NF²模型 集合映射 数据依赖 数据库 18

计算机科学1995 Vol. 22 №. 3

76-78

用集合映射构造 NF² 中的新依赖

范志新 左万历 周长林

（吉林大学计算机科学系 长春130023）

TP311.13

摘 要 In this paper, we study NF² relations by applying set-mapping instead of data-corresponding. We consider set-mapping be a fundamental concept as a way of data dependency construction in representing NF² relations. A series of mappings is given to build a complete mapping framework. Several mappings are reconstruted as data dependencies under simple conditions. Set-mappings and their simplifications are shown by M-graph.

关键词 Dependency, Set mapping, Horizontal decomposition.

1. 引 言

NF²的一个公开问题是如何表示关系。我们认为数据依赖是一种很好的表示方法，由依赖表示的数据库模式具有最小数据冗余、上佳分解、高效查询等良好特性。Ozsoyoglu 和 Yuan[20,31,21]采用函数依赖FD和多值依赖MVD给出一类具有非循环、4NF和保持依赖的非冗余分解。Armstrong 和 Delobel[2]，Sagiv 等[25]，Kambayashi 和 Yoshikawa[16]给出了退化多值依赖 DMVD。Jaeschke 和 Schek[15]，Thomas[28]，Fischer 和 Gucht[12]，Gucht[29]引入弱多值依赖 WMVD。Sadri 和 Ullman[24]，Beeri 和 Vardi[4]，Fagin 和 Maier[10]，Takeda[27]讨论了样板依赖 TD。Makinouchi[19]，Gucht 和 Fischer[30]，Fischer 等[11]讨论了嵌套关系中不同依赖的组合。

本文将继续依循上述观点，把关系看作一组数据依赖的集合，从新的角度利用一组由集合映射构造的依赖（包括 FD，DMVD，MVD，WMVD）来描述某些条件下的关系，可以使用集合映射来扩展函数概念。现实世界中集合有时比原子值更自然。我们非正式地讨论一下 MVD，采用的定义与 Fagin[8]的定义等价。U 是属性集，$X\subset U$，$Y\subset U$，$X\cap Y=\varnothing$，$Z=U-XY$。U 上的关系 r 满足 MVD $X\rightarrow\rightarrow Y$，即对 X 的值集中任一 x，存在一个关于 x 的与值集 Z(x) 无关的值集 Y(x)。也可以用另一种方式重述这一概念：对于任何 $y\in Y(x)$，y 与所有 $z\in Z(x)$ 相关。我们采用集合映射描述这一相关：对 X 的值集中的任一 x，存在一个 Y(x) 到 Z(x) 的映射。

2. 基本概念和定义

U 是属性集，X，Y，Z 是 U 的子集，假定 X，Y，Z 不相交，$X\cup Y\cup Z=U$。泛关系模式是属性集 U 和一组依赖 D。一个数据库模式是 U 的子集的集合，其中 U 的每个属性至少包含在一个关系里。

$W\subset U$ 上的约束 $t\in r$（写成 t[W]），是一个元组，其中 t 包含 W 中所有属性的元素。如果 t 只包含 W 中部分属性，则称为子元组。

我们只考虑两种类型的域，一种域是由原子值组成的简单域，另一种域是由一些简单域 D 的幂集 2^D 构成的复合域[15,5,18]。

W(x) 表示 W 上关于 x 的一个值集。$W_i(x)$ 表示 W(x) 上对应 $V_i(x)$ 的一个子集，其中 $V=U-XW$，$\bigcup W_i(x)=W(x)$，$\bigcup V_i(x)=V(x)$。$(x, Y_i(x), Z_i(x))$ 是一个元组，$Y_i(x)$ 和 $Z_i(x)$ 之间的映射是子元组 $(Y_i(x), Z_i(x))$。设 $\bigcup W_k(x)=W(x)$，其中 $W_k(x)\neq\varnothing$，即不允许映射集为空集。为简单起见，用 W_k 代替 $W_k(x)$，于是，W_i 和 W_j $(1\leqslant i, j\leqslant n)$ 之间就存在四种关系：①重合——$W_i=W_j$，用 C 表示；②内交——$W_i\subset W_j$，用 I 表示；③外交——$W_i-W_j\neq\varnothing$，$W_j-W_i\neq\varnothing$，$W_i\cap W_j\neq\varnothing$，用 E 表示；④分离——$W_i\cap W_j=\varnothing$，用 D 表示。

为了直观，我们引入 M-图来表示对于给定 $x\in X$，Y(x) 和 Z(x) 之间的映射。假定 $(Y_i(x)\ Z_i(x))$ 和 $(Y_j(x), Z_j(x))$ 是两个映射，它们的 M 图见图2-1。在 M-图中，●表示非空集，○表示空集，连线表示映射。

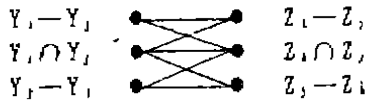


图2.1

这样，用 M-图表示两个元组(Y_i, Z_i)和(Y_j, Z_j)之间的映射，共有16种。如果不考虑对称的话，可归纳为10种映射型(mapping type)；用 Map 表示如下：

(1)Map C-C： 见图2.2。

(2)Map C-I： 见图2.3，Map I-C 是它的对偶。

(3)Map C-E： 见图2.4，Map E-C 是它的对偶。

(4)Map C-D： 见图2.5，Map D-C 是它的对偶。

(5)Map D-D： 见图2.6。

(6)Map I-I： 图2.7a 和图2.7b 是它的两种形式。

(7)Map I-E： 见图2.8，Map E-I 是它的对偶。

(8)Map I-D： 见图2.9，Map D-I 是它的对偶。

(9)Map E-D： 见图2.10，Map D-E 是它的对偶。

(10)Map E-E： 见图2.11。

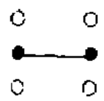


图2.2

图2.3

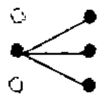


图2.4

图2.5

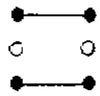


图2.6

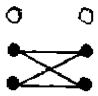


图2.7a

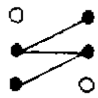


图2.7b

图2.8

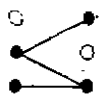


图2.9

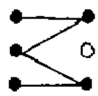


图2.10

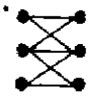


图2.11

这些型构成两个元组间的一个完全映射系，我们称它们为基本型。应用这些型构造数据依赖，在某些约束下可以获得简单形式。第三节将讨论这样一些具有良好性质的依赖形式。

3. 特殊依赖

引入水平分解的概念[9,2]，可以将一个关系分解成一些较小的关系，从而使那些在全局中不能满足的良好特性，可以在局部得到满足。有关水平分解的工作见文[22][13][16][7][12][23][14]。与竖向规范化不同，水平规范化建立在“选择”和“并”的基础上。这里，我们将试图分解集合以适应水平分解。

按照第二节中的假定，有如下结构：

	X	Y	Z
t_{11}:	x_1	Y_{11}	Z_{11}
t_{1n_1}:	x_1	Y_{1n_1}	Z_{1n_1}
t_{k1}:	x_k	Y_{k1}	Z_{k1}
t_{kn_k}:	x_k	Y_{kn_k}	Z_{kn_k}

其中 t_{ij}是一个具有集合值的元组。

给定 i，我们希望 $Y_{ij}(1\leqslant j\leqslant n_i)$之间是不相交的，即划分成等价类，$Z_{ij}$也同样，这样可以获得具有最小冗余的分解。对第二节中讨论的不同映射来说，有着不同的冗余度。下面将讨论第二节中的一些型在某些条件下的形式。为了简化，我们将不考虑对偶型。下文中，X，Y，Z，W 也可分别用来表示相应属性的值集。

假定对于给定的 $x\in X$，$W(x)$中的任意两个集合 W_i 和 W_j 之间都满足相同关系：C，D，I，E，我们有

命题3.1 给定 $x\in X$，$W_i\subset W$，$1\leqslant i\leqslant n$，如果 W(x)满足关系 I，则有：

$$W_{i1}\subset W_{i2}\subset\cdots\subset W_{in}$$

其中$(W_{i1},\cdots,W_{in})$是$(W_1,\cdots,W_n)$的一个重排。

证明 我们非正式地证明。假定 $W_i, W_j, W_k\subset W$，$W_i\subset W_k$，$W_j\subset W_k$，根据前面假设，$W_i\subset W_j$，于是

$W_i\subset W_j\subset W_k$。 □

命题3.2 设有映射(Y_i, Z_i)，$1\leqslant i\leqslant n$，$Y_1\subset\cdots\subset Y_n$，$Z_{i1}\subset\cdots\subset Z_{in}$，满足 Dep I-I 的映射必为：

$$(Y_1, Z_{i1}),\cdots,(Y_k, Z_{ik}),\cdots,(Y_n, Z_{in})$$

证明 设 Y_1映射 Z_{ik}。根据图2.7a，如果有 Z_{it} 满足 $Z_{it}\subset Z_{ik}$，则存在 Y_t 映射 Z_{it} 满足 $Y_t\subset Y_1$。但 Y_1为最小，从而 $Z_{ik}=Z_{i1}$。除去映射(Y_1, Z_{i1})，考虑余下映射，重复上述过程，同上。 □

下面我们给出某些特殊情形下，相应于第二节中某些映射的依赖及其最终化简形式，Dep 表示依赖，其中一些依赖还给出示例。

(1)Dep C-C：最终形式为：$Y=Y_1$，$Z=Z_1$，见例3.

1。

(2)Dep C-I：最终形式为：$Y=Y_1, Z=\max\{Z_i\}$，见例3.2。

(3)Dep C-E：最终形式为：$Y=Y_1, Z=\bigcup Z_i$，见例3.3。

(4)Dep C-D：最终形式为：$Y=Y_1, Z=\bigcup Z_i$，见例3.4。

(5)Dep I-I(图2.7a)：根据命题3.1和命题3.2，最终形式为：$Y=\max\{Y_i\}, Z=\max\{Z_i\}$，见例3.5。

上述五种依赖都可以转化为一类，此类满足多值依赖 MVD X→→Y|Z，我们把该类称作全映射依赖。

例3.1

X	Y	Z
a	{f,t}	{r}
a	{f,t}	{r}
b	{u,g,l}	{h,n}

→

X	Y	Z
a	{f,t}	{r}
b	{u,g,l}	{h,n}

例3.2

X	Y	Z
a	{e}	{k}
a	{e}	{p,r,k}
a	{e}	{p,k}
b	{d,i}	{u,s}
b	{d,i}	{s}

→

X	Y	Z
a	{e}	{p,r,k}
b	{d,i}	{u,s}

例3.3

X	Y	Z
a	{e}	{p,k}
a	{e}	{p,q}
b	{d,i}	{s,t}
b	{d,i}	{r,s}

→

X	Y	Z
a	{e}	{k,p,q}
b	{d,i}	{s,t,r}

例3.4

X	Y	Z
a	{e}	{k}
a	{e}	{p,q}
b	{d,i}	{s}
b	{d,i}	{r}

→

X	Y	Z
a	{e}	{k,p,q}
b	{d,i}	{s,r}

例3.5

X	Y	Z
a	{e}	{p}
a	{r,e}	{p,q}
a	{r,k,e}	{v,p,q}
b	{d}	{r}
b	{d,h}	{r,s}

→

X	Y	Z
a	{r,e,k}	{v,p,q}
b	{d,h}	{r,s}

(6)Dep D-D：最终形式恰好满足 WMVD X-w→Y，具有水平分解的最简形式，分离的等价类也非常自然。

(7)Dep I-I(图2.7b)：我们通过分解 Y_i 和 Z_i 以减少冗余。设有映射(Y_i, Z_i)，$1\leqslant i\leqslant n$，$Y_1\subset\cdots\subset Y_n$，$Z_{i1}\subset\cdots\subset Z_{in}$，我们有：

命题3.3 满足 Dep I-I 的映射必为

$(Y_1, Z_{in}), \cdots, (Y_h, Z_{i_{n+1-h}}), \cdots, (Y_n, Z_{i1})$。

证明 设 Y_1映射 Z_{ik}。根据图2.7b，如果有 Z_{it}，满足 $Z_{ik}\subset Z_{it}$，则存在 Y_t 映射 Z_{it} 满足 $Y_t\subset Y_1$。但 Y_1 为最小，从而 $Z_{ik}=Z_{in}$。除去映射(Y_1, Z_{in})，考虑余下映射，重复上述过程，同上。 □

依据命题3.3，按如下方法可将 Y_i 和 Z_i 分解成不相交集。令 $Y_0=Z_{i0}=\varnothing$，$Y_{k-1,k}=Y_k-Y_{k-1}$，$Z_{i_{k-1},i_k}=Z_{ik}-Z_{i_{k-1}}$，$1\leqslant k\leqslant n$。于是，原映射转化为：

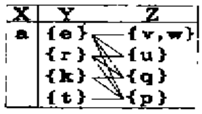
$(Y_{h-1,h}, Z_{i_k,i_{k-1}})$，$1\leqslant h\leqslant n$，$1\leqslant k\leqslant n+1-h$

在原型系统中，使用指针代替实际集合的存储。不相交集提供了较好的映射，从而把关系 r 经水平分解分裂为较小的子关系。例3.6就是一个示例。

例3.6

X	Y	Z
a	{e}	{p,q,u,v,w}
a	{r,e}	{p,q,u}
a	{r,k,e}	{p,q}
a	{r,e,t,k}	{p}

→



(8)Dep I-D：我们用与 Dep I-I 中同样的方法分解 Y_i。令 $Y_0=\varnothing$，$Y_{k-1,k}=Y_k-Y_{k-1}$，$1\leqslant k\leqslant n$。原映射转化为：

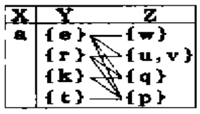
$(Y_{k-1,k}, Z_k)$，$1\leqslant k\leqslant n$

例3.7给出示例结果。这个形式与图3.6中的 Dep I-I 形式相似，我们可以将它们归结为一类。

例3.7

X	Y	Z
a	{e}	{w}
a	{r,e}	{u,v}
a	{r,k,e}	{q}
a	{r,e,t,k}	{p}

→



我们讨论了几种简单条件下由第二节中的基本型构造的依赖。我们试图将 Y(x)和 Z(x)的域分解成不相交的集，以便经水平分解得到的子关系具有最小冗余。这样的分解在特殊的联接下是无损的。

4. 结论

本文采用一种新方法给出 NF^2关系中数据依赖的概述，通过扩充传统的数据相关，我们提出了一系列集合映射类型，引入 M-图来表示这些基本型，比较详细地探讨了由上述基本型构造的一些依赖。从这一新观点出发，可以构造各种简单和复杂条件下的依赖。我们没有考虑空集映射，Levene[17]详尽地讨论了带空值的扩展嵌套关系和数据依赖。使用集合映射代替单值对应，可以扩展到无穷集映射，本文不详细讨论。

有些问题尚待进一步研究：考虑 NF^2的多层映射；尝试将属性集 U 分成四个或更多部分；使用集合映射的分解如何联接；数据更新；无穷集映射等。

鸣 谢 NF^2讨论班提供了热烈而富有启发的建议。感谢崔亨洙教授提供某些原始材料。本文经施伯乐教授审阅，并提出有价值的意见。(参考文献共31篇略)