

可能性分布-模糊关系模型的数据库探讨*

洪 国 彬

〔计算机科学(电脑)系〕

摘要 本文利用变域 Fuzzy 集来表示可能性分布-模糊关系模型,并对由此产生的若干模型构成数据库进行讨论.

关键词 可能性分布,模糊数据库,变域 Fuzzy 集,模糊查询

0 引言

传统的数据库只能用来处理精确而又良构的数据.现实世界,特别是复杂大系统中存在着大部分模糊数据,甚至是不完整的数据,为了收集、检索和处理现实世界中的模糊数据,提出了模糊数据库.自从 L. A. Zadeh 于 1964 年提出模糊集合概念,以及后期的近似推理理论以来,许多国家的部分计算机工作者,都在可能性分布理论的基础上,致力于模糊关系数据库的理论探讨,有的人还实现了简单、实用,但功能和结构又不完整的模糊关系数据库系统.但是,所有的模糊关系数据库系统都是基于模糊关系模型,或按类似关系的集中模型,或可能性分布-关系模型之一来实现的,可是表达现实世界的数据库能力却不够.然而,基于可能性分布-模糊关系模型的数据库表达和处理现实世界的信息,其能力是充分的,但是理论的处理或实现都很难,特别是可能性分布-模糊关系模型的表达也莫衷一是.因此,对此种模型,以及由此模型建立的数据库的理论探讨是必要的.本文着重对这种模型的建造进行分析及定义,并简单地对模糊查询进行设想.

1 预备知识

可能性分布-模糊关系模型,就是用可能性分布来表示数据本身以及数据之间所具有的模糊性的一种数据模型. L. A. Zadeh 的 Fuzzy 集是对论域 U 而言的, Zadeh 没有研究论域 U 对 Fuzzy 集的影响,以及 U 本身为 Fuzzy 集的情况.论域 U 为模糊集,基于 U 的 Fuzzy 即建立在模

* 本文 1991-04-23 收到.

糊概念上的 Fuzzy 集,数据之间关系的模糊性,体现的一个方面就是变域 Fuzzy 集.下面介绍变域 Fuzzy 集及其性质.

定义 1 设 U 是论域, x 是 U 的 Fuzzy 子集, 定义 T 是 x 的 Fuzzy 集, 如果对 T 的讨论是在 Fuzzy 集 x 上进行的, 记 $T(x)$, x 称为 T 的直接论域.

定义 2 设 U 是一个论域, x, A 都是 U 的模糊子集. 那么, 关于 x 的 Fuzzy 集 $A(x)$ 可定义为: $A(x) = g(x, A)$, 其中, g 为二元函数. 这样, Fuzzy 集就同论域 x 联系起来了, 称 x 为 A 的直接论域. 定义 2 中, 我们对 g 作如下约定: 若 $B(x) = g(x, B)$, 则 $\forall x \in U$; 若 $\mu_x(u) = 0$, 则 $\mu_{B(x)} = 1$. 当 x 为 U 的分明子集时, 这一约定显然为 $\forall u \in U$, 如果 $u \notin x$, 则 $\mu_{B(x)}(u) = 1$. 这就是说明, 对于直接论域以外的任意元素 u , 我们认为 u 当然具有模糊性质 B .

定义 3 $\forall u \in U, \mu_{T(x)}(u) = 1 \wedge (1 - \mu_x(u) + \mu_T(u))$ 这一定义诚然符合上述的约定.

2 基于可能性分布-模糊关系模型数据库

在 Fuzzy 推理中, 事实通常表示为: “目标的属性是值”. 例如: “高敏的跳水水平很高”, “诸葛亮的用兵如神” 等等, “高敏” “诸葛亮” 是目标, “跳水水平” “用兵” 是属性, “很高” “如神” 是值. 形式上, 上述事实可表示为: “ $V(x)$ 是 A ”, V 是属性, x 是目标, A 是值. 这里 A 可以是 Fuzzy 子集, 而目标 x 必须是精确概念. 因此, Zadeh 的 Fuzzy 集无法表示诸如 “红西红柿是熟的”, “浓茶水是苦的” 这类命题. 这是由于在上述命题中, 目标 “红西红柿” “浓茶水” 是一模糊概念. 因此, 本文就利用变域 Fuzzy 集来处理这类问题, 从而构造出这类可能性分布-模糊关系模型.

假定 U 是所有值的集合, 命题 “ $V(x)$ 是 A ” 在 U 上的可能性分布为: $\pi_{V(x)}(u) = \mu_{A(x)}(u)$. 式中, Fuzzy 集 x 的论域是 U , $\mu_{A(x)}(u)$ 是 u 在 $A(x)$ 中的隶属程度. $\pi_{V(x)}(u)$ 是由事实 “ $V(x)$ 是 A ” 给出的, $V(x)$ 取值 u 的可能性.

为了描述模糊数据之间关系的模糊性, 事实表示的规则一般可表示为, 如果 $V_1(x_1)$ 是 A_1 , 那么 $V_2(x_2)$ 是 A_2 “这类形式的复合命题. 假定 $V_1(x_1)$ 和 $V_2(x_2)$ 分别取值于集合 (论域) U_1 和 U_2 . 则上述规则可导致在 $U_1 \times U_2$ 上有如下的可能性分布

$$\pi_{V_2(x_2)|V_1(x_1)}(u_1, u_2) = \min(1, 1 - \mu_{A_1(x_1)}(u_1) + \mu_{A_2(x_2)}(u_2)).$$

模型(1)

对于复合规则的可能性分布可以采用下述方式求得: 如果 $V_1(x_1), V_2(x_2), \dots, V_n(x_n)$ 分别是在基集 U_1, U_2, \dots, U_n 上取值, 则复合规则 “ $V_1(x_1)$ 是 A_1 且 $V_2(x_2)$ 是 A_2 且 \dots 且 $V_n(x_n)$ 是 A_n ” 可导致在 $U_1 \times U_2 \times \dots \times U_n$ 上的联合可能性分布. 即

$$\pi_{V_1(x_1)V_2(x_2)\dots V_n(x_n)}(u_1, u_2, \dots, u_n) = \min_{i=1, \dots, n} [\mu_{A_i(x_i)}(u_i)].$$

模型(2)

同样, 复合规则前件 “ $V_1(x_1)$ 是 A_1 或 $V_2(x_2)$ 是 A_2 或 \dots 或 $V_n(x_n)$ 是 A_n ” 可导致在 $U_1 \times U_2 \times \dots \times U_n$ 上如下的联合可能性分布

$$\pi_{V_1(x_1)V_2(x_2)\dots V_n(x_n)}(u_1, u_2, \dots, u_n) = \max_{i=1, \dots, n} [\mu_{A_i(x_i)}(u_i)].$$

模型(3)

借助于模型(1), (2), (3), 可以导出复合规则的可能性分布, 从而定义出可能性分布-模糊

关系的新模型. 假定 $V_1(x_1), V_2(x_2), \dots, V_n(x_n)$ 分别取值于 $Z_1, Z_2, \dots, Z_n; W_1(Y_1), W_2(Y_2), \dots, W_m(Y_m)$ 分别取值于 U_1, U_2, \dots, U_m .

1 模型 F_1 . 复合规则“如果 $V_1(x_1)$ 是 A_1 或 $V_2(x_2)$ 是 A_2 或……或 $V_n(x_n)$ 是 A_n , 那么 $W_1(Y_1)$ 是 B_1 ”可导 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1$ 上有如下的条件可能性分布. 即

$$\begin{aligned} & \pi_{W_1(Y_1) | V_1(x_1) V_2(x_2) \dots V_n(x_n)}(Z_1, Z_2, \dots, Z_n, U_1) \\ &= \min(1, 1 - H(Z_1, Z_2, \dots, Z_n) + \mu_{B_1(Y_1)}(u_1)), \end{aligned}$$

其中 $H(Z_1, Z_2, \dots, Z_n) = \max_{i=1, 2, \dots, n} [\mu_{A_i(x_i)}(Z_i)]$. 此时, 可进行以下定义.

定义 2.1 可能性-模糊关系 R 的一个模糊元组 t 是 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1$ 的一个元素 $(z_1, z_2, \dots, z_n, \mu_1)$ 与一个隶属度 μ 构成的二元序偶 $\langle (z_1, z_2, \dots, z_n, \mu_1), \mu \rangle$, 其中 z_i 取值于 $Z_i, i=1, 2, \dots, n, u_1$ 取值于 U_1 , 条件是 $W_1(Y_1)$ 是 B_1 .

2 模型 F_2 . 复合规则“如果 $V_1(x_1)$ 是 A_1 且 $V_2(x_2)$ 是 A_2 且……且 $V_n(x_n)$ 是 A_n , 那么, $W_1(Y_1)$ 是 B_1 或 $W_2(Y_2)$ 是 B_2 或……或 $W_m(Y_m)$ 是 B_m ”可导致在 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 上有如下的条件可能性分布

$$\begin{aligned} & \pi_{W_1(Y_1) W_2(Y_2) \dots W_m(Y_m) | V_1(x_1) V_2(x_2) \dots V_n(x_n)}(Z_1, \dots, Z_n, u_1, \dots, u_m) \\ &= \min(1, 1 - H(Z_1, \dots, Z_n) + G(u_1, \dots, u_m)). \end{aligned}$$

其中, $H(Z_1, \dots, Z_n) = \min_{i=1, \dots, n} [\mu_{A_i(x_i)}(Z_i)], G(u_1, \dots, u_m) = \max_{j=1, \dots, m} [\mu_{B_j(Y_j)}(u_j)]$. 此时, 可进行以下定义.

定义 2.2 可能性分布-模糊关系 R 的一个模糊元组 t 是 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 的一个元素 $(z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m)$ 与一个隶属度 μ 构成的二元序偶 $\langle (z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m), \mu \rangle$, 其中 z_i 取值于 $Z_i, i=1, 2, \dots, n, \mu_j$ 取值于 $U_j, j=1, 2, \dots, m$, 条件是 $W_1(Y_1)$ 是 B_1 或 $W_2(Y_2)$ 是 B_2 或……或 $W_m(Y_m)$ 是 B_m .

3 模型 F_3 . 复合规则“如果 $V_1(x_1)$ 是 A_1 或 $V_2(x_2)$ 是 A_2 或……或 $V_n(x_n)$ 是 A_n , 那么, $W_1(Y_1)$ 是 B_1 且 $W_2(Y_2)$ 是 B_2 且……且 $W_m(Y_m)$ 是 B_m ”可导致如下的条件可能性分布

$$\begin{aligned} & \pi_{W_1(Y_1) \dots W_m(Y_m) | V_1(x_1) V_2(x_2) \dots V_n(x_n)}(Z_1, \dots, Z_n, u_1, \dots, u_m) \\ &= \min(1, 1 - H(Z_1, \dots, Z_n) + G(u_1, \dots, u_m)). \end{aligned}$$

其中, $H(Z_1, Z_2, \dots, Z_n) = \max_{i=1, \dots, n} [\mu_{A_i(x_i)}(Z_i)], G(u_1, u_2, \dots, u_m) = \min_{j=1, \dots, m} [\mu_{B_j(Y_j)}(\mu_j)]$. 此时, 可进行以下可定义.

定义 2.3 可能性分布-模糊关系 R 的一个模糊元素 t 是 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 的一个元素 $(z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m)$ 与一个隶属度 μ 构成的二元序偶 $\langle (z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m), \mu \rangle$, 其中 z_i 取值于 $Z_i, i=1, 2, \dots, n, \mu_j$ 取值于 $U_j, j=1, 2, \dots, m$, 条件是 $W_1(Y_1)$ 是 B_1 且 $W_2(Y_2)$ 是 B_2 且……且 $W_m(Y_m)$ 是 B_m .

4 模型 F_4 . 复合规则“如果 $V_1(x_1)$ 是 A_1 且 $V_2(x_2)$ 是 A_2 且……且 $V_n(x_n)$ 是 A_n , 那么 $W_1(Y_1)$ 是 B_1 且 $W_2(Y_2)$ 是 B_2 且……且 $W_m(Y_m)$ 可导致在 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 上有如下条件可能性分布

$$\begin{aligned} & \pi_{W_1(Y_1) W_2(Y_2) \dots W_m(Y_m) | V_1(x_1) V_2(x_2) \dots V_n(x_n)}(Z_1, Z_2, \dots, Z_n, u_1, \dots, u_m) \\ &= \min(1, 1 - H(Z_1, Z_2, \dots, Z_n) + G(u_1, u_2, \dots, u_m)). \end{aligned}$$

其中, $H(Z_1, Z_2, \dots, Z_n) = \min_{i=1, \dots, n} [\mu_{A_i}(Z_i)]$,

$$G(u_1, u_2, \dots, u_m) = \min_{j=1, \dots, m} [\mu_{B_j}(u_j)].$$

此时, 可进行以下定义.

定义 2.4 可能性分布-模糊关系 R 的一个模糊元素 t 是 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 的一个元素 $(z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m)$ 与一个隶属度 μ 构成的二元序偶 $\langle (z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m), \mu \rangle$, 其中 z_i 取值于 $Z_i, i=1, 2, \dots, n, \mu_j$ 取值于 $U_j, j=1, 2, \dots, m$, 条件是 $W_1(Y_1)$ 是 B_1 且 $W_2(Y_2)$ 是 B_2 且 \dots 且 $W_m(Y_m)$ 是 B_m .

按照上述方法, 其它复合规则的表示所构成的可能性分布-模糊关系模型也可构造出来了. 对这类模型的进一步推广, 也可得出可能性分布-按类似关系的集中模型. 同时, 综合前述关于事实表示的复合规则, 可以得到下列基于可能性分布-模糊关系模型的定义. 令 Ω 是所有 Fuzzy 子集的最大论域, 是一个分明集, $V_i(x_i)$ 是 Fuzzy 集合 $i=1, 2, \dots, n$ 表示属性的论域, $W_j(Y_j)$ 也是 Fuzzy 集合, $j=1, 2, \dots, m$, 表示模糊数据之间的条件. 则关系模式 $Z_1, Z_2, \dots, Z_n, u_1, u_2, \dots, u_m$ 上的一个模糊关系 R 被定义为 $n+m$ 之隶属函数 $u, \mu \rightarrow [0, 1]$. 此时, 模糊关系 R 的一个模糊元组 t 是 $Z_1 \times Z_2 \times \dots \times Z_n \times U_1 \times U_2 \times \dots \times U_m$ 的一个元素 $(z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m)$ 与一个隶属度 μ 构成的二元序偶 $\langle (z_1, z_2, \dots, z_n, u_1, u_2, \dots, u_m), \mu \rangle$, 其中 z_i 取值于 $Z_i, i=1, 2, \dots, n, u_j$ 取值于 $U_j, j=1, 2, \dots, m$, 条件是 $W_j(Y_j), j=1, 2, \dots, m$.

在传统的关系数据库中, 常用二维表格来表示关系, 其中列表示属性, 行表示元组, 由此出发, 在上述的定义下, 可以把可能性分布-模糊关系视为三维表, 因为它是二维的通卡尔积制 $[0, 1]$ 闭区间的条件映射来定义的.

3 模糊查询处理

对于通常的关系模型, 可以考虑在查询检索条件中使用模糊集. V. Yahani 以 SEQUEL 为基础揭示了模糊查询的处理方法, 基础语言以 Prolog 为代表. 对于模糊数据之间的依赖关系问题, 模糊查询块的结构也可以采用类似情况.

SELECT<List of attributes>

FROM<List of relations>

WHERE<Conditions>

WITH<degree>

一个模糊查询块的执行结果是一个模糊关系, 它的结构和内容都由该模糊查询块决定, 结果关系的属性由 SELECT 语句的<List of attributes>说明, 这些属性值从 FROM 语句中的<List of relations>指明的关系中选择; WHERE 语句中的<Conditions>是个模糊逻辑表达式, 包括属性名、常数(模糊的或非模糊的)和模糊逻辑运算符, WITH 语句中<degree>是一个阈值.

模糊关系模型, 按类似关系的集中, 可能性分布-关系模型, 都可看作是可能性分布-模糊关系模型的特殊情况; 然而, 与近似推理密切相关的模糊关系代数运算, 如投影和特指的公式化等, 还未能解决, 而作为模糊关系数据库构成中, 冗余就伴随着数据库的构成而存在, 还有待于广大的计算机工作者的努力.

4 结束语

可能性分布-模糊关系模型是按照现实世界的数据库表示建立的数据模型,因此,同样存在着和另外三种模糊数据模型类似问题,甚至更加严重.作为理论处理或实现,模型的模糊函数依赖和关系计算将是一个非常困难的问题.但是,机器模仿人的智能,就是要模拟人类对模糊数据的处理能力,在传统的关系数据库基础上,发展模糊数据库是没有潜力的,而模糊信息处理的研究才刚刚开始,有待于今后的大力发展.

参 考 文 献

- [1] 马野元秀,模糊集合与计算机软件,计算机科学,4(1990).
- [2] Zadeh, L. A. 近似推理的理论,计算机科学,2—4(1990).
- [3] 李 凡、高雅卿,可能性理论在模糊推理中的应用,模式识别与人工智能,12,3(1989).
- [4] 刘东波、李德毅,模糊关系数据库系统 FPDB,小型微型计算机系统,11,2(1990).
- [5] 高雅卿,变域 Fuzzy 集及其在知识工程中的应用,计算机工程,5(1990).
- [6] 李 凡,知识工程中各种模糊度量方法及其相互关系,计算机工程,3(1990).

An Inquiry into the Distribution of possibility and the Database Derived from Some Models of Fuzzy Relationship

Hong Guobin

(Department of Computer Science)

Abstract By making use of the fuzzy set of metadomain, the author represents the distribution of possibility as the model of fuzzy relationship; and inquires into the database derived from some of these models.

Key words distribution of possibility, fuzzy database, fuzzy set of metadomain, fuzzy finding