

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

## Actividades resueltas

### Apriori - Aprendizaje de reglas de asociación

#### Descripción de la actividad

Una tienda online tiene la siguiente base de datos con seis registros correspondientes al contenido de cestas de la compra.

	leche	queso	pan	tarta_ manzana	pastel	pastas_ te
<b>1</b>	1	1	1	0	0	0
<b>2</b>	1	1	1	1	1	0
<b>3</b>	1	0	1	1	1	0
<b>4</b>	1	1	1	0	0	0
<b>5</b>	0	0	0	1	1	1
<b>6</b>	0	1	1	1	0	1

Un valor de 1 para un determinado atributo quiere decir que el producto correspondiente ha sido adquirido y forma parte de la cesta de la compra, mientras que un valor de 0 indica lo contrario, es decir que el producto no ha sido adquirido y no forma parte de la cesta de la compra. Por ejemplo, la cesta de la compra con identificador 1 contendría leche, queso y pan.

Se ejecuta el **algoritmo apriori** para generar reglas de asociación con un soporte mínimo de 0.67 y una confianza mínima es 0.8.

Aplica el algoritmo apriori para obtener las reglas que cumplen los requisitos mínimos de interés y frecuencia. Describe claramente los pasos que se realizan en la ejecución del algoritmo. Además, identifica las mejores reglas obtenidas (aquellas que cumplen con los requisitos mínimos de interés y frecuencia).

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

## Resolución de la actividad

En la fase 0 del algoritmo apriori se determina la cobertura mínima a partir del soporte mínimo. En este caso el enunciado define un soporte mínimo de 0.67 y sabemos que la base de datos tiene 6 registros, por lo tanto, la cobertura mínima es de 4 y se calcularía tal que así  $6 \times 0.67 = 4$ . Para que una regla se considere suficientemente frecuente el número de ejemplos que cumplen la regla (antecedente y consecuente) debe ser de 4.

En la fase 1 del algoritmo se generan los ítem-sets que tengan una cobertura mínima de 4 para que las reglas que se vayan a generar en la fase 2 a partir de estos ítem-sets también tengan una cobertura mínima de 4 (soporte mínimo de 0.67).

Para generar los ítem-sets iteramos en el número de k, es decir en el número de pares atributo-valor que tiene el ítem-set.

En la primera iteración y para  $k = 1$ , generamos los ítem-sets con un único ítem, calculamos su cobertura (contamos cuantos ejemplos/registros en la base de datos tienen los valores definidos en el ítem-set) y nos quedamos sólo con aquellos que tienen una cobertura mínima de 4:

leche=1	cobertura = 4
<del>leche=0</del>	<del>cobertura = 2</del>
queso=1	cobertura = 4
<del>queso=0</del>	<del>cobertura = 2</del>
pan=1	cobertura = 5
<del>pan=0</del>	<del>cobertura = 1</del>
tarta_manzana=1	cobertura = 4
<del>tarta_manzana=0</del>	<del>cobertura = 2</del>
pastel=1	cobertura = 3
<del>pastel=0</del>	<del>cobertura = 3</del>
pastas_te=1	cobertura = 2
pastas_te=0	cobertura = 4

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

En la segunda iteración ( $k=2$ ), generamos los ítem-sets de dos pares atributo-valor combinando los ítem-sets encontrados en la primera iteración. Entonces calculamos la cobertura para esos ítem-sets. Por ejemplo, hay 4 ejemplos en la base de datos que tienen el par atributo leche con valor 1 y el par atributo pastas\_te con valor 0, por lo que el ítem-set  $\text{leche}=1, \text{pastas\_te}=0$  tiene cobertura 4. Por último, escogemos los ítem-sets que tienen cobertura mínima de 4:

```

leche=1, queso=1 cobertura = 3
leche=1, pan=1          cobertura = 4
leche=1, tarta_manzana=1 cobertura = 2
leche=1, pastas_te=0     cobertura = 4
queso=1, pan=1          cobertura = 4
queso=1, tarta_manzana=1 cobertura = 2
queso=1, pastas_te=0 cobertura = 3
pan=1, tarta_manzana=1 cobertura = 3
pan=1, pastas_te=0      cobertura = 4
tarta_manzana=1, pastas_te=0 cobertura = 2

```

En la tercera iteración ( $k=3$ ), generamos los ítem-sets de tres pares atributo-valor combinando los ítem-sets de dos pares encontrados en la iteración anterior. Entonces calculamos la cobertura para esos ítem-sets y escogemos los que tienen cobertura mínima de 4:

```

leche=1, pan=1, pastas_te=0      cobertura = 4
leche=1, queso=1, pan=1 cobertura = 3
queso=1, pan=1, pastas_te=0 cobertura = 3

```

Como en la tercera iteración se ha generado un único ítem-set de tres elementos no va a ser posible generar más ítem-sets de 4 o más elementos, por lo que se finaliza la fase 1 del algoritmo.

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

En la fase 2 del algoritmo se van a generar reglas a partir de los ítem-sets generados en la fase 1. Debemos tener en cuenta que con ítem-sets de 1 elemento no se pueden generar reglas porque no tenemos un ítem para situarlo en el antecedente y otro en el consecuente. Por lo tanto, partimos sólo de los ítem-sets de 2 y 3 ítems que serían los siguientes:

leche=1, pan=1	cobertura = 4
leche=1, pastas_te=0	cobertura = 4
queso=1, pan=1	cobertura = 4
pan=1, pastas_te=0	cobertura = 4
leche=1, pan=1, pastas_te=0	cobertura = 4

A partir del ítem-set leche=1, pan=1 se pueden generar dos reglas:

SI leche=1 ENTONCES pan=1

SI pan=1 ENTONCES leche=1

Entonces tenemos que calcular la confianza de estas dos reglas, es decir la proporción de ejemplos que satisfacen la regla (cumplen el antecedente y consecuente) respecto a aquellos ejemplos que satisfacen el antecedente (independientemente de que satisfagan el consecuente o no). El valor en el numerador es el número de ejemplos que satisfacen la regla y se corresponde con el valor de la cobertura. Por lo tanto, las dos reglas tendrán una confianza de  $4/x$ , donde  $x$  será el número de ejemplos que satisfacen el antecedente. Para la primera regla, hay 4 ejemplos que tienen para el atributo leche un valor de 1, es decir que satisfacen el antecedente leche=1. Para la segunda regla, hay 5 ejemplos que tienen para el atributo pan un valor de 1, es decir que satisfacen el antecedente pan=1. Haciendo los cálculos obtenemos las siguientes confianzas:

SI leche=1 ENTONCES pan=1	confianza = $4/4 = 1$
SI pan=1 ENTONCES leche=1	confianza = $4/5 = 0.80$

Por último, descartamos aquellas reglas que no superen el valor mínimo de confianza que es 0.8 según se ha definido en el enunciado. En este caso nos quedamos con las dos reglas ya que tienen la confianza mínima y por tanto son interesantes.

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

Se repite el proceso para el resto de ítem-sets.

A partir del ítem-set leche=1, pastas\_te=0 se generan las siguientes reglas con confianza mínima:

SI leche=1 ENTONCES pastas_te=0	confianza = $4/4 = 1$
SI pastas_te=0 ENTONCES leche=1	confianza = $4/4 = 1$

A partir del ítem-set queso=1, pan=1 se generan las siguientes reglas con confianza mínima:

SI queso=1 ENTONCES pan=1	confianza = $4/4 = 1$
SI pan=1 ENTONCES queso=1	confianza = $4/5 = 0.80$

A partir del ítem-set pan=1, pastas\_te=0 se generan las siguientes reglas con confianza mínima:

SI pan=1 ENTONCES pastas_te=0	confianza = $4/5 = 0.80$
SI pastas_te=0 ENTONCES pan=1	confianza = $4/4 = 1$

Por último, a partir del ítem-set de tres elementos leche=1, pan=1, pastas\_te=0 se generan las siguientes reglas con confianza mínima:

SI leche=1 AND pan=1 ENTONCES pastas_te=0	confianza = $4/4 = 1$
SI leche=1 AND pastas_te=0 ENTONCES pan=1	confianza = $4/4 = 1$
SI pan=1 AND pastas_te=0 ENTONCES leche=1	confianza = $4/4 = 1$
SI leche=1 ENTONCES pan=1 AND pastas_te=0	confianza = $4/4 = 1$
SI pan=1 ENTONCES leche=1 AND pastas_te=0	confianza = $4/5 = 0.80$
SI pastas_te=0 ENTONCES leche=1 AND pan=1	confianza = $4/4 = 1$

En este punto finalizaría la fase 2 del algoritmo apriori.

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

El conjunto final de reglas de asociación generadas por el algoritmo apriori sería el siguiente:

SI leche=1 ENTONCES pan=1	cobertura = 4	confianza = 1
SI leche=1 ENTONCES pastas_te=0	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES leche=1	cobertura = 4	confianza = 1
SI queso=1 ENTONCES pan=1	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES pan=1	cobertura = 4	confianza = 1
SI leche=1 AND pan=1 ENTONCES pastas_te=0	cobertura = 4	confianza = 1
SI leche=1 AND pastas_te=0 ENTONCES pan=1	cobertura = 4	confianza = 1
SI pan=1 AND pastas_te=0 ENTONCES leche=1	cobertura = 4	confianza = 1
SI leche=1 ENTONCES pan=1 AND pastas_te=0	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES leche=1 AND pan=1	cobertura = 4	confianza = 1
SI pan=1 ENTONCES leche=1	cobertura = 4	confianza = 0.80
SI pan=1 ENTONCES queso=1	cobertura = 4	confianza = 0.80
SI pan=1 ENTONCES pastas_te=0	cobertura = 4	confianza = 0.80
SI pan=1 ENTONCES leche=1 AND pastas_te=0	cobertura = 4	confianza = 0.80

**Nota:** Si en lugar de definir una confianza de 0.80 se hubiera utilizado un valor más alto, por ejemplo 0.90, y se hubiera mantenido el valor de soporte del 0.67, la fase 1 del algoritmo generaría los mismos ítem-sets pero en la fase 2 se habrían descartado todas las reglas con una confianza del 0.80. Por lo tanto, en este caso el conjunto final de reglas sería el siguiente:

SI leche=1 ENTONCES pan=1	cobertura = 4	confianza = 1
SI leche=1 ENTONCES pastas_te=0	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES leche=1	cobertura = 4	confianza = 1
SI queso=1 ENTONCES pan=1	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES pan=1	cobertura = 4	confianza = 1
SI leche=1 AND pan=1 ENTONCES pastas_te=0	cobertura = 4	confianza = 1
SI leche=1 AND pastas_te=0 ENTONCES pan=1	cobertura = 4	confianza = 1
SI pan=1 AND pastas_te=0 ENTONCES leche=1	cobertura = 4	confianza = 1
SI leche=1 ENTONCES pan=1 AND pastas_te=0	cobertura = 4	confianza = 1
SI pastas_te=0 ENTONCES leche=1 AND pan=1	cobertura = 4	confianza = 1