

Reglas de Asociación

EP7144 - Técnicas de Minería de Datos

Mg. Enver Gerald Tarazona Vargas
enver.tarazona@pucp.edu.pe

Escuela de Post-Grado

Universidad Nacional Agraria La Molina (UNALM)



- 1 Reglas de Asociación
 - Introducción
 - Reglas de Asociación

Analítica Descriptiva

- El objetivo es describir patrones en el comportamiento de las instancias (p. ej. comportamiento del cliente).
- Usualmente se le conoce también como aprendizaje no supervisado debido a que no existe una variable respuesta que dirija el proceso de aprendizaje.
- Las tres técnicas más usadas son: Reglas de Asociación, Reglas de Secuencias y Clustering.

Técnicas de Analítica Descriptiva

Table 4.1 Examples of Descriptive Analytics

Type of Descriptive Analytics	Explanation	Example
Association rules	Detect frequently occurring patterns between items	Detecting what products are frequently purchased together in a supermarket context Detecting what words frequently co-occur in a text document Detecting what elective courses are frequently chosen together in a university setting
Sequence rules	Detect sequences of events	Detecting sequences of purchase behavior in a supermarket context Detecting sequences of web page visits in a web mining context Detecting sequences of words in a text document
Clustering	Detect homogeneous segments of observations	Differentiate between brands in a marketing portfolio Segment customer population for targeted marketing

Reglas de Asociación

- Las reglas de asociación son útiles para descubrir relaciones que sean de interés en conjuntos de datos.
- Estas relaciones pueden ser descritas en la forma de reglas de asociación.
- Las reglas de asociación usualmente parten de una base de datos de transacciones, D .
- Cada transacción consiste de un identificador y un conjunto de items (p. ej. productos, páginas web, cursos) $\{i_1, i_2, \dots, i_n\}$ seleccionados de todos los ítems posibles (I).

Reglas de Asociación I

- Por ejemplo, la siguiente tabla es un ejemplo de una base de datos de transacciones:

<i>TID</i>	<i>Items</i>
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- Una regla de asociación es un implicancia de la forma $X \Rightarrow Y$, tal que $X \subset I$, $Y \subset I$ y $X \cap Y = \emptyset$.
- X es conocida como la regla antecedente, mientras que Y como la regla consecuente.
- De la tabla anterior, la siguiente regla de asociación puede ser encontrada:

$$\{\text{Diaper}\} \rightarrow \{\text{Beer}\}$$

Reglas de Asociación II

- Esta regla sugiere que existe una fuerte asociación entre la venta de pañales y de cerveza porque muchos clientes que compran pañales también compran cerveza.
- Es importante notar que las reglas de asociación son de naturaleza estocástica, lo cual implica que no deben ser interpretados como verdades universales y son caracterizadas por medidas estadísticas que cuantifican la fuerza de la asociación.
- Las reglas miden asociaciones y no deben ser interpretadas de forma causal.

Confianza y Soporte

- **Itemset:** es la colección de uno o más ítems.

$$\{\text{Milk, Bread, Diaper}\}$$

- **Soporte:** Es la frecuencia con la que ocurre un itemset.

$$s(\{\text{Milk, Bread, Diaper}\})=2/5$$

- **Regla de Asociación:** es una relación entre dos itemset.

$$X \rightarrow Y$$

representa el patrón de que cuando X ocurre también ocurre Y

$$\{\text{Milk, Diaper}\} \rightarrow \{\text{Beer}\}$$

Confianza y Soporte

Medidas de Evaluación de una Regla $X \rightarrow Y$

- Soporte:**

$$s(X \rightarrow Y) = s(X \cup Y) = \frac{\# \text{ de trans. que contienen a } (X \cup Y)}{\# \text{ total de transacciones}}$$

- Confianza:** $c(X \rightarrow Y) = P(Y|X) = \frac{s(X \cup Y)}{s(X)}$

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

$\{\text{Milk, Diaper}\} \Rightarrow \text{Beer}$

$$s = s(\text{Milk, Diaper, Beer}) = \frac{2}{5} = 0.4$$

$$c = \frac{s(\text{Milk, Diaper, Beer})}{s(\text{Milk, Diaper})} = \frac{2}{3} = 0.67$$

Reglas de Asociación

- Dado un conjunto de datos, el objetivo es encontrar todas la reglas de asociación que cumplan:
- Soporte \geq minsup
- Confianza \geq mincon

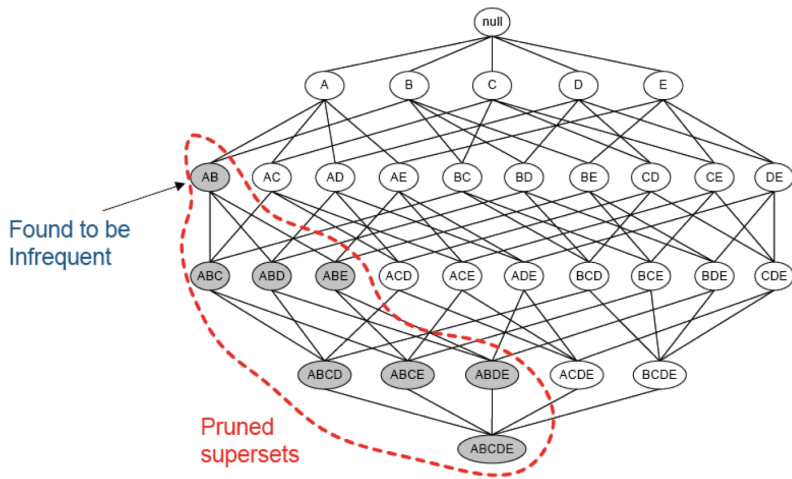
Reglas de Asociación

- Generalmente se siguen 2 pasos:
 - 1 **Generacion de itemsets frecuentes:** Encontrar todos los itemsets que cumplan con $\text{Soporte} \geq \text{minsup}$
 - 2 **Derivacion de reglas de asociacion:** Generar reglas de asociación a partir de los itemset frecuentes encontrados que cumplan con $\text{Confianza} \geq \text{mincon}$.

Reglas de Asociación

- Algoritmo A priori.
- Si un itemset es frecuente, entonces todos sus subconjuntos deben ser frecuentes también.

Reglas de Asociación



Reglas de Asociación

Item	Count
Bread	4
Coke	2
Milk	4
Beer	3
Diaper	4
Eggs	1



Itemset	Count
{Bread,Milk}	3
{Bread,Beer}	2
{Bread,Diaper}	3
{Milk,Beer}	2
{Milk,Diaper}	3
{Beer,Diaper}	3



Itemset	Count
{Bread,Milk,Diaper}	3

Soporte Mínimo = 3

Reglas de Asociación

- Algoritmo A priori.
- Una vez encontrados los itemset frecuentes, se encuentran las reglas de asociación:
- Por ejemplo si $\{A, B, C, D\}$ es un itemset frecuente, entonces las posibles reglas de asociación serian:

$ABC \rightarrow D,$	$ABD \rightarrow C,$	$ACD \rightarrow B,$	$BCD \rightarrow A,$
$A \rightarrow BCD,$	$B \rightarrow ACD,$	$C \rightarrow ABD,$	$D \rightarrow ABC$
$AB \rightarrow CD,$	$AC \rightarrow BD,$	$AD \rightarrow BC,$	$BC \rightarrow AD,$
$BD \rightarrow AC,$	$CD \rightarrow AB,$		

- Se encontrarían las reglas que cumplan con $\text{Confianza} \geq \text{mincon}$

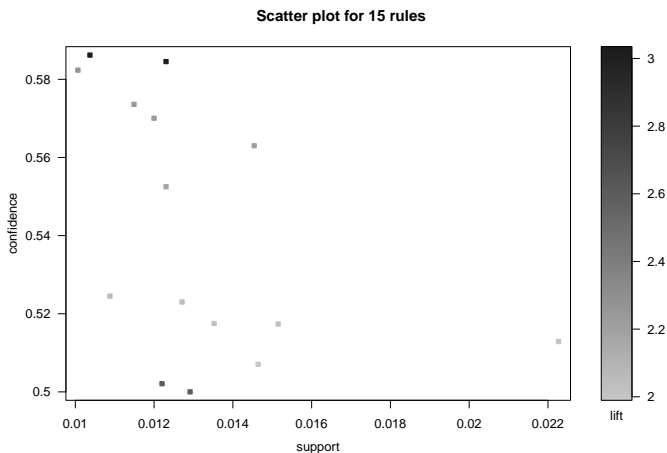
Reglas de Asociación

```
library(arules)
data("Groceries")
summary(Groceries)
labels(Groceries)$items

rules <- apriori(Groceries, parameter =
                  list(supp = 0.01, conf = 0.5, target = "rules"))
inspect(rules)
```


Reglas de Asociación

```
library(arulesViz)  
plot(rules)
```

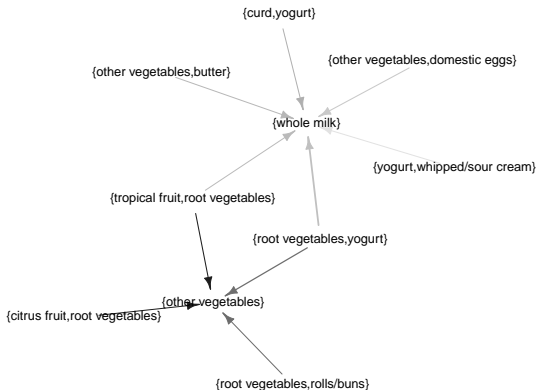


Reglas de Asociación

```
subrules <- head(sort(rules, by="lift"), 10)  
plot(subrules,method="graph",control=list(alpha=1))
```

Graph for 10 rules

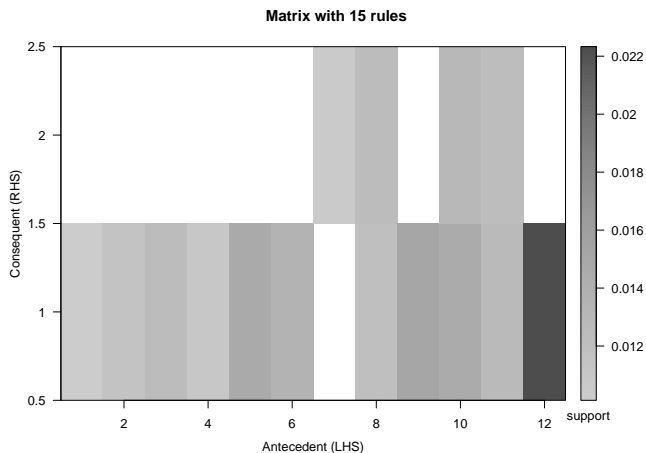
width: support (0.01 – 0.015)
color: lift (2.053 – 3.03)



Reglas de Asociación

```
> plot(rules,method="matrix",measure="support")
Itemsets in Antecedent (LHS)
[1] "{curd,yogurt}"
[2] "{other vegetables,butter}"
[3] "{other vegetables,domestic eggs}"
[4] "{yogurt,whipped/sour cream}"
[5] "{other vegetables,whipped/sour cream}"
[6] "{pip fruit,other vegetables}"
[7] "{citrus fruit,root vegetables}"
[8] "{tropical fruit,root vegetables}"
[9] "{tropical fruit,yogurt}"
[10] "{root vegetables,yogurt}"
[11] "{root vegetables,rolls/buns}"
[12] "{other vegetables,yogurt}"
Itemsets in Consequent (RHS)
[1] "{whole milk}"      "{other vegetables}"
```

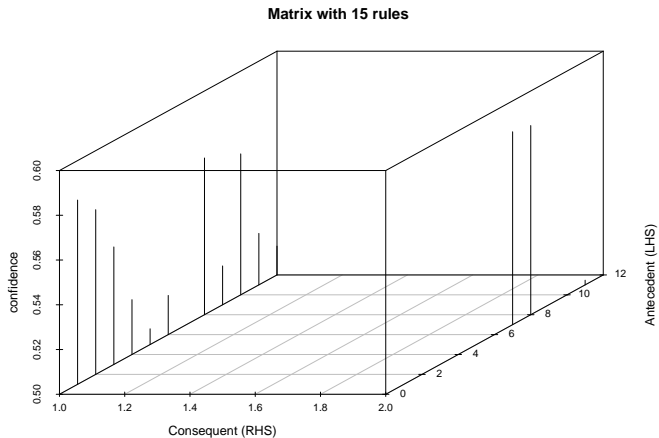
Reglas de Asociación



Reglas de Asociación

```
> plot(rules,method="matrix3D",measure="confidence")
Itemsets in Antecedent (LHS)
[1] "{curd,yogurt}"
[2] "{other vegetables,butter}"
[3] "{other vegetables,domestic eggs}"
[4] "{yogurt,whipped/sour cream}"
[5] "{other vegetables,whipped/sour cream}"
[6] "{pip fruit,other vegetables}"
[7] "{citrus fruit,root vegetables}"
[8] "{tropical fruit,root vegetables}"
[9] "{tropical fruit,yogurt}"
[10] "{root vegetables,yogurt}"
[11] "{root vegetables,rolls/buns}"
[12] "{other vegetables,yogurt}"
Itemsets in Consequent (RHS)
[1] "{whole milk}"          "{other vegetables}"
```

Reglas de Asociación



Reglas de Asociación

- se puede escoger solo algunas reglas
`subrules<-rules[1:2]`
- Con la función `sample` se puede tomar una muestra de las reglas
`subrules<-sample(rules,2)`

Reglas de Asociación

```
library(discretization)
load("bupa.rda")
disc.bupa=chiM(bupa)
dbupa=disc.bupa$Disc.data

for (i in 1:7){dbupa[,i]=as.factor(dbupa[,i])}

dbupa<-as.data.frame(dbupa)
dbupa.ar<-as(dbupa, "transactions")
```


Reglas de Asociación

```
rules <- apriori(dbupa.ar,parameter =  
  list(supp = 0.20, conf = 0.9,target = "rules"))  
summary(rules)  
inspect(rules)  
  
library(arulesViz)  
plot(rules, measure=c("support", "lift"),  
  shading="confidence", interactive=TRUE)  
plot(rules,method="graph",control=list(alpha=1))  
plot(rules, method="graph",  
  control=list(alpha=1,type="items"))
```