LAS DEPENDENCIAS MULTIVALUADAS COMO REGLAS DE INFERENCIA

Marta Elena Millán González*
María Covadonga Fernandez Baizán**

* Departamento de Ciencias de la Computación , Universidad del Valle, Cali-Colombia
e-mail: millan@ borabora.univalle.edu.co

** Departamento de Lenguajes y Sistemas Informáticos e Ingenieria de Software, Universidad Politécnica de Madrid , España
e-mail : cfbaizan @fi.upm.es

Resumen

Bajo el enfoque conjuntista del modelo relacional [Spyratos & Lecluse, 1986] , las dependencias multivaluadas (dmv) se consideran reglas de inferencia y se incorporan al proceso de inferencia.

El conjunto de axiomas de deducción propuesto por Laurent y Spyratos [1987b] se extiende, de manera natural, para considerar las dmv como reglas de inferencia cuando estas se aplican únicamente a tuplas positivas.

1.INTRODUCCION

La lógica ha sido utilizada ampliamente en las bases de datos como sistema de inferencia o como lenguaje de representación, debido principalmente a la facilidad que ofrece para representar de manera uniforme datos, programas, interrogaciones y restricciones de integridad [Dahl, 1982; Fagin, 1977b; Gaillard-Ferrand, 1990; Rybinski, 1987; Sagiv, 1980; Sagiv, Delobel, Parker, Fagin, 1981; Nicolas, 1978].

Por otra parte , un enfoque basado en construcciones de la teoría de conjuntos, el Modelo Particional [Spyratos, 1987] es un modelo deductivo de bases de datos construído sobre el concepto de Retículo de Particionales, que incorpora, de manera natural, un mecanismo deductivo basado en propiedades de conjuntos. Bajo este enfoque, los esquemas de relación, las restricciones y la base de datos se consideran cadenas de símbolos sin interpretar y en cuya interpretación se utilizan subconjuntos de una población de objetos determinada [Spyratos y Lecluse, 1986]. En el proceso de deducción de nuevos hechos se utiliza información adicional, que normalmente es externa a la base de datos. La información semántica, al igual que las dependencias funcionales, se expresa como restricciones sobre las interpretaciones y la componente deductiva se introduce, básicamente, por medio de la operación de inclusión de conjuntos.

En este articulo se presenta, en la sección 2 el enfoque conjuntista del modelo relacional propuesto por Lecluse [1987] . En la sección 3 se extiende el modelo para incorporar las dmvs como reglas de inferencia y finalmente en la sección 4 se presenta un conjunto de axiomas de deducción que incorpora dependencias multivaluadas como reglas de inferencia a partir del propuesto Laurent y Spyratos [1987b].

2. EL ENFOQUE CONJUNTISTA DEL MODELO RELACIONAL

El enfoque conjuntista del modelo relacional incorpora tres componentes básicas: una componente **sintáctica**: el modelo relacional, un dominio **semántico**: el conjunto de partes de una población de objetos definida y una **interpretación** que permite ligar la componente sintáctica y el dominio semántico [Spyratos y Lecluse, 1986].

Bajo este enfoque (semántica conjuntista) las tuplas del modelo relacional se ven como cadenas de símbolos cuyas interpretaciones corresponden a subconjuntos de una población de objetos. Una tupla ab es una cadena compuesta por los símbolos a y b. Considerando un mundo posible ω, una interpretación de a, I(a), corresponde al conjunto de los individuos de ω que satisfacen la propiedad a y por tanto la tupla ab corresponde al conjunto de los individuos de ω que satisfacen, al mismo tiempo, las propiedades a y b.

Intuitivamente, surge la noción de verdad bajo la cual una tupla t es verdadera siempre que su interpretación corresponda a un conjunto diferente de vacío. Por otro lado, esta semántica conjuntista y particularmente la operación de inclusión de conjuntos, permite introducir también, de manera natural, el concepto de **inferencia** [Gaillard-Ferrand, 1990; Laurent y Spyratos, 1987a, 1987b; Lecluse, 1987].

Sintaxis

La sintaxis corresponde básicamente a la del modelo relacional.

El **universo** es un conjunto finito, no vacío de **atributos** $U=\{ A_1, A_2, \ldots, A_n \}$ cada uno de los cuales tiene asociado un **dominio** que corresponde a un conjunto infinito contable de valores o símbolos. Bajo el enfoque conjuntista estos dominios deben ser disjuntos, es decir, $A_i \cap A_j = \phi$ y $U \cap A_i = \phi$ para cada i.

De igual manera un **esquema de relación** R es un subconjunto no vacío de U. El conjunto Sch(U) está formado por todos los esquemas de relación.

Una **tupla** t definida sobre un esquema de relación R es una función definida en R tal que $t(Ai) \in dom(Ai)$ para todo Ai de R. El dominio de R está conformado por todas las tuplas sobre R.

Una **relación** definida sobre R es un conjunto de tuplas de R, es decir, es un subconjunto de dom(R).

Finalmente, una **base de datos** sobre U es un par $D=(\delta, \Sigma)$ tal que:

- δ es una función que asigna a cada esquema de relación una relación finita sobre R
- Σ es el conjunto de dependencias funcionales constituídas por pares ordenados de subconjuntos de U (X,Y) denotados por $X \rightarrow Y$.

Dada una base de datos $D=(\delta, \Sigma)$, el **esquema de la base de datos** es el conjunto de todos los esquemas de relación que tiene asignado relaciones no vacías bajo δ, es decir el conjunto

$$\{ U \supseteq R \mid R \neq \varnothing , \delta(R) \neq \varnothing \}.$$

Semántica

La semántica, uno de los aportes más importantes del modelo particional, considera el mundo real como un conjunto contable e infinito de objetos que se identifican con el conjunto ω de los enteros no-negativos [Spyratos, 1987].

El conjunto de todos los subconjuntos de ω, $P\omega = \{ \tau \mid \omega \supseteq \tau \}$ se considera el dominio semántico en el que los esquemas de relación, las relaciones y las dependencias toman significado [Spyratos y Lecluse, 1986].

Interpretaciones

Se considera un universo fijo de atributos $U=\{ A_1, A_2, \ldots, A_n \}$ y sus dominios asociados $dom(A_i)$. Se denota por SIMBOLOS al conjunto unión de los dominios de los atributos y por TUPLAS a la unión de todos los dominios.

$$SIMBOLOS = \cup \{ dom(A_i) \mid A_i \in U \}$$

$$TUPLAS = \cup \{ dom(R) \mid U \supseteq R, R \neq \emptyset \}$$

Definición 1

Una interpretación de U es una función I de SIMBOLOS en el conjunto de partes $P\omega$ tal que,

$$\forall A \in U, \forall a, a' \in dom(A), \quad (I(a) \cap I(a') = \phi \quad v \quad a = a')$$

De acuerdo con Spyratos [1987], símbolos diferentes de un mismo dominio se asocian con conjuntos disjuntos de enteros. Este mismo concepto extendido a tuplas significa que:

$\forall U \supseteq R, \forall a_1, a_2, \ldots, a_k \in dom(R), \ I(a_1, a_2, \ldots, a_k) = I(a_1) \cap I(a_2) \cap \ldots\ldots \cap I(a_k).$

Definición 2

Sea I una interpretación de U. Una tupla t de TUPLAS es **verdadera** si su interpretación corresponde a un conjunto diferente de vacío, es decir, $I(t) \neq \phi$; en caso contrario es **falsa**.

Dependencias Funcionales

Una **dependencia funcional** $X \rightarrow Y$ se satisface en una interpretación I si para toda propiedad X, todos los individuos que verifican X, verifican también la propiedad Y [Lecluse, 1987], es decir, una dependencia $X \rightarrow Y$ se satisface en una interpretación I si y solamente si

$$\forall x \in dom(X), \exists y \in dom(Y), I(y) \supseteq I(x)$$

Definición 3

Sea D= (δ , Σ) una base de datos definida sobre U. Una interpretación I de U es un modelo de D si:

1. $\forall\ U \supseteq R \neq \phi, \forall\ t \in \delta(R),\ \ I(t) \neq \phi$

2. $\forall\ X \rightarrow Y \in \Sigma\ ,\ \forall\ x \in \mathrm{dom}(X)\ \ \forall\ y \in \mathrm{dom}(Y)$

$$[\, I(x) \cap I(y) \neq \phi \Rightarrow I(y) \supseteq I(x)\,]$$

Consistencia de una Base de Datos y Deducción

La definición de consistencia de una base de datos requiere la definición de modelo. De esta manera, una base de datos es semánticamente consistente si y solo si tiene al menos un modelo, en caso contrario es semánticamente inconsistente [Spyratos y Lecluse, 1986].

Por definición, todas las tuplas que aparecen en una base de datos consistente D son verdaderas, puesto que existe una interpretación I que es un modelo de D.

Sinembargo, pueden existir tuplas que sean también verdaderas bajo esa interpretación y que no aparecen explícitamente en la base de datos, lo que implica la necesidad de definir un mecanismo de deducción de todas las tuplas verdaderas que se derivan de la base de datos.

Definición 4

Sea D = (δ,Σ) una base de datos sobre un universo U. Sea t cualquier tupla de TUPLAS. Se dice que D implica t, D |= t , si t es verdadera en cada modelo de D. Si T es un conjunto de tuplas entonces D implica T, D |= T si D |= t para toda tupla t en T.

Dado un modelo D, se denota por T(m) al conjunto de todas las tuplas en TUPLAS que son verdaderas en m. Se define [Spyratos y Lecluse, 1986] el conjunto de todas las tuplas que son verdaderas en cada modelo de D como sigue:

$$T(D) = \cap \{ T(m) \mid m \text{ es un modelo de } D \}$$

Por tanto,

$$\forall\ t \in TUPLAS\ D \models t \Leftrightarrow t \in T(D)$$

A este mismo enfoque se incorpora la semántica necesaria para que en la base de datos pueda manejar información negativa [Laurent y Spyratos 1987b] . Con este propósito el concepto de tupla verdadera se "refina" introduciendo dos tipos diferentes de tuplas: positivas y negativas. Las primeras corresponden a las tuplas usuales del modelo relacional y las últimas a la ausencia de individuos, en el mundo real, satisfaciendo una propiedad.

Con el propósito de incorporar información negativa se introduce la siguiente definición para un esquema de relación R:

$$\sim dom(R) = \{ \sim t \mid t \in dom(R) \}$$

En cada dominio de relación R a cada tupla t sobre R se le asocia el símbolo ~t. Una tupla t es una **tupla positiva** y ~t es una **tupla negativa**.

Definición 5

Sea I una interpretación. Una tupla positiva t es **verdadera** si $I(t) \neq \phi$ y una tupla negativa t es **verdadera** si $I(t) = \phi$.

A partir de esta base y de los conceptos de verdad y consistencia Laurent y Spyratos [1987b] proponen un conjunto de axiomas para el procesamiento de consultas incorporando información negativa, que permite decidir cuándo una tupla t se deriva de una base de datos $D=(\delta,\Sigma)$. Este conjunto de axiomas de inferencia considera únicamente dependencias funcionales Σ.

3. INCORPORANDO DMV COMO REGLAS DE INFERENCIA

Cuando se introducen reglas de inferencia como parte de la información externa es importante reconsiderar los conceptos de tupla derivada e implicada ya que si una regla de inferencia es aplicable entonces deben existir tuplas verdaderas que satisfacen el patrón de entrada de dicha regla. El que existan tales tuplas significa que la regla se aplica a tuplas implicadas que están en la base de datos ó que han sido derivadas por aplicación de axiomas de deducción y por tanto las tuplas del patrón de salida serán también tuplas implicadas.

Extensión del concepto de Base de Datos, bajo el enfoque conjuntista a Reglas de inferencia representando dependencias multivaluadas

Definición 7

Una dependencia multivaluada considerada como una regla de inferencia se representa explícitamente por medio de dos triplas complementarias (X,Y,U) y (X,Z,U) donde $Z = U - (X \cup Y)$ que se denotan por $X \twoheadrightarrow Y$, $X \twoheadrightarrow Z$.

Semántica

$\forall\ (X \twoheadrightarrow Y) \in \Phi, \forall x \in \mathrm{dom}(X), \forall y \in \mathrm{dom}(Y), \forall z \in \mathrm{dom}(Z)$ con $Z = U - (X \cup Y)$

$I(x) \cap I(y) \cap I(z) \neq \varnothing \ \wedge \ I(x) \cap I(y') \cap I(z') \neq \varnothing \ \Rightarrow \ I(x) \cap I(y') \cap I(z) \neq \varnothing$

Definición 8

Una base de datos sobre U es una tripla $D = (\delta, \Sigma, \Phi)$ tal que:

- δ es una función que asigna a cada esquema de relación una relación finita sobre R
- Σ es el conjunto de dependencias funcionales constituídas por pares ordenados de subconjuntos de U (X,Y) denotados por $X \rightarrow Y$ y
- Φ es un conjunto de triplas (X,Y,U) representando reglas de inferencia donde X, Y son subconjuntos de U.

Extensión del concepto de Interpretación

Definición 9

Dada una interpretación I de la base de datos , I es un modelo de D ssi:

- Cada tupla de δ es verdadera en I
- Cada $X \rightarrow Y$ de Σ es verdadera en I en el sentido siguiente:

  $\{(x,y) \quad \forall x \in dom(X) \;\; \forall y \in dom(Y) \quad I(x) \cap I(y) \neq \emptyset \}$ es una función.
- Cada tupla generada por la aplicación de una regla de inferencia en Φ es verdadera en I.

4. AXIOMAS DE DEDUCCION

Incorporación de Dependencias Multivaluadas, representadas como Reglas de Inferencia, al Sistema Axiomático para Procesamiento de Interrogaciones

Interrogar una base de datos D significa determinar si dada una tupla t la base de datos implica esa tupla t. Es decir, se trata de resolver un problema de inferencia que permita definir cuándo $D \models t$.

Dada una base de datos consistente $D = (\delta , \Sigma, \Phi)$ sobre un universo U el siguiente conjunto de axiomas permite derivar todas las tuplas positivas y negativas que son verdaderas en D. Se asume que las reglas de inferencia, que representan dependencias multivaludas, se aplican únicamente a tuplas positivas verdaderas y que el conjunto de dependencias funcionales y de dependencias multivaluadas representadas como reglas de inferencia es cerrado bajo los axiomas de Beeri, Fagin y Howard [1977].

Las tuplas se denotan con letras minúsculas y los esquemas de relación con letras mayúsculas. Dadas dos tuplas x e y, se denota por xy el "join " de x e y.

Conjunto de Axiomas

A1. xy |- x

A2. ~x |- ~xy

A3. xy, yz, Y → X o Y → Z |- xyz

A4. xy, y=y', X → Y |- ~xy'

A5. ~xyz, xy, Y → X o Y → Z |- ~yz

A6. xyz, xy'z' , X→ →Y |- xy'z

Definición 11

Dada una base de datos D =(δ , Σ , Φ) y una tupla t, positiva o negativa, se dice que t se deriva de D, D |- t, ssi a partir de las tuplas registradas en D se puede generar t aplicando los axiomas A1-A6.

El conjunto de tuplas derivadas corresponde entonces al cierre de la base de datos con respecto al conjunto de axiomas A1-A6.

Teorema 1

Si (Σ ∪ Φ), el conjunto de dependencias funcionales y reglas de inferencia es consistente entonces $(\Sigma \cup \Phi)^+$ es consistente.

Prueba

Cada $f \in (\Sigma \cup \Phi)^+$ está implicado por (Σ ∪ Φ), es decir, f es válida en cada interpretación de $(\Sigma \cup \Phi)^+$. Particularmente, f es válida en la interpretación m y por lo tanto existe un modelo para $(\Sigma \cup \Phi)^+$. $(\Sigma \cup \Phi)^+$ es entonces consistente.

Teorema 2

Si una tupla t se deriva por aplicación de una regla de inferencia R en Φ entonces t es verdadera en todo modelo de D.

Prueba

Si una regla de inferencia es aplicable, las tuplas que satisfacen el patrón de entrada de la reglasde inferencia pueden ser tuplas verdaderas que aparecen explícitamente en la base de datos, es decir tuplas implicadas ó tuplas verdaderas que se han derivado por la aplicación de cualquiera de los axiomas de inferencia .

Sean t_1 y t_2 dos tuplas de D que satisfacen una regla de inferencia R . Como t_1 y t_2 están en D son verdaderas en todo modelo de D, es decir, son tuplas implicadas $D \models t_1$ y $D \models t_2$ y por tanto t_1 y $t_2 \in T(D)$. Si las dos tuplas satisfacen R y considerando que la regla de inferencia R es válida, entonces las tuplas derivadas por la regla t_3 y t_4 son verdaderas en todos los modelos en los caules son verdaderas t_1 y t_2. Por tanto, $D \models t_3$ y $D \models t_4$.

Por otra parte, si t_1 y t_2 son tuplas derivadas por la aplicación de cualquiera de los axiomas, ellas son también verdaderas en todo modelo de D y por lo tanto t_3 y t_4 serán también verdaderas en todo modelo que satisface a t_1 y t_2, t_3 y $t_4 \in T(D)$.

Teorema 3

$$D \models t \text{ si y solo si } D \vdash t$$

para cada base de datos consistente $D=(\delta, \Sigma, \Phi)$ y para cada tupla t positiva o negativa.

Prueba

Consistencia

Se requiere probar que si $D \vdash t$ entonces $D \models t$.

Teniendo en cuenta que al conjunto de axiomas propuesto por Laurent y Spyratos [1987b] se le ha añadido el axioma 6, únicamente se prueba que este axioma produce tuplas verdaderas (ver Teorema 2).

Completitud

Como las reglas de inferencia que representan dmv se aplican solamente a tuplas positivas, la prueba de completitud [Millán, 1995], al igual que en el caso en el que únicamente se consideran dependencias funcionales [Laurent y Spyratos ,1987b] consiste en probar la siguiente implicación:

$$\forall\ t \in \mathrm{TUPLAS}\ ,\ t \in P(D) \quad \Rightarrow \quad t \in P$$

donde P(D) y P son respectivamente conjuntos de tuplas positivas implicadas y derivadas por D.

5.REFERENCIAS BIBLIOGRAFICAS

Beeri, C. [1980]. On the membership problem for multivalued dependencies. ACM Transaction on Database Systems,Vol. 5, Nº 3.

Beeri, C., Fagin, R. y Howard, J. [1977]. A complete axiomatization for functional and multivalued dependencies in database relations. Proc. ACM-SIGMOD Int. Conf. on Management of Data. Canadá.

Beeri, C. y Vardi, M.Y. [1984b]. Formal Systems for tuple and equality generating dependencies. SIAM Journal of Computing, Vol.13, Nº 1.

Dahl, Verónica [1982]. On Database Systems Development through Logic. ACM TODS,Vol. 7, Nº1, March.

Fagin, Ronald [1977a]. Multivalued Dependencies and a new normal form for relational databases. ACM TODS, Vol. 2, Nº 3.

Fagin, Ronald [1977b]. Functional dependencies in a relational database and propositional logic. IBM J. Res. and Develop. ,Vol. 21, Nº6.

Fernández, M.C., García, A., Millán, S. Pérez, C.,Portaencasa, R.,Santos, E. [1994]. Mathematical Foundations of Deductive Databases. Proc. of AMSE SYS'94 International Conference: Systems Analysis, Control & Design, Lyon, France, Julio.

Gaillard-Ferrand, I. [1990]. Semantique partitionnelle des bases de donnes et programmation en logique. Universidad de Orleans.

Gallaire, H. [1983]. Logic Databases vs. Deductive Databases. Logic Programming Workshop, Lisboa ,Portugal.

Gallaire, H., Minker, J. (eds.) [1978]. Logic and Databases. Plenum Press, New York.

Laurent, D. [1987]. La Logique des Partitions: Application a L'information disjonctive dans les Bases de Donnees. Thèse de troisième cycle, Université D'Orleans.

Laurent, D. y Spyratos, N. [1987a]. Partition semantics for query-answering in relational databases with incomplete information. INRIA, Paris.

Laurent, D. y Spyratos, N. [1987b]. Introducing negative information in relational databases. INRIA, Paris.

Lecluse, C. [1987]. Une sémantique Ensembliste pour les Bases de Données. Application au Modèle Relationnel. Thèse de troisième cycle, Université de Paris-Sud.

Maier, David [1983]. The Theory of Relational Databases. Computer Science Press, Rockville, Maryland.

Mendelzon, A. [1979]. On axiomatizing multivalued dependencies in relational databases. Journal of the ACM, Vol.26, N° 1.

Millán, Marta [1995] Las dependencias multivaluadas en el proceso deductivo bajo una semántica conjuntista. Tesis Doctoral, Universidad Politécnica de Madrid, España.

Nicolas, J.M. [1978]. First order logic formalization for functional, multivalued and mutual dependencies. Proceedings ACM-SIGMOD, Austin, Texas.

Pervin, William [1987]. Inference rules for multivalued dependencies. ACM SIGCSE Bulletin, Vol. 19, N° 3.

Rybinski, Henry [1987]. On first-order-logic databases. ACM transactions on Database Systems. Vol. 12, N°3 , Sept.

Sagiv, Y. [1980]. An algorithm for inferring multivalued dependencies with an application to propositional logic. Journal of the ACM, Vol. 27, N° 2.

Sagiv, Y., Delobel, C., Parker, S. y Fagin, R. [1981]. An Equivalence between relational databases dependencies and a Fragment of Propositional Logic. Journal of ACM, Vol. 28, N° 3.

Spyratos, Nicolas [1987]. The Partition Model: A Deductive Database Model. ACM Transactions on Database Systems,Vol. 12, Nº 1, March.

Spyratos, N. y Lecluse, C. [1986]. The Semantics of queries and updates in Relational Databases. INRIA Research Report, Nº 561, France.

Spyratos, N. y Lecluse, C. [1987]. Incorporating functional dependencies in deductive query answering. Proc. International Conference on Data Engeneering, Los Angeles.

Ullman, Jeffrey [1988]. Principles of database and Knowledge-base systems, Vol.1, Computer Science Press, Rockville, Maryland.