

# 不确定关系的数据依赖问题研究

周 宇 刘国华 叶杰敏

(东华大学计算机科学与技术学院 上海 201620)

**摘 要** 数据依赖是属性间取值相互约束条件的形式化表示,函数依赖是一种对数据库模式设计及查询具有重要影响的数据依赖。对于不确定关系,由于一条元组中属性取值存在多种可能,属性间取值相互约束的情况要比一般关系中的复杂。为进一步研究不确定关系中数据间的联系,给出了不确定关系模式的形式化定义,提出了 3 类不确定函数依赖,并证明了其推导规则。此 3 类不确定函数依赖可以检查不确定关系的可能世界中是否存在违反规范化设计要求的风险,从而避免数据冗余,更新异常操作。

**关键词** 不确定关系,不确定函数依赖,规范化设计,推导规则

**中图法分类号** TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.08.045

## Data Dependence Research on Uncertain Relation

ZHOU Yu LIU Guo-hua YE Jie-min

(School of Computer Science and Technology, Donghua University, Shanghai 201620, China)

**Abstract** Data dependencies are the formal representation for the mutual constraints condition between the values of attributes, and functional dependencies are some of important influence data dependencies for the schema design and query in the database. In the uncertain relation, there are several possible values for every attribute in a tuple, so the mutual constraints for the values of attributes are more complicated than in the certain relation. We proposed a formal definition for uncertain relation and three kind of uncertain function dependencies, and then proved their inference rules. We would illustrate that above uncertain function dependences have important meaning in the guidance of normal form design in the uncertain relation.

**Keywords** Uncertain relation, Uncertain function dependence, Standardized design, Inference rules

## 1 引言

函数依赖能够提高数据存储中索引的效率及查询速度,优化数据库模式设计,有效地减小数据冗余和更新异常<sup>[1-6]</sup>。然而在不确定数据库中,由于数据的不确定性,使得在其上的数据依赖比经典的函数依赖复杂得多,不确定数据库设计的好坏也不像经典关系数据库那样有成熟的规范化成果。以往关于不确定数据库的研究主要集中在数据模型表示和不确定数据查询方式,几乎没有人对不确定关系模式提出形式化定义,而多以不确定关系实例为研究基础,从关系模式上无法看出传统关系与不确定关系的区别。为了揭示二者之间的本质区别,亟需提出一种适合描述不确定关系的模式,并在此基础上,进一步讨论传统关系和不确定关系在数据依赖上的差异。

本文首先给出了不确定关系模式的形式化定义,接着在此基础上提出了 3 类不确定函数依赖,并对这 3 类不确定函数依赖的推导规则进行证明,最后简单举例说明所提不确定函数依赖在不确定关系规范化设计中的作用。

## 2 问题描述

不确定关系在语义上都可以看作由许多确定的可能实例组成的集合,这些可能实例也称为可能世界(possible worlds)。每一个可能世界实际上都是一个传统关系,因而不确定关系可以简单看作是传统关系(数据不存在不确定性)的集合,每个不确定关系在某个时刻的取值只能是这些传统关系集合中的一个元素。在经典的关系数据库设计中,一般都将关系设计成满足某种范式(例如 3NF 或 BCNF)的要求。然而在不确定关系中,却会出现这样的情况:不确定关系的某些可能世界满足范式,而其他可能世界却不满足范式。可能世界是不确定关系在具体某一时刻的具体体现,这意味着不确定关系在某时刻可能出现不满足范式设计的情况。

现在以病人个人信息表  $R$  为例,医生给病人进行诊断时有时不能完全确定患者到底患何种疾病,因而对患者所要采取的治疗方法也局限在一个范围内,简单起见,假设患者只患一种疾病,治疗方案也只需一套。在设计数据库时,可以将其设计为不确定数据库,如表 1 所列。

到稿日期:2013-10-06 返修日期:2013-12-30 本文受国家自然科学基金项目(61070032)资助。

周 宇(1990-),女,硕士,主要研究领域为不确定关系的数据依赖、数据库查询;刘国华(1966-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库理论、数据库安全、Web 数据管理、业务过程管理;叶杰敏(1989-),男,硕士,主要研究领域为不确定关系的数据依赖。

表1 病人个人信息表 R

Patient	Diagnose	Treatment
Mary	{Cancer, AIDS}	T1
John	{Fever, Flu}	T2
Mark	{Flu, HIV}	T2

简单起见,假设医生只需要知道某位患者患何种病,就可以确认用什么治疗方法给患者治疗,亦即上述关系表中的主键为(Patient,Diagnose)。进一步观察还可以发现,如果表中不存在不确定数据,那么此关系应该为 BC 范式(BCNF)。然而将这个不确定关系展开成可能世界表示时,BCNF 的性质不再满足,例如 R 的一个可能世界如表 2 所列。

表2 R 的一个可能世界

Patient	Diagnose	Treatment
Mary	Cancer	T1
John	Flu	T2
Mark	Flu	T2

在表 2 中, Treatment 完全函数依赖于 Diagnose,而非 (Patient,Diagnose),因此存在非主属性部分函数依赖于候选键,不满足 BCNF,实际上,它连第二范式(2NF)都不满足。换言之,此不确定关系中可能存在某些可能世界违反规范化设计,潜藏着数据冗余和更新异常的风险。

### 3 相关工作

无论是传统关系还是不确定关系,数据依赖都是一个非常重要的性质,它们在关系分解、查询效率和索引建立上有着重要作用。目前已有许多研究者对各种不同的数据依赖进行了研究。

Bohanon 在文献[7]中提出了条件函数依赖(Conditional Function Dependencies CFDS),它可以用于清洗数据,提高数据质量。条件函数依赖在经典函数依赖的基础上增加了一个条件,它要求部分元组组成的集合需要满足所要求的函数依赖。条件函数依赖是近年来研究的热点,Cong 在文献[8]中探讨了检查关系表中违法 CFDS 的方法及将其修复的方法。Fan 在文献[9]中探讨了 CFDS 的传播问题,即研究了在原关系表中成立的 CFDS 在它的视图图中是否依然成立。此后,又出现了许多 CFDS 的变体,如 Bravo 在文献[10]中又提出了条件包含依赖,它不仅可以用于数据清洗,而且可用于上下文模式匹配,但上述条件函数依赖都未讨论有关数据不确定时的情况。

对于不确定关系中的数据依赖,Sarma 在文献[6]中提出了水平函数依赖和垂直函数依赖,主要用于对不确定关系进行分解。随后,Sushovan 在文献[11]中提出了几种面向概率数据库的数据依赖:概率函数依赖、概率近似函数依赖、条件概率函数依赖和条件概率近似函数依赖。它们分别是对经典函数依赖、近似函数依赖、条件函数依赖和条件近似函数依赖的逻辑扩展,即在数据依赖的基础上添加概率这个特征值。这些数据依赖都只考虑了具体不确定关系表,而未从不确定关系模式的角度进行讨论。

Song 在文献[12]中提出一种可比较依赖,可比较依赖包括了经典函数依赖、匹配依赖、度量函数依赖的特性,对于数据空间中多样数据类型之间数据依赖关系的表示起到显著的作用。文献[13]提出了一种基于差异比较的依赖,它不同于基于相同语义的数据依赖,而是一种基于差异函数的数据依赖。基于差异的依赖可以很好地用于数据分区、数据完整性

检测、查询优化、记录联动等。

### 4 基本定义

**定义 1(域<sup>[6]</sup>)** 域是一组具有相同数据类型的值的集合, $Dom = \{a_1, a_2, \dots, a_n\}, a_i (1 \leq i \leq n)$  为常量,  $Dom$  既可为有限集,也可为无限集。 $Dom(A)$  表示属性 A 的域。

**定义 2(域集合)** 域集合是一组由集合组成的域。

**定义 3(关系模式)** 关系模式 R 为一个五元组  $R(U, D, Dom, F)$ , 其中, R 为关系名; U 为 R 中属性的集合; D 为各个属性的域的集合; Dom 是一个定义域为 U、值域为 D 的映射函数; F 为属性间数据依赖的集合。

**定义 4(不确定关系模式)** 不确定关系模式 R 为一个五元组  $R(U, D', Dom, F')$ , 其中, R 为关系名; U 为 R 中属性的集合; D' 为各个属性的域组成的集合, D' 中元素为 D 中元素的幂集; Dom 是一个定义域为 U、值域为 D' 的函数; F' 为属性间数据依赖的集合。

传统的关系模式与不确定关系模式的主要区别在于它们属性的域和数据依赖不同,不确定关系模式中属性的域为域集合,它表示不确定关系上的属性的可能取值集合,但在具体的某一时刻,不确定关系上各个元组的取值都是确定的。

**定义 5** 给定一组域集合  $D_{S1}, D_{S2}, \dots, D_{Sn}$ , 这些域集合可以相同。 $D_{S1}, D_{S2}, \dots, D_{Sn}$  的笛卡尔积为  $D_{S1} \times D_{S2} \times \dots \times D_{Sn} = \{d_{S1} \times d_{S2} \times \dots \times d_{Sn} \mid d_i \in D_i, i = S1, S2, \dots, Sn\}$ , 域集合的笛卡尔积的子集称为不确定关系实例。值为真的不确定关系实例即为不确定关系。每一个元素  $(d_{S1}, d_{S2}, \dots, d_{Sn})$  称为不确定关系上的一条元组。

例 1 设  $R(\text{Patient}, \text{Diagnose}, \text{Treatment})$  为不确定关系模式,表 3 是 R 的任意一个不确定关系 r。

表3 关系 r

Patient	Diagnose	Treatment
Mary	{Cancer, AIDS}	T1

表 3 中只有一条数据,它表示不确定关系的一条元组,“{Cancer, AIDS}”表示这条元组在 Diagnose 属性上的取值是在 Cancer、AIDS 中二选一,元组存在两种取值可能性。不确定关系 r 的可能世界可以表示为表 4 和表 5 中两个传统关系  $r_1$  和  $r_2$ 。换句话说,在某一时刻, r 是一个确定的关系,其上的所有数据都是确定的,或取为表 4,或取为表 5。

表4 关系  $r_1$ 

Patient	Diagnose	Treatment
Mary	Cancer	T1

表5 关系  $r_2$ 

Patient	Diagnose	Treatment
Mary	AIDS	T1

**定义 6(属性集的可能取值)** 设不确定关系模式为 R(U), X 是 U 的一个非空子集,  $X = \{A_1, A_2, \dots, A_m\}$ , r 为 R 的一个任意不确定关系, t 为 r 上任意一个元组,则元组 t 在属性集 X 上的各个属性的可能取值为集合  $d_{A1}, d_{A2}, \dots, d_{Am}$ , 则记 t 在 X 上的可能取值集合为  $t.[X] = d_{A1} \times d_{A2} \times \dots \times d_{Am}$ 。

例如表 3 中,  $t.[\text{Diagnose}] = \{\text{Cancer}, \text{AIDS}\}$ 。

**定义 7(不确定函数依赖)** 设 r 为不确定关系模式 R(U) 的任意一个不确定关系, X, Y 为 U 上的非空子集,如果 r 中可能存在某些可能世界满足  $X \rightarrow Y$ , 则称在 R 上有 Y 不确定函数依赖于 X。

现提出在不确定关系模式下的3类不确定函数依赖,设  $R(U)$  为一不确定关系模式,  $r$  为  $R$  的一个不确定关系,  $X, Y$  是  $U$  上的两个非空子集,  $t_1, t_2$  为  $r$  上的任意两元组,  $t_1.[X] \neq \phi, t_2.[X] \neq \phi$ .

### 定义 8(3类不确定函数依赖)

(1) 如果  $t_1.[X] \cap t_2.[X] \neq \phi$ , 有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 否则  $t_1.[X] \cap t_2.[X] = \phi$ , 则称  $Y$  第一类不确定函数依赖于  $X$ , 记为  $X \rightarrow_{U1} Y$ .

(2) 如果  $t_1.[X] \subseteq t_2.[X]$ , 有  $t_1.[Y] \subseteq t_2.[Y]$ , 否则  $t_1.[X] \cap t_2.[X] = \phi$ , 则称  $Y$  第二类不确定函数依赖域  $X$ , 记为  $X \rightarrow_{U2} Y$ .

(3) 如果  $t_1.[X] \cap t_2.[X] \neq \phi$ , 有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 或者  $t_1.[X] \subseteq t_2.[X]$  时, 有  $t_1.[Y] \subseteq t_2.[Y]$ , 否则  $t_1.[X] \cap t_2.[X] = \phi$ , 则称  $Y$  第三类不确定函数依赖于  $X$ , 记为  $X \rightarrow_{U3} Y$ .

根据不确定函数依赖的定义, 可以得到不确定函数依赖与函数依赖有如下联系:

① 对于不确定关系模式  $R$  的任意不确定关系  $r$ , 如果  $r$  中任意两元组在  $X$  属性上都不存在交集, 那么该不确定关系任意一个可能世界都满足函数依赖  $X \rightarrow Y$ .

② 对于不确定关系模式  $R$  的任意不确定关系  $r$ , 如果  $r$  满足上述3类不确定函数依赖, 则表示该不确定关系在某时刻可能存在可能世界满足  $X \rightarrow Y$ . 实际上, 当不确定关系各个属性的可能取值集合都只有唯一的元素时, 这3类不确定函数依赖便转化为函数依赖.

## 5 不确定函数依赖推导规则

### 5.1 第一类不确定函数依赖的推导规则

(1) 自反性(Reflexivity) 若  $Y \subseteq X \subseteq U$ , 则  $X \rightarrow_{U1} Y$  成立.

证明: 因为  $Y \subseteq X$ , 所以当  $t_1.[X] \cap t_2.[X] \neq \phi$  时, 一定有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 故  $X \rightarrow_{U1} Y$  成立, 自反性得证.

(2) 传递性(Transitivity) 若  $X \rightarrow_{U1} Y$  且  $Y \rightarrow_{U1} Z$  都成立, 则  $X \rightarrow_{U1} Z$  也成立.

证明: 因为  $X \rightarrow_{U1} Y$ , 故当  $t_1.[X] \cap t_2.[X] \neq \phi$  时, 有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 又因为  $Y \rightarrow_{U1} Z$ , 同样地, 当  $t_1.[Y] \cap t_2.[Y] \neq \phi$  时, 有  $t_1.[Z] \cap t_2.[Z] \neq \phi$ , 综上所述, 当  $t_1.[X] \cap t_2.[X] \neq \phi$  时, 亦有  $t_1.[Z] \cap t_2.[Z] \neq \phi$ , 即  $X \rightarrow_{U1} Z$ , 传递性得证.

(3) 增广性(Augmentation) 若  $X \rightarrow_{U1} Y$  成立且  $Z \subseteq U$ , 则  $XZ \rightarrow_{U1} YZ$  也成立.

证明: 对于任意两条元组  $t_1, t_2, t_1.[XZ] = t_1.[X] \times t_1.[Z], t_2.[XZ] = t_2.[X] \times t_2.[Z], t_1.[YZ] = t_1.[Y] \times t_1.[Z], t_2.[YZ] = t_2.[Y] \times t_2.[Z]$ , 根据  $X \rightarrow_{U1} Y$ , 当  $t_1.[X] \cap t_2.[X] \neq \phi$  时, 必有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 进而当  $t_1.[XZ] \cap t_2.[XZ] \neq \phi$  时, 有  $t_1.[YZ] \cap t_2.[YZ] \neq \phi$ , 根据定义,  $XZ \rightarrow_{U1} YZ$  成立, 增广性得证.

### 5.2 第二类不确定函数依赖的推导规则

(1) 自反性(Reflexivity) 若  $Y \subseteq X \subseteq U$ , 则  $X \rightarrow_{U2} Y$  成立.

证明: 若  $t_1.[X] \subseteq t_2.[X]$ , 由于  $Y \subseteq X \subseteq U$ , 因此  $t_1.[Y] \subseteq t_2.[Y]$ ,  $X \rightarrow_{U2} Y$  成立, 自反性得证.

(2) 传递性(Transitivity) 若  $X \rightarrow_{U2} Y$  且  $Y \rightarrow_{U2} Z$  都成立, 则  $X \rightarrow_{U2} Z$  也成立.

证明: 若  $t_1.[X] \subseteq t_2.[X]$ , 由于  $X \rightarrow_{U2} Y$ , 因此  $t_1.[Y] \subseteq$

$t_2.[Y]$ ; 再由  $Y \rightarrow_{U2} Z$ , 可知  $t_1.[Z] \subseteq t_2.[Z]$ , 亦即当  $t_1.[X] \subseteq t_2.[X]$  时,  $t_1.[Z] \subseteq t_2.[Z]$ , 根据定义,  $X \rightarrow_{U2} Z$  成立, 传递性得证.

(3) 增广性(Augmentation) 若  $X \rightarrow_{U2} Y$  成立且  $Z \subseteq U$ , 则  $XZ \rightarrow_{U2} YZ$  也成立.

证明: 对于任意两条元组  $t_1, t_2, t_1.[XZ] = t_1.[X] \times t_1.[Z], t_2.[XZ] = t_2.[X] \times t_2.[Z], t_1.[YZ] = t_1.[Y] \times t_1.[Z], t_2.[YZ] = t_2.[Y] \times t_2.[Z]$ , 根据  $X \rightarrow_{U2} Y$ , 当  $t_1.[X] \subseteq t_2.[X]$  时, 必有  $t_1.[Y] \subseteq t_2.[Y]$ , 进而当  $t_1.[XZ] \subseteq t_2.[XZ]$  时, 有  $t_1.[YZ] \subseteq t_2.[YZ]$ , 根据定义,  $XZ \rightarrow_{U2} YZ$  成立, 增广性得证.

### 5.3 第三类不确定函数依赖的推导规则

(1) 自反性(Reflexivity) 若  $Y \subseteq X \subseteq U$ , 则  $X \rightarrow_{U3} Y$  成立.

证明: 因为  $Y \subseteq X \subseteq U$ , 对于任意两条元组  $t_1, t_2$ , 若  $t_1.[X] \cap t_2.[X] \neq \phi$ , 则有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 若  $t_1.[X] \subseteq t_2.[X]$ , 则有  $t_1.[Y] \subseteq t_2.[Y]$ , 故  $X \rightarrow_{U3} Y$  成立, 自反性得证.

(2) 传递性(Transitivity) 若  $X \rightarrow_{U3} Y$  且  $Y \rightarrow_{U3} Z$  都成立, 则  $X \rightarrow_{U3} Z$  也成立.

证明: 对于任意两条元组  $t_1, t_2$ , 若  $t_1.[X] \cap t_2.[X] \neq \phi$ , 根据  $X \rightarrow_{U3} Y$ , 有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 而由  $Y \rightarrow_{U3} Z$ , 又有  $t_1.[Z] \cap t_2.[Z] \neq \phi$ ; 若  $t_1.[X] \subseteq t_2.[X]$ , 根据  $X \rightarrow_{U3} Y$ , 有  $t_1.[Y] \subseteq t_2.[Y]$ , 而由  $Y \rightarrow_{U3} Z$ , 又有  $t_1.[Z] \subseteq t_2.[Z]$ . 综上所述可知, 当  $t_1.[X] \cap t_2.[X] \neq \phi$  时,  $t_1.[Z] \cap t_2.[Z] \neq \phi$ ; 当  $t_1.[X] \subseteq t_2.[X]$  时,  $t_1.[Z] \subseteq t_2.[Z]$ . 根据定义,  $X \rightarrow_{U3} Z$ , 传递性得证.

(3) 增广性(Augmentation) 若  $X \rightarrow_{U3} Y$  成立且  $Z \subseteq U$ , 则  $XZ \rightarrow_{U3} YZ$  也成立.

证明: 对于任意两条元组  $t_1, t_2, t_1.[XZ] = t_1.[X] \times t_1.[Z], t_2.[XZ] = t_2.[X] \times t_2.[Z], t_1.[YZ] = t_1.[Y] \times t_1.[Z], t_2.[YZ] = t_2.[Y] \times t_2.[Z]$ . 若  $t_1.[X] \cap t_2.[X] \neq \phi$ , 根据  $X \rightarrow_{U3} Y$ , 必有  $t_1.[Y] \cap t_2.[Y] \neq \phi$ , 进而当  $t_1.[XZ] \cap t_2.[XZ] \neq \phi$  时, 有  $t_1.[YZ] \cap t_2.[YZ] \neq \phi$ . 若  $t_1.[X] \subseteq t_2.[X]$ , 根据  $X \rightarrow_{U3} Y$ , 有  $t_1.[Y] \subseteq t_2.[Y]$ , 进而当  $t_1.[XZ] \subseteq t_2.[XZ]$  时, 有  $t_1.[YZ] \subseteq t_2.[YZ]$ , 根据定义,  $XZ \rightarrow_{U3} YZ$  成立, 增广性得证.

## 6 不确定函数依赖在指导规范化设计中的作用

不确定函数依赖可以用于检查不确定关系的设计过程中是否存在违法关系模式规范化要求的风险, 即检查出不确定关系是否存在某些可能世界不满足规范化要求, 这些风险将导致数据冗余和更新异常. 为了阐述这3类不确定函数依赖给不确定关系带来的潜在风险, 现简单举例说明.

例2 设不确定关系为  $R(ABC)$ , 如表6所列.

表6 不确定关系  $R$

	A	B	C
t1	{10, 11}	{20, 21}	{30, 31}
t2	{11, 12}	{21, 22}	{31, 32}
t3	{13, 14}	{23, 24}	{33, 34}
t4	{15, 16}	{24, 25}	{34, 35}

例2中,  $t1$  和  $t2$  元组在  $A$  属性上存在交集, 同时在  $B$  属性上也存在交集, 不确定关系满足第一类不确定函数依赖  $A \rightarrow_{U1} B, B \rightarrow_{U1} C, A \rightarrow_{U1} C$ , 某一时刻这个不确定关系存在如表7所列的一个可能世界.

表7 R的一个可能世界

	A	B	C
t1	11	21	31
t2	11	21	31
t3	13	24	34
t4	15	24	34

可能世界中满足  $A \rightarrow B, B \rightarrow C$ , 并且  $A \rightarrow C$ , 表明这个可能世界存在传递函数依赖, 不满足 3NF 的要求, 存在数据冗余和更新异常的风险。所以当不确定关系中存在第一类不确定函数依赖时, 该不确定关系则可能存在某些可能世界不满足规范化设计的要求。

例3 设不确定关系为  $R(ABC)$ , 如表8所列。

表8 不确定关系 R

	A	B	C
t1	{10}	{20}	{31}
t2	{10,11}	{20,21}	{31,32}
t3	{13}	{23}	{33}
t4	{13,14}	{23,24}	{33,34}
t5	{13,14,15}	{23,24,25}	{33,34,35}

在表8中,  $t_1, A \subseteq t_2, A, t_3, A \subseteq t_4, A \subseteq t_5, A$  存在第二类不确定函数依赖  $A \rightarrow_{U2} B, AB \rightarrow_{U2} C$ , 不确定关系  $R$  存在如表9所列的可能世界。

表9 R的一个可能世界

	A	B	C
t1	10	20	31
t2	10	20	31
t3	13	23	33
t4	14	23	34
t5	14	24	35

这个可能世界满足函数依赖  $A \rightarrow B, AB \rightarrow C$ , 其中  $AB$  作为候选键, 而  $B$  函数依赖于  $A$ , 存在主属性部分函数依赖于候选键, 因而这个可能世界不满足 BCNF 的设计要求, 存在数据冗余和更新的风险。所以当不确定关系中存在第二类不确定函数依赖时, 该不确定关系同样可能存在某些可能世界不满足规范化设计的要求。

例4 设不确定关系为  $R(ABCD)$ , 如表10所列。

表10 不确定关系 R

	A	B	C	D
t1	{10,11}	{20,21}	{30,31}	{40,41}
t2	{11,12}	{21,22}	{31,32}	{41,42}
t3	{13}	{23}	{33}	{43}
t4	{13,14}	{23,24}	{33,34}	{43,44}
t5	{13,14,15}	{23,24,25}	{33,34,35}	{43,44,45}

在表10中,  $t_1, A \cap t_2, A \neq \phi, t_3, A \subseteq t_4, A \subseteq t_5, A$ , 不确定关系  $r$  存在第三类不确定函数依赖  $AB \rightarrow_{U3} C, A \rightarrow_{U3} D$ , 该不确定关系存在如表11所列的一个可能世界。

表11 R的一个可能世界

	A	B	C	D
t1	10	20	31	41
t2	10	20	31	41
t3	13	23	33	43
t4	14	23	34	43
t5	14	24	35	43

可能世界满足函数依赖  $AB \rightarrow C, A \rightarrow D$ , 其中  $AB$  为候选键, 可能世界中存在非主属性  $D$  部分函数依赖于候选键  $AB$ , 因而不满足 2NF 的设计要求, 存在数据冗余和更新异常的风

险。所以当不确定关系中存在第三类不确定函数依赖时, 该不确定关系可能存在可能世界不满足规范化设计的要求。

综上所述, 当不确定关系中存在上述3类不确定函数依赖时, 会导致不确定关系的某些可能世界不满足规范化设计的要求。通过这3类不确定函数依赖, 可以检查出不确定关系设计中的缺陷, 从而指导设计者对不确定关系进行进一步的规范化设计和模式分解操作, 避免数据冗余及更新异常。

**结束语** 在不确定关系中面临着如何进行规范化设计的问题。现有的数据依赖无法指导不确定关系进行规范化设计。为此, 本文提出了3类不确定函数依赖并证明了其推导规则。不确定函数依赖可以检查出在不确定关系的设计中是否存在不满足规范化要求的情况, 从而避免数据冗余和数据更新异常, 为不确定关系模式分解的问题奠定了理论基础。本文中的不确定关系是基于等概率数据, 后续工作将针对非等概率数据进行研究。

## 参考文献

- [1] Armstrong W W. Dependency structures of data base relationships [C]//Proceedings of the IFIP Congress. Stockholm, Sweden, 1974; 580-583
- [2] Fagin R. Multivalued Dependencies and a New Normal Form for Relational Databases[J]. ACM Transactions on Database Systems(TODS), 1977, 2(3): 262-278
- [3] Abiteboul S, Hull R, Vianu V. Foundations of Databases[M]. New York, USA: Addison-Wesley, 1995; 1-34, 157-269
- [4] Armstrong W W, Delobel C. Decompositions and Functional Dependencies in Relations[J]. ACM Transactions on Database Systems(TODS), 1980, 5(4): 404-430
- [5] Elmasri R, Navathe S B. Fundamentals of Database System(3th Edition)[M]. Addison Wesley Publishing Company, 1999; 20-34
- [6] Sarma A, Ullman J, Widom J. Schema design for uncertain databases[J]. Proceedings of the 3rd Alberto Mendelzon Workshop on Foundations of Data Management, 2009, 22(3): 277-302
- [7] Bohannon P, Fan W, Geerts F, et al. Conditional functional dependencies for data cleaning[C]//Proceedings of the International Conference on Data Engineering (ICDE'07). Istanbul, Turkey, 2007; 746-755
- [8] Fan W. Dependencies revisited for improving data quality[C]//Proceedings of the ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS'08). Vancouver, Canada, 2008; 159-170
- [9] Koudas N, Saha A, Srivastava D, et al. Metric functional dependencies[C]//Proceedings of the International Conference on Data Engineering (ICDE'09). Shanghai, China, 2009; 1275-1270
- [10] Bravo L, Fan W, Geerts F, et al. Increasing the expressivity of conditional functional dependencies without extra complexity [C]//Proceedings of the International Conference on Data Engineering (ICDE'08). Cancun, Mexico, 2008; 516-525
- [11] Fan W, Ma S, Hu Y, et al. Propagating functional dependencies with conditions [C]//Proceedings of the International Conference on Very Large Databases (VLDB'08). Auckland, New Zealand, 2008; 391-407
- [12] Song S X, Chen L. Differential Dependencies: Reasoning and Discovery[J]. ACM Transactions on Database Systems (TODS), 2011, 36(3): 16
- [13] Bravo L, Fan W, Ma S. Extending dependencies with conditions [C]//Proceedings of the International Conference on Very Large Databases (VLDB'07). Vienna, Austria, 2007; 243-254