

EJEMPLO DE REGLAS DE ASOCIACIÓN

Productos a vender

I = {alcohol, algodón, cubreboca, Gel antibacterial, guantes}

Tabla de transacciones

No. Transacción	Ítems comprados
1	{algodón, cubreboca, Gel antibacterial}
2	{cubreboca, guantes}
3	{alcohol, cubreboca}
4	{cubreboca, Gel antibacterial, guantes}
5	{alcohol, Gel antibacterial}
6	{alcohol, cubreboca}
7	{alcohol, Gel antibacterial}
8	{alcohol, cubreboca, , Gel antibacterial}
9	{alcohol, cubreboca, Gel antibacterial}

Representación tabular

Transacción	alcohol	algodón	cubreboca	Gel antibacterial	guantes
1	0	1	1	1	0
2	0	0	1	0	1
3	1	0	1	0	0
4	0	0	1	1	1
5	1	0	0	1	0
6	1	0	1	0	0
7	1	0	0	1	0
8	1	1	1	1	0
9	1	0	1	1	0
Suma	6	2	7	6	2

Se establece $fi(\varphi) = 2$, $min_sup = 2$, $min_conf = 50\%$.

Si $fi(\varphi) = 2$, significa que un conjunto de datos es frecuente si ocurre 2 o más veces en D.

PASO NO. 1. GENERAR LOS CONJUNTOS DE ELEMENTOS FRECUENTES

PRIMERA ITERACIÓN (L1)

Se refiere (1-itemsets) a los conjuntos frecuentes de 1 ítem, esto es, los productos individuales.

C1 =

Itemset	Soporte
alcohol	6
algodón	2
cubreboca	7
Gel antibacterial	6
Guantes	2

Se analiza si las ocurrencias cumplen con $\min_sup = 2$. $\min_sup \geq f_i(\phi) = 2$. Todos cumplen.

Por lo tanto, el resultado es:

L1 =

Itemset	Soporte
alcohol	6
algodón	2
cubreboca	7
Gel antibacterial	6
Guantes	2

SEGUNDA ITERACIÓN (L2)

Se generan los 2-itemset candidatos para el conjunto C2, lo cual se realiza a partir de L1.

Itemset	Soporte
alcohol	6
algodón	2
cubreboca	7
Gel antibacterial	6
Guantes	2



Itemset	Soporte
alcohol	6
algodón	2
cubreboca	7
Gel antibacterial	6
Guantes	2

Itemset	Soporte
{alcohol, algodón}	1
{alcohol, cubrebocas}	4
{alcohol, gel antibacterial}	4
{alcohol, guantes}	0
{algodón, cubreboca}	2
{algodón, gel antibacterial}	2
{algodón, guantes}	0
{cubreboca, gel antibacterial}	4
{cubreboca, guantes}	2
{gel antibacterial, guantes}	1

Se analiza si las ocurrencias cumplen con $\text{min_sup} = 2$. $\text{min_sup} \geq \text{fi}(\varphi) = 2$.

L2 =

Itemset	Soporte
{alcohol, cubrebocas}	4
{alcohol, gel antibacterial}	4
{algodón, cubreboca}	2
{algodón, gel antibacterial}	2
{cubreboca, gel antibacterial}	4
{cubreboca, guantes}	2

TERCERA ITERACIÓN (L3)

Se generan los 3-itemset candidatos para el conjunto C3, lo cual se realiza a partir de L2.

Itemset	Soporte
{alcohol, cubrebocas}	4
{alcohol, gel antibacterial}	4
{algodón, cubreboca}	2
{algodón, gel antibacterial}	2
{cubreboca, gel antibacterial}	4
{cubreboca, guantes}	2



Itemset	Soporte
{alcohol, cubrebocas}	4
{alcohol, gel antibacterial}	4
{algodón, cubreboca}	2
{algodón, gel antibacterial}	2
{cubreboca, gel antibacterial}	4
{cubreboca, guantes}	2

Itemset	Soporte
{alcohol, cubrebocas, gel antibacterial}	2
{algodón, cubrebocas, gel antibacterial}	2
{cubreboca, gel antibacterial, guantes}	1

Paso número 2. Poda, hay que revisar que se cumpla la propiedad a priori que establece: Si un conjunto de elementos Z no es frecuente, entonces para cualquier elemento A, Z U A no será frecuente (Larose & Larose, 2015).

Hay que revisar si los conjuntos creados tienen subconjuntos frecuentes.

Subconjunto	Frecuencia	Decisión
{alcohol, cubrebocas, gel antibacterial}		No se poda
{alcohol, cubrebocas}	4	
{alcohol, gel antibacterial}	4	
{cubrebocas, gel antibacterial}	4	
{algodón, cubrebocas, gel antibacterial}		No se poda
{algodón, cubrebocas}	2	
{algodón, gel antibacterial}	2	
{cubrebocas, gel antibacterial}	4	

L3 =	Itemset
	{alcohol, cubrebocas, gel antibacterial}
	{algodón, cubrebocas, gel antibacterial}

Hasta aquí se termina el proceso de creación de conjuntos frecuentes debido a que no existen k-1 elementos en común en los conjuntos generados en L3.

PASO NO. 2. GENERAR LAS REGLAS DE ASOCIACIÓN FUERTES

Se puede extraer una regla de asociación dividiendo el conjunto de elementos Y en dos subconjuntos no vacíos, X e Y-X, de modo que $X \rightarrow Y-X$ satisface el umbral de confianza (Tan, Steinbach & Kumar, 2014).

Por ejemplo: Sea $X = \{1, 2, 3\}$ un conjunto de elementos frecuente. Hay seis reglas de asociación candidatas que se pueden generar a partir de X: $\{1, 2\} \rightarrow \{3\}$, $\{1, 3\} \rightarrow \{2\}$, $\{2, 3\} \rightarrow \{1\}$, $\{1\} \rightarrow \{2, 3\}$, $\{2\} \rightarrow \{1, 3\}$ y $\{3\} \rightarrow \{1, 2\}$. Como cada uno de sus soportes es idéntico al soporte de X, las reglas deben satisfacer el umbral de soporte (Tan, Steinbach & Kumar, 2014).

Las reglas de asociación fuertes deben satisfacer dos condiciones: soporte mínimo y confianza mínima (Han, Kamber & Pei, 2012).

Considerando las siguientes fórmulas:

$$\text{Confianza } (A \Rightarrow B) = P(B | A) = \frac{\text{support_count}(A \cap B)}{\text{support_count}(A)}$$

Esto significa lo siguiente:

Support_count ($A \cap B$) es el número de transacciones que contiene los itemsets $A \cap B$ y support_count(A) es el número de transacciones que contienen el itemset A.

Regla	Soporte	Confianza
alcohol, cubrebocas \Rightarrow gel antibacterial	$2/9 = 0.22$	$2/4 = 50\%$
alcohol, gel antibacterial \Rightarrow cubrebocas	$2/9 = 0.22$	$2/4 = 50\%$
cubrebocas, gel antibacterial \Rightarrow alcohol	$2/9 = 0.22$	$2/4 = 50\%$
alcohol \Rightarrow cubrebocas, gel antibacterial	$2/9 = 0.22$	$2/6 = 33\%$
cubrebocas \Rightarrow alcohol, gel antibacterial	$2/9 = 0.22$	$2/7 = 29\%$
gel antibacterial \Rightarrow alcohol, cubrebocas	$2/9 = 0.22$	$2/6 = 33\%$
algodón, cubrebocas \Rightarrow gel antibacterial	$2/9 = 0.22$	$2/4 = 50\%$
algodón, gel antibacterial \Rightarrow cubrebocas	$2/9 = 0.22$	$2/2 = 100\%$
cubrebocas, gel antibacterial \Rightarrow algodón	$2/9 = 0.22$	$2/4 = 50\%$
Algodón \Rightarrow cubrebocas, gel antibacterial	$2/9 = 0.22$	$2/2 = 100\%$
Cubrebocas \Rightarrow algodón, gel antibacterial	$2/9 = 0.22$	$2/7 = 29\%$
gel antibacterial \Rightarrow algodón, cubrebocas	$2/9 = 0.22$	$2/6 = 33\%$

Reglas de asociación finales considerando una Confianza del 50%

Regla	Soporte	Confianza
alcohol, cubrebocas \Rightarrow gel antibacterial	$2/9 = 0.22$	$2/4 = 50\%$
alcohol, gel antibacterial \Rightarrow cubrebocas	$2/9 = 0.22$	$2/4 = 50\%$
cubrebocas, gel antibacterial \Rightarrow alcohol	$2/9 = 0.22$	$2/4 = 50\%$
algodón, cubrebocas \Rightarrow gel antibacterial	$2/9 = 0.22$	$2/4 = 50\%$
algodón, gel antibacterial \Rightarrow cubrebocas	$2/9 = 0.22$	$2/2 = 100\%$
cubrebocas, gel antibacterial \Rightarrow algodón	$2/9 = 0.22$	$2/4 = 50\%$
Algodón \Rightarrow cubrebocas, gel antibacterial	$2/9 = 0.22$	$2/2 = 100\%$

EVALUACIÓN DE LAS REGLAS DE ASOCIACIÓN

Para evaluar las reglas de asociación generadas, se aplica el Lift.

El Lift calcula el *ratio* entre la confianza de la regla y el soporte del itemset en el consecuente de la regla. Para las variables binarias, el Lift es equivalente a otra medida objetiva llamada **factor de interés** (Tan, Steinbach & Kumar, 2014.:372).

No todas las reglas de asociación son igualmente útiles. Existe una medida que puede cuantificar la utilidad de una regla de asociación llamada carga (lift) que se define de la siguiente manera (Larose & Larose, 2015):

La fórmula presentada por Larose & Larose (2015):

$$Lift = \frac{\text{confianza de la regla}}{\text{proporción a priori del consecuente}}$$

Lift de las reglas generadas:

Número total de transacciones = 9

$$Lift = \frac{\text{confianza de la regla}}{\text{proporción a priori del consecuente}}$$

1) alcohol, cubrebocas \Rightarrow gel antibacterial

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{0.5}{0.22} = 2.72$$

Esto significa que los clientes que compraron alcohol y cubre boca tienen 2.72 veces más la probabilidad de comprar gel antibacterial que los clientes de todo el conjunto de datos.

2) alcohol, gel antibacterial \Rightarrow cubrebocas

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{0.5}{0.22} = 2.72$$

Esto significa que los clientes que compraron alcohol y gel antibacterial tienen 2.72 veces más la probabilidad de comprar cubrebocas que los clientes de todo el conjunto de datos.

3) cubrebocas, gel antibacterial \Rightarrow alcohol

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{0.5}{0.22} = 2.72$$

Esto significa que los clientes que compraron cubreboca y gel antibacterial tienen 2.72 veces más la probabilidad de comprar alcohol que los clientes de todo el conjunto de datos.

4) algodón, cubrebocas \Rightarrow gel antibacterial

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{0.5}{0.22} = 2.72$$

Esto significa que los clientes que compraron algodón y cubrebocas tienen 2.72 veces más la probabilidad de comprar gel antibacterial que los clientes de todo el conjunto de datos.

5) algodón, gel antibacterial \Rightarrow cubrebocas

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{1.0}{0.22} = 4.54$$

Esto significa que los clientes que compraron algodón y gel antibacterial tienen 4.54 veces más la probabilidad de comprar cubrebocas que los clientes de todo el conjunto de datos.

6) cubrebocas, gel antibacterial \Rightarrow algodón

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{0.5}{0.22} = 2.27$$

Esto significa que los clientes que compraron algodón tienen 2.27 veces más la probabilidad de comprar cubrebocas y gel antibacterial que los clientes de todo el conjunto de datos.

7) algodón \Rightarrow cubrebocas, gel antibacterial

Proporción a priori = $2/9 = 0.22$

$$Lift = \frac{1.0}{0.22} = 4.54$$

Esto significa que los clientes que compraron algodón tienen 4.54 veces más la probabilidad de comprar cubrebocas y gel antibacterial que los clientes de todo el conjunto de datos.

OTRO EJEMPLO:

Ejemplo propuesto por Tan Pang-Ning; Steinbach, Michael & Kumar, Vipin. (2014). Introduction to Data Mining. Pearson.

En este ejemplo se plantea $\text{min_sup} = 3$,

Table 6.1. An example of market basket transactions.

TID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

Table 6.2. A binary 0/1 representation of market basket data.

TID	Bread	Milk	Diapers	Beer	Eggs	Cola
1	1	1	0	0	0	0
2	1	0	1	1	1	0
3	0	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	0	1

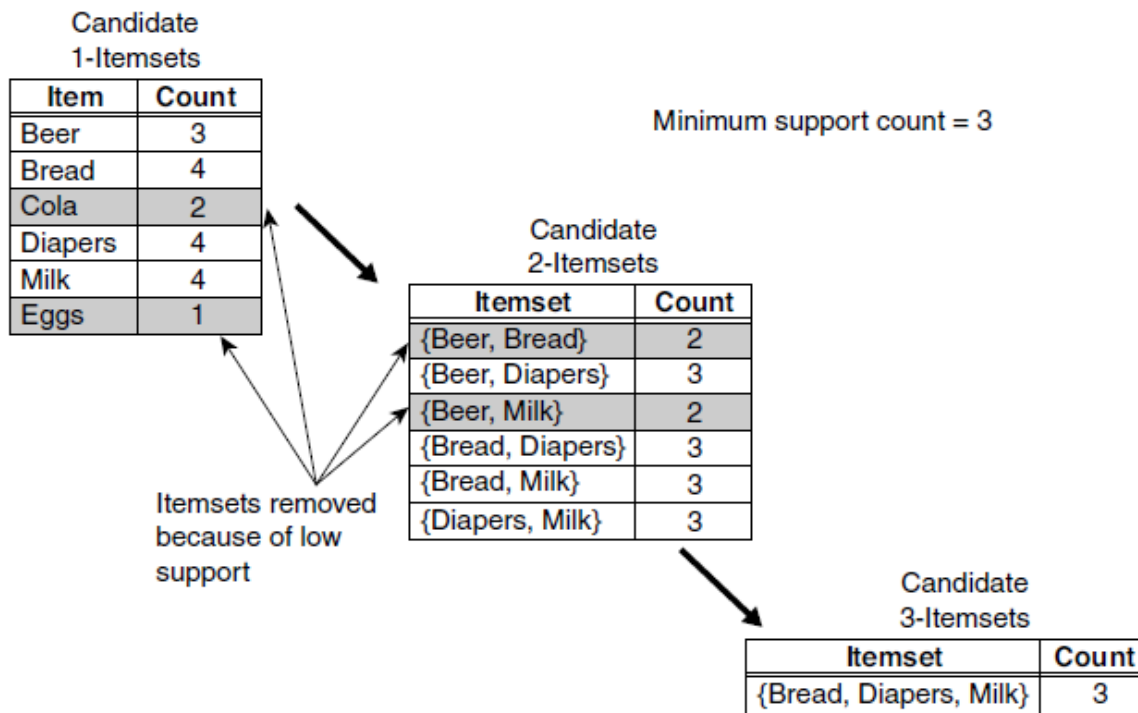


Figure 6.5. Illustration of frequent itemset generation using the *Apriori* algorithm.