

面向对象数据库中的模糊类层次建模

郑学军 王春森

(复旦大学计算机科学系 上海 200433)

摘 要 本文提出了一种建立面向对象数据库中模糊类层次模型的方法。利用属性与类的相关和对象与类间的语义距离来建立对象与类的关系；利用属性与类的相关和类与子类间的语义距离来建立类与子类的关系。同时，讨论了模糊类层次中多重继承的问题。这种模型扩充了现有的面向对象数据模型，并与其兼容。

关键词 面向对象数据库, 模糊, 类层次, 模糊等价, 语义距离。

FUZZY CLASS HIERARCHY MODELLING IN OBJECT-ORIENTED DATABASE

ZHENG XUEJUN WANG CHUNSEN

(Department of Computer Science, Fudan University, Shanghai 200433)

Abstract This paper presents an approach to establishing fuzzy class hierarchy model in object-oriented database. The relationship between the object and the class is established through the relevance of the attribute to the fuzzy class and the semantic distance between the object and the class. The relationship between the class and the sub-class is established through the relevance of the attribute to the fuzzy class and the semantic distance between the class and the sub-class. Simultaneously, the problem of multiple inheritance in class hierarchy is considered. This model extends existing object-oriented data models and is compatible with it.

Keywords Object-oriented database, fuzzy, class hierarchy, fuzzy equal, semantic distance.

一、引 言

当前, 计算机的应用领域在不断扩大, 数据库技术的应用已从 MIS 发展到非 MIS 以外的许多领域, 如 CAD、CAM、GIS、GIMS、CASE 等等。它们所涉及的数据已从简单的字符、

本文 1995 年 9 月 13 日收到。郑学军, 硕士研究生, 研究方向: 面向对象、软件测试。王春森, 副教授, 研究方向: 软件工程。

整型、浮点型数据发展到图形、图象、声音、超文本、空间几何数据和知识型数据等复杂数据类型,并且它们之间还增加了继承、包含等复杂的关系。这就使得曾在 MIS 和相关领域中获得了成功的关系数据库很难处理,而最近兴起的面向对象数据库则能较好地解决这些问题。数据库研究的主要目标之一是在模型中不断增加语义,传统数据模型擅长于表示精确的和有良好结构的数据。然而在现实世界中,特别是在使用知识比较多的应用系统中(如人工智能系统、数据库的耦合等),最突出的一个问题就是范式间知识的表示不匹配。这主要是由于大多数知识的表达是模糊的,用传统的方法不易表示清楚。解决这个问题一个办法是利用模糊逻辑来扩充数据模型。本文以模糊逻辑中的模糊等价关系为基础来扩充面向对象数据模型。

本文考虑到模糊数据对面向对象数据模型中类层次的影响,提出了一种模糊的面向对象数据模型。这种数据模型与现有的面向对象数据模型兼容,保留了逻辑范式中的基本属性,同时包含了类层次中的不确定性。对象属性值之间通过象“逻辑与”、“逻辑或”、“逻辑异或”这样的逻辑操作符连接起来,数据的语义因不同的数据库操作而不同。面向对象数据模型中的类与对象(或子类)的关系是通过属性与类的相关和类与对象(或子类)的语义(概念)距离来建立的。假定模糊定义域中的所有属性都遵循模糊等价关系。

定义 模糊等价关系

如果模糊属性定义域 Z 上的任意两个值 x, y 满足:

$$[1] \text{ 自反性 } \forall x: x \in Z \quad S(x, x) = 1$$

$$[2] \text{ 对称性 } \forall x, \forall y: x, y \in Z \quad S(x, y) = S(y, x)$$

$$[3] \text{ 传递性 } \forall x, \forall y, \forall z: x, y, z \in Z \quad S(x, y) \geq \text{MAX}(\text{MIN}(S(x, z), S(z, y)))$$

则属性满足 Z 上的模糊等价关系。这里, $S(x, y)$ 表示同一属性的两个属性值的模糊等价(相似)程度,最大值为 1,即两个属性值相同。

二、模糊类层次

面向对象数据模型把现实世界表示成对象的集合,其基本的构造是对象,由标识符和状态来表示。不同的对象由不同的标识符区分,对象的状态由它的属性值组成,具有相似属性的一组对象形成一个对象类(简称类)。类由它的数据和实现方法定义,把它们封装在一起。对象属性可以是简单属性,也可以是组合属性。组合属性又由简单属性和组合属性组成,这就形成了类组合层次,即类聚集。

一个类可以派生出子类,子类又可以派生新的子类,这就形成了类-子类关系(类继承),即类层次。面向对象数据模型的范式就是通过使用类层次和类聚集来建立现实世界的模型的。

属性的定义域 Z 是属性可能的取值范围,属性的值域是属性允许的取值范围。例如,年龄 age 是一个属性,假定它的定义域是 $0 < Z(\text{age}) < 200$, 如果有一个类 student , 这个类中的一个属性 age 的值域可能是: $4 < D(\text{age}) < 40$, 很显然,定义域包含值域。所以,一个属性在其定义域内满足模糊等价关系必在其值域内满足模糊等价关系。

模糊类层次和传统类层次的一个显著区别就是在类间没有精确的界限划分。这种模糊性是由形成类的对象属性的不精确性引起的,例如若考虑“年轻学生”类,很显然,“年轻”是

模糊的。我们可以用模糊逻辑来解决这个问题,假如“年轻学生”类的隶属函数为:

当 $4 < x \leq 25$ 时, $\mu(x) = 1$;

当 $25 < x < 40$ 时, $\mu(x) = 1 / (1 + ((x - 25) / 10)^2)$ 。

从上式可以看出,当年龄在 4 至 25 岁时,属于“年轻学生”类的程度为 1; 当年龄为 25, 26, 27, ..., 40 岁时,属于“年轻学生”类的程度分别约为 1, 0.999, 0.996, ..., 0.307。如果取阈值为 0.300, 就能说年龄在 4 至 40 岁的学生都属于“年轻学生”类,只是年龄不同,属于该类的隶属程度有所不同。

类中的每个属性都有一个形式化的值域以便区分,即属性与类相关。因为定义域中的属性值是通过模糊等价关系连在一起的,即使一个对象的属性值与类的值域中的值不同,也可能是类中的成员,所以类中的实际属性值可能与形式化值域中的理论值不同。这些值模糊等价的程度就反映了这个对象属于这个类的隶属度。这样,对象与对象之间、类与类之间没有精确的界限划分,在面向对象数据库中就形成了模糊类层次。所以,在建立模糊类层次模型时,应考虑属性与类的相关。

除了考虑属性与类的相关外,还应考虑对象与类、子类与父类间的语义距离。在模糊类层次中,一个对象是否属于一个类,一个类是否是一个类的子类,考虑两者间存在的语义距离是很重要的。一个类属于一个类的子类的隶属程度随着继承层次的增加而增加,称为“强 ISA 关系”。相反,随着继承层次的增加而减少,称为“弱 ISA 关系”。

三、建立模糊类层次的模型

为了叙述清楚起见,采用下列表示方式:

$A(O) = \{a_1, a_2, \dots, a_n\}$ 表示类 O 的属性为: a_1, a_2, \dots, a_n ;

$FC(O) = \{C_1, C_2, \dots, C_n\}$ 表示类 O 的直接父类为: C_1, C_2, \dots, C_n ;

$MAX()$, $MIN()$ 分别为最大、最小函数。

1. 对象-类关系

考虑类 O 中具有属性 a_1, a_2, \dots, a_n 的对象 O_i , 从前述的属性与类的相关和对象与类的语义距离来看,我们给出对象 O_i 属于类 O 的隶属函数:

$$\mu_C(O_i) = f(g(R(a_j, O), I(D(a_j)/O_i(a_j)))) \quad (1)$$

这里, $R(a_j, O)$ 为相关程度,即属性 a_j 和类 O 的相关程度,通常由用户提供;

$D(a_j)$ 为类 O 的属性 a_j 的值域;

$O_i(a_j)$ 为对象 O_i 的属性 a_j 的值域;

$I(D(a_j)/O_i(a_j))$ 为包含程度,即对同一属性 a_j , 类 O 包含对象 O_i 的程度(对象 O_i 的属性 a_j 的值域上的值和类的属性 a_j 的值域上的值的模糊等价程度)。下面将讨论计算包含程度的几种情况;

函数 f 反映了对象与类间存在的连接类型;

函数 g 为类中 n 个属性的集成。

函数 f 和 g 既可从父类继承,也可在类中定义。

现在来讨论包含程度的几种情况:

(1) 如果 $O_i(a_j) \subseteq D(a_j)$, 则 $I(D(a_j)/O_i(a_j)) = 1$;

如果 $O_i(a_j)$ 为空, 则 $I(D(a_j)/O_i(a_j)) = 0$ 。

(2) 如果 $O_i(a_j)$ 的值唯一, 且 $O_i(a_j) \not\subseteq D(a_j)$, 则 $I(D(a_j)/O_i(a_j)) = \text{MAX}(S(O_i(a_j), y)), \forall y: y \in D(a_j)$ 。

(3) 如果 $O_i(a_j)$ 的值多于一个, 且 $O_i(a_j) \not\subseteq D(a_j)$, 则关于 $O_i(a_j)$ 可能有三种不同的解释。当对象的一个属性有多个值时, 这和数据库中出现的不同语义是一致的, 属性值间通过“逻辑与”、“逻辑或”和“逻辑异或”连接起来, 包含程度 $I(D(a_j)/O_i(a_j))$ 就依赖于这三种属性语义。

① “逻辑与”语义

在“逻辑与”中, 一个属性有多个值, 这些值都同时为真, “逻辑与”语义一般由数据嵌套引起。包含程度为:

$$I(D(a_j)/O_i(a_j)) = \text{MIN}(\text{MAX}(S(x, y)), T(O_i(a_j)))$$

这里, $\forall x: x \in O_i(a_j), \forall y: y \in D(a_j)$; T 是一个阈值, 表示对象属性值之间模糊等价程度的最小值, 即 $\text{MIN}(S(x, z)), \forall x, \forall z: x, z \in O_i(a_j)$ 。这说明无论属性值中元素与定义域中元素等价程度如何, 包含程度都不可能超出对象属性值自身元素间存在的模糊等价程度。

② “逻辑或”语义

在“逻辑或”中, 一个属性有多个值, 这些值中的部分或全部可能同时为真, 当属性值是通过“逻辑或”连接起来时, 和“逻辑与”相比, 数据的模糊性增加了。假如对象的属性 a 有两个值, 如 v_1, v_2 , 在“逻辑或”中, 属性值域可能的情况为 $\{v_1\}, \{v_2\}, \{v_1, v_2\}$, 所以包含程度为:

$$I(D(a_j)/O_i(a_j)) = \text{MAX}(\text{MAX}(S(x, y)), \text{MIN}(\text{MAX}(S(x, z)), T(O_i(a_j))))$$

这里, $\forall x: x \in O_i(a_j), \forall y, \forall z: y, z \in D(a_j)$ 。

③ “逻辑异或”语义

在“逻辑异或”中, 一个属性有多个值, 这些值中的一个为真。当属性值是通过“逻辑异或”连接起来时, 和“逻辑与”相比, 数据的模糊性增加了。包含程度为:

$$I(D(a_j)/O_i(a_j)) = \text{MAX}(S(x, y))$$

这里, $\forall x: x \in O_i(a_j), \forall y: y \in D(a_j)$ 。

2. 类-子类关系

假定类 C 和它的父类 C_i 的属性分别为:

$$A(C) = \{a_1, a_2, \dots, a_k, a_{k+1}, \dots, a_n\} \quad A(C_i) = \{a_1, a_2, \dots, a_k\}$$

类 C 的前 k 个属性是从父类 C_i 中继承而来的, 其余的为类 C 所特有的。类 C 是类 C_i 子类的隶属函数:

$$\mu_{c_i}(C) = f(g(R(a_j, C_i), I(D_{c_i}(a_j)/D_c(a_j)))) \quad (2)$$

这里, $R(a_j, C_i)$ 为相关程度, 即属性 a_j 和类 C_i 的相关程度, 通常由用户提供;

$D_{c_i}(a_j)$ 为父类 C_i 的属性 a_j 的值域;

$D_c(a_j)$ 为类 C 的属性 a_j 的值域;

$I(D_{c_i}(a_j)/D_c(a_j))$ 为包含程度, 即对同一属性 a_j , 父类 C_i 包含类 C 的程度(类 C_i 的属性 a_j 的值域上的值和类 C 的属性 a_j 的值域上的值的模糊等价程度)。下面将讨论计算包含程度的几种情况;

函数 f 反映了类与父类间存在的连接类型;

函数 g 为类中 n 个属性的集成。

(函数 f 和 g 既可从父类继承,也可在类中定义。)

求解包含程度 $I(D_{c_i}(a_j)/D_c(a_j))$ 时,与建立对象-类关系模型一样,也存在三种情况:

(1) 如果 $D_c(a_j) \subseteq D_{c_i}(a_j)$, 则 $I(D_{c_i}(a_j)/D_c(a_j)) = 1$; 如果 $D_c(a_j)$ 空, 则

$$I(D_{c_i}(a_j)/D_c(a_j)) = 0$$

(2) 如果 $D_c(a_j)$ 的值唯一, 且 $D_c(a_j) \not\subseteq D_{c_i}(a_j)$, 则

$$I(D_{c_i}(a_j)/D_c(a_j)) = \text{MAX}(S(x, y)), \forall x: x \in D_c(a_j), \forall y: y \in D_{c_i}(a_j)$$

(3) 如果 $D_c(a_j)$ 的值多于一个, 且 $D_c(a_j) \not\subseteq D_{c_i}(a_j)$, 则

$$I(D_{c_i}(a_j)/D_c(a_j)) = \text{MAX}(S(x, y)), \forall x: x \in D_c(a_j), \forall y: y \in D_{c_i}(a_j)$$

3. 多重继承的考虑

在面向对象方法中,多重继承是很重要的。多重继承是指一个类继承了多个父类的属性,这样就有可能引起继承冲突。为了解决这种冲突,提出了一种建立多重继承模型的方法。前面已提到,属性与类相关,这里再次利用这一特征来派生多重继承的成员值,引进几种表示形式:

$$IA(O) = TA(C_1) \cup TA(C_2) \cup \dots \cup TA(C_n)$$

表示类 O 从它的父类 C_1, C_2, \dots, C_n 继承而来的属性。

$$TA(O) = A(O) \cup TA(C_1) \cup TA(C_2) \cup \dots \cup TA(C_n)$$

表示类 O 的属性,包括它自身的属性和它从父类继承而来的属性。

$$RA(O) = \{r_1, r_2, \dots, r_n\}$$

表示类 O 中重新定义的继承属性的集合。

现在假设类 O 的直接父类 $FC(O) = \{C_1, C_2, \dots, C_n\}$, $C_i, C_j \in FC(O)$, $C_i \notin FC(C_j)$, $C_j \notin FC(C_i)$, $IA(O) = \{a_1, a_2, \dots, a_n, a_i, a_j\}$, $RA(O) = \{a_i, a_j\}$, $a_i \in A(C_i)$, $a_j \in A(C_j)$, 属性 a_i, a_j 的定义域相同,分两种情况讨论:

① 如果 $D_{c_i}(a_i) = D_{c_j}(a_j)$ 和 $D_c(a_i) = D_{c_i}(a_i)$, 则多重继承层次中不存在冲突,类 O 为类 C_i 子类的隶属函数 $\mu_{c_i}(O)$, 由属性 a_i 与类 O 的相关和类 O 与父类 C_i 间的语义距离确定。类 O 为类 C_j 子类的隶属函数 $\mu_{c_j}(O)$, 由属性 a_j 与类 O 的相关和类 O 与父类 C_j 间的语义距离确定。

② 如果 $RA(O) = \{a_k\}$, 且属性 a_i, a_j, a_k 的定义域相同,这里仅重新定义一个属性。不失一般性,假定 $D_c(a_k) \neq D_{c_i}(a_i) \neq D_{c_j}(a_j)$, 这是多重继承中的典型情况。属性在类中定义,也在父类中定义,重新定义的值域发生了冲突。为了解决这种冲突,除了考虑概念距离和属性语义以外, $\mu_{c_i}(O)$ 和 $\mu_{c_j}(O)$ 还分别依赖于包含程度 $I(D_{c_i}(a_i)/D_c(a_k))$ 和 $I(D_{c_j}(a_j)/D_c(a_k))$ 。

数据模型中的多重继承问题主要是由于属性的值域与类的相关引起。因为类层次中每个类都有一个属性的值域与类联系,类中的对象与从不同类层次中继承来的定义域或值域的不兼容无关,在多重继承中只关心值域,而不关心冲突的定义域,因为有冲突定义域的属性可以看作不同的类。

4. 模型中隶属函数的分析

前面讨论的类层次中的模糊性可能由各种原因引起,现在我们分析在一般情况下隶属函数中的函数 f 和 g 可能出现的几种情况。如果类层次中有多个类 $C_k (1 \leq k \leq m)$, 重新表达 (1) 式为:

$$\mu_{c_k}(O_i) = f_k(g_k(R(a_j, C_k), I(D_{c_k}(a_j)/O_i(a_j)))) \quad (3)$$

可以分两种情况讨论(3)式:

(1) 假定 $f_k(1 \leq k \leq m) = f$, $g_k(1 \leq k \leq m) = g$, $R(a_j, C_k) = 1(1 \leq k \leq m)$, 则

$$\mu_{c_k}(O_i) = f(I(D_{c_k}(a_j)/O_i(a_j)))$$

由此可见,在相关程度相同的情况下,模糊性是由于数据值的不精确性引起,根据数据库操作中的语义就可以使用(1)式来建立类层次中的对象-类关系。

(2) 假定所有的包含程度 $I(D_{c_k}(a_j)/O_i(a_j)) = 1$, (3) 式可改写为:

$$\mu_{c_k}(O_i) = f_k(g_k(R(a_j, C_k), 1))$$

① 如果 $g_k(1 \leq k \leq m) = g$, $R(a_j, C_k)(1 \leq k \leq m) = R$, 则 $\mu_{c_k}(O_i) = f_k(\text{CON})$

这里, CON 是一个常量, 仅利用函数 f 就可以建立类层次中两个不精确类的语义距离模型。

② 如果 $f_k(1 \leq k \leq m) = f$, $g_k(1 \leq k \leq m) = g$, 则 $\mu_{c_k}(O_i) = f(g(R(a_j, C_k), 1))(1 \leq k \leq m)$ 。

这样就可以建立仅依赖于属性与类相关的类层次模型。

四、在中医儿科诊断中的具体应用

现在,计算机广泛应用于各个方面,在医学上也有成功的应用,特别是医学方面的专家系统。鉴于中医知识表达的模糊性,使用本文提出的模型来建立一个诊断小儿肺部疾病的中医专家系统,限于篇幅,仅以极小的一部分为例来说明前面所述的方法。咳嗽是小儿肺部疾患的一种常见症候,有外感咳嗽和内伤咳嗽之分。外感咳嗽分为风寒咳嗽和风热咳嗽等;内伤咳嗽又分为痰热咳嗽、痰湿咳嗽、阴虚咳嗽和肺虚咳嗽等。针对每一种症候有不同的症候分析、治疗方法和方药。每种咳嗽的症候很多,比如:痰的多少、稀稠、发热程度、面、目、唇、口、舌、苔、情绪、大小便等,下面仅以三个属性作为例子来说明本文提出的建立模糊类层次模型的方法。数据库的概念模型如图1所示;属性痰、舌质和苔的模糊等价矩阵分别如表1、表2和表3所示。

对于类“痰热咳嗽”,假定它的属性值域为:

$$D_{\text{痰热咳嗽}}(\text{痰}) = \{\text{很多, 多}\}, D_{\text{痰热咳嗽}}(\text{舌质}) = \{\text{红}\}, D_{\text{痰热咳嗽}}(\text{苔}) = \{\text{黄}\}$$

1. 计算对象属性值的包含程度

在派生类中的对象之前,应考虑对象属性值的包含程度,其计算方法应分以下几种

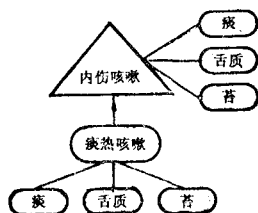


图1 数据库的概念模型

表1 痰的模糊等价矩阵

痰	很多	多	少	无
很多	1.0	0.9	0.85	0.75
多	0.9	1.0	0.95	0.75
少	0.85	0.95	1.0	0.75
无	0.75	0.75	0.75	1.0

痰 = {很多, 多, 少, 无}

表2 舌质的模糊等价矩阵

苔 质	红	淡红	淡嫩
红	1.0	0.8	0.6
淡红	0.8	1.0	0.6
淡嫩	0.6	0.6	1.0

舌质={红, 淡红, 淡嫩}

表3 苔的模糊等价矩阵

苔	很黄	黄	白腻	红
很黄	1.0	0.81	0.75	0.25
黄	0.81	1.0	0.75	0.25
白腻	0.75	0.75	1.0	0.25
红	0.25	0.25	0.25	1.0

苔={很黄, 黄, 白腻, 红}

情况:

(1) 如果一个对象, 它的属性“痰”值为“多”, 即 $O(\text{痰}) = \{\text{多}\}$, 因为 $O(\text{痰}) \subseteq D_{\text{痰热咳嗽}}(\text{痰})$, 所以 $I(D_{\text{痰热咳嗽}}(\text{痰})/O(\text{痰})) = 1$ 。

(2) 如果 $O(\text{痰}) = \{\text{少}\}$, 即一个对象的属性“痰”值只有一个, 并且 $O(\text{痰}) \not\subseteq D_{\text{痰热咳嗽}}(\text{痰})$, 所以 $I(D_{\text{痰热咳嗽}}(\text{痰})/O(\text{痰})) = \text{MAX}(S(\text{“少”, } x)) = \text{MAX}(0.85, 0.95) = 0.95$, 这里, $\forall x: x \in D_{\text{痰热咳嗽}}(\text{痰})$ 。

(3) 如果 $O(\text{痰}) = \{\text{少, 无}\}$, 即一个对象的属性“痰”值多于一个, 并且 $O(\text{痰}) \not\subseteq D_{\text{痰热咳嗽}}(\text{痰})$ 。

① “逻辑与”语义

阈值 $T(O(\text{痰})) = \text{MIN}(x, y) = \text{MIN}(1.0, 0.75, 0.75, 1.0) = 0.75$, 这里, $\forall x, \forall y: x, y \in O(\text{痰})$ 。所以, $I(D_{\text{痰热咳嗽}}(\text{痰})/O(\text{痰})) = \text{MIN}(\text{MAX}(S(x, y)), 0.75) = \text{MIN}(\text{MAX}(0.85, 0.75, 0.95, 0.75), 0.75) = 0.75$, 这里, $\forall x, \forall y: x \in O(\text{痰}), y \in D_{\text{痰热咳嗽}}(\text{痰})$ 。

② “逻辑异或”语义

$I(D_{\text{痰热咳嗽}}(\text{痰})/O(\text{痰})) = \text{MAX}(S(x, y)) = \text{MAX}(0.85, 0.75, 0.95, 0.75) = 0.95$, 这里, $\forall x, y: x \in O(\text{痰}), y \in D_{\text{痰热咳嗽}}(\text{痰})$ 。

③ “逻辑或”语义

阈值 $T(\text{痰}) = 0.75$, 所以 $I(D_{\text{痰热咳嗽}}(\text{痰})/O(\text{痰})) = \text{MAX}(\text{MAX}(S(x, y)), \text{MIN}(\text{MAX}(S(x, z)), T(O(\text{痰})))) = \text{MAX}(0.95, 0.75) = 0.95$, 这里, $\forall x: x \in O(\text{痰}), \forall y, \forall z: y, z \in D_{\text{痰热咳嗽}}(\text{痰})$ 。

2. 建立类层次的模型

(1) 对象-类关系

假定一个小儿的症状为: $O(\text{痰}) = \{\text{少}\}$, $O(\text{舌质}) = \{\text{淡红}\}$, $O(\text{苔}) = \{\text{白腻}\}$, 假定这三种属性与类“痰热咳嗽”的相关程度分别为:

$$R(\text{痰}, \text{痰热咳嗽}) = 2.0, R(\text{舌质}, \text{痰热咳嗽}) = 3.0, R(\text{苔}, \text{痰热咳嗽}) = 1.0$$

最大相关程度为 $R_{\text{MAX}} = R(\text{舌质}, \text{痰热咳嗽}) = 3.0$, 假定函数 f 是取所有值的平均值 $f(x_1, x_2, \dots, x_n) = (x_1 + x_2 + \dots + x_n)/n$, 函数 g 是 $g(x, y) = y * (x/R_{\text{MAX}})$, 因此给出对象属于类“痰热咳嗽”的隶属度为:

$$\begin{aligned} \mu_{\text{痰热咳嗽}}(O) &= f(g(2.0, 0.95), g(3.0, 0.8), g(1.0, 0.75)) \\ &= f(0.95 * (2.0/3.0), 0.8 * (3.0/3.0), 0.75 * (1.0/3.0)) \\ &= 0.57 \end{aligned}$$

这说明给出对象(痰少, 舌质淡红, 苔白腻)属于类“痰热咳嗽”的隶属度为 57%。

(2) 类-子类关系

假定另有一个类, 它的值域为: $D_{\text{内伤咳嗽}}(\text{痰}) = \{\text{少}\}$, $D_{\text{内伤咳嗽}}(\text{舌质}) = \{\text{红}\}$, $D_{\text{内伤咳嗽}}(\text{苔}) = \{\text{黄}\}$, 使用两种方式来建立类-子类关系模型。一是作为一种“强 ISA 关系”, 子类是父类的真正成员, 假定 $f(x_1, x_2, \dots, x_n) = \text{MAX}(x_1, x_2, \dots, x_n)$; 另外就是作为一种“弱 ISA 关系”, 子类很少依赖父类, 假定 $f(x_1, x_2, \dots, x_n) = \text{MIN}(x_1, x_2, \dots, x_n)$ 。同时, 假定相关程度分别为: $R(\text{痰}, \text{内伤咳嗽}) = 2.5$, $R(\text{舌质}, \text{内伤咳嗽}) = 0.5$, $R(\text{苔}, \text{内伤咳嗽}) = 1.5$, g 函数同 (1) 的假定。

$$\begin{aligned}\mu_{\text{内伤咳嗽}}(\text{痰热咳嗽}) &= f(g(2.5, \text{MAX}(0.75, 0.75)), g(0.5, 1.0), g(1.5, 1.0)) \\ &= f(0.75 * (2.5/2.5), 1.0 * (0.5/2.5), 1.0 * (1.5/2.5)) \\ &= f(0.75, 0.20, 0.60)\end{aligned}$$

对于“强 ISA 关系”:

$$\mu_{\text{内伤咳嗽}}(\text{痰热咳嗽}) = \text{MAX}(0.75, 0.20, 0.60) = 0.75$$

对于“弱 ISA 关系”:

$$\mu_{\text{内伤咳嗽}}(\text{痰热咳嗽}) = \text{MIN}(0.75, 0.20, 0.60) = 0.20$$

五、结 论

从前面的讨论中可以看出, 引起类层次模糊性的根源有两个: 一个是数据值(属性值), 另一个是语义。数据值的模糊性表明对象属性值不是属性的理论值域的子集, 仅与这些值模糊等价, 隶属于这些值的程度小于 1。模糊性的另一个重要根源是数据和类层次类型的语义。从数据的观点看, 模糊性由属性的多值引起; 从模型的观点看, 模糊性表明属性与类的相关程度和父类与类、类与对象间的语义距离。本文提出的数据模型较有效地处理了这两大根源所造成的模糊性。

总之, 本文提出了一种建立面向对象数据模型的方法, 通过这种方法, 可以更好地建立现实世界的模型。对象成员值可以动态地计算, 并给出一个应用实例说明具体的计算方法。

参 考 文 献

- [1] 王松年编,《模糊系统理论及其应用》讲义, 复旦大学计算机系, 1994.
- [2] B. P. Buckles, R. George and F. E. Petry, "Towards a Fuzzy Object Model," Proc. of the NAFIPS-91 Workshop on Uncertainty Modelling in the 90's, pp. 73~77, 1991.
- [3] D. Dubois, H. Prade and J. P. Rossazza, "Vagueness, Typicality and Uncertainty in Class Hierarchies", Int. J. of Intelligent Systems, pp. 167~183, 1991.