

可能世界内数值型不确定数据匹配模型^{*}

申德荣, 于 戈, 寇 月, 聂铁铮
(东北大学 信息科学与工程学院, 沈阳 110004)

摘 要: 针对数值型不确定数据的匹配进行研究。将存在的数值型不确定数据归纳为 range、or_set probability、unknown、negative 五种类型, 并给出了形式化定义; 提出匹配度和完备度概念, 可精确地描述不确定数据的匹配情况, 并给出了匹配模型; 提出基于 range 数据类型的匹配度和完备度为最小粒度, 定义了其他类型的不确定数据类型的匹配规则, 并给出了相应的匹配规则定义。最后通过实例分析, 验证了本匹配模型的有效性。

关键词: 不确定数据; 匹配模型; 数据库

中图分类号: TP39 **文献标志码:** A **文章编号:** 1001-3695(2008)09-2607-03

Matching model on numerical uncertain data in possible world

SHEN De-rong, YU Ge, KOU Yue, NIE Tie-zheng

(School of Information Science & Engineering, Northeastern University, Shenyang 110004, China)

Abstract: This paper focused on the matching model of uncertain data. Firstly, classified existing uncertain data into 5 types, and gave normal definitions, proposed the concepts of matching degree and completeness degree to describe their matching effectively and gave the matching model. Then, presented the idea of matching degree and completeness degree between range types as minimum granularity to define the matching rules between others, and defined their detail matching rules. Lastly, cases demonstrated the availability of the matching model.

Key words: uncertain data; matching model; database

在可能世界中, 不确定信息普遍存在, 如人们只知道某一属性取值的范围或可能的取值, 无法确定该属性的确切值。目前流行的数据管理一直是针对完备而精确的数据管理, 即普遍采用关系数据库管理数据信息。到目前为止, 已有许多有关不确定数据的管理, 典型的研究可分为两个阶段: a) 在 20 世纪 80 年代末到 90 年代初, 针对扩展关系数据库模型的研究, 重点面向 null 值和 maybe 值的不确定数据信息管理^[1~3]; b) 最近几年针对特定的不确定性如基于概率的面向 sensor network 的 interval 数据的研究^[4,5]、基于三维的移动对象的研究^[6,7]、针对不确定数据的聚集操作的研究^[8~11]和数据流 (data lineage) 管理^[12,13]等。尽管有关不确定信息的研究起步较早, 但已有研究都具有一定的局限性。因为大多是针对特定一种或几种不确定数据类型进行相对独立的研究, 没有可共享性和通用性, 也没有统一的规范标准, 导致不确定数据没有真正得到应用。另外, 由于目前支持的数据管理还主要局限于确定数据的管理, 约束了其在实际场景中的数据描述, 最终导致无法发现数据内更多的潜在信息。

随着计算机技术和网络技术的发展, 数据的丰富, 人们都得益于来自数据的潜在知识, 如应用人工智能推理和数据挖掘技术可挖掘数据深层的语义知识。但就数据本身表示能力来说, 已有精确的数据表达能力以及处理能力都显得苍白无力。因为在现实世界中, 数据本身是不确定的, 而目前所依赖的数据只是具有确定性的很少一部分, 或者说只是在确定的前提下

实现的数据处理。显然, 已有的确定的数据处理能力是不够的。数据表示和处理能力的局限性同样也体现在 Web 搜索过程中, 如目前 Web 提供的还处于只支持关键字的模糊查询范围内, 而普遍存在的不确定数据并没有得到有效利用。例如: 查询“支持率高于 50% 领导 的信息”, 只能得到“有 50% 的信息”, 而 40% ~ 60% 支持率的信息却没有被发现。可见, 面向广泛的具有不确定性的数据处理已迫在眉睫, 并且不确定数据处理的成熟和应用将推动数据处理更上一个新台阶。

本文针对数值型不确定数据描述与匹配进行研究, 目的是为用户提供更灵活精确的实体描述形式, 使其更接近于现实世界的抽象, 为人们有效地描述不确定数据提供一定的参考。同时, 通过实现不确定数据的匹配操作, 有效地发现潜在的知识, 并加以利用, 达到为人们提供更准确的决策知识的目的。

数值型不确定数据定义

目前, 有关非确定数据的定义和分类还没有严格的定义, Trio 中将数据分为 exact 和 inexact 两种^[14]。有关 inexact 数据的描述又有多种, 如不确定的数据、概率数据、模糊集数据、近似数据、不完备数据和不精确数据等。Motro^[15]将不确定信息分为不确定和不精确两类。不确定是指属性值的可信性, 如根据疑难病人的各项检测信息得出可能的病症, 其可信度不是 100%; 概率是指属性取某一值的概率, 如心脏病人中吸烟者占 75%, 非吸烟者占 25%, 其概率和为 1; 肥胖者得心脏病的可能

收稿日期: 2007-09-12; 修回日期: 2007-12-19 基金项目: 国家自然科学基金资助项目 (60673139, 60573090)

作者简介: 申德荣 (1964-), 女, 辽宁铁岭人, 教授, 博士, 主要研究方向为数据库、Web 数据管理、数据网格 (shenderong@ise.neu.edu.cn); 于戈 (1962-), 男, 教授, 博士, 主要研究方向为数据库、数据挖掘、数据流; 寇月 (1980-), 女, 助教, 博士研究生, 主要研究方向为 Web 数据管理、数据网格; 聂铁铮 (1980-), 男, 助教, 博士研究生, 主要研究方向为 Web 数据管理、Web 服务。

性为 0.7,但没有模糊度和为 1 的约束;该病人的年龄在 20~25 岁为近似数据;不完备的数据是指有信息丢失,如一部分病人的病例中没有记录病人的血型;不精确数据是指数据的取值可能是集合中的数据之一,等等。

文献 [15] 中除了将非确定的数据定义为不确定和不精确数据外,还包括不完备、模糊、不一致、不明确。其中除了不明确为语义模糊概念外,其他都涵盖了 Trio 中的定义。

归纳已有文章中讨论的数据不确定性,本文把不确定数据分为如下几类:

- a) 范围值 (range), 如 John 的年龄为 [37, 43];
- b) Or-set 值, 如 John 的年龄或者为 37 或为 43;
- c) 否定值 (negative), 如 John 的年龄不是 37;
- d) 未知值 (unknown), 如 null 值;
- e) 模糊值 (vague), 如 John 为年轻人;
- f) 概率值 (probability), 如 John 的年龄为 20 岁的概率是 0.6, 为 22 岁的概率是 0.4;
- g) 模糊集值 (fuzzy), 如 John 的工资是 high 的可能性为 0.6, low 的可能性为 0.3。

在可能世界中,这些不确定数据普遍存在,并交叠在一起,如 John 的年龄为 [37, 43] 的概率为 0.8,而取 [40, 45] 的概率为 0.2,这是将概率值与范围值结合表示属性值;再如 John 为年轻人的可能性为 0.6,而为老年人的可能性为 0.3,这是将 vague 值与 fuzzy 值结合表示数据的不确定性。

本文只针对数值型的不确定数据进行定义,并基于相应的定义给出相应的匹配规则。

定义 1 数值型不确定数据类型 (UN_M_Element_Type)

```
UN_M_Element_Type ::= or_set | range | probability | negative | unknown
range ::= rd - rd
p_element ::= rd | range
probability ::= p_element , p [ p_element , p ]
or_set ::= p_element [ , p_element ]
negative ::= O p_element
unknown ::= or_set | range | probability | negative |
p ::= 0..d *
rd ::= d * .. d *
O ::= | not in |
d ::= 0|1|2|3|4|5|6|7|8|9
```

定义 1 中包含了五种不确定数据类型 (range, or_set, probability, unknown, negative)。因为 vague 和 fuzzy 通常是描述非数值型的数据,其不包括在定义 1 中。Unknown 类型可以是确定数据类型中的任何类型。在数据的实际匹配处理过程中,将依据其预匹配的数据动态地指定其数据类型。

数值型不确定数据匹配模型

在已有数值型数据的匹配中,典型的为精确匹配,不适合不确定数据间的匹配。例如,查询价格 100 元的 U 盘,查询结果为包含有“100 元”“价格”“U 盘”的信息,而无法发现价格为 80~120 元的信息。为此,本章针对定义 1 中定义的五种特定数值型不确定数据类型间的匹配规则进行研究。首先给出匹配度定义,之后以 range 类型间的匹配度为基础,定义其他数值型不确定数据类型间的匹配规则。

数据类型间的匹配度

从外观上看,上面给出的五种不确定数据类型的定义差别很大,很难给出统一的匹配规则。但究其实质,它们都是由单值数据组成的。为此,本文提出将各种不确定数据类型进行转换,即均转换为由最小粒度的数据组成,使它们具有可比性,如均由单数据值组成。然而,由于单数据值粒度最小,势必影响数据间的匹配效率。为提高匹配效率,确定以 range 粒度为数据原子单位,单数据值看做是上界和下界相同的 range 类型值。这样,通过将所有数据离散化为 range 类型,并以 range 数据类型的匹配度为基本匹配单元,计算其他数据类型的匹配度。本文依据两数据值的交叠概率计算其匹配度。

由于不确定数据间的匹配结果也一定具有不确定性。为有效地描述提供者提供的信息满足请求的匹配程度,本文提出匹配度和完备度的概念,用来描述不确定数据的匹配结果的不确定性。匹配度描述提供者提供的数据与请求的数据信息的匹配程度,分为完全匹配、不匹配和部分匹配。如图 1 所示, V_r 和 V_p 分别为 range 类型的三种匹配关系。完备度表示提供者提供的数据满足请求者需求的数据信息的程度。两者匹配值越大越好。匹配度与完备度具体见定义 2。

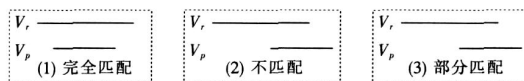


图 1 V_r 和 V_p 均为 range 类型的匹配示意图

定义 2 Range 数据类型间匹配度 ($m(V_r, V_p)$) 与完备度 ($c(V_r, V_p)$)。令 V_r 为用户的请求值, V_p 为提供者提供的数据值, V_r 和 V_p 都是 range 类型, 则:

a) V_r 与 V_p 的匹配度定义为 $m(V_r, V_p) = (\text{ran}(V_r) \cap \text{ran}(V_p)) / \text{ran}(V_p)$;

b) V_r 与 V_p 的完备度定义为 $c(V_r, V_p) = (\text{ran}(V_r) \cap \text{ran}(V_p)) / \text{ran}(V_r)$ 。

其中: $\text{ran}()$ 为区间函数; $\text{ran}(V_r)$ 和 $\text{ran}(V_p)$ 分别描述 V_r 和 V_p 的区间。例如, $V_r = [10 \sim 20]$, $V_p = [15 \sim 30]$, 则 $m(V_r, V_p) = 5/15 = 0.33$, $c(V_r, V_p) = 5/10 = 0.55$ 。

数值型不确定数据匹配模型

针对本文定义的五种不确定数据类型,基于定义 2 中的匹配度和完备度定义,定义各不确定数据类型间的匹配规则,为实现其不确定数值型数据间的匹配提供相应的遵循准则。

定义 3 数值型不确定数据匹配模型 (MM)。令 \mathcal{U} 为不确定数据类型集合, RU 为匹配规则集, 则不确定数据匹配模型定义为 $MM = \{ \mathcal{U}, RU \}$ 。其中: $RU = \{ (R_{(T_i, T_j)}, m(V_{T_i}, V_{T_j}), c(V_{T_i}, V_{T_j})) \mid T_i, T_j, V_{T_i} \text{ 和 } V_{T_j} \text{ 分别为类型 } T_i \text{ 和 } T_j \text{ 的数据实例 } j; R_{(T_i, T_j)} \text{ 为数据类型 } T_i \text{ 与 } T_j \text{ 间的匹配规则名称; } m(V_{T_i}, V_{T_j}) \text{ 为 } V_{T_i} \text{ 与 } V_{T_j} \text{ 的匹配度; } c(V_{T_i}, V_{T_j}) \text{ 为 } V_{T_i} \text{ 与 } V_{T_j} \text{ 的完备度} \}$ 。

本文中, $\mathcal{U} = \{ \text{range, or_set, probability, unknown, negative} \}$, $RU = \{ R_{(\text{range-range})}, R_{(\text{range-single})}, R_{(\text{range-or_set})}, R_{(\text{range-probability})}, R_{(\text{or_set-range})}, R_{(\text{or_set-or_set})}, R_{(\text{or_set-probability})}, R_{(\text{probability-range})}, R_{(\text{probability-or_set})}, R_{(\text{probability-probability})} \}$ 。

以 range 数据类型匹配度和完备度为基础,各种数据类型的匹配规则具体定义如下:

1) V_r 为 range 类型的匹配规则

规则 1 $R_{(range-range)}$ 。令 V_r, V_p 均为 range 类型,则 V_r 和 V_p 的匹配度和完备度具体定义同定义 2。其中包括 V_p 为单值类型,即 $ran(V_p) = 1$ 。

规则 2 $R_{(range-or_set)}$ 。令 V_r 为 range 类型, V_p 为 or_set 类型, $V_p = \{V_{p_1}, V_{p_2}, \dots, V_{p_n}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n m(V_r, V_{p_i}) / n$, $c(V_r, V_p) = \sum_{i=1}^n c(V_r, V_{p_i}) / n$ 。

规则 3 $R_{(range-probability)}$ 。令 V_r 为 range 类型, V_p 为 probability 类型, $V_p = \{V_{p_1}, P_{p_1}, V_{p_2}, P_{p_2}, \dots, V_{p_n}, P_{p_n}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n (m(V_r, V_{p_i}) \times P_{p_i})$, $c(V_r, V_p) = \sum_{i=1}^n (c(V_r, V_{p_i}) \times P_{p_i})$ 。

2) V_r 为 or_set 类型

规则 4 $R_{(or_set-range)}$ 。令 V_r 为 or_set 类型, V_p 为 range 类型, $V_r = [V_{r_1}, V_{r_2}, \dots, V_{r_m}]$, 则 $m(V_r, V_p) = \sum_{i=1}^m m(V_{r_i}, V_p)$, $c(V_r, V_p) = \sum_{i=1}^m c(V_{r_i}, V_p)$ 。

规则 5 $R_{(or_set-or_set)}$ 。令 V_r 为 or_set 类型, V_p 为 or_set 类型, $V_r = [V_{r_1}, V_{r_2}, \dots, V_{r_n}]$, $V_p = \{V_{p_1}, V_{p_2}, \dots, V_{p_m}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n (\sum_{j=1}^m m(V_{r_i}, V_{p_j}) / m)$, $c(V_r, V_p) = \sum_{i=1}^n (\sum_{j=1}^m c(V_{r_i}, V_{p_j}) / m)$ 。

规则 6 $R_{(or_set-probability)}$ 。令 V_r 为 or_set 类型, V_p 为 probability 类型, $V_r = [V_{r_1}, V_{r_2}, \dots, V_{r_m}]$, $V_p = \{V_{p_1}, P_{p_1}, V_{p_2}, P_{p_2}, \dots, V_{p_n}, P_{p_n}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n (\sum_{j=1}^m (m(V_{r_i}, V_{p_j}) \times P_{p_j}))$, $c(V_r, V_p) = \sum_{i=1}^n (\sum_{j=1}^m (c(V_{r_i}, V_{p_j}) \times P_{p_j}))$ 。

3) V_r 为 probability 类型

令 $V_r = \{V_{r_1}, P_{r_1}, V_{r_2}, P_{r_2}, \dots, V_{r_n}, P_{r_n}\}$, 通常其适合于描述发布信息,而请求描述大多采用单值概率,即 $V_r = V_r, P_r$ 。因此,本部分只讨论 probability 类型的特例 ($V_r = (V_r, P_r)$) 的情况。

规则 7 $R_{(probability-range)}$ 。令 V_r 为 probability 类型, V_p 为 range 类型, $V_r = V_r, P_r$, 则 $m(V_r, V_p)$ 和 $c(V_r, V_p)$ 定义同规则 1。若 $m(V_r, V_p) > P_r$, 则为匹配的结果。

规则 8 $R_{(probability-or_set)}$ 。令 V_r 为 probability 类型, V_p 为 or_set 类型, $V_r = V_r, P_r$, $V_p = \{V_{p_1}, V_{p_2}, \dots, V_{p_m}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n (m(V_r, V_{p_i})) / n$, $c(V_r, V_p) = \sum_{i=1}^n (c(V_r, V_{p_i})) / n$ 。若 $m(V_r, V_p) > P_r$, 则 V_p 为匹配的结果。

规则 9 $R_{(probability-probability)}$ 。令 V_r 为 probability 类型, V_p 为 probability 类型, $V_r = V_r, P_r$, $V_p = \{V_{p_1}, P_{p_1}, V_{p_2}, P_{p_2}, \dots, V_{p_m}, P_{p_m}\}$, 则 $m(V_r, V_p) = \sum_{i=1}^n (m(V_r, V_{p_i}) \times P_{p_i})$, $c(V_r, V_p) = \sum_{i=1}^n (c(V_r, V_{p_i}) \times P_{p_i})$ 。若 $m(V_r, V_p) > P_r$, 则 V_p 为匹配的结果。

4) V_r 为 negative 类型

将 negative 类型转换为 range 类型,如 negative(100), 转换为 (100,)。之后,遵循 V_r 为 range 类型的匹配规则实现匹配。

实例介绍

假设存在如表 1 所示的汽车售价信息,请求“查找汽车为 Mazda6 2.0 或 Passat1.8T 的车的信息。a)若匹配谓词为“price in [10~20]”,即 V_r 为 range 数据类型;b)若匹配谓词为“price in (19,20)”,即 V_r 为 or_set 数据类型;c)若匹配谓词为“price in 18,0.5”,即 V_r 为 probability 数据类型。具体匹配结果见表 2。

表 1 汽车的售价信息表

编号	车名称	生产商	经销商	售价/万元
	Mazda6	长春一汽	经销商 1	20.5
	Mazda6	长春一汽	经销商 2	19~21
	Mazda6	长春一汽	经销商 3	(19,20,21)
	Passat1.8T	上海大众	经销商 4	(20,0.5)(21,0.5)
	Passat1.8T	上海大众	经销商 5	(18.5,19.5,20)
	Passat1.8T	上海大众	经销商 6	18~20

表 2 匹配结果表

编号	售价/万元	数据类型	$V_r=[19\sim20]$ (range)		$V_r=(19,20)$ (or_set)		$V_r=(18,0.5)$ (probability)	
			匹配度	完备度	匹配度	完备度	匹配度	完备度
①	20.5	single	0	0	0	0	0	0
②	19~21	range	0.67	0.2	0.67	0.2	0	0
③	(19,20,21)	or-set	0.67	0.06	0.67	0.2	0	0
④	(20,0.5)(21,0.5)	probability	0.5	0.05	0.5	0.05	0	0
⑤	(18,20)	or-set	1	0.2	0.5	0.1	0.5	0.1
⑥	19~20	range	1	0.33	1	0.2	0	0

根据请求 a):若按匹配度、完备度降序排列,则得到的匹配结果集为 { }。分析表中的实际信息可知:是最好的,属于完全匹配,它们提供的服务一定能满足请求;而为部分匹配,其中的匹配度大于 的匹配度,因为 能比 提供更多的满足请求的资源; 为不满足的服务。进一步观察 可知, 比 能提供更广泛的服务资源,因为 中只提供了离散的三个报价信息,并且是或的关系,实际上,相当于只提供了一个满足服务请求的服务。而 提供的是连续的值,范围比 大。所以, 的完备度比 大。同理, 也是如此。

根据请求 b):匹配结果集为 { }。因为 为完全匹配,完全满足请求; 为部分匹配,并且匹配度和完备度都相同,按 range 类型优先; 和 虽然匹配度一致,但 的完备度高于 ,因为 中满足部分的概率为 0.5,所以 好于 。

根据请求 c):匹配结果集为 { }。只有 满足服务请求。实际也是如此。

从以上结果分析可知:实际分析结果与采用匹配规则计算所获得的结果是一致的,可见该匹配模型是有效的。

结束语

本文针对可能世界中广泛存在的不确定信息进行研究。虽然已有针对不确定信息管理的相关研究,但大多是针对只支持几种特定的不确定数据类型的模型研究,而有关不确定信息的匹配问题的研究还没有见到。本文分析了可能世界中广泛存在的不确定数据类型,给出了统一的形式化定义,提出了不确定数据类型间的匹配模型,并详细定义了匹配规则。最后通过实例验证,本匹配模型是可行的。(下转第 2612 页)

算法 1 skeleton_identify(V)

```

 $V_s \leftarrow V, T \leftarrow \emptyset, i \leftarrow |V_s|$ 
while( $i > 0$ )
     $i \leftarrow |V_s|$ 
    for all  $v \in V_s$  且  $\text{degree}(v) = 1$ 
         $V_s \leftarrow V_s - \{v\}$ 
         $T \leftarrow T \cup \{v\}$ 
    end for
     $i \leftarrow i - |V_s|$ 
end while

```

算法结束后, V_s 中存放的是所有主干子图上的节点。

当整个图等价于树且其深度最大时, 算法迭代次数最多; 节点数多于 1 的树最少有两个度为 1 的节点 (叶子节点), 所以对于图 G 来说, 最多迭代次数为 $\lceil n/2 \rceil$ 。主干子图生成过程的执行次数最多为 $(n^2/2)$, 即最坏计算复杂度为 $O(n^2)$ 。

桩树生长算法

桩树生长算法对于给定的任意树根节点 r , 通过递归方式生成以 r 为根的整个桩树, 用于识别子树。由于该算法是一个递归算法, 调用时先令辅助集 $T = \emptyset$ 。具体描述如下:

算法 2 grow_stub_tree(r)

```

 $N \leftarrow \text{neighbor}(r)$ 
 $N \leftarrow N - T$ 
 $T \leftarrow T \cup N$ 
for all  $v \in N$  且  $v \notin V_s$ 
    add_child( $r, v$ )
    grow_stub_tree( $v$ )
end for

```

根据主干子图的性质, 图 G 上所有不属于其主干子图的 $(n-s)$ 个节点一定都属于桩树节点, 它们都将在桩树生长过程中被处理。桩树生长算法相当于树的广度优先遍历算法, 最坏时间复杂度为 $O(n-s)$ 。

结束语

幂律特征为复杂网络的研究拓展了视野。主干子图是幂律特征图中度较高且相对重要的节点的集合, 而子树则正好相反。幂律特征保证了节点度分布的非均匀性。本文正是基于这一特性定义了主干子图的相关概念, 证明了其若干性质, 并在此基础上给出了基于主干子图的聚类算法。主干子图 $G_s(V_s, E_s)$ 与原图 $G(V, E)$ 的同态等价保证了 G_s 继承 G 的多数性质, 这使得该聚类算法可应用于有幂律特征的大型图的混合布局, 很多对幂律特征图的分析研究可以转换为对其主干子图的分析研究。后续的主要工作可归结为两方面: 丰富主干子图理论, 如利用主干子图识别幂律特征图的关键节点; 将基于主干子图的聚类算法与绘图算法结合, 寻求更高效的幂律特征图布局方法。

参考文献:

- [1] FALOUTSOS M, FALOUTSOS P, FALOUTSOS C. On power-law relationships of the Internet topology[C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication New York: ACM Press, 1999: 251-262.
- [2] WAXMAN B M. Routing of multipoint connections[J]. IEEE Journal on Selected Areas in Communications, 1988, 6(9): 1617-1625.
- [3] DOAR M B. A better model for generating test networks[C]//Proc of Global Telecommunications Conference 1996: 86-93.
- [4] BRODO A, CLAFFY K C. Internet topology: connectivity of IP graphs[C]//Proc of SPIE 2001: 172-187.
- [5] ANDERSON R, CHUNG F, LU L. Analyzing the small world phenomenon using a hybrid model with local network flow[C]//Proc of the 3rd International Workshop on Algorithms and Models for the Web-Graph 2004: 19-30.
- [6] ANDERSEN R, CHUNG F, LU L. Drawing power-law graphs using a local/global decomposition[J]. Algorithmica, 2007, 47(4): 397.
- [7] ZMANYIE. Incomplete and uncertain information in relational databases[EB/OL]. (2005-03-01). <http://code.ulb.ac.be/dbfiles/>.
- [8] ABITEBOUL S, PARIS K, GOSTA G. On the representation and querying of sets of possible worlds[C]//Proc of ACM SIGMOD International Conference on the Management of Data New York: ACM Press, 1987: 34-48.
- [9] BARGA R S, PU C. Accessing imprecise data: an approach based on intervals[J]. IEEE Data Engineering Bulletin, 1993, 16(2): 12-15.
- [10] CHENG R, PRABHAKAR S. Managing uncertainty in sensor databases[J]. SIGMOD Record Issue on Sensor Technology, 2003, 32(4): 41-46.
- [11] CHENG R, SINGH S, PRABHAKAR S. U-DBMS: a database system for managing constantly-evolving data[C]//Proc of the 31st VLDB Conference on Very Large Databases 2005: 1271-1274.
- [12] TRAJCEVSKI G, WOLFSON O, HNICHS K, et al. Managing uncertainty in moving objects databases[J]. ACM Transactions on Database Systems, 2004, 29(3): 463-507.
- [13] CHENG R, PRABHAKAR S, KALASHNIKOV D V. Querying imprecise data in moving object environments[J]. IEEE Transactions on Knowledge and Data Engineering Archive, 2004, 16(9): 1112-1127.
- [14] ROSS R, SUBRAHMANYAN V S. Aggregate operators in probabilistic databases[J]. Journal of the ACM, 2005, 52(1): 54-101.
- [15] SCOTNEY B, McCLEAN S. Database aggregation of imprecise and uncertain evidence[J]. Journal of Information Sciences, 2003, 155(3-4): 245-263.
- [16] MORRISSEY J M. Imprecise information and uncertainty in information systems[J]. ACM Transactions on Information Systems, 1990, 8(2): 159-180.
- [17] HONG X, McCLEAN S, SCOTNEY B, et al. Evidential integration of semantically heterogeneous aggregates in distributed databases with imprecision[C]//LNCS4224. 2006: 961-969.
- [18] WIDOM J. Trio: a system for integrated management of data, accuracy, and lineage[C]//Proc of the 2nd Biennial Conference on Innovative Data Systems Research 2005.
- [19] MUTSUZAKIM, THEOBALD M, DeKEIJZER A, et al. Trio-one: layering uncertainty and lineage on a conventional DBMS[C]//Proc of CDR Monterey [series], 2007: 269-274.
- [20] SARMA A D, BENJELLOUN O, HALEVY A, et al. Working models for uncertain data[EB/OL]. (2006-03-02). <http://wikiedlab.cs.umass.edu/pub/>.
- [21] MOTRO A. Management of uncertainty in database systems[EB/OL]. (2005-03-15). <http://ise.gnu.edu/~ami/research/publications/pdf/modem94.pdf>

(上接第 2609 页)

参考文献:

- [1] ZMANYIE. Incomplete and uncertain information in relational databases[EB/OL]. (2005-03-01). <http://code.ulb.ac.be/dbfiles/>.
- [2] ABITEBOUL S, PARIS K, GOSTA G. On the representation and querying of sets of possible worlds[C]//Proc of ACM SIGMOD International Conference on the Management of Data New York: ACM Press, 1987: 34-48.
- [3] BARGA R S, PU C. Accessing imprecise data: an approach based on intervals[J]. IEEE Data Engineering Bulletin, 1993, 16(2): 12-15.
- [4] CHENG R, PRABHAKAR S. Managing uncertainty in sensor databases[J]. SIGMOD Record Issue on Sensor Technology, 2003, 32(4): 41-46.
- [5] CHENG R, SINGH S, PRABHAKAR S. U-DBMS: a database system for managing constantly-evolving data[C]//Proc of the 31st VLDB Conference on Very Large Databases 2005: 1271-1274.
- [6] TRAJCEVSKI G, WOLFSON O, HNICHS K, et al. Managing uncertainty in moving objects databases[J]. ACM Transactions on Database Systems, 2004, 29(3): 463-507.
- [7] CHENG R, PRABHAKAR S, KALASHNIKOV D V. Querying imprecise data in moving object environments[J]. IEEE Transactions on Knowledge and Data Engineering Archive, 2004, 16(9): 1112-1127.
- [8] ROSS R, SUBRAHMANYAN V S. Aggregate operators in probabilistic databases[J]. Journal of the ACM, 2005, 52(1): 54-101.
- [9] SCOTNEY B, McCLEAN S. Database aggregation of imprecise and uncertain evidence[J]. Journal of Information Sciences, 2003, 155(3-4): 245-263.
- [10] MORRISSEY J M. Imprecise information and uncertainty in information systems[J]. ACM Transactions on Information Systems, 1990, 8(2): 159-180.
- [11] HONG X, McCLEAN S, SCOTNEY B, et al. Evidential integration of semantically heterogeneous aggregates in distributed databases with imprecision[C]//LNCS4224. 2006: 961-969.
- [12] WIDOM J. Trio: a system for integrated management of data, accuracy, and lineage[C]//Proc of the 2nd Biennial Conference on Innovative Data Systems Research 2005.
- [13] MUTSUZAKIM, THEOBALD M, DeKEIJZER A, et al. Trio-one: layering uncertainty and lineage on a conventional DBMS[C]//Proc of CDR Monterey [series], 2007: 269-274.
- [14] SARMA A D, BENJELLOUN O, HALEVY A, et al. Working models for uncertain data[EB/OL]. (2006-03-02). <http://wikiedlab.cs.umass.edu/pub/>.
- [15] MOTRO A. Management of uncertainty in database systems[EB/OL]. (2005-03-15). <http://ise.gnu.edu/~ami/research/publications/pdf/modem94.pdf>