第六章 关系数据理论

多值依赖与公理系统

一. **多值依赖及第四范式**

二. **数据依赖的公理系统**

- 逻辑蕴涵
- Armstrong公理系统
- 函数依赖集的闭包
- 函数依赖集的等价及最小函数依赖集

1）多值依赖（MultiValued Dependency，简记MVD）

- **定义6.9**

设R（U）是一个属性集U上的一个关系模式， X、 Y和Z是U的子集，并且Z＝U－X－Y。关系模式R（U）中**多值依赖** X→→Y成立，当且仅当对R（U）的**任一关系r**，给定的一对（x，z）值，有一组Y的值，这组值仅仅决定于x值而与z值无关；

若X→→Y，而Z=Φ，则称X→→Y为平凡的多值依赖，否则为非平凡的多值依赖。

例 Teaching（课程C, 教师T, 参考书B）

对于C的每一个值，T有一组值与之对应，而不论B取何值；

因此C →→T。

多值依赖具有下列性质：

- 多值依赖具有对称性。若X→→Y，则X→→Z，其中Z=U-X-Y；
- 函数依赖可以看作特殊的多值依赖。若 X→Y，则X→→Y；

多值依赖和函数依赖的基本区别：

- 在关系模式R（U）中，函数依赖X→Y的有效性决定于X，Y这两个属性集的值。只要在R（U）的任何一个关系r中，元组在X和Y上的值满足函数依赖的定义，则函数依赖X→Y在任何属性集W（XY⊆W ⊆ U）上成立。
- 对于多值依赖来说，若X→→Y在W（W ⊂ U）上成立，而在U上则不一定成立。所以多值依赖的有效性与属性集的范围有关。X→→Y在U上成立，则在W（XY ⊆ W ⊆ U）上成立；反之，则不然。
- 若函数依赖XY在R（U）上成立，则对于任何Y` ⊂ Y均有X→Y`成立。对于多值依赖X→→Y，若在R（U）上成立，我们却不能断言对于任何 Y` ⊂ Y有X→→Y成立。

2）第四范式

- 定义6. 10

  关系模式R<U,F>∈1NF，如果对于R的**每个非平凡多值依赖**X→→Y（$Y \nsubseteq X$），**X都含有候选码**，则R∈4NF。

- 如果R ∈ 4NF， 则R ∈ BCNF
  - 对于每一个非平凡的多值依赖X→→Y， X都含有候选码，则有X→Y；
  - 允许的是函数依赖（是非平凡多值依赖），不允许的是非平凡且非函数依赖的多值依赖；
  - 如果一个关系模式是BCNF，而且至少有一个码是由一个属性组成，此关系模式就是4NF;

规范化小结

- **规范化目的**：使结构更合理，**消除**插入、修改、删除**异常**，使**数据冗余尽量小**，便于插入、删除和更新。
- **规范化方法**：将关系模式**投影分解**成两个或两个以上的关系模式。
- **规范化原则**：遵从概念单一化 “一事一地”原则，即一个关系模式描述一个实体或实体间的一种联系。规范的实质就是**概念的单一化**。
- **规范化要求**：分解后的关系模式集合应当与原关系模式“**等价**”，即经过自然联接可以恢复原关系而不丢失信息，并保持属性间合理的联系。

二、数据依赖的公理系统

- 问题的提出：

  在关系模式规范化处理过程中，不仅要知道一个由语义决定的函数依赖集合，还要知道由这个已知的函数依赖集合所**蕴含**（或推导出）的所有函数依赖集合。为此，需要一个有效而完备的公理系统，Armstrong公理系统即是这样的一个系统。

- 相关定义：

  一个函数依赖可以通过已知的函数依赖**推导**出来；如利用X→Y和Y→Z可以推导出X →Z，可以说，函数依赖X→Y和Y→Z**逻辑蕴含**（**Logical Implication**）了X →Z。

主要内容：

1、逻辑蕴涵

2、Armstrong公理系统

3、函数依赖集的闭包

4、函数依赖集的等价及最小依赖集

1、逻辑蕴涵

- 定义6.11

  对于满足一组函数依赖 F 的关系模式R <U，F>，其任何一个关系r，若函数依赖X→Y都成立（即r中任意两元组t、s，若t[X]=s[X]，则t[Y]=s[Y]），则称F逻辑蕴含X→Y，记为F⊨X→Y；

- 例如，设F={ A→B，B→C }，则函数依赖A→C被F逻辑蕴含，记作：F ⊨A→C。即函数依赖集 F 逻辑蕴含函数依赖A→C。

2、Armstrong公理系统

- Armstrong公理系统：一套推理规则，是模式分解算法的理论基础；
- 关系模式$R<U, F>$来说有以下的推理规则：
  - A1.自反律（Reflexivity）：若$Y \subseteq X \subseteq U$，则$X \rightarrow Y$为$F$所蕴含。
  - A2.增广律（Augmentation）：若$X \rightarrow Y$为F所蕴含，且$Z \subseteq U$，则$XZ \rightarrow YZ$为F所蕴含。
  - A3.传递律（Transitivity）：若$X \rightarrow Y$及$Y \rightarrow Z$为F所蕴含，则$X \rightarrow Z$为F所蕴含。

根据推理规则导出的规则

1.根据A1，A2，A3这三条推理规则可以得到下面三条推理规则：

- 合并规则：由$X \to Y$，$X \to Z$，有$X \to YZ$。（A2， A3）
- 伪传递规则：由$X \to Y$，$WY \to Z$，有$XW \to Z$。（A2， A3）
- 分解规则：由$X \to Y$及 $Z \subseteq Y$，有$X \to Z$。（A1， A3）

2.根据合并规则和分解规则，可得引理6.1；

引理6.l $X \to A_1 A_2 \ldots A_k$成立的充分必要条件是$X \to A_i$成立（i=l，2，...，k）。

3、函数依赖集的闭包

- **定义6.l2**

  在关系模式R<U，F>中为*F*所逻辑蕴含的**函数依赖的全体**叫作**F的闭包（closure）**，记为F^+，即$F^+=\{X\rightarrow Y|记为F|=X\rightarrow Y\}$。

  $F^+=\{F_i|从阿氏公理中导出的所有函数依赖的集合\}$

- 一般情况下，$F \subseteq F^+$，如果$F=F^+$，则称F为一个函数依赖的完备集；

例1： 已知关系模式R（XYZ），*F*={X→Y, Y→Z}，求F⁺；

根据阿氏公理可得出：

F^+ ={X→φ, Y→φ, Z→φ, XY→φ, XZ→φ, YZ→φ, XYZ→φ,

X→X, Y→Y, Z→Z, XY→X, XZ→X, YZ→Y, XYZ→X,

X→Y, Y →Z, XY→Y, XZ→Y, YZ→Z, XYZ→Y,

X→Z, Y→YZ, XY→Z, XZ→Z, YZ→YZ, XYZ→Z,

X→XY, XY→XY, XZ→XY, XYZ→XY,

X→XZ, XY→YZ, XZ→XZ, XYZ→YZ,

X→YZ, XY→XZ, XZ→XY, XYZ→XZ,

X→ZYZ, XY→XYZ, XZ→XYZ, XYZ→XYZ }

共计42个FD；

- Armstrong公理系统是有效的、完备的：
  - **有效性**：由F出发根据Armstrong公理推导出来的每一个函数依赖一定在F^+中；
  - **完备性**：F^+中的每一个函数依赖，必定可以由F出发根据Armstrong公理推导出来。

属性集的闭包

- 定义6.13

  设F为属性集U上的一组函数依赖，$X \subseteq U$，$X_F^+ = \{A \mid X \rightarrow A$能由F根据Armstrong公理导出$\}$，$X_F^+$称为**属性集X关于函数依赖集F的闭包。**

- 在实际使用中，判定$X \rightarrow Y$是否能从已知的F根据阿氏公理导出的问题，就转化为求出X_F^+，然后判定Y是否为X_F^+的子集的问题。

- 引理6.2

  设F为属性集U上的一组函数依赖，X，Y $\subseteq$ U，X→Y能由F根据阿氏公理导出的**充分必要条件**是Y $\subseteq X_F^+$ 。

- 用途

  将判定X→Y是否能由F根据Armstrong公理导出的问题，**转化为求出X_F^+ 、判定Y是否为X_F^+的子集的问题**。

算法6.1 求属性集X（$X\subseteq U$）关于U上的函数依赖集F 的闭包X_F^+ ；

输入：X，F　　输出：X_F^+

步骤：

（1）令$X^{(0)}=X$，$i=0$

（2）求B，这里B = { A |($\exists$ V)($\exists$ W)($V\rightarrow W\in F\wedge V\subseteq X^{(i)}\wedge A\in W$)}；

（3）$X^{(i+1)}=B\cup X^{(i)}$

（4）判断$X^{(i+1)}= X^{(i)}$吗?

（5）若相等或$X^{(i)}=U$，则$X^{(i)}$就是X_F^+，算法终止。

（6）若否，则 $i=i+l$，返回第（2）步。

例1： 已知关系模式$R<U，F>$，其中$U=\{A，B，C，D，E\}$；

$F=\{AB \rightarrow C，B \rightarrow D，C \rightarrow E，EC \rightarrow B，AC \rightarrow B\}$。求$(AB)_F^+$。

解 设$X^{(0)}=AB$；

(1) $X^{(1)}=AB \cup CD=ABCD$ ；

(2) $X^{(0)} \neq X^{(1)}$

$X^{(2)}=X^{(1)} \cup BE=ABCDE$。

(3) $X^{(2)}=U$，算法终止

$\rightarrow (AB)_F^+ =ABCDE$。

例2：已知关系模式$R<U，F>$，其中$U=\{A，B，C，D\}$；$F=\{A\rightarrow B，B\rightarrow C，D\rightarrow B\}$。求$A_F^+$、（AD）$_F^+$及（BD）$_F^+$；

解： 设$X^{(0)}=A$；

(1) $X^{(1)}=A\cup B=AB$；

(2) $X^{(0)}\neq X^{(1)}$

$X^{(2)}=X^{(1)}\cup C=ABC$；

(3) $X^{(3)}=X^{(2)}$，算法终止；

→A_F^+ =ABC，（AD）$_F^+$ = ABCD，（BD）$_F^+$ =BCD

例3：已知关系模式$R<U，F>$，其中$U=\{A，B，C，D，E，F\}$；

$F=\{AB\to C, BC\to AD, D\to E, CF\to B\}$。求$(AB)_F^+$。

解　设$X^{(0)}=AB$；

(1)　$X^{(1)}=AB\cup C=ABC$　；

(2)　$X^{(0)}\neq X^{(1)}$

$X^{(2)}=X^{(1)}\cup D=ABCD$　；

(3)　$X^{(3)}=ABCD\cup E=ABCDE$　；

(3)　$X^{(4)}=X^{(3)}$　，算法终止

→$(AB)_F^+=ABCDE$。

例4：已知关系模式R(A,B,C,D)，F={AB→C,C →D,D →A}

求：蕴含于给定函数的所有非平凡函数依赖；

(1)求关系R的所有单属性的闭包；

$A_F^+ = A$ ，$B_F^+ = B$ ，$C_F^+ = ACD$，$D_F^+ = AD$

(2)求关系R的所有双属性的闭包；

$(AB)_F^+ = ABCD$，$(AC)_F^+ = ACD$，$(AD)_F^+ = AD$，

$(BC)_F^+ = ABCD$，$(BD)_F^+ = ABCD$，$(CD)_F^+ = ACD$

(3)求关系R的所有三属性的闭包；

$(ABC)_F^+ = ABCD$，$(ABD)_F^+ = ABCD$

$(ACD)_F^+ = ACD$，$(BCD)_F^+ = ABCD$

(4)求关系R的所有四属性的闭包；

$(ABCD)_F^+ = ABCD$

根据以上计算，寻找关系R的超码和候选码，列出所有属性闭包：

$A^+=A$	$AB^+=ABCD$	$ABC^+=ABCD$
$B^+=B$	$AC^+=ACD$	$ABD^+=ABCD$
$C^+=ACD$	$AD^+=AD$	$ACD^+=ACD$
$D^+=AD$	$BC^+=ABCD$	$BCD^+=ABCD$
	$BD^+=ABCD$	$ABCD^+=ABCD$
	$CD^+=ACD$	

超码：AB,BC,BD,ABC,ABD,BCD,ABCD

候选码：AB,BC,BD

4、函数依赖集的等价及最小依赖集

- 定义6.14　设F和G是两个函数依赖集：

  ①如果$F^+ \subseteq G^+$，则称G是F的一个覆盖，或称G覆盖F；

  ②如果$F^+ \subseteq G^+$和$G^+ \subseteq F^+$同时成立，即$F^+ = G^+$，则称F和G等价。

- 引理6.3　$F^+ = G^+$的充分必要条件是$F \subseteq G^+$ 和$G \subseteq F^+$ ；

  要判定$F \subseteq G^+$，只须逐一对F中的函数依赖$X \rightarrow Y$，考察Y是否属于$X_G{}^+$就行了；引理6.3给出了判断两个函数依赖集等价的可行算法。

- 研究函数依赖集等价的目的：

　　是为了对指定函数依赖集找出它的最小函数依赖等价集，即找出包含函数依赖尽可能少、甚至最少的函数依赖等价集，从而使模式分解简化，分解出最简单的关系模式。

- 定义6.15

如果函数依赖集F满足下列条件，则称F为一个极小函数依赖集，亦称为**最小依赖集或最小覆盖**。

(1) F中任一函数依赖的**右部仅含有一个属性**。

(2) F中不存在这样的函数依赖$X \to A$，使得F**与**$F-\{X \to A\}$等价。

(3) F中不存在这样的函数依赖$X \to A$，X有真子集Z使得

$F-\{X \to A\} \cup \{Z \to A\}$**与**F等价。

- **计算最小覆盖的算法：**

给定函数依赖集F，求其最小覆盖的过程如下：

  - 逐一检查F 中各函数依赖$X \to Y$，若$Y=A_1 \ldots A_k$， $k>=2$，则用$\{X \to A_j \mid j=1,\ldots,k\}$来取代它；**（分解规则）**
  - 逐一检查F中各函数依赖$X \to A$，令$G=F-\{X \to A\}$，如果$A \in X_G^+$ ，则F与G等价，故从F中去掉$X \to A$。 **（FD不包含无关属性）**
  - 逐一取出F中各函数依赖$X \to A$，若$X=B_1B_2 \ldots B_m$，$m>=2$，则逐一考查Bi（$i=1,\ldots,m$），如果$A \in (X-B_i)_F^+$ ，则以$X-B_i$取代X; **（FD的左部都是唯一出现）**

例1：设F是关系模式R（ABC）的FD集，F=｛ A→BC，B→AC，C→A｝，试求F_{min}。

① 先把F中的FD分解为右边是单属性形式：

F=｛ A→B，A→C，B →A ，B→C，C→A ｝

②去掉F中冗余的函数依赖：

判断A→B，设：G_1=｛ A→C，B→A，B→C，C→A｝，得：A_{G1}^+=AC ∵ B∉A_{G1}^+ ∴ A→B不冗余

判断A→C，设：G_2=｛ A→B，B→A，B→C，C→A｝，得：A_{G2}^+=ABC ∵ C∈A_{G2}^+ ∴ A→C冗余

判断B→A，设：G_3=｛ A→B，B→C，C→A｝，得：B_{G3}^+=BCA ∵ A∈B_{G3}^+ ∴ B→A冗余

判断B→C，设：G_4=｛ A→B，C→A｝，得：B_{G4}^+=B ∵ C∉B_{G4}^+ ∴ B→C不冗余

判断C→A，设：G_5=｛ A→B，B→C ｝，得：C_{G5}^+=C ∵ A∉C_{G5}^+ ∴ C→A不冗余

F_m=｛ A→B，B→C，C→A｝

例2：求F={AB→C，A→B，B→A}的最小函数依赖集F_{min}。

解：(1)去掉F中冗余的函数依赖：

判断AB→C是否冗余。设：G_1={ A→B，B→A}，得：

$(AB)_{G1}^+=AB$ ∵ $C \notin (AB)_{G1}^+$ ∴ AB→C不冗余

判断A→B是否冗余。设：G_2={ AB→C，B→A}，得：$A_{G2}^+=A$

∵ $B \notin A_{G2}^+$ ∴ A→B不冗余

判断B→A是否冗余。设：G_3={ AB→C，A→B }，得：$B_{G3}^+=B$

∵ $A \notin B_{G3}^+$ ∴B→A不冗余

函数依赖集仍然为F={AB→C，A→B，B→A}；

(2) 去掉各函数依赖左部冗余的属性（本题只需考虑AB→C的情况）

方法1：在决定因素中去掉B，若$C \in A_F^+$，则以A→C代替AB→C。

求得：$A_F^+=ABC$ $\because$ $C \in A_F^+$ $\therefore$ 以A→C代替AB→C

故：$F_m=\{A\to C，A\to B，B\to A\}$

方法2：在决定因素中去掉A，若$C \in B_F^+$，则以B→C代替AB→C。

求得：$B_F^+=ABC$ $\because$ $C \in B_F^+$ $\therefore$ 以B→C代替AB→C

故：$F_m=\{B\to C，A\to B，B\to A\}$

例3：设F是关系模式R（ABC）的FD集，F=｛ A→BC，B→C，A→B，AB→C ｝，试求F_{min}。

解：

① 先把F中的FD分解为右边是单属性形式：

F=｛ A→B，A→C，B→C，AB→C ｝

② 去掉F中冗余的函数依赖：得F=｛ A→B，AB→C ｝

③ 去掉各函数依赖左部冗余的属性：F中AB→C可从A→B和B→C推出，得F=｛ A→B，B→C ｝或F=｛ A→B，A→C ｝。

- 应当指出：F的最小依赖集F_m不一定是唯一的，它与对各函数依赖Fd_i及X→A中X各属性的处置顺序有关。
- $F=\{A \rightarrow B, B \rightarrow A, B \rightarrow C, A \rightarrow C, C \rightarrow A\}$

  $F_{m1}=\{A \rightarrow B, A \rightarrow C, C \rightarrow A\}$

  若$F = \{A\rightarrow B, B\rightarrow A, A\rightarrow C, B\rightarrow C, C\rightarrow A\}$

  则$F_{m2}=\{A \rightarrow B, B\rightarrow C, C \rightarrow A\}$

  若$F = \{C\rightarrow A, B \rightarrow A , A \rightarrow B, A \rightarrow C, B \rightarrow C\}$

  则$F_{m3}=\{B \rightarrow A, A \rightarrow B, B \rightarrow C\}$

  ……

判断：假设有属性集U={A,B,C,D,E}，

函数依赖集F={A→B,B→C,AD→E}

函数依赖集G={A→B,A→C,B→C,AD→E}，

问F和G是否是最小函数依赖集？

答案：F是最小依赖集，G不是最小依赖集。

快速求候选码的方法

- 首先对于给定的R(U)和函数依赖集F,可以将它的属性划分为4类:
- **L类**,仅出现在F的函数依赖左部的属性。
- **R类**,仅出现在F的函数依赖右部的属性。
- **N类**,在F的函数依赖左部和右部均未出现的属性。
- **LR类**,在F的函数依赖左部和右部两部均出现的属性。

根据以下定理和推论来求解候选码。

- **定理1**:对于给定的关系模式R及其函数依赖集F,若X($X \in R$)是L类属性,则X必为R的任一候选码的成员。
- **推论1**:对于给定的关系模式R及其函数依赖集F,若X($X \in R$)是L类属性,且X^+包含了R的全部属性,则X必为R的唯一候选码。
- **定理2**:对于给定的关系模式R及其函数依赖集F,若X($X \in R$)是R类属性,则X不在任何候选码中。
- **定理3**:设有关系模式R及其函数依赖集F,如果X是R的N类属性,则X必包含在R的任一候选码中。
- **推论2**:对于给定的关系模式R及其函数依赖集F,如果X是R的N类和L类组成的属性集,且X^+包含了R的所有属性,则X是R的唯一候选码。

练习题

1、设关系模式R（ABCDE）上FD集为F，且F={A→BC，CD →E，B →D,E →A}，试求：

1） B_F^+的值；

2） $(ABC)_F^+$的值；

3） R的候选键；

4）计算F的最小函数依赖集；

2、关系模式R（ A ，B，C，M，T），根据语义有如下函数依赖集：F={B→C，MT→B，MC→T，MA→T，AB→C}，关系模式R的码是（ C ）

A. MT　　B. MC　　C. MA　　D. AB