维普资讯 http://www.cqvip.com

模糊数据库 代数理论 数据库 (15)

计算机科学 1999 Vol. 26 №. 1

一个模糊数据库的代数理论*)

An Algebraic Theory of Fuzzy Database

59-61

张师超 蒋运承 袁鼎荣

(广西师范大学数学与计算机系 桂林 541004)

TP311.13

Abstract In this paper, we set up an 1NF algebraic theory of fuzzy database, and proved that the fuzzy first normal forms are closed under operations of the fuzzy relational algebra. We also proved that fuzzy relational algebra and fuzzy relational calculus are equivalet.

Keywords Fuzzy database, Relational calculus, Relational algebra

近几年,国内许多学者对模糊数据库进行了有效的研究,取得了很多重要的研究成果,但是,模糊数据是一个复杂的对象,现有的模糊数据库模型中仍存在不同程度的缺陷。例如,尚缺乏模糊数据库的关系演算语言方面的研究工作。为了支撑模糊数据库的应用,本文建立了一个 1NF 模糊数据库的代数理论,即:建立了一个模糊数据库模型,研究了相应的模糊关系演算语言及模糊关系代数,并证明了模糊的第一范式在该理论中定义的模糊关系代数操作下是封闭的,以及模糊关系代数与模糊演算语言的等价性。

一、模糊数据库模型 FRM 与概念

1.1 模糊数据的表示

模糊数据有很多种表示方法,本文采用隶属函数表示。隶属函数是由 Zadeh 提出的一种表示模糊数据的方法,它把模糊数据 X 定义为论域 D 上的一个隶属函数:

$$f: D \to [0,1]$$

即把 D 上的每个元素映射到[0,1]区间上的一个值,表示该元素隶属于 X 的程度,值越大表示隶属的程度越高。在论域 D 为有限集时,可用列表表示,否则可用一个函数过程来计算隶属函数。

例如,设论域 $D=\{1,2,3,4,5,6\}$,X 表示"靠近 4 的数"用隶属函数表示为:

$X=\{0/1, 0.2/2, 0.8/3, 1/4, 0.8/5, 0.2/6\}$。

精确数也可以用隶属函数表示,如 4 表示为$\{1/4\}$。

定义 1 设 X_1 和 X_2 是两个模糊数据,它们的语义距离定义为它们之差 X_1-X_2 的某种范数 $\|X_1-X_2\|$。如文[1]中的切比雪夫距离、欧几里得距离及加权明科夫基距离等等都是这样定义的。

1.2 模型 FRM 与概念

模糊关系模式 R 是由属性 $A_1, A_2, \cdots, A_n$ 组成的一个有限集合。对于每一个属性 A_i,它有一个对应的值域 $dom(A_i)$,并设 K 是 R 的主键集合。

由以上模糊数的表示可知,对于每一个属性 A_i,它的所有可能性分布组成的集合 $Poss(A_i)$为:

$$Poss(A_i)=\{\pi: dom(A_i)\to[0,1]\}$$

我们称这个集合 $Poss(A_i)$为属性 A_i 的潜在的模糊值域。

显然,只要 $dom(A_i)$是无限的,则 $Poss(A_i)$也是无限的,这不适合数据库的应用,我们有必要定义一个有限子集 $u_i\subset\{\pi\in Poss(A_i)\}$。

定义 2 属性 A_i 的潜在的模糊值域 $Poss(A_i)$的一个有限子集 u_i,称为属性 A_i 的模糊值域。记为 $fdom(A_i)$,即:

$$fdom(A_i)=u_i\subset Poss(A_i)=\{\pi: dom(A_i)\to[0,1]\}$$

例如,"大约 30"岁的模糊值域可以表示为:

$\{0.75/28, 0.8/29, 1/30, 0.9/31, 0.85/32\}$

定义 3 设 A_i 和 A_j 是两属性,如果它们的语义距离 $\|A_i-A_j\|<\varepsilon$,则称 A_i ε-等于 A_j,记为:$A_i=_\varepsilon A_j$,其中 ε 是专家给定的充分小的正数。

对于定性模糊数据,我们只进行"$=_\varepsilon$"和"$\neq_\varepsilon$"的比较,而对于定量模糊数据,我们还要进行大小比较,而确定定量模糊数大小比较的方法有许多种,本文采用王士同所提出的模糊数质心的概念,并据此确定定量模糊数的大小比较。

定义 4 设定量模糊数据 $X=\{\mu_1/u_1, \mu_2/u_2, \cdots, \mu_n/u_n\}$,则模糊数据 X 的质心 $m(X)$定义为:

$$m(X)=\left(\sum_{i=1}^{n}\mu_i u_i\right)/\sum_{i=1}^{n}u_i$$

定义 5 设定量模糊数据 X_1, X_2 均取于同一论域,若 $m(X_1)>m(X_2)$,则称 $X_1>X_2$;若 $m(X_1)<m(X_2)$,则称 $X_1<X_2$

定义 6 设 $U=\{A_1, A_2, \cdots, A_n\}$是属性集合,$U$

*)本文由国家 863 计划项目和广西教委基金项目资助。

维普资讯 http://www.cqvip.com

的模糊值域 $fdom(U)$ 为：

$$fdom(U)=fdom(A_1)\times fdom(A_2)\times\cdots\times fdom(A_n)$$

U 上的一个元组是一个映射 $\tau:U\rightarrow fdom(U)$，使得：

$$\forall x\in U(\tau(x)\in fdom(U))$$

即：元组 τ 可用 $(a_1,a_2,\cdots,a_n)$ 表示，其中 $(a_1,a_2,\cdots,a_n)\in fdom(U)$。

对于 $Y=\{Y_1,Y_2,\cdots,Y_k\}$ 是 U 的子集合，τ 在 Y 上的投影为：

$$\tau(Y)=(\tau(Y_1),\tau(Y_2),\cdots\tau(Y_k))$$

模糊关系模式 R 上的一个关系 r 是 R 上的一些元组的集合，这种关系称为模糊关系。一个模糊数据库是一些模糊关系的集合。

定义 7 设元组 $\alpha=\langle a_1,a_2,\cdots,a_n\rangle$，$\beta=\langle b_1,b_2,\cdots,b_n\rangle$。则 α 与 β 的语义距离定义为：

$$\|\alpha-\beta\|=(\sum_{i=1}^{n}\|a_i-b_i\|)/n$$

定义 8 设元组 $\alpha=\langle a_1,a_2,\cdots,a_n\rangle$，$\beta=\langle b_1,b_2,\cdots,b_n\rangle$。如果它们的语义距离 $\|\alpha-\beta\|<\varepsilon$，则称 α ε-等于 β，记作：$\alpha=_\varepsilon\beta$，其中 ε 是专家给定的充分小的正数。

定义 9 设 R 是一个模糊关系模式，K 是 R 的主键集合。R 上的模糊关系 r 是 R 上有限个非空元素的集合，它满足：(1)对于 r 中的每一个元组 α 及 $A\in R$，$\alpha(A)$ 是原子值，即 $\alpha(A)$ 是唯一的；(2)对于 $\forall\alpha,\beta\in r$，$\forall A\in K(\alpha(A)=_\varepsilon\beta(A))$ 当且仅当 $\alpha=_\varepsilon\beta$，我们称满足这两个条件的关系为模糊的第一范式。

定义 10 设 R 是一个模糊关系模式，r_1 和 r_2 是 R 上的两上模糊关系。r_1 ε-等于 r_2 当且仅当：(1)对于 r_1 上的任意元组 α，在 r_2 中存在元组 β，使得 $\alpha=_\varepsilon\beta$，并且(2)对于 r_2 上的任意元组 α，在 r_1 中存在元组 β，使得 $\alpha=_\varepsilon\beta$。

r_1 ε-等于 r_2 记为：$r_1=_\varepsilon r_2$。

二、关系演算

2.1 元组关系演算

我们定义模糊的 1NF 元组关系演算(1FTRC)为 $\{\alpha|\Phi(\alpha)\}$，其中 α 是一个元组变量，Φ 是一个公式。下面先定义原子公式，然后再定义合式公式 Φ。

定义 11 一个原子公式是如下式子之一：

(1)$R(\beta)$ 是一个原子公式，其中谓词 R 是模糊关系名，β 是元组变量。$R(\beta)$ 表示 β 属于 R，或 $\beta\in R$；

(2)$\beta=\gamma(i)$ 是一个原子公式，其中 β 是元组 γ 在第 i 个属性的投影 $\gamma(i)$ 的属性值；

(3)$a\ \theta\ \beta(i)$，$\beta(i)\ \theta\ a$；$\beta(i)\ \theta\ \gamma(i)$ 是原子公式，其中 a 是一已知的模糊数，θ 是一个关系比较符号，对于定性模糊数据，$\theta=(=_\varepsilon,\neq_\varepsilon)$；对于定量模糊数据，$\theta=(=_\varepsilon,>)$；

(4)$\beta(i)=\{\gamma|\Phi'(\gamma)\}$ 是原子公式，其中 Φ' 是一个带有变量 γ 的公式。

定义 12 我们递归定义合式公式 Φ 如下：

(1)每一个原子公式是合式公式；

(2)如果 Φ_1 和 Φ_2 是合式公式，则 $\Phi_1\wedge\Phi_2$，$\Phi_1\vee\Phi_2$ 和 $\neg\Phi_1$ 是合式公式；

(3)如果 Φ 是一个合式公式，则 $\forall\beta(\Phi)$，$\exists\beta(\Phi)$ 是合式公式；

(4)空 Φ 也是一个合式公式。

显然，1FTRC 中允许使用 $\{\alpha|\neg(\alpha\in r)\}$，从而 1FTRC 必须满足安全限制条件。

2.2 域关系演算

我们定义模糊的域关系演算为 $\{X_1X_2\cdots X_n|\Phi(X_1,X_2,\cdots,X_n)\}$，其中 $X_1,X_2,\cdots,X_n$ 为元组变量 X 的各个域变量，Φ 是一个公式。下面先定义原子公式。然后定义合式公式 Φ。

定义 13 原子公式有下列几种：

(1)$R(X_1,X_2,\cdots,X_n)$，其中谓词 R 是一个模糊关系名，X_i 是一个已知的模糊数或域变量。$R(X_1,X_2,\cdots,X_n)$ 表示由分量 $X_1,X_2,\cdots,X_n$ 组成的元组属于 R；

(2)$X_i\ \theta\ a$，或 $a\ \theta\ X_i$，其中 a 为一个已知的模糊数，X_i 为域变量，θ 是一个关系比较符号，$X_i\ \theta\ a$ 表示 X_i 与 a 之间满足 θ，对于定性模糊数据，$\theta=(=_\varepsilon,\neq_\varepsilon)$；对于定量模糊数据，$\theta=(=_\varepsilon,>)$；

(3)$X_i\ \theta\ Y_j$，X_i 和 Y_j 分别为元组 X 和 Y 的域变量，θ 为一个关系比较符号，$X_i\ \theta\ Y_j$ 表示 X_i 和 Y_j 应满足 θ 关系，对于定性模糊数据，$\theta=(=_\varepsilon,\neq_\varepsilon)$；对于定量模糊数据，$\theta=(=_\varepsilon,>)$。

定义 14 域关系演算中合式公式 Φ 的定义，只需将定义 3.2 中的(3)改为：如果 Φ 是一个合式公式，则 $(\forall X_i)(\Phi)$，$(\exists X_i)(\Phi)$ 是合式公式，其它不变。

显然，这种域关系演算语言满足安全限制条件。

注意：上面定义的元组关系演算或域关系演算的合式公式 Φ 除在"$=_\varepsilon$"处的差异外，与传统的关系演算语言中的公式相同。因此，它们的安全限制条件与传统的关系演算的安全限制条件类似，其差异是"="被置换成"$=_\varepsilon$"。

三、关系代数

下面给出模糊数据库的基本的关系代数操作。

设 r,s 是模糊模式 $R=\{U_1\times U_2\times\cdots\times U_n\}$ 上的模糊关系，设 K 是 R 的关键字集。

(1)并($\cup$)

$$\begin{aligned}r\cup s=\{\tau|&\exists\tau_1\in r,\exists\tau_2\in s(K'\in K(\tau_1(K')=_\varepsilon\tau_2(K')))\wedge\tau=\{\max(\pi_r(u_i),\pi_s(u_i))/u_i:u_i\in U_i,i=1,2,\cdots,n.\}\\&\vee(\tau\in r\wedge(\forall\tau'\in s(\exists K'\in K(\tau(K')\neq_\varepsilon\tau'(K')))))\\&\vee(\tau\in s\wedge(\forall\tau'\in r(\exists K'\in K(\tau(K')\end{aligned}$$

维普资讯 http://www.cqvip.com

$\neq_t \tau'(K')))\}$。

(2)交($\cap$)

$r\cap s=\{\tau|\exists\ \tau_1\in r,\exists\ \tau_2\in s(K'\in K(\tau_1(K')=_t\tau_2(K')))\wedge\tau=\{\min(\pi_r(u_i),\pi_s(u_i))/u_i : u_i\in U_i,i=1,2,\cdots,n.\}\}$

(3)差(−)

$r-s=\{\tau|\tau\in r\wedge(\forall\ \tau'\in s(\exists\ K'\in K(\tau(K')\neq_t\tau'(K')))\}$.

(4)笛卡尔积($\times$)

设 s' 是模糊模式 $S=\{V_1/V_2\times\cdots/V_m\}$ 的模糊关系，则

$r\times s'=\{t(a_1,a_2,\cdots,a_n,b_1,\cdots,b_m) : a_i\in U_i,b_j\in V_j,1\leqslant i\leqslant n,1\leqslant j\leqslant m\}$

(5)选择(σ)

设 r' 是模糊模式 R 上的一个模糊关系，Q 是一个 k 元模糊谓词，则选择操作定义如下：

$r=\sigma_Q(r')$，其中 Q 与传统谓词的区别是"="被置换成"$=_t$"

(6)投影(π)

投影是使用并操作将传统投影合成一个关系。设 r 是模糊模式 R 上的一个模糊关系，它的属性是 $A_1,A_2,\cdots,A_m$，$S=\{A_{i1},A_{i2},\cdots,A_{ik}\}$ 是 R 的子集，则 r 在 S 上的投影定义为：

$\pi_{A_{i1},A_{i2},\cdots,A_{ik}}(r)=\bigcup\{S(A_{i1},A_{i2},\cdots,A_{ik})\}$

(7)自然连接(И)

设 r 是模糊模式 $R=\{U_1\times U_2\times\cdots\times U_m\}$ 上的一个模糊关系，s 是模糊模式 $S=\{V_1/V_2/\cdots/V_n\}$ 的一个模糊关系。$R\cap S=P$ 且 $P\neq\Phi$，则 r 和 s 的自然连接定义为：

$r\text{И}s=\{\tau|\exists\ \tau_1\in r,\exists\ \tau_2\in s(\forall\ K'\in P(\tau_1(K')=_t\tau_2(K')))\wedge\tau(R-P)=\tau_1(R-P)\wedge\tau(P)=(\min(\pi_r(u_i),\pi_s(u_i))/u_i : u_i\in P)\wedge\tau(S-P)=\tau_2(S-P)\}$。其中 $i=1,2,\cdots,t$，这里 t 是 U_i 的模糊值域的维数。

四、结论

上面，我们已经给出了模糊的关系演算语言和关系代数操作的定义，从这些定义来看，显然可以得到如下定理：

定理 1 模糊的第一范式(1NF)在并、交、差、笛卡尔积、投影、选择和自然连接操作下是封闭的。

定理 2 如果 E 是一个模糊的 1NF 关系代数表达式，则必定存在一个与之等价的安全模糊的 1NF 元组关系演算表达式 E'。

定理 3 如果 E 是一个安全模糊的 1NF 元组关系演算表达式，则存在一个与之等价的模糊的 1NF 关系代数表达式 E'。

定理 4 对于每一个安全模糊的 1NF 元组关系表达式 E，都存在一个与之等价的安全模糊域关系演算表达式 E'。

定理 5 对于每一个安全的模糊域关系演算表达式 E，都存在着与之等价的模糊的 1NF 关系代数表达式 E'。

说明：这些定理的证明可以由传统的关系数据库理论中类似的定理的证明进行构造，其差异是"="被置换成"$=_t$"。

参 考 文 献

1 何新贵．模糊数据库中的语义距离及模糊视图．计算机学报，1989，10：757～764

2 Takahashi Y. Fuzzy Database Query Languages and Their Relational Completeness Theorem. IEEE Trans. on Knowledge and Data Engineering, 1993, 5(1): 122～125

3 Ullman J D. Principle of Database Systems. Computer Science Press, 1980

4 Yu Hairong, Ramer A. A Combined Approach to Uncertain Data Analysis. Advance in Intelligent Data Analysis (IDA-97), LNCS 1280, 1997

(上接第 58 页)

4 Pawlak Z, Orlowsak E. Expressive power of knowledge representation. Intl. J. of Man-Machine Studies, 1984(20)

5 Slowinski R, Stefanowski J. 'RoughDAS' and 'RoughClass' software implementations of the rough sets approach. In: Slowinski R ed. Intelligent Dicision Support: Handbook of Applications and Advances of rough sets Theory. Kluwer Acasemic Pub., 1992

6 Kerber R. ChiMerge: Discretization of Numeric Attributes. In: Proceedings of 10th National Conf. on Artificial Intelligence (AAAI 92). MIT Press, 1992

7 史忠植．高级人工智能．科学出版社，1997