A \rightarrow V$, 并且保证 $\forall x \in X, \forall a \in A$, 则有 $P(x, a) \in V_a$ 。

定义2 信息系统 $S = \langle X, A, V, P \rangle$, $x \in X$, 函数 P 关于 x 的映射 P_x 称为 S 中关于对象 x 的信息: $P_x = \{ \langle a, V \rangle \mid a \in A, v \in V, P(x, a) = v \}$ 。

本文1992-07-10收到。

定义3 称 $\tilde{S} = \langle X, A, \tilde{V}, \tilde{P} \rangle$ 为一属性值模糊信息系统. 其中 X, A 同定义1; $\tilde{V} = \bigcup_{a \in A} P_{\tilde{V}_a}, \tilde{V}_a$ 表示 V_a 的模糊子集, $P_{\tilde{V}_a}$ 表示关于 V_a 的所有可能的模糊子集; $\tilde{P}: X \times A \rightarrow \tilde{V}$, 且保证有 $\tilde{P}(x, a) \in \tilde{V}_a$.

定义4 称 $\tilde{S} = \langle X, /A, \tilde{V}, \tilde{P} \rangle$ 为一属性模糊信息系统, 其中, X, \tilde{V} 的解释同定义3; $/A = P_{/A}$ 为 A 的所有可能性的模糊集合的集合, 而 $\tilde{P}: X \times /A \rightarrow \tilde{V}$, 此外还要保证, $\forall x \in X, \forall \tilde{A} \in /A$, 则有 $\tilde{P}(x, \tilde{A}) \in \tilde{V}_{/\tilde{A}}, \tilde{V}_{/\tilde{A}} = \{\tilde{V}_{/\tilde{A}} | \tilde{A} \in /A\}$, 且 $\tilde{V}_{/\tilde{A}} = \{\tilde{V}_a | a \in A, \text{并由 } a \text{ 在 } \tilde{A} \text{ 中的隶属度 } \tilde{a} \text{ 所对应}\}$.

定义5 $\tilde{S} = \langle X, A, \tilde{V}, \tilde{P} \rangle, \tilde{P}_x$ 称为 \tilde{S} 中关于对象 X 的属性值模糊信息, $x \in X$, 而 $\tilde{P}_x = \{\langle a, \tilde{V}_a \rangle | P(x, a) = \tilde{V}_a\}$.

定义6 $\tilde{S} = \langle X, /A, \tilde{V}, \tilde{P} \rangle, \tilde{P}_x$ 称为 \tilde{S} 中关于 x 的属性模糊信息, $x \in X$, 而 $\tilde{P}_x = \{\langle \tilde{A}, \tilde{V}_{/\tilde{A}} \rangle | \tilde{P}_x(x, \tilde{A}) = \tilde{V}_{/\tilde{A}}, \tilde{A} \in /A\}$.

定义7 信息系统 $S = \langle X, A, V, P \rangle$ 称为属性值模糊信息系统 $\tilde{S} = \langle X, A, \tilde{V}, \tilde{P} \rangle$ 与属性模糊信息系统 $\tilde{S} = \langle X, /A, \tilde{V}, \tilde{P} \rangle$ 的基系统.

由一述的定义可以发现, 模糊信息实际上是一些模糊集合的集合, 包括属性值模糊集合和属性模糊集合, 而普通信息则只是包括具体属性的元素和这些元素对应值域中的某个值, 即普通信息由普通集合映射所决定, 而模糊信息则是由模糊映射所确定的. 从表示的角度来看, 对于普通信息 P , 可以用一张二维表的形式表示, 对于模糊信息 \tilde{P} , 它的表示形式又是如何呢? 一种比较有效的办法是利用矩阵的形式.

例1. 设 $S = \langle X, A, V, P \rangle$, 其中 $X = \{x, y, z\}, A = \{a, b, c\}, V = \bigcup_{a \in A} V_a, V_a = \{r, s\}, V_b = \{t, u\}, V_c = \{v, w\}$. (1) 对于 $S = \langle X, A, V, P \rangle$ 中的信息表示如下, 即 P 的定义为

$$\begin{matrix} & a & b & c \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{pmatrix} s & t & w \\ r & \mu & v \\ s & t & v \end{pmatrix} \end{matrix},$$

这就完全符合一张二维表格的形式. (2) 对于 $\tilde{S} = \langle X, A, \tilde{V}, \tilde{P} \rangle$ 中的模糊信息 \tilde{P} , 用矩阵 Q 为表示, 则表示属性对应取值的模糊性, 或者说是属性值模糊信息 \tilde{V}_A . 举例如下

$$\begin{aligned} \tilde{P}_x: x \rightarrow A &= \{a, b, c\}, \\ \tilde{V}_A: \tilde{V}_a &= \tilde{V}_a = (0.5/r, 0.6/s), \tilde{V}_b = (0.3/t, 0.9/u), \\ &\tilde{V}_c = (0/v, 1/w), \\ \tilde{P}_y: y \rightarrow A &= \{a, b, c\}, \\ \tilde{V}_A: \tilde{V}_a &= \tilde{V}_a = (0.1/r, 0.25/s), \tilde{V}_b = (0.8/t, 0.9/u), \\ &\tilde{V}_c = (0.3/v, 0.3/w); \end{aligned}$$

$$\begin{aligned} \tilde{P}_z: Z \rightarrow A &= \{a, b, c\}, \\ \tilde{V}_a: &= \tilde{V}_s = (1/R, 0.2/S), \tilde{v}b = (0.3/t, 0.7/u), \\ \tilde{V}_c &= (0.5/v, 0.7/w). \end{aligned}$$

于是 Q 矩阵表示为

$$\begin{matrix} & \begin{matrix} r & s \end{matrix} \\ \begin{matrix} x_a \\ y_a \\ z_a \end{matrix} & \begin{bmatrix} 0.5 & 0.6 \\ 0.1 & 0.25 \\ 1 & 0.2 \end{bmatrix} \end{matrix}, \quad \begin{matrix} & \begin{matrix} t & u \end{matrix} \\ \begin{matrix} x_b \\ y_b \\ z_b \end{matrix} & \begin{bmatrix} 0.3 & 0.9 \\ 0.8 & 0.9 \\ 0.3 & 0.7 \end{bmatrix} \end{matrix}, \quad \begin{matrix} & \begin{matrix} v & w \end{matrix} \\ \begin{matrix} x_c \\ y_c \\ z_c \end{matrix} & \begin{bmatrix} 0 & 1 \\ 0.3 & 0.3 \\ 0.5 & 0.7 \end{bmatrix} \end{matrix}.$$

(3)对于 $\tilde{S} = \langle X, /A, \tilde{V}, \tilde{P} \rangle$ 中的模糊信息 \tilde{P} 用 \tilde{P} 矩阵和 \tilde{Q} 矩阵来表示, 其中 \tilde{P} 矩阵表示某个对象 x 与 \tilde{A} 的对象, 或者称为对象 x 的属性模糊信息, 记为 $x \rightarrow \tilde{A}$, \tilde{Q} 矩阵的意义同 (2). 举例如下, 设: $\tilde{P}_x: x \rightarrow \tilde{A} = (0.1/a, 0.8/b, 0.4/c)$, \tilde{V}_A 同 (2); $\tilde{P}_y: y \rightarrow \tilde{A} = (0.2/a, 0.3/b, 0.88/c)$, \tilde{V}_A 同 (2); $\tilde{P}_z: z \rightarrow \tilde{A} = (0.8/a, 0.1/b, 0.3/c)$, \tilde{V}_A 同 (2). 于是, \tilde{P}, \tilde{Q} 矩阵表示为 $\tilde{Q}_a, \tilde{Q}_b, \tilde{Q}_c$ 同 (2),

$$\begin{matrix} & \begin{matrix} a & b & c \end{matrix} \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} 0.1 & 0.8 & 0.4 \\ 0.2 & 0.3 & 0.88 \\ 0.8 & 0.1 & 0.3 \end{bmatrix} \end{matrix}.$$

根据上述例子的表达, 有以下的 (1)-(3). (1)构成的关系是一个二维表. (2)每一元素增加了一个隶属度, 因此从本质上说, 它们构成的模糊关系是 Codd 关系模型的进一步拓广, 由此出发, 可以把模糊关系视为三维表. 因为它是由二维表的迪卡尔积到 $[0, 1]$ 闭区间的映射来定义的, 但对于值模糊, 关系是明确的, 还是二维表. (3)是在 (2) 的基础上, 对于对象的每个属性又增加了一个模糊度, 因此, 相应构成的属性模糊的关系可视为一种加权叉积形式, 即 $W_1 * F(D_1) \times W_2 * F(D_2) \times W_3 * F(D_3) \times \cdots \times W_n * F(D_n)$, 其中 D_1, D_2, \cdots, D_n 为 n 个论域, $F(D_1), F(D_2), \cdots, F(D_n)$ 分别为论域 $D_i (i=1, \cdots, n)$ 上的模糊子集构成的集合, W_i 分别表示第 i 个属性在关系中的模糊集. 故此种关系下的元组 r , 可表示成为 $r = (W_1 * x_1, W_2 * x_2, \cdots, W_n * x_n)$.

2 模糊信息系统与相应的模糊关系数据库

上述已经说明模糊信息系统有两种不同性质的系统, 即属性值模糊信息系统及属性模糊信息系统, 同时也可设知, 属性值模糊信息系统是属性模糊信息系统的一种特例.

2.1 属性值模糊信息系统与相应的关系数据库

属性值模糊信息系统下的关系数据库, 一般情况下可分为四种类型, 即模糊关系数据库, 按类似关系的集中, 可能性分布——关系模型, 可能性分布——模糊关系模型.

2.1.1 模糊关系模型——第一类型 这是一种最简单的推广, 也就是在通常关系中加上隶属度而变成模糊关系, 虽然这种系统实现起来比较容易, 但这种系统不是模糊关系数据库研究的核心.

2.1.2 按类似关系的集中——第二类型 这种类型是为了检查与查询的方便而提出的一种系统模型,即在检查与查询条件相符的组后,将相同的组汇集在一起.为此,可以把集合看成元素,从而推广了关系模型.如文献[7].

2.1.3 可能性分布——关系模型——第三类型 用可能性分布表示数据本身所具有的模糊性,提出了可以具有这些模糊性作为属性值的可能性分布——关系模型.这种类型在众多的文献中可以查到.如文献[2],[7].

2.1.4 可能性分布——模糊关系模型——第四类型 可能性分布——模糊关系模型,除了第三类型所述的数据值本身所具有的模糊性以外,还有关于数据之间关系的模糊性.这种模型,目前只是公式化了大部分的关系代数数据运算和关系计算,真正实现的系统还不理想,比较有代表性的如文献[3]所述.

2.2 属性模糊信息系统与相应模糊关系数据库

属性模糊信息系统比属性值模糊信息系统有更大的能力去表达现实世界的信息,但基于属性模糊所构成的模糊关系的数据库形式,目前还处于理论性的探讨阶段,更确切一点地说,只能处于对属性模糊的模糊关系的操作与运算的程度上,因为此种数据库,有关于信息的存贮与检查的处理还未能解决.为了更清楚地说明问题,利用前面的例子一个信息系统 \tilde{S} 是非常简单的,其属性的个数很少,只有 a, b, c 三个,利用矩阵表示用了四个矩阵(即 $3+1$).为转化压缩,矩阵 \tilde{P} 弃掉, $\tilde{Q}_a, \tilde{Q}_b, \tilde{Q}_c$ 矩阵相应地转化成 $Q^{\sim}_a, Q^{\sim}_b, Q^{\sim}_c$,即

$$\begin{matrix} & U_a & r & s \\ Q^{\sim}_a: y_a & \begin{bmatrix} x_a & \begin{bmatrix} 0.1 & 0.5 & 0.6 \\ 0.2 & 0.1 & 0.25 \\ z_a & \begin{bmatrix} 0.8 & 0.1 & 0.2 \end{bmatrix} \end{bmatrix} \end{matrix}, & \begin{matrix} & U_b & r & s \\ Q^{\sim}_b: y_b & \begin{bmatrix} x_b & \begin{bmatrix} 0.8 & 0.3 & 0.9 \\ 0.3 & 0.8 & 0.9 \\ z_b & \begin{bmatrix} 0.1 & 0.3 & 0.7 \end{bmatrix} \end{bmatrix} \end{matrix}, & \begin{matrix} & U_c & r & s \\ Q^{\sim}_c: y_c & \begin{bmatrix} x_c & \begin{bmatrix} 0.4 & 0 & 1 \\ 0.88 & 0.3 & 0.3 \\ z_c & \begin{bmatrix} 0.3 & 0.5 & 0.7 \end{bmatrix} \end{bmatrix} \end{matrix} \end{matrix} \end{matrix}$$

$Q^{\sim}_a, Q^{\sim}_b, Q^{\sim}_c$ 矩阵的 U_a, U_b, U_c 列是 \tilde{P} 矩阵相应的 a, b, c 列. 这种形式便于存贮和检索,如利用贴近度的检索、查询是方便的,但相应构成的关系操作与运算就比较麻烦.而且,最关键的问题在于,这种情况下,建立的关系模型的数据库,只能表示属性个数较少的情况,因为这情况下属性个数等于相应建立的数据库文件数,那么对于一个能相应地表示现实世界能力的信息系统,这种数据库是不理想的.反之,如果不进行矩阵压缩的工作,构成的文件数也会多于上述情况,而且从存贮和检索的角度还更难实现.

综合上述四种不同性质的模糊信息系统及其相应的模糊数据库(特别是属性模糊信息系统下的模糊关系数据库),若以近似通常关系模型的水平建立数据模型,则理论的处理或实现都比较容易,可是表示现实世界的数据库能却不够.在属性值模糊的信息系统下,第一、二、三、四类型的数据库就是如此.若按照现实世界的数据库表示建立数据模型,理论的处理或实现就难了,在属性模糊的信息系统下,属于这种类型的模糊数据库(说得更准确一点就是数据模型),就是这种情况.因此,研究模糊关系数据库最根本的问题是,寻找协调两者的数据模型.

3 结束语

人类所进行的几乎是模糊信息的处理,模糊信息处理为人工智能的研究发展开辟了一条

崭新的道路,但是模糊信息处理的研究才刚刚开始.就目前所实现的模糊关系数据库却是利用成熟的传统数据库技术进行拓广来处理,如利用 dbase-Ⅲ 来管理,由于 dbase-Ⅲ 是局限于 INF 的情况,这给模糊关系数据库的发展和实现带来种种不利因素.因此,重新研制模糊数据库管理系统是一条必由之路,只有这样,才能实现描述复杂和更接近现实世界的各种现实对象,这个问题,还有待于今后的大力研究和发展.

参 考 文 献

- [1] 李景洲,模糊信息系统的一个数学模型,计算机工程与设计,1 (1992).
- [2] 马野元秀,模糊集合与计算机软件,计算机科学,4 (1990).
- [3] 刘东波,李德毅,模糊关系数据库系统库系统 FRDB,小型微型计算机系统科学院计算所正规学院,2 (1990).
- [4] Zadeh, L. A.,近似推理论[1-Ⅲ],计算机科学,2-4 (1990).
- [5] 何新贵,属性模糊的模糊关系型数据库,计算机工程与设计,1 (1990).
- [6] 何新贵,模糊关系型数据库的数据模型,计算机学报,2 (1990).
- [7] 杨泽民、陈国权,模糊自动推理软件 FAIC 的设计与实现,模糊系统与数学,2,1 (1988).

Fuzzy Information System and Its Corresponding Relational Database

Hong Guobin

(Department of Computer Science)

Abstract Starting from fuzzy information system, the author makes a systematic classification and analysis on various types of fuzzy relational databases according to their present conditions; and exemplifies the difficulties and problems for realizing fuzzy relational database by making use of RDB.

Key words fuzzy information system, relational database, fuzzy relational database