第33卷 第17期 计 算 机 工 程 2007年9月
Vol.33 No.17 Computer Engineering September 2007
·软件技术与数据库· 文章编号：1000—3428(2007)17—0086—04 文献标识码：A 中图分类号：TP311

XML多值依赖的推理规则集问题

荣凌燕，刘国华
(燕山大学信息科学与工程学院，秦皇岛 066004)

摘 要：XML多值依赖的推理规则集问题是解决XML数据依赖的蕴涵问题的基础，是XML规范化理论的关键问题之一。该文对XML树、树元组等进行了重新定义，与Vincent等人不同，提出了基于DTD的XML多值依赖的概念，通过对XML的关系化表示给出了其形式化定义，定义了XML多值依赖集的闭包、XML多值依赖路径依赖基以及XML多值依赖路径集的闭包等概念，给出了一个有效且完备的推理规则集，并对其有效性及完备性进行了证明。

关键词：DTD；多值依赖；XML树；推理规则

Inference Rules for Multivalued Dependency in XML

RONG Ling-yan, LIU Guo-hua
(College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

【Abstract】The problem of inference rules for multivalued dependency is the key to solve implication problem between dependencies in XML and the key problem of XML normalization theory. In this paper, the definition of XML tree and tree tuples etc. are given in a new way. Different from Vincent, the concept of multivalued dependency for XML based on DTD is given. The definitions of closure, path dependency basis and the closure of paths of multivalued dependencies for XML are also proposed. A sound and complete set of inference rules is presented, and soundness and completeness of the inference rules are proved.

【Key words】DTD; multivalued dependency; XML tree; inference rule

1 概述

XML自推出以来，由于具有跨平台、简单易用等特性，已成为Web上进行数据传输与交换的标准。虽然XML很容易表达来自不同源的数据，但是其所能表示的语义信息却相对有限。同关系数据库相似，设计不好的XML数据模式会引起更新异常。

例1 某公司开设了培训班，对新进人员进行培训。该公司(*Company*)有多名新进人员(*Staffer*)，根据个人情况，一个人可参加多个培训班(*Training*)，一个培训班可有多名学员，每个培训班开设固定的课程(*Course*)。下面是符合这个语义约束的一个DTD D_1，图1是一个符合D_1的XML文档实例。

```
<! ELEMENT Company (Staffer*)>
  <! ATTLIST Company
     Cname CDATA #REQUIRED>
<! ELEMENT Staffer (Sname, Training*)>
  <! ATTLIST Staffer
     Sno CDATA #REQUIRED>
<! ELEMENT Sname (#PCDATA)>
<! ELEMENT Training (Tname, Course*)>
  <! ATTLIST Training
     Tno CDATA #REQUIRED>
<! ELEMENT Tname (#PCDATA)>
<! ELEMENT Course (#PCDATA)>
```

显然，模式D_1存在数据冗余，因为在XML文档中必须为每个职员重复存储培训班的信息，否则就会出现数据的不一致性以及操作异常现象：(1)更新异常：如果要更新一个培训班的信息，那么必须更新XML文档中所有参加了此培训班的职员中相关的信息，否则就造成信息的不一致。(2)插入、删除异常：如果要在一个培训班中增加一门课程，那么必须在XML文档中所有参加了这个培训班的职员中插入这门课程的信息。否则就会造成该文档的数据不一致现象。同样要删除一个培训班，必须在每个参加了此项培训的职员中删除此信息，否则也会造成信息的不一致。

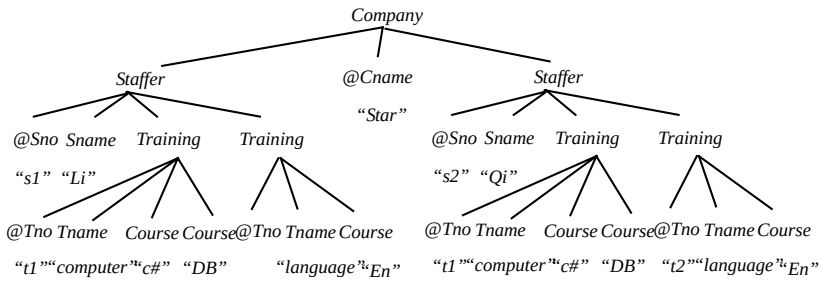

图1 一个符合DTD D_1的XML文档实例

可以看出，图1所示的XML文档存在数据冗余和操作异常正是由于DTD D_1在设计上存在问题。在D_1中，公司和职员有直接联系，公司和培训班也有直接联系，但职员和培训班所开设的课程没有直接联系，而模式中把两个没有直接联系的实体嵌套在一起，造成了数据冗余和操作异常。

根据上面的分析，本文基于DTD提出了XML多值依赖的概念，并通过对XML的关系化表示给出了XML多值依赖

基金项目：教育部科学技术研究基金资助重点项目(205014)；河北省教育厅自然科学指令计划基金资助项目(2005102)
作者简介：荣凌燕(1981－)，女，硕士研究生，主研方向：半结构化数据及XML数据库；刘国华，教授、博士生导师
收稿日期：2006-09-17 **E-mail：**yiran214@sina.com

的形式化定义。在此基础上给出了一个 XML 多值依赖的有效和完备的推理规则集，并对其有效性和完备性进行了证明，为进一步完善 XML 数据库模式设计奠定了基础。

2 相关工作

虽然有关XML数据模式设计的研究已经取得了一些初步成果，但还没有形成统一的规范和完整的理论体系。Marcelo Arenas等人引入XML函数依赖表达语义，定义XML范式——XNF，提出了DTD转换为XNF的无损连接算法，并证明XML函数依赖是非可公理化的[1]。但文章模型中的“相等”有二义性。文献[2]基于路径、路径表达式的方法，在不存在DTD的情况下提出一组XML函数依赖的推理规则，证明是正确的，并在一元函数依赖的前提下证明是完备的[3]。文献[4]在存在DTD的情况下提出一组正确且完备的函数依赖推理规则集。文献[5]讨论了函数依赖和规范化在关系及XML间的传播问题，通过给出转换方法，证明了满足BCNF的关系和规范化XML文档间的对应关系。M. W. Vincent等人初步研究了XML中的多值依赖，基于路径表达式给出了多值依赖的定义，并基于该定义，提出一个XML的范式[6]。但文中没有解决多值依赖的推理规则集和逻辑蕴涵等问题。文献[7]在文献[6]的基础上给出了一个扩展的哈希算法来检测XML文档对给定多值依赖集的满足性，并给出了算法的性能分析。

以上方法对 XML 多值依赖问题虽有涉及，但对其研究仅仅停留在文档一级，脱离了模式的定义。而关于 XML 多值依赖的推导公理问题还尚未给出有效的解决方法。这样就无法解决 XML 数据中多值依赖的蕴涵问题，直接影响着规范化算法的有效性和完备性。

3 基本定义

本文采用文献[1]中关于 DTD 及 DTD 路径的定义。用 $Paths(D)$ 表示 D 中所有路径的集合。

为统一比较结点“值相等”，本文重新给出 XML 树的定义，这是在文献[1]基础上修改得到的。

定义 1(XML 树) 一棵 XML 树定义为 $T=(V,lab,ele,att,val,root)$

其中，V 表示 T 中结点的有限集；lab 表示 V 到 $El\cup Att\cup S$ 的函数；ele 表示从 V 到一系列 V 结点的偏序函数，使得对于任意 $v\in V$，若 $ele(v)$ 有定义，则 $lab(v)\in El$；att 为从 $V\times Att$ 到 V 的偏序函数，使得对任意 $v\in V$ 且 $@l\in Att$，若 $att(v,@l)=v_1$，则 $lab(v)\in El$ 且 $lab(v_1)=@l$；val 为使得对任意 $v\in V$，若 $lab(v)=El$，则 $val(v)=El$；若 $lab(v)=S$ 或 $lab(v)\in Att$，则 $val(v)$ 是一个字符串；$root$ 为 V 中一个唯一的结点，叫作 T 的根。

定义 2(XML 树路径) 给定 XML 树 T，则对字符串 $w_1...w_n$，($w_1...w_n\in El$，$w_n\in El\cup Att\cup S$)，如果 T 中存在节点 $v_1...v_n$，使得 $v_1=root$，v_{i+1} 是 v_i 的孩子 $(i\in[1,n-1])$，$lab(v_i)=w_i$ $(i\in[1,n])$，称 $w_1...w_n$ 为 T 中的路径。

定义 3(T 符合 D) 给定 DTD $D=(E,A,P,R,r)$ 和 XML 树 $T=(V,lab,ele,att,val,root)$，如果满足下列条件，称 T 符合 D(记作 $T\vDash D$)。其中，lab 是从 V 到 $E\cup A\cup\{S\}$ 的映射；对于 $v\in V$，若 $ele(v)=[v_1,...,v_n]$，则 $lab(v_1)...lab(v_n)$ 定在 $P(lab(n))$ 定义的正则表达式中；对于任意 $v\in V$，$@l\in A$，如果 $att(v,@l)$ 有定义，则 $@l\in R(lab(v))$；$lab(root)=r$。

在文献[1]的树元组定义中，把以元素和属性或文本结点结束的路径映射为不同的类型，这样在进行“相等”比较时产生了二义性。本文把 $Paths(D)$中的路径全部映射到结点上，从而解决了该问题。

定义 4(树元组) 给定 DTD $D=(E,A,P,R,r)$ 和符合 D 的 XML 树 $T=(V,lab,ele,att,val,root)$，树元组 t 定义为 $Paths(D)$ 到 $V\cup\{\perp\}$ 的映射，满足：

(1)若 $t[p_1]=t[p_2]$ 且 $t[p_1]\in Vert$，则 $p_1=p_2$；

(2)若 $t[p_1]=\perp$ 且 p_1 是 p_2 的前缀，则 $t[p_2]=\perp$；

(3) $\{p\in Paths(D)\mid t[p]\neq\perp\}$ 是有限的。

树元组 $t[p]$ 可记作 $t.p$。用 T(D)表示 D 中所有树元组的集合。

4 XML 多值依赖

下面给出 XML 多值依赖的形式化定义。

定义 5(XML 多值依赖，XMVD) 给定 DTD D，任意 XML 文档树 $T\vDash D$。XML 模式 D 上的一个多值依赖(XMVD)是形如形如 $p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m$ 的一个命题，其中 $\{p_1,...,p_k, q_1,..,q_m, r_1,...,r_s\}\subseteq Paths(D)$。XML 树 T 满足多值依赖，如果存在树元组 t, s$\in T$(D)，满足 $val(t.p_i)=val(s.p_i)$，$\forall i\in[1,...,k]$，则必定也存在树元组 u$\in T$(D)，使得 $val(u.p_i)=val(t.p_i)=val(s.p_i)$，$val(u.q_i)=val(t.q_i)$ $\forall j\in[1,...,m]$ 且 $val(u.r_x)=val(s.r_x)$ $\forall x\in[1,...,s]$ 成立。

由 t 和 s 的对称性可知，同样存在树元组 w$\in T$(D)，使得 $val(w.p_i)=val(t.p_i)=val(s.p_i)$，$val(w.q_i)=val(s.q_i)$ 且 $val(w.r_x)=val(t.r_x)$，$\forall i\in[1,...,k]$，$\forall j\in[1,...,m]$，$\forall x\in[1,...,s]$ 成立。

例 2 考虑图 1 中符合D_1的XML文档实例，XML多值依赖可以表示为：

$Company.@Cname\rightarrow\rightarrow Company.Staffer.@Sno$

$Company.@Cname\rightarrow\rightarrow Company.Staffer.Training.@Tno$

$Company.Staffer.Training.@Tno\rightarrow\rightarrow Company.Staffer.@Sno$

$Company.Staffer.Training.@Tno\rightarrow\rightarrow Company.Staffer.Training.Course.S$

定义 6(XMVD 集的闭包) 设 Σ 是路径集 $Paths(D)$ 上的一个 XMVD 集，$\{p_1,...,p_k\}\subseteq Paths(D)$，则 XMVD 集Σ的闭包 Σ^+ 可定义如下：

$$\Sigma^+=\{p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m\mid\Sigma\vdash p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m\}$$

引理 1 设 DTD D，任意的 XML 文档树 $T\vDash D$，$\{p_1,...,p_k, q_1,..,q_m\}\subseteq Paths(D)$，则 $p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m$ 当且仅当 $p_1,...,p_k\rightarrow\rightarrow Paths(D)-\{p_1,...,p_k\cup q_1,..,q_m\}$。

证明 XMVD 定义表明它的对称性。$p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m$ 和 $p_1,...,p_k\rightarrow\rightarrow Paths(D)-\{p_1,...,p_k\cup q_1,..,q_m\}$ 总是满足的。

在引理 1 中，当 $\{p_1,...,p_k\}\cap\{q_1,..,q_m\}\neq\varnothing$ 时，可以简单地从右部 $q_1,..,q_m$ 中删除相交的路径集 $\{p_1,...,p_k\}\cap\{q_1,..,q_m\}$ 而不影响 XMVD 的满足性。可以表述为下面的引理。

引理 2 设 DTD D，任意的 XML 文档树 $T\vDash D$，$q_1',...,q_m'=q_1,..,q_m-p_1,...,p_k$，则 $p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m$ 当且仅当 $p_1,...,p_k\rightarrow\rightarrow q_1',...,q_m'$。

证明 由引理 1 $p_1,...,p_k\rightarrow\rightarrow q_1,..,q_m$ 当且仅当 $p_1,...,p_k\rightarrow\rightarrow Paths(D)-\{p_1,...,p_k\cup q_1,..,q_m\}$，又因为 $Paths(D)-p_1,...,p_k\cup q_1,..,q_m=Paths(D)-p_1,...,p_k\cup q_1',...,q_m'$，由引理 1 有 $p_1,...,p_k\rightarrow\rightarrow q_1',...,q_m'$ 当且仅当 $p_1,...,p_k\rightarrow\rightarrow Paths(D)-\{p_1,...,p_k\cup q_1',...,q_m'\}$。

5 XML 多值依赖推理规则集

为解决逻辑蕴涵的判定问题，需要从一组已知的多值依赖成立，推导出其它多值依赖成立。这就需要一个推理规则集，且它是有效和完备的。有效性是指从 XMVD 集 Σ 使用推理规则集 M 推出的 XMVD 必定在 Σ^+，完备性是指 Σ^+ 中的 XMVD 都能从 XMVD 集 Σ 使用推理规则集推出。下面给出 XML 上多值依赖的推理规则集。

给定 DTD D，XML 文档树 $T \models D$，$\{ p_1,...,p_k, q_1,..,q_m, r_1,...,r_s, w_1,...,w_l \} \subseteq Paths(D)$，则

XMVD0（补规则）：$p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \vdash p_1,...,p_k \rightarrow\rightarrow r_1,...,r_s$，其中 $r_1,...,r_s = Paths(D) - \{ p_1,...,p_k \cup q_1,..,q_m \}$；

XMVD1（包含规则）：$\{ q_1,..,q_m \subseteq p_1,...,p_k \subseteq Paths(D) \} \vdash p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$；

XMVD2（扩展规则）：$\{ p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m, r_1,...,r_s \subseteq w_1,...,w_l \} \vdash p_1,...,p_k \cup w_1,...,w_l \rightarrow\rightarrow q_1,..,q_m \cup r_1,...,r_s$；

XMVD3（传递规则）：$\{ p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m, q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s \} \vdash p_1,...,p_k \rightarrow\rightarrow \{ r_1,...,r_s - q_1,..,q_m \}$；

XMVD4（并规则）：$\{ p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m, p_1,...,p_k \rightarrow\rightarrow r_1,...,r_s \} \vdash p_1,...,p_k \rightarrow\rightarrow \{ q_1,..,q_m \cup r_1,...,r_s \}$；

XMVD5（投影规则）：$\{ p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m, p_1,...,p_k \rightarrow\rightarrow r_1,...,r_s \} \vdash \{ p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cap r_1,...,r_s, p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m - r_1,...,r_s \}$。

定理 1 XMVD0～XMVD5 对于 XMVD 之间的逻辑蕴涵关系的推导是有效的。

证明 为证明方便，这里用 $val(t.p_i = s.p_i)$ 表示 $val(t.p_i) = val(t.s_i)$，$t.(p_i \cup w_y)$ 表示 $t.p_i \cup t.w_y$。其余表示同。且在下面的证明中 $\forall i \in [1,...,k]$，$\forall j \in [1,...,m]$，$\forall x \in [1,...,s]$，$\forall y \in [1,...,l]$。

(1) XMVD0 与 XMVD1 的有效性可由 XMVD 的定义和引理 1 直接导出。

(2)用反证法来证明 XMVD2 的有效性。假定在某个文档树 $T \square D$ 上存在树元组 t，s，满足 $val(t.p_i = s.p_i)$。但不存在树元组 u，使得

$$val(u.(p_i \cup w_y) = t.(p_i \cup w_y)), val(u.(q_j \cup r_x) = t.(q_j \cup r_x),$$
$$val(u.(Paths(D) - \{p_i \cup q_j \cup r_x \cup w_y\})$$
$$= s.(Paths(D) - \{p_i \cup q_j \cup r_x \cup w_y\}))$$

由于 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$ 成立，存在树元组 v，使得

$$val(v.p_i = t.p_i = s.p_i), val(v.q_j = t.q_j), val(v.(Paths(D) - \{p_i \cup q_j\})$$
$$= s.(Paths(D) - \{p_i \cup q_j\}))$$

成立。又由 $r_1,...,r_s \subseteq w_1,...,w_l$，根据 XMVD1，对任意树元组有 $w_1,...,w_l \rightarrow\rightarrow r_1,...,r_s$ 成立。所以有

$$val(v.w_y = t.w_y), val(v.r_x = t.r_x),$$
$$val(v.(Paths(D) - r_x \cup w_y) = s.(Paths(D) - r_x \cup w_y))$$

从而有

$$val(v.(p_i \cup w_y) = t.(p_i \cup w_y)), val(v.(q_j \cup r_x) = t.(q_j \cup r_x)$$

由 $Paths(D) - \{p_i \cup q_j \cup r_x \cup w_y\} \subseteq Paths(D) - \{p_i \cup q_j\}$ z，得

$$val(v.(Paths(D) - \{p_i \cup q_j \cup r_x \cup w_y\}) =$$
$$s.(Paths(D) - \{p_i \cup q_j \cup r_x \cup w_y\}))$$

至此，证明了 v 和 u 是完全相同的，与假设 u 不存在相矛盾，因此原假设不成立。XMVD2 的有效性得证。

(3)用反证法证明 XMVD3 的有效性。假定在某个文档树 $T \square D$ 上树元组 t，s，使得 $val(t.p_i = s.p_i)$。而不存在 u，使得

$$val(u.p_i = t.p_i), val(u.(r_x - q_j) = t.(r_x - q_j)),$$
$$val(u.(Paths(D) - \{p_i \cup (r_x - q_j)\}) = s.(Paths(D) - \{p_i \cup (r_x - q_j)\}))$$

成立。由假设，存在树元组 w，使

$$val(w.p_i = t.p_i), val(w.q_j = s.q_j),$$
$$val(w.(Paths(D) - \{p_i \cup q_j\}) = t.(Paths(D) - \{p_i \cup q_j\}))$$

由于 $q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s$，而树元组 w 与 s 在 $q_1,..,q_m$ 上相等，因此存在树元组 v，使得

$$val(v.q_i = s.q_i = w.q_i), val(v.r_x = w.r_x),$$
$$val(v.(Paths(D) - r_x \cup q_j) = s.(Paths(D) - r_x \cup q_j))$$

下面考察树元组 v。

1)在路径 $p_1,...,p_k$ 上，分 $p_1,...,p_k \cap r_1,...,r_s$ 和 $p_1,...,p_k - r_1,...,r_s$ 两部分考虑。显然 $val(v.(p_i \cup r_x) = t.(p_i \cup r_x))$。由已知，根据图 2 知，$val(v.(p_i - r_x) = w.(p_i - r_x) = s.(p_i - r_x) = t.(p_i - r_x))$。所以 $val(v.p_i = t.p_i)$。

2) 在路径 $r_1,...,r_s - q_1,..,q_m$ 上，由图 2 知，$val(v.(r_x - q_j) = t.(r_x - q_j))$。

3) 在 $Paths(D) - \{ p_1,...,p_k \cup \{ r_1,...,r_s - q_1,..,q_m \}\}$ 上，令 $w_1,...,w_l = Paths(D) - \{ p_1,...,p_k \cup \{ r_1,...,r_s - q_1,..,q_m \}\}$。由 $w_1,...,w_l \cap r_1,...,r_s \subseteq q_1,..,q_m$，得 $val(v.(r_x \cup w_y) = s.(r_x \cup w_y))$。同理得 $val(v.(w_y - r_x) = s.(w_y - r_x))$。至此得 v 和 s 在 $Paths(D) - \{ p_1,...,p_k \cup \{ r_1,...,r_s - q_1,..,q_m \}\}$ 上的值相同。

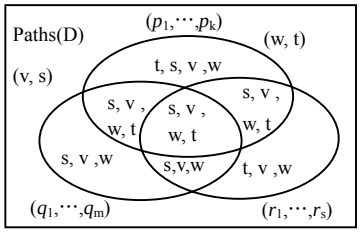

图 2 文氏图

综上 v 与 u 完全相同，这与假设 u 不存在矛盾，假设不成立。XMVD3 有效性得证。

(4) XMVD4 的证明过程如下。

1) $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$，$p_1,...,p_k \rightarrow\rightarrow r_1,...,r_s$ (已知)

2) $p_1,...,p_k \rightarrow\rightarrow p_1,...,p_k \cup q_1,..,q_m$，$p_1,...,p_k \cup q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s \cup q_1,..,q_m$ (1)，XMVD2)

3) $p_1,...,p_k \cup q_1,..,q_m \rightarrow\rightarrow Paths(D) - \{ p_1,...,p_k \cup q_1,..,q_m \cup r_1,...,r_s \}$ (2)，XMVD0)

4) $p_1,...,p_k \rightarrow\rightarrow Paths(D) - p_1,...,p_k \cup q_1,..,q_m \cup r_1,...,r_s$ (2)，3)，XMVD3)

5) $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cup r_1,...,r_s$ (4)，XMVD0)

在证明过程中，由 2)得出的 XMVD 的左部和右部存在交集，先应用补规则，再对 2)、3)应用传递规则，再一次应用补规则得到结论。从而 XMVD4 的有效性得证。

(5)由已知和 XMVD4，有 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cup r_1,...,r_s$ 成立。令 $w_1,...,w_l = Paths(D) - \{ p_1,...,p_k \cup q_1,..,q_m \cup r_1,...,r_s \}$，由 XMVD0，$p_1,...,p_k \rightarrow\rightarrow w_1,...,w_l$。再由 XMVD4，$p_1,...,p_k \rightarrow\rightarrow w_1,...,w_l \cup r_1,...,r_s$，同理 $p_1,...,p_k \rightarrow\rightarrow Paths(D) - \{p_1,...,p_k \cup w_1,...,w_l \cup r_1,...,r_s\}$。对右部化简，得

$$Paths(D) - \{ p_1,...,p_k \cup w_1,...,w_l \cup r_1,...,r_s \}$$

$= \{ p_1,...,p_k \cup q_1,..,q_m \cup r_1,...,r_s \} - \{ p_1,...,p_k \cup r_1,...,r_s \}$

$= q_1,..,q_m - p_1,...,p_k \cup r_1,...,r_s$

$= \{ q_1,..,q_m - r_1,...,r_s \} - p_1,...,p_k$

从而有 $p_1,...,p_k \rightarrow\rightarrow \{ q_1,..,q_m - r_1,...,r_s \}$成立。

同理 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cap r_1,...,r_s$ 成立。XMVD5 的有效性得证。

至此， XMVD 推理规则集的有效性得证。

在证明完备性之前，首先给出几个概念。

定义 7(XMVD 路径集的闭包) 设 Σ 是路径集 $Paths(D)$ 上的 XMVD 集， $p_1,...,p_k \subseteq Paths(D)$,则 $p_1,...,p_k$ 关于 Σ 的闭包 $\{p_1,...,p_k\}^+$可定义如下：

$\{p_1,...,p_k\}^+ = \{ q_1,..,q_m \mid p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$ 可用 XMVD 推理规则从 Σ 中导出$\}$。

定义 8(XMVD 路径依赖基) 设 Σ 是路径集 $Paths(D)$ 上的 XMVD 集， $\{p_1,...,p_k\} \subseteq Paths(D)$，定义 $P = \{ q_1,..,q_m \mid \Sigma \models p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \}$，则称 P 的最小基集 MB(P)为 $\{p_1,...,p_k\}$，关于 XMVD 集 Σ 的路径依赖基，记作 $\mathrm{DEPATH}_\Sigma(p_1,...,p_k)$。

定理 2 XMVD0～XMVD5 对于 XMVD 之间的逻辑蕴涵的推理是完备的。

证明 根据逆否命题的等价性，要证明完备性，即要证明对于某个不能由 Σ 根据 XMVD 推理规则导出的 XMVD σ ：$p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$，则 σ 一定不为 Σ 所逻辑蕴涵，或者说，至少存在一个 XML 文档树实例 T，使 $T \in \mathrm{SAT}(\Sigma)$ 而 $T \notin \mathrm{SAT}(\sigma)$。

令 $w_1,...,w_l \in \mathrm{DEPATH}_\Sigma(p_1,...,p_k)$ ，$w_1,...,w_l$ 覆盖 $Paths(D) - \{p_1,...,p_k\}$。采用这样的方法来构造文档树实例T：T 中共有 2^m个树元组，每个树元组在 $\{p_1,...,p_k\}$ 中的路径上的值都为 1，每个树元组对应一个$<a_1,..,a_m>$序列 $(a_i \in \{0,1\})$，每个树元组在 W_i 中路径上的值都相等，为 a_i。

则文档树 T 有这样的性质：

(1)每个以 W_i 为右部的 XMVD 在 T 中有效。

(2)若某 XMVD 的右部是 W_i 的非空子集，则 XMVD 在 T 中有效当且仅当它的左部与 W_i 相交。

先证明这两个性质。

(1)根据 T 的构造，可知 XMVD $\varnothing \rightarrow\rightarrow W_i$ 成立，根据扩展规则，对每个路径集 Y，有 $Y \rightarrow\rightarrow W_i$ 在 T 中成立，性质 1 得到证明。

(2)令 Y 和 Z 是 $Paths(D)$ 中的两个路径集，且 Y 与 W_i 不相交，Z 是 W_i 的非空真子集。令 l 为 $W_i - Z$ 中的一个路径，并定义 $Y_T(x)$ 表示 $\{y \mid u \in \mathrm{T(D)}, val(u[\mathrm{X}]) = x$ 且 $val(u[\mathrm{Y}]) = y\}$。则根据 T 的构造，可知，对于任意树元组 $u \in$T (D)，$Z_T(y)$ 有两个值 0 和 1，但由于 Z 必与 l 分配相同的值，因此 $Z_T(yl)$的值只有一个。所以 $Z_T(y) \neq Z_T(yl)$。由此可知 XMVDY→→Z 在 T 中不成立。

下面，首先证明 XML 文档树实例 $T \in \mathrm{SAT}(\Sigma)$。

设 $q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s$ 为 Σ^+ 中的任意 XMVD。现证明 $q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s$ 在 T 上成立。由 T 的构造，易知 $q_1,..,q_m \rightarrow\rightarrow \{r_1,\ldots,r_s\} \cap \{p_1,...,p_k\}$ 成立。若 $\{r_1,\ldots,r_s\} \cap W_i = \varnothing$或 $\{r_1,\ldots,r_s\} \cap W_i = W_i$，由性质 1，$q_1,..,q_m \rightarrow\rightarrow \{r_1,\ldots,r_s\} \cap W_i$ ($q_1,..,q_m \rightarrow\rightarrow \varnothing$总成立)。若 $\{ q_1,..,q_m \} \cap W_i = \varnothing$，由 XMVD2，得 $\{ Paths(D) - W_i \} \rightarrow\rightarrow r_1,...,r_s$ 在 Σ^+ 中。 $p_1,...,p_k \rightarrow\rightarrow \{ Paths(D) - W_i \}$ 在 Σ^+ 中，由 XMVD3，$p_1,...,p_k \rightarrow\rightarrow \{r_1,\ldots,r_s\} - \{ Paths(D) - W_i \}$，即有 $p_1,...,p_k \rightarrow\rightarrow \{r_1,\ldots,r_s\} \cap W_i$ 在 Σ^+ 中。这与假设$W_i \in \mathrm{DEPATH}_\Sigma(p_1,\ldots,p_k)$ 矛盾，所以 $q_1,..,q_m$ 必与 W_i 相交。由性质 2，$q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s \cap W_i$ 成立。再根据并规则，有 $q_1,..,q_m \rightarrow\rightarrow r_1,...,r_s$ 在 T 上成立。$T \in \mathrm{SAT}(\Sigma)$ 得证。

其次，证明 XML 文档树实例 $T \notin \mathrm{SAT}(\sigma)$。

假定 σ 为 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$ 不在Σ^+中。则对给定i，必有 $\{ q_1,..,q_m \} \cap W_i$ 是 W_i 的非空真子集，否则因为对任意i，有 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cap \{p_1,...,p_k\}$ 以及 $p_1,...,p_k \rightarrow\rightarrow \{ q_1,..,q_m \} \cap W_i$ 成立，则 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$ 也在 Σ^+中。与假设矛盾。所以根据T的性质 2，$p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cap W_i$ 不成立。假定 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m$ 在 T 上成立，则由于 $p_1,...,p_k \rightarrow\rightarrow W_i$ 在 T 上成立，根据 XMVD5，可得 $p_1,...,p_k \rightarrow\rightarrow q_1,..,q_m \cap W_i$ 在T上成立，这与已证结论矛盾，因此假设不成立。$T \notin \mathrm{SAT}(\sigma)$ 得证。

6 结论

多值依赖语义的引入对于 XML 文档是非常重要的。本文给出的 XML 多值依赖的有效且完备的推理规则集为进一步完善 XML 数据库模式设计奠定了基础。以后的工作将在此基础上对 XML 多值依赖的蕴涵问题、范式及规范化算法问题进行研究。

参考文献

1 Arenas M, Libkin L. A Normal Form for XML Documents[J]. ACM Transactions on Database Systems, 2004, 29(1): 195-232.

2 Vincent M, Liu J, Liu C. Strong Functional Eependencies and Their Application to Normal Forms in XML[J]. ACM Transactions on Database Systems, 2004, 29(3): 445-462.

3 Vincent M, Liu Jixue. Completeness and Decidability Properties for Functional Dependencies in XML[Z]. CoRR cs.DB/0301017: 2003.

4 谈子敬, 庞引明, 施伯乐. XML 上的函数依赖推理[J]. 软件学报, 2003, 14(9): 1564-1570.

5 谈子敬, 施伯乐. 函数依赖和规范化在关系和 XML 间的传播[J]. 软件学报, 2005, 16(4): 533-539.

6 Vincent M, Liu J. Multivalued Dependencies and a 4nf for XML[C]//Proc. of the 15th Conference on Advanced Information Systems Engineering, Klagenfurt, Austria. 2003: 14-29.

7 Liu J, Vincent M W, Liu Chengfei, et al. Checking Multivalued Dependencies in XML[C]//Proceedings of the 7th Asia-Pacific Web Conference on Technologies Research and Development, Shanghai, China. 2005: 320-332.