

Reglas de Asociación

Eduardo Morales, Hugo Jair Escalante

INAOE

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

- 1 **Introducción**
- 2 **Apriori**
- 3 **Extensiones**
- 4 **Atributos Continuos**
- 5 **Otros Aspectos**
- 6 **Clasificación y Asociación**

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Objetivo: encontrar asociaciones o correlaciones entre los elementos u objetos de bases de datos transaccionales, relacionales o *data-warehouses*
- Las reglas de asociación tienen diversas aplicaciones como:
 - Soporte para la toma de decisiones
 - Diagnóstico y predicción de alarmas en telecomunicaciones
 - Análisis de información de ventas
 - Distribución de mercancías en tiendas
 - Segmentación de clientes con base en patrones de compra

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Son parecidas a las reglas de clasificación
- Se encuentran también usando un procedimiento de *covering*, sin embargo, en el lado derecho de las reglas, puede aparecer cualquier par o pares atributo-valor
- Para encontrar este tipo de reglas se debe de considerar cada posible combinación de pares atributo-valor del lado derecho.
- Para posteriormente poderlas usando:
 - Cobertura: número de instancias predichas correctamente
 - Precisión: proporción de número de instancias a las cuales aplica la regla

Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Transacción	Elementos Comprados
1	A,B,C
2	A,C
3	A,D
4	B,E,F

Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Encontrar las reglas de asociación $X \Rightarrow Z$ de la tabla anterior con:
 - Cobertura mínima de 50%
 - Precisión mínima de 50%
- Las reglas que cumplen con estas restricciones son:
 - $A \Rightarrow C$ (50%, 66.6%)
 - $C \Rightarrow A$ (50%, 100%)

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Una regla de asociación es una expresión de la forma $X \Rightarrow Z$ donde X y Z son conjuntos de elementos.
- El significado intuitivo:

Las transacciones de la base de datos que contienen X tienden a contener Z

Definiciones

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- $I = \{i_1, i_2, i_3, \dots, i_m\} \Rightarrow$ un conjunto de literales, atributos
- $D \Rightarrow$ un conjunto de transacciones $T, T \subseteq I$
- $TID \Rightarrow$ un identificador asociado a cada transacción
- $X \Rightarrow$ un conjunto de elementos $X \in I$
- Una *regla de asociación* es una implicación:
 - $X \Rightarrow Z, X \subseteq I, Z \subseteq I$ y $X \cap Z = \emptyset$

Definiciones

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- *Soporte* (o *cobertura*), s , es la probabilidad de que una transacción contenga $\{X, Z\}$
- *Confianza* (o *eficiencia*), c , es la probabilidad condicional de que una transacción que contenga $\{X\}$ también contenga $\{Z\}$.

Reglas de Asociación

- Evaluamos las reglas de acuerdo al soporte y la confianza de las mismas.
- En reglas de asociación, la cobertura se llama soporte (*support*) y la precisión se llama confianza (*confidence*).
- Se pueden leer como:

$$\text{soporte}(X \Rightarrow Z) = P(X \cup Z)$$

Número de transacciones que tienen X y Z , entre el número total de transacciones

$$\text{confianza}(X \Rightarrow Z) = P(Z|X) = \frac{\text{soporte}(X \cup Z)}{\text{soporte}(X)}$$

Número de transacciones que tienen X y Z , entre el número de transacciones que contienen X

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Buscamos reglas con un mínimo soporte (soporte $\geq \text{sop_min}$) y confianza (confianza $\geq \text{conf_min}$)
- Inicialmente buscamos (independientemente de que lado aparezcan), pares atributo-valor que cubran una gran cantidad de instancias.
- A los conjuntos de pares atributo-valor, se les llama *item-sets* y a cada par atributo-valor *item*.

Ejemplo: Análisis de Canasta de Mercado

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación



Ejemplo: Análisis de Canasta de Mercado

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Un ejemplo típico de reglas de asociación es el análisis de la canasta de mercado
- Esto es, encontrar asociaciones entre los productos de los clientes, que pueden impactar a las estrategias mercadotécnicas
- Ya que tenemos todos los conjuntos de *itemsets*, los transformamos en reglas con la confianza mínima requerida
- Algunos *itemsets* producen más de una regla y otros no producen ninguna

Reglas de Asociación

- Por ejemplo, si seguimos con los datos de la tabla de “Jugar Golf”, el *itemset*:

humedad=normal, viento=no, clase=P

- Puede producir las siguientes posibles reglas:

If humedad=normal and viento=no Then clase=P 4/4

If humedad=normal and clase=P Then viento=no 4/6

If viento=no and clase=P Then humedad=normal 4/6

If humedad=normal Then viento=no and clase=P 4/7

If viento=no Then clase=P and humedad=normal 4/8

If clase=P Then viento=no and humedad=normal 4/9

If true Then humedad=normal and viento=no and clase=P
4/12

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Si pensamos en 100% de éxito, entonces sólo la primera regla cumple.
- De hecho existen 58 reglas considerando la tabla completa que cubren al menos dos ejemplos con un 100% de exactitud (*accuracy*)

Apriori (Agrawal et al. '94)

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

El proceso es más o menos el siguiente y sigue dos pasos:

- 1 Genera los *itemsets*
 - Genera todos los *itemsets* con un elemento
 - Usa estos para generar los de dos elementos, y así sucesivamente
 - Toma todos los que cumplen con el mínimo soporte (esto permite eliminar posibles combinaciones)
- 2 Genera las reglas revisando que cumplan con el criterio mínimo de confianza.

Algoritmo (1)

Apriori()

$L_1 = \text{find-frequent-1-itemsets}(D)$

for ($k = 2; L_{k-1} \neq \text{NULL}; k++$)

 % generate-&-prune candidate k-itemsets

$C_k = \text{AprioriGen}(L_{k-1})$

forall transactions $t \in D$

$C_t = \text{subset}(C_k, t)$

forall candidates $c \in C_t$

$c.\text{count}++$

$L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$

Return $\cup_k L_k$

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Algoritmo (2): AprioriGen

```

AprioriGen( $L$ ) % Assume transactions in lexicographic order
insert into  $C_k$  all  $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-1}$ 
    from  $p, q \in L$ 
where  $p.item_1 = q.item_1, p.item_2 = q.item_2, \dots,$ 
     $p.item_{k-1} < q.item_{k-1}$ 
% Prune itemsets s.t. some  $(k-1)$ -subset of  $c$  is  $\notin L$ 
% A  $(k-1)$  itemset that is not frequent cannot be a subset of
% a frequent  $k$ -itemset, then it is removed
forall itemsets  $c \in C_k$ 
    forall  $(k-1)$ -subsets  $s$  of  $c$  do
        if ( $s \notin L_{k-1}$ ) then
            delete  $c$  from  $C_k$ 

```

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

Otros

Aspectos

Clasificación y

Asociación

Algoritmo (3): AssocRules y GenRules

AssocRules()

forall large itemsets $l_k, k \geq 2$

GenRules(l_k, l_k)

GenRules(l_k, a_m) % Generate all valid rules $a \rightarrow (l_k - a)$,
for all $a \subset a_m$

$A = \{(m-1) - \text{itemsets } a_{m-1} \mid a_{m-1} \subset a_m\}$

forall $a_{m-1} \in A$

$conf = \mathbf{support}(l_k) / \mathbf{support}(a_{m-1})$

if($conf \geq min_conf$) then

output rule $a_{m-1} \rightarrow (l_k - a_{m-1})$ with
confidence $conf$, $support = \mathbf{support}(l_k)$

if($m - 1 > 1$) then

GenRules(l_k, a_{m-1}) % Generate rules with
subsets of a_{m-1} as antecedents

Propiedades

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Si una conjunción de consecuentes de una regla cumple con los niveles mínimos de soporte y confianza, sus subconjuntos (consecuentes) también los cumplen.
- Por el contrario, si algún *item* no los cumple, no tiene caso considerar sus superconjuntos.
- Esto da una forma de ir construyendo reglas, con un solo consecuente, y a partir de ellas construir de dos consecuentes y así sucesivamente.

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Este método hace una pasada por la base de datos para cada conjunto de *items* de diferente tamaño.
- El esfuerzo computacional depende principalmente de la cobertura mínima requerida, y se lleva prácticamente todo en el primer paso.
- El proceso de iteración del primer paso se llama *level-wise* y va considerando los superconjuntos nivel por nivel.

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Lo que se tiene es una propiedad anti-monótona: Si un conjunto no pasa una prueba, ninguno de sus superconjuntos la pasan.
- Si un conjunto de *items* no pasa la prueba de soporte, ninguno de sus superconjuntos la pasan. Esto se aprovecha en la construcción de candidatos para no considerar todos.

Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

Otros

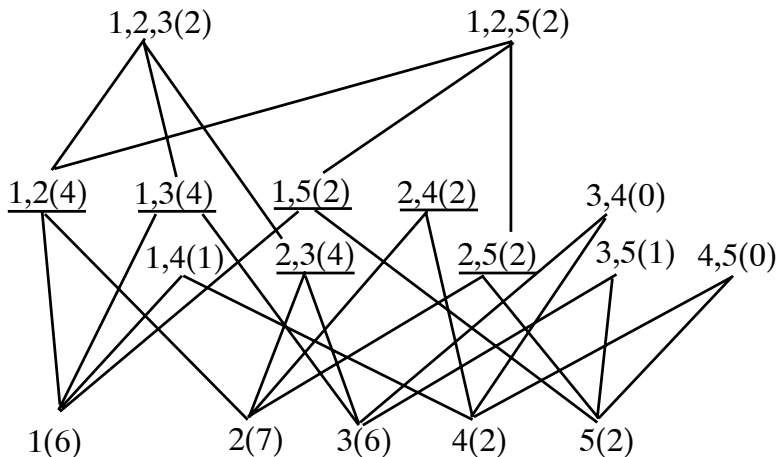
Aspectos

Clasificación y

Asociación

id1	p1,p2,p5
id2	p2,p4
id3	p2,p3
id4	p1,p2,p4
id5	p1,p3
id6	p2,p3
id7	p1,p3
id8	p1,p2,p3,p5
id9	p1,p2,p3

Ejemplo



Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Una vez que tenemos los conjuntos de *items*, generar las reglas es relativamente sencillo.
 - Para cada conjunto I de *items*, genera todos sus subconjuntos.
 - Para cada subconjunto $s \subset I$, genera una regla: $s \Rightarrow (I - s)$ si:

$$\frac{\text{soporte}(I)}{\text{soporte}(s)} \geq \text{nivel_confianza}$$

- Todas las reglas satisfacen los niveles mínimos de soporte.

Extensiones

Se han hecho algunas mejoras al algoritmo básico de reglas de asociación (Apriori) para hacerlo más eficiente:

- Usar tablas hash para reducir el tamaño de los candidatos de los *itemsets*
- Eliminar transacciones (elementos en la base de datos) que no contribuyan en superconjuntos a considerar
- Dividir las transacciones en particiones disjuntas, evaluar *itemsets* locales y luego, con base en sus resultados, estimar los globales.
- Hacer aproximaciones con muestreos en la lista de productos, para no tener que leer todos los datos

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Extensiones: FP-Growth

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Posiblemente la extensión más exitosa es FP-Growth
- Transforma la base de datos en una estructura compacta FP-trees (*Frequent Pattern tree*)
- Tiene un algoritmo eficiente para encontrar patrones frecuentes
- Usa un método que descompone las tareas en tareas más pequeñas
- No genera candidatos y evita leer la base de datos repetidamente
- Es un orden de magnitud más rápido que Apriori

Algoritmo FP-Growth

- Encuentra los *itemsets* de tamaño 1 (igual que Apriori), cuenta su soporte y ordenalos
- Toma la primera transacción, obtén su *itemset* y construye una rama del árbol con los *items* en orden descendente
- Toma la siguiente transacción y repite el proceso, pero si existe un *item* en alguna rama del árbol, se comparte ese nodo y se aumenta su soporte
- Examina los nodos más profundos y sus caminos y cuenta su soporte y de ahí construye los patrones frecuentes

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Transacción	Items	Contamos el soporte de cada <i>item</i> y los ordenamos:	
		Item	Soporte
T1	I1,I2,I3		
T2	I2,I3,I4	I2	5
T3	I4,I5	I1	4
T4	I1,I2,I4	I3	4
T5	I1,I2,I3,I5	I4	4
T6	I1,I2,I3,I4	I5	2

Si el soporte mínimo es 2, eliminamos a I5

Ejemplo

- Tomamos la primera transacción T1: I2,I1,I3 (en orden de frecuencia), con 3 *items*: {I2:1}, {I1:1}, {I3:1}, I2 se liga a la raíz, I1 a I2 e I3 a I1
- Con la segunda transacción: T2: I2,I3,I4, I2 ya está ligado a la raíz, I3 se liga a I2, e I4 a I3, como se comparte I2, se incrementa el contador de I2: {I2:2}, {I3:1}, {I4:1}
- Con la tercera transacción: T3: I4, I5, I4 se liga a la raíz e I5 se liga a I4
- Con la cuarta: T4: I2, I1, I4, I2-I1 ya existen, se liga I4 a I1 y se aumenta el contador de I2 e I1: {I2:3}, {I1:2}, {I4:1}
- Con T5: I2, I1, I3, I5, I2-I1-I3 ya existe, se liga I5 a I3 y se aumentan contadores: {I2:4}, {I1:3}, {I3:2}, {I5:1}
- Con T6: I2, I1, I3, I4, nos queda: {I2:5}, {I1:4}, {I3:3}, {I4:1}

Contenido

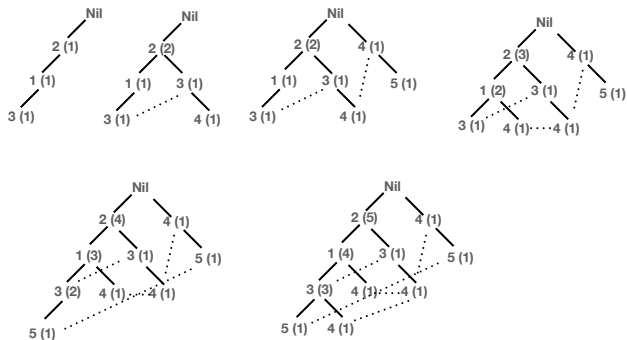
Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Ejemplo



Ejemplo 2

TID	Items Bought	(Ordered) Frequent Items
100	<i>f, a, c, d, g, i, m, p</i>	<i>f, c, a, m, p</i>
200	<i>a, b, c, f, l, m, o</i>	<i>f, c, a, b, m</i>
300	<i>b, f, h, j, o</i>	<i>f, b</i>
400	<i>b, c, k, s, p</i>	<i>c, b, p</i>
500	<i>a, f, c, e, l, p, m, n</i>	<i>f, c, a, m, p</i>

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

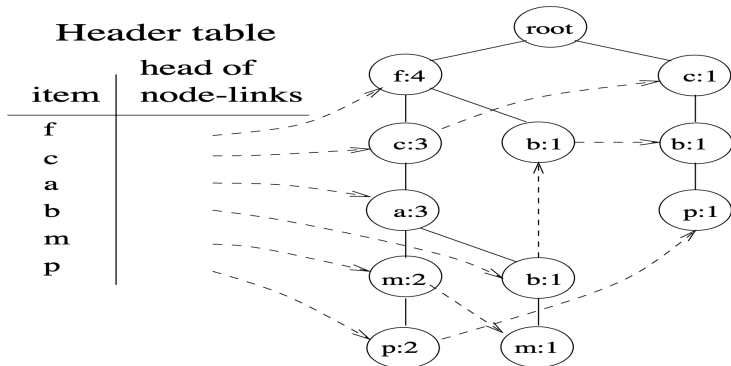
Otros

Aspectos

Clasificación y

Asociación

Ejemplo



Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Para p , el patrón $(p : 3)$ tiene dos caminos:
 $\langle f : 4, c : 3, a : 3, m : 2, p : 2 \rangle$ y $\langle c : 1, b : 1, p : 1 \rangle$,
en el 1o dice que (f, c, a, m, p) ocurre dos veces, y el
prefijo $\langle f : 2, c : 2, a : 2, m : 2 \rangle$ es lo que cuenta.
- En el otro caso, sería el prefijo $\langle c : 1, b : 1 \rangle$. Los dos
forman parte de una base de patrones condicionales
- Estos nos lleva a una sólo rama $(c : 3)$, y el único
patrón frecuente es: $(cp : 3)$

Ejemplo

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

Otros

Aspectos

Clasificación y

Asociación

- Para m , hay dos caminos $\langle f : 4, c : 3, a : 3, m : 2 \rangle$ y $\langle f : 4, c : 3, a : 3, b : 1, m : 1 \rangle$ (p ya no se toma en cuenta)
- La base de patrones condicionales son $(f : 2, c : 2, a : 2)$ y $(f : 1, c : 1, a : 1, b : 1)$
- De aquí se deriva un patrón frecuente:
 $\langle f : 3, c : 3, a : 3 \rangle$

Ejemplo

Contenido

Introducción

Apriori

Extensiones

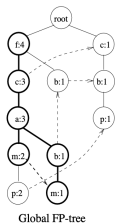
Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

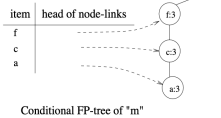
- Esto detona la generación de posibles reglas considerando f , c y a dado m
- De aquí se consideran todas las combinaciones derivadas de m ; $\{(m : 3), (am : 3), (cm : 3), (fm : 3), (cam : 3), (fam : 3), (fcam : 3), (fcm : 3)\}$
- Similarmente, para b hay 3 caminos, pero como no se generan patrones frecuentes, se desechan

Ejemplo



(f:2, c:2, a:2)
 (f:1, c:1, a:1, b:1)
 Conditional pattern base of "m"

Header table



Conditional pattern base of "am": (f:3, c:3)

Conditional FP-tree of "am"



Conditional pattern base of "cam": (f:3)

Conditional FP-tree of "cam"



Ejemplo

item	conditional pattern base	conditional FP-tree
p	$\{(f:2, c:2, a:2, m:2), (c:1, b:1)\}$	$\{(c:3)\} p$
m	$\{(f:4, c:3, a:3, m:2), (f:4, c:3, a:3, b:1, m:1)\}$	$\{(f:3, c:3, a:3)\} m$
b	$\{(f:4, c:3, a:3, b:1), (f:4, b:1), (c:1, b:1)\}$	\emptyset
a	$\{(f:3, c:3)\}$	$\{(f:3, c:3)\} a$
c	$\{(f:3)\}$	$\{(f:3)\} c$
f	\emptyset	\emptyset

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

ECLAT

- Es un método para encontrar patrones frecuentes usando un formato vertical de los datos
- Esto permite no leer otra vez la base de transacciones porque el soporte está dado

Item	Soporte
I1	{T1,T4,T5,T6}
I2	{T1,T2,T4,T5,T6}
I3	{T1,T2,T5,T6}
I4	{T2,T3,T4,T5}
I5	{T3,T5}

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Extensiones: Diferentes abstracciones

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

Otros

Aspectos

Clasificación y

Asociación

- Encontrar reglas de asociación a diferentes niveles de abstracción.
- Normalmente se empieza con las clases superiores, y los resultados pueden servir para filtrar clases inferiores.
- Por ejemplo, considerar reglas de asociación sobre computadoras e impresoras, y luego sobre laptops y estaciones de trabajo, por un lado, y sobre impresoras laser y de punto por otro, etc.

Extensiones: Diferentes abstracciones

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

Al proceder a las subclases se puede considerar:

- Un criterio de soporte uniforme
- Reduciendo el criterio para las subclases
- Considerar todas las subclases independientemente del criterio de soporte

Extensiones: Diferenes abstracciones

Contenido

Introducción

Apriori

Extensiones

Atributos

Continuos

Otros

Aspectos

Clasificación y

Asociación

- Tomando en cuenta el criterio de soporte de una de las superclases de un *item* o k superclases de k *items*
- Considerar *items* aunque el nivel de soporte de sus padres no cumplan con el criterio de soporte, pero que sea mayor que un cierto umbral.

Al encontrar reglas de asociación a diferentes niveles de abstracción es común generar reglas redundantes o reglas que no nos dicen nada nuevo (e.g., la regla más general, ya decía lo mismo), por lo que es necesario incorporar mecanismos de filtrado.

Extensiones: Combinaciones de tablas

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Encontrar reglas de asociación combinando información de múltiples tablas o reglas de asociación multidimensionales.
- Los DataCubes pueden servir para encontrar reglas de asociación multidimensionales.

Atributos continuos

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Las reglas de asociación, al igual que los árboles de decisión y las reglas de clasificación que hemos visto, funcionan, en su forma original, con atributos discretos.
- Al igual que en las otras técnicas se han propuesto mecanismos para manejar atributos continuos.

Atributos continuos

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

Los enfoques más comunes son:

- Discretizar antes de minar en rangos usando posiblemente jerarquías predefinidas.
- Discretizar dinámicamente durante el proceso tratando de maximizar algún criterio de confianza o reducción de longitud de reglas.
- Por ejemplo, ACRS (Association Rule Clustering System), mapea atributos cuantitativos a una rejilla y luego utiliza *clustering*.

Atributos continuos

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Primero asigna datos a “contenedores” delimitados por rangos (que después pueden cambiar).
- Los esquemas más comunes son: Contenedores del mismo tamaño, contenedores con el mismo número de elementos, y contenedores con elementos uniformemente distribuidos.
- Después se encuentran reglas de asociación utilizando los contenedores.
- Una vez que se tienen las reglas, éstas se agrupan si forman rectángulos más grandes dentro de la rejilla.

Atributos continuos

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Discretizar utilizando información semántica, i.e., formar grupos con elementos cercanos (posiblemente haciendo *clustering* sobre los atributos).
- Una vez establecidos los *clusters*, encontrar las reglas de asociación con esos *clusters* basados en distancias o similaridades.

Reglas de Asociación

- El que se encuentre una regla de asociación no necesariamente quiere decir que sea útil.
- Por ejemplo, si se analizan 10,000 compras, de las cuales 6,000 compraron videojuegos, 7,500 videos y 4,000 las dos, posiblemente se genere una regla: compra videojuegos => compra videos [soporte= $4,000/10,000 = 40\%$ y confianza= $4,000/6,000 = 66\%$].
- Sin embargo, el 75% de los clientes compran videos por lo que el comprar videojuegos reduce las posibilidades de comprar videos.

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Otras medidas

- Se han propuesto otras medidas que nos pueden dar más información

- *Lift*:

$$\text{lift}(X \Rightarrow Z) = \frac{\text{soporte}(X \cup Z)}{\text{soporte}(X)\text{soporte}(Z)}$$

- Si $\text{lift} = 1$, las probabilidades del antecedente y consecuente son independientes, por lo que se pueden eliminar esas reglas
- Si $\text{lift} > 1$ nos indica cierta dependencia y por lo tanto utilidad de la regla
- Si $\text{lift} < 1$ nos indica que la presencia de uno tiene efecto negativo en la presencia del otro (son substitutos)
- Existen otras medidas como: *conviction*, *all-confidence*, *collective strength*, *leverage*

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Con esto, se pueden encontrar reglas de asociación correlacionadas.
- Se puede estimar si la correlación es estadísticamente significativa usando una χ^2 .
- Si un conjunto de elementos está correlacionado, cualquier superconjunto de este también lo está.
- Esto puede ayudar a buscar los conjuntos mínimos correlacionados y construir a partir de ahí sus superconjuntos.

Meta-Reglas

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Las meta-reglas permiten especificar la forma de las reglas.
- Podemos buscar por reglas de asociación que tengan formas específicas:
$$P_1(X, Y) \wedge P_2(X, W) \Rightarrow compra(X, \text{libros de ML})$$
donde P_i es un predicado variable que se instancia con algún atributo de la base de datos, y las X , Y y W son posibles valores de los atributos.

Uso de Restricciones

- Se pueden usar restricciones sobre los tipos de datos, jerarquías, o formas posibles de las reglas a encontrar para reducir el espacio de búsqueda.
- Las restricciones pueden ser:
 - ① Antimonótonas: Si un conjunto no satisface una condición, entonces tampoco la satisfacen sus superconjuntos
 - ② Monótonas: Si un conjunto satisface una restricción, entonces también la satisfacen todos sus superconjuntos
 - ③ Suscintas *succint*: Podemos enumerar todos los conjuntos que satisfacen una restricción
 - ④ Convertibles: Podemos convertir una restricción a alguna de las clases anteriores
 - ⑤ No convertibles

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Reglas de Asociación, de Clasificación y árboles de Decisión

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación

Exploración de dependencias vs. Predicción enfocada

Diferentes combinaciones de atributos dependientes e independientes vs. Predice un atributo (clase) a partir de otros

Búsqueda completa (todas las reglas encontradas) vs. búsqueda heurística (se encuentra un subconjunto de reglas)

Reglas de Asociación

Contenido

Introducción

Apriori

Extensiones

Atributos
Continuos

Otros
Aspectos

Clasificación y
Asociación

- Los árboles usan heurísticas de evaluación sobre un atributo, están basados en *splitting*, y normalmente realizan sobreajuste seguido de podado.
- Las reglas de clasificación utilizan una heurística de evaluación de condición (par atributo-valor), están basados en *covering*, y utilizan sobre todo criterios de paro (y a veces sobreajuste y podado).
- Las reglas de asociación se basan en medidas de confianza y soporte, consideran cualquier conjunto de atributos con cualquier otro conjunto de atributos.

Otros Temas Relacionados

- Se ha trabajado para hacer algoritmos más eficientes, que consideren ruido, valores faltantes, atributos mixtos, etc.
- El mecanismo de construcción de reglas de asociación también se ha utilizado para construir reglas de clasificación
 - Lo “único” que se tiene que asegurar es que se tenga a la clase como el único consecuente
- Otros temas relacionados a clasificadores, reglas de clasificación y de asociación son: (i) el enfoque *Lógico Combinatorio* (bloque socialista) y (ii) *Subgroup Discovery* (busca patrones/reglas entre objetos con respecto a una variable de interés)

Contenido

Introducción

Apriori

Extensiones

Atributos
ContinuosOtros
AspectosClasificación y
Asociación