软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(5):962−971 [doi: 10.3724/SP.J.1001.2011.03807]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

# 紧致依赖与内涵亏值*

马 垣[1+], 张学东[2], 迟呈英[1]

[1](辽宁科技大学 软件学院,辽宁 鞍山 114051)
[2](辽宁科技大学 电子信息工程学院,辽宁 鞍山 114051)

## Compact Dependencies and Intent Waned Values

MA Yuan[1+], ZHANG Xue-Dong[2], CHI Cheng-Ying[1]

[1](School of Software Engineering, Liaoning University of Science and Technology, Anshan 114051, China)
[2](School of Electronics and Information Engineering, Liaoning University of Science and Technology, Anshan 114051, China)
+ Corresponding author: E-mail: mayuanas@sina.com.cn

**Ma Y, Zhang XD, Chi CY. Compact dependencies and intent waned values. *Journal of Software*, 2011,22(5): 962−971.** http://www.jos.org.cn/1000-9825/3807.htm

**Abstract**: In this paper, Intent Waned Values and Compact Dependency are proposed as new concepts. It has strictly been proven that a dependent foundation that consisted of compact dependencies is an irredundant and complete dependency basis on an axiom system that includes only one rule: attributes are increased at the left side and reduced at the right side. In this way, another irredundant and complete dependency basis is found, which is distinguished from the Guigues-Duquenne basis. This basis changes the traditional idea that there exists a sole irredundant and complete dependency basis and discloses the prospect of finding multiple kinds of irredundant and complete dependency basis for a variety of requirements.
**Key words**: value dependency; Guigues-Duquenne basis; compact dependency; intent waned value; database

**摘 要**: 提出了“内涵亏值”与“紧致依赖”的概念,证明了由“紧致依赖”组成的依赖基对于“左部加属性、右部减属性”这一规则的公理系统是无冗余而完整的.由此发现了除 Guigues-Duquenne 基以外还有其他无冗余完整依赖基,改变了只有唯一的一个无冗余完整依赖基的传统观念,揭开了寻找多种无冗余完整依赖基以满足多样化需求的序幕.
**关键词**: 值依赖;Guigues-Duquenne 基;紧致依赖;内涵亏值;数据库
**中图法分类号**: TP311 **文献标识码**: A

值依赖是数据库理论、粗糙集理论、形式概念分析等很多领域共同研究的课题.在数据库理论中,它被称为值依赖[1];在粗糙集理论中,它被称为规则[2];在形式概念分析中,它被称为属性蕴含[3].本文采用数据库中的名称:值依赖.文献[4]中已证明,在一个关系中成立的值依赖的个数是随关系的尺寸而呈指数级增长的,因此要把它们全部表示出来十分困难.这样,Guigues 和 Duquenne 提出了用他们名字命名的 Guigues-Duquenne 基[5],证明

* 基金项目: 国家自然科学基金(60775036)
收稿时间: 2009-03-31; 修改时间: 2009-06-01; 定稿时间: 2009-12-02
CNKI 网络优先出版: 2010-11-26 16:35, http://www.cnki.net/kcms/detail/11.2560.TP.20101126.1635.000.html

了这个基对 Armstrong 公理系统(或称推导规则)是完整的、无冗余的.完整是指所有在该关系中成立的值依赖都可由这个基中的值依赖用 Armstrong 公理系统推导出来,无冗余是指这个基中的每个值依赖都不能由这个基中的其他值依赖用 Armstrong 公理系统推导出来.给出对 Armstrong 公理系统的无冗余完整依赖基是很自然的,因为 Armstrong 公理系统对值依赖是有效而完备的.这里,有效是指从某关系中成立的一个值依赖的集合出发,用 Armstrong 公理系统推导出的每个值依赖都一定在该关系中成立;完备是指在任何关系中每个与这个值依赖集合同时成立的值依赖都可以从这个值依赖集合出发,用 Armstrong 公理系统推导出来.由于 Armstrong 公理系统的这种有效性和完备性,所以多年来人们一直没有考虑是否还有其他公理系统也能有无冗余完整依赖基.人们一直认为,只有有效而完备的公理系统才能有无冗余完整依赖基.然而,我们近年来研究发现,原来还存在另外的公理系统(该系统有效,但不完备)也存在着对这个公理系统的无冗余完整依赖基.这个依赖基由本文提出的“紧致依赖”组成,这种紧致依赖是由本文提出的“内涵亏值”求出的.

另外,由于实际应用的问题是多种多样的,单一的一种依赖基很难方便地满足这些多种多样的需求.例如,在关联规则的挖掘中,这里的紧致依赖的依赖基就完成了 Guigues-Duquenne 基所不能完成的工作(见本文第 3 节).因而,我们的发现还为寻找多种无冗余完整依赖基、满足多种实际问题的需要、实现个性化服务揭开了序幕.

本文第 1 节提出基本定义及基本原理.第 2 节提出紧致依赖及内涵亏值.第 3 节说明紧致依赖在求置信度为 1 的关联规则方面的应用.第 4 节进行分析与展望.

## 1 基本定义及基本定理

本文将用形式概念的方法解决这个问题.

**定义 1**[3]. 设 $U$ 是对象的集合,$M$ 是属性的集合,$I\subseteq U\times M$ 是 $U$ 与 $M$ 间的关系,则称三元组$\mathbb{K}=(U,M,I)$为一个形式背景(简称背景).对于 $X\subseteq U,Y\subseteq M$,定义两个函数:

- $f(X)=\{m\in M|\forall u\in X,(u,m)\in I\}$
- $g(Y)=\{u\in U|\forall m\in Y,(u,m)\in I\}$

而且,若 $f(X)=Y,g(Y)=X$,则称$(X,Y)$是一个形式概念(简称概念).$X$ 是这个概念的外延,$Y$ 是这个概念的内涵.$\mathbb{K}$的全部概念的集合记作 $\mathcal{B}(\mathbb{K})$ .

任何一个关系数据库中的关系都很容易改写成一个背景[1].本文默认关系数据库中的关系都已作了这种转化,于是,本文以下将用背景来论述.

**定义 2**[1]. 设$\mathbb{K}=(U,M,I)$是一个背景,$A,B\subseteq M$,则表达式 $A\rightarrow B$ 表示一个值依赖,$A$ 称为左部,$B$ 称为右部.若按定义 1 中的 $g$ 函数,有 $g(A)\subseteq g(B)$,则称 $A\rightarrow B$ 在$\mathbb{K}$中成立.设$\Sigma$是值依赖的集合.如果$\Sigma$中每个值依赖都在$\mathbb{K}$中成立,则称$\Sigma$在$\mathbb{K}$中成立.

**定义 3**[3]. 设$\mathbb{K}=(U,M,I)$是一个背景,$(X_1,Y_1)$及$(X_2,Y_2)$是它的两个概念.若 $X_1\subseteq X_2$(这时必有 $Y_2\subseteq Y_1$),则称$(X_1,Y_1)$是$(X_2,Y_2)$的子概念,$(X_2,Y_2)$是$(X_1,Y_1)$的父概念.若 $X_1\subset X_2$ 且不存在概念$(X_3,Y_3)$使得 $X_1\subset X_3\subset X_2$,则称$(X_1,Y_1)$是$(X_2,Y_2)$的直接子概念,$(X_2,Y_2)$是$(X_1,Y_1)$的直接父概念.记作$(X_1,Y_1)\prec(X_2,Y_2)$.

**定义 4**. 设$\mathbb{K}=(U,M,I)$是一个背景,令 $V=\mathcal{B}(\mathbb{K}), E=\{(\alpha,\beta)\in\mathcal{B}(\mathbb{K})\times\mathcal{B}(\mathbb{K})\mid\alpha\prec\beta\}$ ,则称以 $V$ 中的元素为顶点、以 $E$ 中的元素为边的图$(V,E)$为$\mathbb{K}$概念集合的 Hasse 图.

**引理 1**[3]. 设$\mathbb{K}=(U,M,I)$是一个形式背景,$X_1,X_2\subseteq U,Y_1,Y_2\subseteq M$,则

(1) $X_1\subseteq X_2\Rightarrow f(X_1)\supseteq f(X_2)$

(2) $Y_1\subseteq Y_2\Rightarrow g(Y_1)\supseteq g(Y_2)$

(3) $f(X_1\cup X_2)=f(X_1)\cap f(X_2)$

(4) $g(Y_1\cup Y_2)=g(Y_1)\cap g(Y_2)$

(5) $X_1\subseteq g(f(X_1))$

(6) $Y_1\subseteq f(g(Y_1))$

(7) $f(X_1)=f(g(f(X_1)))$

(8) $g(Y_1)=g(f(g(Y_1)))$

(9) $X_1\subseteq X_2$ 则 $g(f(X_1))\subseteq g(f(X_2))$

(10) $Y_1\subseteq Y_2$ 则 $f(g(Y_1))\subseteq f(g(Y_2))$

**定义 5**. 背景$\mathbb{K}$的 Guigues-Duquenne 基[5]是这样一些值依赖的集合,它们在$\mathbb{K}$中成立,而且$\mathbb{K}$中成立的每个值依赖都可以从这个集合用 Armstrong 公理系统[6](或称推导规则):(A1) 若 $A\supseteq B$,则 $A\to B$;(A2) 若 $A\to B$,则 $AC\to BC$;(A3) 若 $A\to B,B\to C$,则 $A\to C$ 推导出来(完整),但这个集合中的任何一个值依赖都不能从这个集合中的其他值依赖用以上推导规则(A1)~(A3)推导出来(无冗余).

Guigues-Duquenne 基可通过$\mathbb{K}$的伪内涵求出[3,5,7].

例 1:企业竞争情况的背景$\mathbb{K}=(U,M,I)$,其中,$U=\{1,2,3,4,5\}$是 5 种企业,$M$={有技术创新能力,产品科技含量高,经济实力强,经济信息利用充分,有国家信贷支持,人力资源充分}是关于竞争的属性.为了简化,下面将用 $a$ 表示“有技术创新能力”,$b$ 表示“产品科技含量高”,$c$ 表示“经济实力强”,$d$ 表示“经济信息利用充分”,$e$ 表示“有国家信贷支持”,$h$ 表示“人力资源充分”.$I$ 见表 1.于是,由伪内涵求出的$\mathbb{K}$的 Guigues-Duquenne 基是

$$\{a\to abh,b\to bh,d\to cd,e\to abeh,ce\to abcdeh,abch\to abcdh,bcdh\to abcdh\}.$$

这里,我们用 $ac$ 表示$\{a,c\}$.本文其余地方也采用这种记法.

**Table 1** A formal context

**表 1** 一个形式背景

| | $a$ | $b$ | $c$ | $d$ | $e$ | $h$ |
|---|---|---|---|---|---|---|
| 1 | × | × | | | × | × |
| 2 | | × | × | | | × |
| 3 | | | × | × | | × |
| 4 | × | × | × | × | | × |
| 5 | | | × | × | | |

由于我们发现存在着另外的公理系统,对这个公理系统也有无冗余而完整的值依赖集合,所以将依赖基的定义推广如下:

**定义 6**. 设$\Omega$是一个推导规则的集合(公理系统),$\Sigma$是一个值依赖的集合,$\mathbb{K}$是一个背景.如果$\Sigma$在$\mathbb{K}$中成立,而且$\Sigma$中的每个值依赖都不能由$\Sigma$中的其他值依赖用$\Omega$中的推导规则推出(无冗余),而$\mathbb{K}$中成立的每个值依赖都可由$\Sigma$中的值依赖用$\Omega$中的推导规则推出(完整),则称$\Sigma$是对于$\Omega$的无冗余完整依赖基.

我们发现的推导规则的集合(公理系统)仅包含 1 个推导规则,即若$\Omega=\{A\to B$,则$(A\cup C)\to(B-D)\}$.这里,$C$ 和 $D$ 是任意属性集合.也就是说,我们发现的值依赖集合中的任何一个值依赖都不能由这个集合中的其他值依赖通过左部加属性、右部减属性来得到,而关系中成立的值依赖都可以通过这个集合中的某个值依赖左部加属性、右部减属性来得到.这种依赖集合是由我们提出的“紧致依赖”组成,而紧致依赖可借助我们提出的“内涵亏值”求出.

## 2 紧致依赖与内涵亏值

**定义 7**. 设$\mathbb{K}=(U,M,I)$是一个形式背景,$A,B\subseteq M$,若值依赖 $A\to B$ 在$\mathbb{K}$中成立且对任何 $A_1\subset A,A_1\to B$ 在$\mathbb{K}$中都不成立,对任何 $B_1\supset B,A\to B_1$ 在$\mathbb{K}$中都不成立,则称 $A\to B$ 是$\mathbb{K}$中的紧致依赖.

例 2:对于表 1 所示背景的全部紧致依赖是

$\{c\to c,h\to h,ch\to ch,b\to bh,d\to cd,bc\to bch,dh\to cdh,a\to abh,e\to abeh,ac\to abcdh,ad\to abcdh,bd\to abcdh,$

$ce \to abcdeh, de \to abcdeh\}$.

紧致依赖可以通过内涵亏值及其横截求出.

**定义 8**. 设$\mathbb{K}=(U,M,I)$是一个形式背景,$(X,Y)$是$\mathbb{K}$的一个概念,$(X_1,Y_1)$是$(X,Y)$的直接父概念,则称 $Y-Y_1$ 是$(X,Y)$的一个内涵亏值,简称亏值(概念格的最大元素$(U,f(U))$没有内涵亏值).$(X,Y)$的所有内涵亏值的集合记为$\mathcal{V}(X,Y)$(最大概念的亏值的集合$\mathcal{V}(U,f(U))=\varnothing$),并将$\cup\mathcal{V}(X,Y)$用$\mathcal{P}(X,Y)$表示.这里,$\cup\mathcal{V}(X,Y)$是$\mathcal{V}(X,Y)$中所有集合的并集.

例 3:设背景$\mathbb{K}$如表 1 所示,其概念格的 Hasse 图如图 1 所示.内涵亏值标示在每个概念与其直接父概念之间的边的旁边.例如,对于#11 概念有

$\mathcal{V}(4,abcdh)=\{cd,ad,ab\}$,$\mathcal{P}(4,abcdh)=\cup\mathcal{V}(4,abcdh)=\cup\{cd,ad,ab\}=\{c,d\}\cup\{a,d\}\cup\{a,b\}=\{a,b,c,d\}=abcd$.

再如,对于#8 概念有$\mathcal{V}(24,bch)=\{c,b\}$,$\mathcal{P}(24,bch)=\cup\mathcal{V}(24,bch)=\cup\{c,b\}=bc$.

对于#2 概念有$\mathcal{V}(1234,h)=\{h\}$,$\mathcal{P}(1234,h)=\cup\mathcal{V}(1234,h)=\cup\{h\}=h$ 等.

**定义 9**[8]. 设 $H=(P,V)$是一个超图.这里,$P$ 是点的集合,$V\subseteq 2^P$ 是超边的集合.若 $E\subseteq P$ 满足$\forall v\in V, v\cap E\neq\varnothing$,则称 $E$ 是 $H$ 的一个横截(transversal).如果 $E$ 是 $H$ 的一个横截,而$\forall E'\subset E$,$E'$都不是 $H$ 的横截,则称 $E$ 是 $H$ 的一个最小横截.

**定义 10**. 设$\mathbb{K}=(U,M,I)$是一个形式背景,$(X,Y)$是$\mathbb{K}$的一个概念.把以$\mathcal{P}(X,Y)$为点集、以$\mathcal{V}(X,Y)$为超边集的超图记作$\mathcal{H}(X,Y)$,并称$\mathcal{H}(X,Y)$为$(X,Y)$的亏值超图.

例 4:设$\mathbb{K}$如表 1 所示,其概念格的 Hasse 图如图 1 所示,则#11 概念的亏值超图$\mathcal{H}(4,abcdh)$如图 2 所示,其最小横截有 $ac$,$ad$ 及 $bd$.#8 概念的亏值超图$\mathcal{H}(24,bch)$如图 3 所示,其最小横截是 $bc$.#2 概念的亏值超图$\mathcal{H}(1234,h)$如图 4 所示,其最小横截是 $h$.

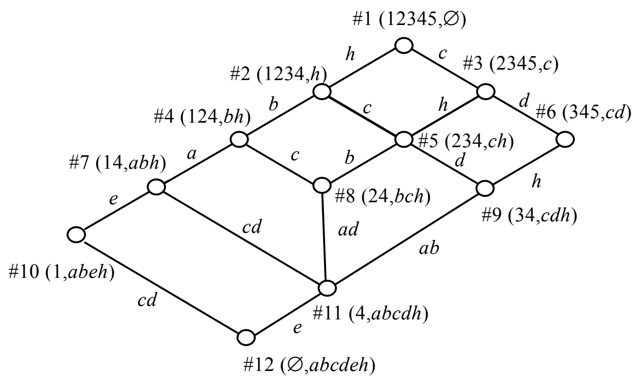


Fig.1 An example of concept lattice and waned values

图 1 一个示例概念格及其亏值

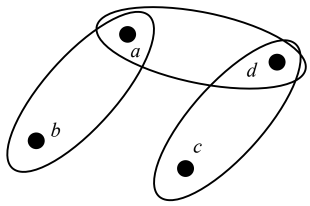


Fig.2 Concept #11 waned values hypergraph

图 2 #11 概念的亏值超图


Fig.3 Concept #8 waned values hypergraph

图 3 #8 概念的亏值超图

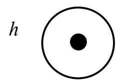


Fig.4 Concept #2 waned values hypergraph

图 4 #2 概念的亏值超图

**引理 2**. 设$\mathbb{K}=(U,M,I)$是一个背景,$B$ 是$\mathbb{K}$的一个内涵,$A$ 是超图$\mathcal{H}(g(B),B)$的一个横截,则 $A\subseteq B$.

证明:因为 $A$ 是超图$\mathcal{H}(g(B),B)$的一个横截,所以 $A$ 是该超图点的集合$\mathcal{P}(g(B),B)$的子集,即所有亏值的并集$\cup\mathcal{V}(g(B),B)$的子集.由于每个亏值都是 $B$ 的子集,所以 $A\subseteq B$. □

**引理 3**. 设$\mathbb{K}=(U,M,I)$是一个背景,$B$ 是$\mathbb{K}$的一个内涵,$A\subseteq B$,则 $g(A)=g(B)$的充要条件是 $A$ 是超图$\mathcal{H}(g(B),B)$的

一个横截.

证明:设$\mathcal{V}(g(B),B)=\{S_1,\ldots,S_m\}$,记产生亏值$S_i$的直接父概念的内涵为$B_i$,即$S_i=B-B_i,1\leqslant i\leqslant m$.

必要性:若$A$是$\mathcal{H}(g(B),B)$的横截,则由引理2可知,$A\subseteq B$.这样,由引理1(2)可知,$g(A)\supseteq g(B)$.下面我们证明$g(A)\supset g(B)$不成立,于是可知,$g(A)=g(B)$.用反证法,若$g(A)\supset g(B)$成立,则由定义3可知,$(g(A),f(g(A)))$是$(g(B),B)$的父概念.于是,必有$(g(B),B)$的某个直接父概念$(g(B_j),B_j)$使得$f(g(A))\subseteq B_j$.这样,由引理1(6)可知,$A\subseteq f(g(A))\subseteq B_j$.然而,$A$是$\mathcal{H}(g(B),B)$的横截,所以对所有$1\leqslant i\leqslant m,A\cap(B-B_i)\neq\varnothing$.于是,对所有$1\leqslant i\leqslant m,A\not\subseteq B_i$.这与$A\subseteq B_j$矛盾.

充分性:若$g(A)=g(B)$,我们证明$A$是$\mathcal{H}(g(B),B)$的横截,用反证法.若不是,则必有某个$S_i$与$A$相交为空,即$A\cap(B-B_i)=\varnothing$.由$g(A)=g(B)$可知,$f(g(A))=f(g(B))$,由引理1(6)可知,$A\subseteq f(g(A))$,所以$A\subseteq f(g(B))$.因为$B$是$\mathbb{K}$的一个内涵,所以由定义1可知,$B=f(g(B))$,所以$A\subseteq B$.这样,由$A\cap(B-B_i)=\varnothing$可知,$A\subseteq B_i$.于是,由引理1(2)可知,$g(A)\supseteq g(B_i)$.由于$(g(B_i),B_i)$是$(g(B),B)$的直接父概念,所以$g(B_i)\supset g(B)=g(A)$.这样就导致$g(A)\supseteq g(B_i)\supset g(B)=g(A)$,即$g(A)\supset g(A)$,出现了矛盾. □

**定理 1**. $A\rightarrow B$是$\mathbb{K}$中紧致依赖的充要条件是:(1) $B$是内涵;(2) $A$是超图$\mathcal{H}(g(B),B)$的一个最小横截.

证明:

必要性:设$A\rightarrow B$是$\mathbb{K}$的紧致依赖,我们证明$B$一定是内涵,$A$一定是$\mathcal{H}(g(B),B)$的最小横截.首先,如果$B$不是内涵,则由定义1可知,$B\neq f(g(B))$.令$B_1=f(g(B))$,则$B\neq B_1$.由于根据引理1(6)可知,$B\subseteq f(g(B))$,所以$B\subset f(g(B))=B_1$.然而,$A\rightarrow B$在$\mathbb{K}$中成立,所以$g(A)\subseteq g(B)$.于是,再由引理1(8)就有$g(A)\subseteq g(B)=g(f(g(B)))$.因为$B_1=f(g(B))$,所以$g(f(g(B)))=g(B_1)$.于是,$g(A)\subseteq g(B)=g(f(g(B)))=g(B_1)$.所以,$A\rightarrow B_1$在$\mathbb{K}$中也成立.这样,$B\subset B_1$将与$A\rightarrow B$是紧致依赖矛盾.所以,$B$是内涵.

其次,我们证明$A\subseteq B$.用反证法,若$A\not\subseteq B$,则对$B_2=A\cup B$有$B\subset B_2$,而且由引理1(4)可知,$g(B_2)=g(A\cup B)=g(A)\cap g(B)$.由于$A\rightarrow B$在$\mathbb{K}$中成立,所以$g(A)\subseteq g(B)$,所以$g(A)\cap g(B)=g(A)$.这样,$g(B_2)=g(A)$,所以$A\rightarrow B_2$在$\mathbb{K}$中也成立.但$B\subset B_2$,这与$A\rightarrow B$是紧致依赖矛盾.于是,$A\subseteq B$.

这样,由引理1(2)就有$g(A)\supseteq g(B)$.再由$A\rightarrow B$在$\mathbb{K}$中成立,所以$g(A)\subseteq g(B)$.于是,$g(A)=g(B)$.这样,因为$B$是内涵,$A\subseteq B$及$g(A)=g(B)$,所以由引理3可知,$A$是$\mathcal{H}(g(B),B)$的一个横截.如果$A$不是最小横截,则有$A$的真子集$A_1$是最小横截.于是,$A_1\subset A\subseteq B$.这样,由引理3可知,$g(A_1)=g(B)$,从而$A_1\rightarrow B$在$\mathbb{K}$中成立.这与$A\rightarrow B$是紧致依赖矛盾.所以,$A$是最小横截.

充分性:设$B$是内涵,而$A$是超图$\mathcal{H}(g(B),B)$的最小横截,我们证明$A\rightarrow B$是紧致依赖.首先,由引理2可知,$A\subseteq B$,再因$A$是横截,所以由引理3可知,$g(A)=g(B)$.于是:① $A\rightarrow B$在$\mathbb{K}$中成立;② 对任何$A_1\subset A$都有$g(A_1)\supseteq g(A)$,并由$A$的最小性可知,$A_1$不是$H$的横截.所以,由引理3可知,$g(A_1)\neq g(B)=g(A)$;所以,$g(A_1)\supset g(A)=g(B)$.这样,$A_1\rightarrow B$在$\mathbb{K}$中不成立;③ 对任何$B_1\supset B$,由引理1(2)有$g(B_1)\subseteq g(B)$.但$g(B_1)\neq g(B)$,这是因为,若$g(B_1)=g(B)$,则$f(g(B_1))=f(g(B))$.再由引理1(6),并注意到$A\cup B$是内涵,就有$B_1\subseteq f(g(B_1))=f(g(B))=B$.这与$B_1\supset B$矛盾,于是可知,$g(B_1)\neq g(B)$.于是,$g(B_1)\subset g(B)=g(A)$.所以,$A\rightarrow B_1$在$\mathbb{K}$中不成立.这样,我们证明了:① $A\rightarrow B$在$\mathbb{K}$中成立;② 对任何$A_1\subset A,A_1\rightarrow B$在$\mathbb{K}$中不成立;③ 对任何$B_1\supset B,A\rightarrow B_1$在$\mathbb{K}$中不成立.因此,按定义7,$A\rightarrow B$是紧致依赖. □

例5:由定理1可知,每个紧致依赖的右部都是内涵,左部都是该内涵的亏值超图的最小横截.由于除最大概念$(U,f(U))$外每个概念都有亏值,所以上例的全部紧致依赖是:由#2概念得到的$h\rightarrow h$;由#3概念得到的$c\rightarrow c$;由#4概念得到的$b\rightarrow bh$;由#5概念得到的$ch\rightarrow ch$;由#6概念得到的$d\rightarrow cd$;由#7概念得到的$a\rightarrow abh$;由#8概念得到的$bc\rightarrow cbh$;由#9概念得到的$dh\rightarrow cdh$;由#10概念得到的$e\rightarrow abeh$;由#11概念得到的$ac\rightarrow abcdh,ad\rightarrow abcdh,bd\rightarrow abcdh$以及由#12概念得到的$ce\rightarrow abcdeh,de\rightarrow abcdeh$.这就是例2中展示的全部紧致依赖.

**定理 2**. 背景$\mathbb{K}$中成立的值依赖都可以由某个紧致依赖用左部加属性、右部减属性的方法得出.

证明:我们只要证明:若 $A\to B$ 在$\mathbb{K}$中成立,则必有$\mathbb{K}$中的紧致依赖 $A_0\to B_0$,使 $A\supseteq A_0,B\subseteq B_0$ 即可.针对 $A\to B$,令 $B_0=f(g(A\cup B))$,则由定义 1 可知,$B_0$ 是内涵.因为由引理 1(6)有 $B\subseteq f(g(B))$,而由引理 1(10)有 $f(g(B))\subseteq f(g(B\cup A))$,所以,① $B\subseteq f(g(B\cup A))=B_0$.又由于 $A\to B$ 成立,所以 $g(A)\subseteq g(B)$,从而 $g(A)=g(A)\cap g(B)$.再由引理 1(4)可知,$g(A)\cap g(B)=g(A\cup B)$,由引理 1(8)可知,$g(A\cup B)=g(f(g(A\cup B)))$,所以 $g(A)=g(f(g(A\cup B)))$.又因 $B_0=f(g(A\cup B))$,于是可知,$g(B_0)=g(f(g(A\cup B)))$,所以有 $g(A)=g(B_0)$.另外,由引理 1(6)还有 $B_0=f(g(A\cup B))\supseteq A\cup B\supseteq A$.这样,由引理 3 可知,$A$ 是 $\mathcal{H}(g(B_0),B_0)$的一个横截.于是,② 有某个 $A$ 的子集 $A_0$ 是 $\mathcal{H}(g(B_0),B_0)$的最小横截.由定理 1 可知,$A_0\to B_0$ 是紧致依赖.而且,由结论①、结论②,还有 $A\supseteq A_0,B\subseteq B_0$. □

**定理 3**. 若 $A_1\to B_1$, $A_2\to B_2$ 是两个不同的紧致依赖,则 $A_1\to B_1$ 不能从 $A_2\to B_2$ 通过左部加属性、右部减属性的方法得出.

证明:用反证法.设 $A_1\to B_1$ 可以从 $A_2\to B_2$ 通过左部加属性、右部减属性的方法得到,则必有 $A_1\supseteq A_2$ 及 $B_1\subseteq B_2$.然而它们都是紧致依赖,所以由定理 1 可知,$B_1$ 及 $B_2$ 都是内涵,所以$(g(B_1),B_1)$是$(g(B_2),B_2)$的父概念.下面我们来证明:如果$(g(B_1),B_1)$是$(g(B_2),B_2)$的父概念,则 $A_1\supseteq A_2$ 是不可能的.

因为 $A_1\to B_1$ 是紧致依赖,所以由定理 1 可知,$A_1$ 是 $\mathcal{H}(g(B_1),B_1)$的最小横截.于是,由引理 2 可知,$A_1\subseteq B_1$.由于$(g(B_1),B_1)$是$(g(B_2),B_2)$的父概念,所以存在概念序列$(X_1,Y_1),(X_2,Y_2),\dots,(X_p,Y_p)$,其中,前面一个是后面一个的直接父概念,而且$(X_1,Y_1)=(g(B_1),B_1),(X_p,Y_p)=(g(B_2),B_2)$.于是,$A_1\subseteq B_1=Y_1\subset Y_2\subset\dots\subset Y_{p-1}\subset Y_p=B_2$.特别是,$A_1\subseteq Y_{p-1}\subset Y_p=B_2$ 就意味着 $B_2$ 与其直接父概念的内涵 $Y_{p-1}$ 中都含有 $A_1$,因而 $B_2$ 对 $Y_{p-1}$ 的亏值中将不含 $A_1$.这样,$\mathcal{H}(g(B_2),B_2)$的任何横截中都要含有不属于 $A_1$ 的元素,因而都不会是 $A_1$ 的子集.因此,作为 $\mathcal{H}(g(B_2),B_2)$最小横截的 $A_2$ 也不会是 $A_1$ 的子集.即,若 $B_1\subseteq B_2$,则 $A_1\not\supseteq A_2$.所以,$A_1\to B_1$ 不能从 $A_2\to B_2$ 通过左部加属性、右部减属性的方法得到. □

这样,由定理 2、定理 3 可知,若设$\Omega=\{A\to B$,则$(A\cup C)\to(B-D)\}$,这里,$C,D$ 是任意属性集合,而$\Sigma$是所有紧致依赖组成的集合,则$\Sigma$是相对于$\Omega$的无冗余完整依赖基.

由于紧致依赖组成的依赖基的无冗余性和完整性不是对 Armstrong 公理系统的,所以在某些方面就弥补了对 Armstrong 公理系统无冗余完整依赖基的不足.例如,在求置信度为 1 的关联规则方面就比 Guigues-Duquenne 基有优越性,显示了不同依赖基可能会有不同特点的服务,我们称为个性化服务.

## 3 紧致依赖在求置信度为 1 的关联规则方面的应用

求置信度为 1 的关联规则用形式概念的“背景”、“函数 $g$”等术语(见定义 1)来叙述,则是这样一个任务:

**定义 11**. 设$\mathbb{K}=(U,M,I)$是一个背景,把 $M$ 中的元素称为项,把 $M$ 的子集称为项集,把 $U$ 中的元素称为事务,则$\mathbb{K}$是挖掘关联规则的事务数据.如果 $A,B\subseteq M,A\cap B=\varnothing$,则表达式 $A\Rightarrow B$ 称为一个关联规则,称$|g(A\cup B)|$为其频率,称 $\frac{|g(A\cup B)|}{|U|}$ 为其支持度,称 $\frac{|g(A\cup B)|}{|g(A)|}$ 为其置信度.如果 $\frac{|g(A\cup B)|}{|U|}\geqslant sup$ 以及 $\frac{|g(A\cup B)|}{|g(A)|}=1$,则称 $A\Rightarrow B$ 为置信度为 1 的强关联规则.这里,$sup$ 是支持度阈值.人们可先求出全部置信度为 1 的关联规则及其频率,然后再从这些关联规则中选出支持度合格的强关联规则.

**定理 4**. 设 $K=(U,M,I)$是一个背景,$A\to C$ 是 $K$ 上的一个紧致依赖,

(1) 若 $B=C-A$,则 $A\Rightarrow B$ 是一个置信度为 1 的关联规则,且其频率为$|g(A)|$.

(2) 若将 $B$ 中一些属性 $A_0$ 移到 $A$ 中形成 $A'$,而 $B$ 中剩余部分为 $B_1$(即 $A'=A\cup A_0,B_1=B-A_0$),则 $A'\Rightarrow B_1$ 也是一个频率为$|g(A)|$、置信度为 1 的关联规则.

(3) 将 $B_1$ 中一些属性 $B_0$ 删除形成 $B'$(即 $B'=B_1-B_0$),则 $A'\to B'$ 仍是频率为$|g(A)|$、置信度为 1 的关联规则.

证明:(1) 由于 $A\to C$ 是$\mathbb{K}$上的一个紧致依赖,所以由定理 1 和引理 2 可知,$A\subseteq C$,再由引理 1(2)可知,$g(A)\supseteq g(C)$.然而 $A\to C$ 在$\mathbb{K}$上成立,所以 $g(A)\subseteq g(C)$.这样,$g(A)=g(C)$.但是 $B=C-A$,即 $A\cup B=C$,所以,$A\Rightarrow B$ 的置信度为 $\frac{|g(A\cup B)|}{|g(A)|}=\frac{|g(C)|}{|g(A)|}=1$.由于按定义 11,其频率为$|g(A\cup B)|$,而现在 $g(A\cup B)=g(C)=g(A)$,所以频率为$|g(A)|$.

(2) 因 $A'=A\cup A_0$ 及 $B_1=B-A_0$,所以 $A'\cup B_1=A\cup B$ 及 $A'\supseteq A$.

由引理 1(2)可知,$g(A')\subseteq g(A)$以及 $\frac{|g(A'\cup B_1)|}{|g(A')|}=\frac{|g(A\cup B)|}{|g(A')|}\geqslant\frac{|g(A\cup B)|}{|g(A)|}=1$,所以 $\frac{|g(A'\cup B_1)|}{|g(A')|}=1$,即 $A'\Rightarrow B_1$ 的置信度为 1.再由 $A'\cup B_1=A\cup B$ 可知,$g(A'\cup B_1)=g(A\cup B)=g(A)$,所以,$A'\Rightarrow B_1$ 的频率仍为$|g(A)|$.

(3) 因为 $B'=B_1-B_0$,所以 $B'\subseteq B_1$,从而 $A'\cup B'\subseteq A'\cup B_1$,由引理 1(2)可知,$g(A'\cup B')\supseteq g(A'\cup B_1)$,所以 $A'\Rightarrow B'$的置信度为 $\frac{|g(A'\cup B')|}{|g(A')|}\geqslant\frac{|g(A'\cup B_1)|}{|g(A')|}=1$,所以 $A'\Rightarrow B'$的置信度为 1.又由于 $A'=A\cup A_0$,所以 $A'\supseteq A$,故 $A'\cup B'\supseteq A'\supseteq A$.于是,由引理 1(2)可知,$g(A)\supseteq g(A'\cup B')\supseteq g(A'\cup B_1)=g(A\cup B)$,由于 $g(A)=g(C)$以及 $A\cup B=C$,所以 $g(A'\cup B')=g(A\cup B)=g(C)=g(A)$,故频率仍为$|g(A)|$. □

**定理 5**. 从紧致依赖按定理 4(1)~定理 4(3)求出的置信度为 1 的关联规则是全部置信度为 1 的关联规则.

证明:我们只要证明以下问题即可:若 $A'\Rightarrow B'$是置信度为 1 的关联规则,则存在紧致依赖 $A\rightarrow C$ 使:(1) $A'=A\cup A_0$,这里 $A_0\subseteq C-A$,而且(2) $B'=B_1-B_0$,这里 $B_1=B-A_0$ 而 $B=C-A$.

令 $C=f(g(A'\cup B'))$,于是,由定义 1 可知,① $C$ 是内涵,而且由引理 1(8)可知,$g(C)=g(f(g(A'\cup B')))=g(A'\cup B')$.由于 $A'\Rightarrow B'$的置信度为 1,所以,② $g(A')=g(A'\cup B')=g(C)$.再由引理 1(6)可知,$A'\cup B'\subseteq f(g(A'\cup B'))$,而 $C=f(g(A'\cup B'))$,所以,③ $A'\subseteq C$.这样,由结论①~结论③,根据引理 3 可知,$A'$是$\mathcal{H}(g(C),C)$的横截.设 $A$ 是满足 $A\subseteq A'$的最小横截,则由定理 1 可知,$A\rightarrow C$ 是一个紧致依赖.对于这个紧致依赖有:

(1) $A\subseteq A'$,令 $A_0=A'-A$,则 $A'=A\cup A_0$.

(2) 由于 $A'\cup B'\subseteq f(g(A'\cup B'))=C$,而 $A'\Rightarrow B'$是关联规则,所以 $A'\cap B'=\varnothing$.所以,若令 $B=C-A$,则有 $B'\subseteq C-A'=C-(A\cup A_0)=C-A-A_0=B-A_0=B_1$,即 $B'\subseteq B_1$.令 $B_0=B_1-B'$,则 $B'=B_1-B_0$. □

例 6:对于表 1 中的背景,由紧致依赖组成的依赖基是

$\{c\rightarrow c,h\rightarrow h,ch\rightarrow ch,b\rightarrow bh,d\rightarrow cd,bc\rightarrow bch,dh\rightarrow cdh,a\rightarrow abh,e\rightarrow abeh,ac\rightarrow abcdh,ad\rightarrow abcdh,bd\rightarrow abcdh,$
$ce\rightarrow abcdeh,de\rightarrow abcdeh\}$.

① 根据定理 4(1),对每个紧致依赖 $A\rightarrow C$ 都产生置信度为 1 的关联规则 $A\Rightarrow C-A$,且频率为$|g(A)|$,它们是(小括号中的数字是该关联规则的频率):

$c\Rightarrow\varnothing(4),h\Rightarrow\varnothing(4),ch\Rightarrow\varnothing(3),b\Rightarrow h(3),d\Rightarrow c(3),bc\Rightarrow h(2),dh\Rightarrow c(2),a\Rightarrow bh(2),e\Rightarrow abh(1),ac\Rightarrow bdh(1),$
$ad\Rightarrow bch(1),bd\Rightarrow ach(1),ce\Rightarrow abdh(0),de\Rightarrow abch(0)$.

去掉右边是空的及频率为 0 的,得到 8 个规则:

$b\Rightarrow h(3),d\Rightarrow c(3),bc\Rightarrow h(2),dh\Rightarrow c(2),a\Rightarrow bh(2),e\Rightarrow abh(1),ac\Rightarrow bdh(1),ad\Rightarrow bch(1),bd\Rightarrow ach(1)$.

② 根据定理 4(2),若 $A\Rightarrow B$ 是上面产生的关联规则,则把 $B$ 的任何真子集 $A_0$ 移到 $A$ 形成的 $A\cup A_0\Rightarrow B-A_0$ 是与 $A\Rightarrow B$ 同频率的置信度为 1 的关联规则.上面的关联规则右部有非空真子集的是:

- $a\Rightarrow bh(2)$,由它得:$ah\Rightarrow b(2),ab\Rightarrow h(2)$.
- $e\Rightarrow abh(1)$,由它得:$ae\Rightarrow bh(1),be\Rightarrow ah(1),eh\Rightarrow ab(1),abe\Rightarrow h(1),aeh\Rightarrow b(1),beh\Rightarrow a(1)$.
- $ac\Rightarrow bdh(1)$,由它得:$abc\Rightarrow dh(1),acd\Rightarrow bh(1),ach\Rightarrow bd(1),abcd\Rightarrow h(1),abch\Rightarrow d(1),acdh\Rightarrow b(1)$.
- $ad\Rightarrow bch(1)$,由它得:$abd\Rightarrow ch(1),[acd\Rightarrow bh(1)],adh\Rightarrow bc(1),[abcd\Rightarrow h(1)],abdh\Rightarrow c(1),[acdh\Rightarrow b(1)]$.
- $bd\Rightarrow ach(1)$,由它得:$bcd\Rightarrow ah(1),[abd\Rightarrow ch(1)],bdh\Rightarrow ac(1),[abcd\Rightarrow h(1)],[abdh\Rightarrow c(1)],bcdh\Rightarrow a(1)$.

其中,有方括号的规则是前面已出现过的.

③ 根据定理 4(3),若 $A'\Rightarrow B_1$ 是上面产生的关联规则,$B_0$ 是 $B_1$ 的非空真子集,$B'=B_1-B_0$,则 $A'\Rightarrow B'$也是同频率的置信度为 1 的关联规则.由上面右部有非空真子集的关联规则产生的新关联规则是:

- $a\Rightarrow b(2),a\Rightarrow h(2)$.
- $e\Rightarrow ab(1),e\Rightarrow ah(1),e\Rightarrow bh(1),e\Rightarrow a(1),e\Rightarrow b(1),e\Rightarrow h(1)$.
- $ae\Rightarrow b(1),ae\Rightarrow h(1),be\Rightarrow a(1),be\Rightarrow h(1),eh\Rightarrow a(1),eh\Rightarrow b(1)$.
- $ac\Rightarrow bd(1),ac\Rightarrow bh(1),ac\Rightarrow dh(1),ac\Rightarrow b(1),ac\Rightarrow d(1),ac\Rightarrow h(1)$.

- $abc \Rightarrow d(1), abc \Rightarrow h(1), acd \Rightarrow b(1), acd \Rightarrow h(1), ach \Rightarrow b(1), ach \Rightarrow d(1)$.
- $ad \Rightarrow bc(1), ad \Rightarrow bh(1), ad \Rightarrow ch(1), ad \Rightarrow b(1), ad \Rightarrow c(1), ad \Rightarrow h(1)$.
- $abd \Rightarrow c(1), abd \Rightarrow h(1), adh \Rightarrow b(1), adh \Rightarrow c(1)$.
- $bd \Rightarrow ac(1), bd \Rightarrow ah(1), bd \Rightarrow ch(1), bd \Rightarrow a(1), bd \Rightarrow c(1), bd \Rightarrow h(1)$.
- $bcd \Rightarrow a(1), bcd \Rightarrow h(1), bdh \Rightarrow a(1), bdh \Rightarrow c(1)$.

根据定理 5,只有以上这些规则是频率不为 0 的置信度为 1 的关联规则.

这样,我们就得到了全部的频率不为 0 的置信度为 1 的关联规则如下:

$b \Rightarrow h(3), d \Rightarrow c(4), bc \Rightarrow h(2), dh \Rightarrow c(2), a \Rightarrow bh(2), e \Rightarrow abh(1), ac \Rightarrow bdh(1), ad \Rightarrow bch(1), bd \Rightarrow ach(1), ah \Rightarrow b(2), ab \Rightarrow h(2), ae \Rightarrow bh(1), be \Rightarrow ah(1), eh \Rightarrow ab(1), abe \Rightarrow h(1), aeh \Rightarrow b(1), beh \Rightarrow a(1), acb \Rightarrow dh(1), acd \Rightarrow bh(1), ach \Rightarrow bd(1), abcd \Rightarrow h(1), abch \Rightarrow d(1), acdh \Rightarrow b(1), abd \Rightarrow ch(1), adh \Rightarrow bc(1), abdh \Rightarrow c(1), bcd \Rightarrow ah(1), bdh \Rightarrow ac(1), bcdh \Rightarrow a(1), a \Rightarrow b(2), a \Rightarrow h(2), e \Rightarrow ab(1), e \Rightarrow ah(1), e \Rightarrow bh(1), e \Rightarrow a(1), e \Rightarrow b(1), e \Rightarrow h(1), ae \Rightarrow b(1), ae \Rightarrow h(1), be \Rightarrow a(1), be \Rightarrow h(1), eh \Rightarrow a(1), eh \Rightarrow b(1), ac \Rightarrow bd(1), ac \Rightarrow bh(1), ac \Rightarrow dh(1), ac \Rightarrow b(1), ac \Rightarrow d(1), ac \Rightarrow h(1), abc \Rightarrow d(1), abc \Rightarrow h(1), acd \Rightarrow b(1), acd \Rightarrow h(1), ach \Rightarrow b(1), ach \Rightarrow d(1), ad \Rightarrow bc(1), ad \Rightarrow bh(1), ad \Rightarrow ch(1), ad \Rightarrow b(1), ad \Rightarrow c(1), ad \Rightarrow h(1), abd \Rightarrow c(1), abd \Rightarrow h(1), adh \Rightarrow b(1), adh \Rightarrow c(1), bd \Rightarrow ac(1), bd \Rightarrow ah(1), bd \Rightarrow ch(1), bd \Rightarrow a(1), bd \Rightarrow c(1), bd \Rightarrow h(1), bcd \Rightarrow a(1), bcd \Rightarrow h(1), bdh \Rightarrow a(1), bdh \Rightarrow c(1)$.

定理 4 中的方法很容易由计算机程序实现.

**算法 1**. 求全部频率不为 0、左部右部都不空、置信度为 1 的关联规则.

输入:形式背景$\mathbb{K}$及它的所有紧致依赖的集合$\Sigma_0$.

输出:全部频率不为 0、左部右部都不空、置信度为 1 的关联规则的集合$\Sigma$.

方法:$\Sigma := \varnothing$

```
For each A→C∈Σ0                    //求按定理 4(1)形成的关联规则
    Begin B:=C−A
        If B≠∅且 g(A)≠∅ Then Σ:=Σ∪{A⇒B(|g(A)|)}
    End
    Σ′:=∅                          //用Σ′存放定理 4(2)形成的关联规则
For each A⇒B(|g(A)|)∈Σ
    Begin If |B|>1 Then            //|B|>1 表示 B 有非空真子集
        For each A0⊂B
            If A0≠∅ Then
                Begin A′:=A∪A0
                    B1:=B−A0
                    Σ′:=Σ′∪{A′⇒B1(|g(A)|)}
                End
    End
    Σ:=Σ∪Σ′
    Σ′:=∅                          //用Σ′存放定理 4(3)形成的关联规则
For each A′⇒B1(|g(A)|)∈Σ
    Begin If |B1|>1 Then           //|B1|>1 表示 B1 有非空真子集
        For each A1⊂B1
            If A1≠∅ Then
            Begin B′:=B1−A1
                Σ′:=Σ′∪{A′⇒B′(|g(A)|)}
```

End

End

$\Sigma:=\Sigma\cup\Sigma'$

输出$\Sigma$,结束.

从定理 4、定理 5 及算法 1 可以看出,由于紧致依赖组成的依赖基对“左部加属性、右部减属性”这一规则的公理系统是无冗余而且完整的,所以在求全部置信度为 1 的关联规则方面非常方便.

## 4 分析与展望

20 多年来,人们一直认为无冗余完整依赖基一定是对有效而完备的 Armstrong 公理系统的.本文提出的由紧致依赖组成的依赖基是第一个对 Armstrong 公理以外的另一个公理系统的无冗余完整依赖基.它展示了对于有效但不完备的公理系统也可以有无冗余完整依赖基.这个发现在理论上有很大意义,它改变了只有 1 种无冗余完整依赖基的观念,展现了寻找多种无冗余完整依赖基是可能的.

寻求多种无冗余完整依赖基在实践上有很大意义,因为在各种实际问题中会遇到各种不同的需求,单纯一种依赖基难以全部满足.比如,一些知识发现的任务需要(或在本质上是需要)置信度为 1 的强关联规则[9−12].众所周知,从 Guigues-Duquenne 基用 Armstrong 公理系统只能推出具体的置信度为 1 的强关联规则.由于 Armstrong 公理系统的推导既没有统一的原则又没有固定的步骤,所以没有从 Guigues-Duquenne 基求出全部置信度为 1 的关联规则的可行算法.因此,没有办法用计算机程序从 Guigues-Duquenne 基求出全部置信度为 1 的强关联规则,这一直是 Armstrong 公理系统的一个很大的缺欠.而我们提出的依赖基推导时使用的不是 Armstrong 公理系统,而是只含 1 个推导规则的公理系统,所以就可以很容易地由本文算法 1 求出所有置信度为 1 的关联规则,弥补了 Armstrong 公理系统依赖基的这一缺欠.这说明,不同的依赖基有可能满足不同的需求,它们将会有不同的具有自己特点的服务.因此,发现能够有多种无冗余完整依赖基这件事的实践意义也是很大的.

另外,本文的发现不仅有较大的理论意义及实践意义,而且还引出了一系列新课题,例如:

(1) 还有哪些公理系统可以有无冗余完整依赖基?

(2) 判定一个公理系统可以有无冗余完整依赖基的准则是什么?

(3) 有无冗余完整依赖基的公理系统一共有多少?

(4) 如何把它们全部找出来?

(5) 每个具有无冗余完整依赖基的公理系统对应的无冗余完整依赖基是否唯一?

(6) 每个依赖基都有哪些特性?都可用于哪些个性化服务?

目前,关于“依赖”的研究仍然很多[13−20],上述这些问题显然也给关于“依赖”的研究提出了新的课题.这些课题的研究解决,将会有更大的理论及实践意义.今后,我们希望在这些问题上继续和众多的老师、专家一起进行研究,争取得到更多、更有意义的新成果.

**References**:


[1] Ma Y. Database Theory. In: New Advancement of Database Technology. 2nd ed., Beijing: Tsinghua University Press, 2007. 24−44 (in Chinese).

[2] Tsumoto S. Medical reasoning and rough sets. In: Kryszkiewicz M, *et al*., eds. Proc. of the Int'l Conf. on Rough Sets and Intelligent Systems Paradigms. New York: Springer-Verlag, 2007. 90−101. [doi: 10.1007/978-3-540-73451-2_11]

[3] Ganter B, Wille R. Formal Concept Analysis Mathematical Foundations. Berlin: Springer-Verlag, 1999. 62−75.

[4] Mannila H, Raibba KJ. On the complexity of inferring functional dependencies. Discrete Applied Mathematics, 1992,40(2): 237−243. [doi: 10.1016/0166-218X(92)90031-5]

[5] Guigues J, Duquenne V. Familles Minimales d'implications Informatives Resultants d'un Tableau de Donnees Binaires. Mathematics and Social Sciences, 1986. 495−518.

[6] Ullman JD. Principles of Database Systens. 2nd ed., New York: Computer Science Press, 1982. 218−220.

[7] Valtchev P, Missaoui R, Godin R. Formal concept analysis for knowledge discovery and data mining: The new challenges. In: Garbonell JG, Siekmann J, eds. Proc. of the Int'l Conf. on Concept Analysis (ICFCA 2004). New York: Springer-Verlag, 2004. 252−271. [doi: 10.1007/978-3-540-24651-0_30]
[8] Eiter T, Gottlob G. Identifying the minimal transversals of a hypergraph and related problems. SIAM Journal on Computing, 1995, 24(6):1278−1340. [doi: 10.1137/S0097539793250299]
[9] Ma Y, Gong X, Tang XM, Dong H. Discovering knowledge of increase or decrease and knowledge of circle in relational database. In: Chinese Association for Artificial Intelligence, ed. Proc. of the 2007 National Conf. on Artificial Intelligence (CAAI-12). Beijing: Beijing University of Posts and Telecommunications Publishing House, 2007. 283−289 (in Chinese with English abstract).
[10] Alachaher LN, Guillaume S. Mining negative and positive influence rules using Kullback-Leibler divergence. In: Proc. of the ICCGI 25. 2007. http://www.lw20.com/201007279930812.html
[11] Boulicaut JF, Bykowski A, Jeudy B. Towards the tractable discovery of association rules with negations. In: Proc. of the FQAS. New York: Springer-Verlag, 2000. 425−434.
[12] Antonie ML, Zaiane OR. Mining positive and negative association rules: An approach for confined rules. In: Boulicaut JF, *et al*., eds. Proc. of the PKDD 2004. LNCS (LNAI) 3202, Heidelberg: Springer-Verlag, 2004. 27−38.
[13] Georgieva T. Discovering branching and fractional dependencies in databases. Data & Knowledge Engineering, 2008,66(2): 311−325. [doi: 10.1016/j.datak.2008.04.002]
[14] Fan WF, Ma S, Hu YL, Liu J, Wu YH. Propagating functional dependencies with conditions. Processing of the VLDB Endowment, 2008,1(1):391−407. [doi: 10.1145/1453856.1453901]
[15] Lee CC, Jiang JHR, Huang CY, Mishchenko A. Scalable exploration of functional dependency by interpolation and incremental SAT solving. In: Proc. of the 2007 IEEE/ACM Int'l Conf. on Computer-Aided Design. 2007. 227−233. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.71.408 [doi: 10.1109/ICCAD.2007.4397270]
[16] Sali A, Schewe KD. Keys and Armstrong databases in trees with restructuring. Acta Cybernetica, 2008,18(3):529−556.
[17] Chen TX, Liu SSQ, Meyer MD, Gotterbarn D. An introduction to functional independency in relational database normalization. In: Proc. of the 45th ACM Annual Southeast Regional Conf. 2007. 221−225. http://www.lw20.com/20110323188486781.html [doi: 10.1145/1233341.1233381]
[18] Hartmann S, Link S, Schewe KD. Functional and multivalued dependencies in nested databases generated by record and list constructor. Annals of Mathematics and Artificial Intelligence, 2006,46(1-2):114−164. [doi: 10.1007/s10472-006-9023-4]
[19] Hartmann S, Link S. On a problem of Fagin concerning multivalued dependencies in relational databases. Theoretical Computer Science, 2006,353(1):53−62. [doi: 10.1016/j.tcs.2005.08.036]
[20] Wei Q, Chen GQ. Optimized algorithm of discovering functional dependencies with degrees of satisfaction. In: Proc. of the 7th Int'l FLINS Conf. on Applied Artificial Intelligence. New York: Springer-Verlag, 2006. 401−415. [doi: 10.1007/978-3-540-70812-4_24]

## 附中文参考文献:

[1] 马垣.数据库理论.见:数据库技术新进展.第 2 版,北京:清华大学出版社,2005.24−44.
[9] 马垣,宫玺,汤新明,董辉.关系数据库中增减型与循环型知识的发现.见:中国人工智能学会,编.中国人工智能学会第 12 届学术年会论文集.北京:北京邮电大学出版社,2007.283−289.

**马垣**(1941－),男,北京人,教授,博士生导师,主要研究领域为数据库理论,知识发现,Rough 集理论,形式概念分析.

**迟呈英**(1963－),女,教授,主要研究领域为中文信息处理,数据库技术.

**张学东**(1963－),男,博士,教授,主要研究领域为多维数字信号处理,计算机视觉理论及应用,生物信息识别理论及应用,嵌入式系统的理论与应用.