

Practica 5

martes, 1 de junio de 2021 11:45

1. Conjuntos de items (itemsets)

- a) Dado un conjunto de datos con los items {A,B,C}, listar todos los posibles **itemsets** (subconjuntos no nulos de A,B,C) que pueden generarse y contarlos.

{A,B,C}, {A,B}, {B,C}, {A,C}, {A}, {B} y {C}

- b) Calcular la cantidad de **itemsets** máxima posible para cada fila de la tabla. Para los items A,B,C ya obtuvo el resultado en a).

Items	#Items	Cantidad de itemsets máxima posible
A	1	
A,B	2	
A,B,C	3	
A,B,C,D	4	

Es como que cada letra es un binario en 1

A = 1

A,B = 3

A,B,C = 7

A,B,C,D = 15

{A,B,C,D}, {A,B,C}, {A,B,D}, {B,C,D}, {A,C,D}, {A,B}, {A,C}, {A,D}, {B,C}, {B,D}, {C,D}, {A}, {B}, {C}, {D}

- c) En base a la tabla de b), determinar la fórmula que calcula la cantidad de **itemsets** que pueden formarse en base a **N** items. Note la rapidez con la que se incrementa la cantidad de **itemsets** con respecto a la cantidad de items ¿qué impacto tiene esto en los algoritmos de generación de **itemsets**?

La fórmula es: $(2^n)-1$

El crecimiento es exponencial, por lo que los algoritmos de generación de itemsets la van a pasar re mal.

- d) En base al conjunto de datos de la tabla de abajo, listar los **itemsets** frecuentes que tienen un soporte mayor a 0.3. Para resolver este ejercicio, realizar una tabla con tantas columnas como **itemsets** posibles, y por cada **itemset** contar la cantidad de veces que aparece en la fila 1, y su soporte en la fila 2.

id	Items
1	A,B,C,D
2	A,B
3	A,B
4	A,D
5	B

Observe el número de **itemsets** resultante ¿cómo se compara con el máximo teórico calculado en a)?

Itemsets	{A}	{B}	{D}	{A,B}	{A,D}
Cantidad	4	4	2	3	2
Soporte	4/5	4/5	2/5	3/5	2/5

Quedó una tabla de 5 elementos, cuando el máximo es 15

- e) En lugar de usar un algoritmo de fuerza bruta como en d), aplique manualmente el algoritmo APriori para generar los itemsets frecuentes con soporte 0.3 o mayor. El resultado debe ser igual al del inciso d).

C1

{A}	4
{B}	3
{C}	1
{D}	2

L1

{A}
{B}
{D}

C2

{A,B}	3
{A,D}	2
{B,D}	1

L2

{A,B}
{A,D}

C3

{A,B,D}	1
---------	---

L3 vacío

Entonces, los itemset con 0,3 o más de soporte son:

{A}
{B}
{D}
{A,B}
{A,D}

2. Reglas de asociación

- e) En base a los itemsets frecuentes generados en el ejercicio 1d) o 1e), genere todas las reglas posibles. Observe el número de reglas resultante. ¿Cómo se compara con el máximo teórico calculado en c)?

A → B

B → A

A → D

D → A

3. Métricas de evaluación de reglas

a) Calcule manualmente el soporte, interés, cobertura y confianza de las reglas obtenidas en el inciso e) del ejercicio 2. Para ello, nuevamente recomendamos hacer una tabla con estos valores como filas y los itemsets como columnas. En base a los resultados, responda:

1. ¿El valor del soporte es el mismo para $X \rightarrow Y$ o para $Y \rightarrow X$?
2. ¿Esto será cierto para cualquier regla? Analice la definición de soporte.
3. ¿Se cumple lo mismo para la cobertura, la confianza y el interés?

Reglas del inciso 2) e)

Regla	Soporte	Interés	Cobertura	Confianza
A → B	3/5	$(3/5) / (4/5 * 4/5) = 0,9375$	$4/5 = 0,8$	$(3/5) / (4/5) = 0,75$
B → A	3/5	$(3/5) / (4/5 * 4/5) = 0,9375$	$4/5 = 0,8$	$(3/5) / (4/5) = 0,75$
A → D	2/5	$(2/5) / (4/5 * 2/5) = 1,25$	$4/5 = 0,8$	$(2/5) / (4/5) = 0,5$
D → A	2/5	$(2/5) / (2/5 * 4/5) = 1,25$	$2/5 = 0,4$	$(2/5) / (2/5) = 1$

1. Si.
2. Si, la union del antecedente y el consecuente es siempre la misma independientemente del orden de la implicación.
3. No, por ejemplo en la cobertura. La cobertura de A → B cuenta la cantidad de veces que aparece A, que no es lo mismo que la cobertura de B → A, que cuenta la cantidad de veces que aparece B. Para la confianza también varía, ya que el denominador es la cantidad de veces que aparece el antecedente, por lo que cambia con el orden de la implicación.
En el interés si se cumple que no importa el orden de la implicación, también es simétrica como con el soporte.