

Tp05.03ReglasDeAsociacion-R

1. Soporte & Confianza. Calcule el soporte y la confianza (cuando corresponda) de los ítemsets del siguiente fragmento del dataset:

a. Calcule el soporte para todos los ítemsets del dataset.

b. Arme todas las reglas resultantes considerando que se solicita un soporte mínimo de 0,3.

| # | A | B | C | ITEMSETS | SOPORTE | soporte minimo >= 0,3 | REGLAS GENERADAS |
|----|---|---|---|----------|---------|-----------------------------|------------------|
| 1 | 0 | 1 | 0 | {B} | 5/10 | 0,5 | {B} |
| 2 | 1 | 0 | 1 | {A,C} | 4/10 | 0,4 | {A}->{C} |
| 3 | 0 | 0 | 1 | {C} | 6/10 | 0,6 | {C} |
| 4 | 1 | 0 | 0 | {A} | 7/10 | 0,7 | {A} |
| 5 | 1 | 1 | 1 | {ABC} | 2/10 | 0,2 | |
| 6 | 0 | 1 | 1 | {BC} | 3/10 | 0,3 | {B}->{C} |
| 7 | 1 | 1 | 0 | {AB} | 3/10 | 0,3 | {A}->{B} |
| 8 | 1 | 0 | 1 | {AC} | 4/10 | 0,4 | {A}->{C} |
| 9 | 1 | 0 | 0 | {A} | 7/10 | 0,7 | {A} |
| 10 | 1 | 1 | 1 | {ABC} | 2/10 | 0,2 | |

c. ¿Cuál es el soporte de A? ¿Cómo es el soporte de AB, AC y ABC con respecto al de A? ¿Por qué?

| ítemset | soporte |
|---------|---------|
| A | 0,7 |
| AB | 0,3 |
| AC | 0,4 |
| ABC | 0,2 |

Los valores de soporte son mas chicos porque a medida que voy agregando elementos, la frecuencia de estos ultimos va a ser como maximo igual al conjunto mayor, en este caso {A}.

d. ¿Cuáles son las reglas de asociación resultantes si establecemos una confianza mínima de 0,7?

| ítemset | confianza |
|------------------------------------|-----------|
| $\sigma(\{A\} \rightarrow \{C\})$ | 0,57 |
| $\sigma(\{B\} \rightarrow \{C\})$ | 0,66 |
| $\sigma(\{A\} \rightarrow \{B\})$ | 0,42 |
| $\sigma(\{AC\} \rightarrow \{B\})$ | 0,5 |

| itemset | confianza |
|------------------------------------|-----------|
| $\sigma(\{AB\} \rightarrow \{C\})$ | 0,6 |
| $\sigma(\{BC\} \rightarrow \{A\})$ | 0,6 |

Se puede observar que ninguna regla obtenida establece la confianza minima requerida.

2. Apriori. Incorpore en una herramienta de data mining el dataset sobre la cesta de compras y responda:

a. ¿Qué parámetros puedo modificar previo a ejecutar el algoritmo sobre el dataset? ¿Qué permite cada uno?

estructura de una REGLA

LHS (parte izquierda de la regla) => RHS (parte derecha de la regla) ejemplo regla:

| LHS (Left Hand Side) | implicacion | RHS (Right Hand Side) |
|----------------------|-------------|-----------------------|
| "butter" | => | "bottled beer" |

```
reglas_beer <- apriori(Groceries, parameter = list(support=0.01, confidence=0.01,
target = "rules"), appearance = list(lhs="bottled beer"))
```

```
parametros = lista(
  support = frecuencia relativa de una regla sobre el total de transacciones
  confidence = veces que RHS se presenta cuando se presenta LHS por cada regla
  target = resultado solicitado {
    'rules': todas las reglas resultantes;
    'frequent itemsets': itemsets mas frecuentes }
)
```

b. ¿Es posible ejecutar el algoritmo apriori con el dataset tal como se encuentra? Realice las operaciones necesarias para permitirlo.

c. Ejecute el algoritmo Apriori sobre los datos y detalle cuáles son las mejores reglas encontradas. ¿Cómo determina cuales son las mejores?

d. ¿Qué nota al ejecutar el algoritmo con el dataset actual? ¿Cuál es la complejidad computacional del mismo? ¿Cómo puede resolverse?

e. Si tuviera que analizar los resultados desde el punto de vista de un analista especializado en marketing, ¿Cuáles son las asociaciones encontradas que le parecen más interesantes? ¿Qué políticas podría implementar a partir de estas?

f. Explore alguna visualización⁴ para el análisis de reglas generadas, explique brevemente.

- g. Utilizando el mismo punto de vista, ¿Cuáles son los ítems marcan la presencia de cerveza?
¿Encuentra una relación lógica en estas asociaciones?
- h. ¿Qué parámetros ajustaría a efectos de modificar la cantidad de cantidad de reglas de asociación generadas? ¿Qué efecto generan esos parámetros? Ejemplifique en función del dataset actual.
- i. Documente todas las actividades desarrolladas y exprese sus conclusiones en cada caso.

3. Incorpore el dataset del Banco de Portugal utilizado en el TP de árboles de decisión y realice las siguientes operaciones:

- a. Aplique las transformaciones necesarias a efectos de poder correr el algoritmo apriori sobre el dataset.
- b. Ejecute el algoritmo apriori y explique los resultados más importantes.
- c. ¿Cuáles son las asociaciones más importantes para determinar si el cliente aceptó o no el producto bancario?
- d. Compare los resultados obtenidos en este punto con respecto a los resultados obtenidos a través de la aplicación de árboles de decisión.
- e. Documente todas las actividades desarrolladas y exprese sus conclusiones en cada caso.