



## Medir la utilidad de una regla de asociación

### Lift

El Lift es la medida que permite cuantificar cuál es la utilidad de una regla de asociación. La medida está definida como:

$$\text{Lift} = \frac{\text{Confianza de la regla}}{\text{Proporción a priori del consecuente}}$$

Lo que es lo mismo decir, que el Lift es la confianza de la regla sobre el soporte del consecuente de la misma.

En el clásico ejemplo donde hay 1000 clientes, 200 de ellos han comprado pañales y de esos 200 además hay 50 que también compraron cerveza. La proporción a priori de los que compraron cerveza es  $50/1000 = 5\%$ , mientras que la confianza de la regla es  $50/200 = 25\%$ .

El Lift para la regla de asociación: *Si compra pañales entonces compra cerveza* es:

$$\text{Lift} = \frac{0.25}{0.05} = 5$$

La interpretación de Lift sería, *Los clientes que compran pañales son cinco veces más propensos a comprar cerveza que el resto de los clientes presentes en el conjunto de transacciones.*

Ahora, supongamos que 40 de los 1000 clientes compraron un maquillaje de los caros mientras que de los 200 que compraron pañales solo 5 compraron dicho maquillaje. En este caso el Lift de la regla de asociación: *Si compra pañales entonces compra maquillaje del caro*, es:

$$\text{Lift} = \frac{5/200}{40/1000} = \frac{0.025}{0.04} = 0.625$$

Entonces, los clientes que han comprado pañales en un 62.5% de las veces compran también maquillaje caro en todas las transacciones.

Las reglas de asociación con valores de Lift cercano a uno son consideradas menos interesantes. ¿Por qué ocurre esto?

Partiendo de la definición de Confianza de una regla:



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**  
Noviembre de 2017

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

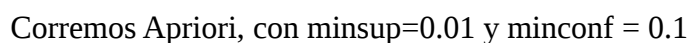
Los eventos A y B serán independientes cuando  $P(A \cap B) = P(A)P(B)$ . Así, que cociente  $P(A \cap B)/P(A)P(B)$  sea cercano a 1 implica que A y B son eventos independientes. Esto significa que el conocimiento de la ocurrencia de A no altera la probabilidad de la ocurrencia de B. Esta relaciones no son útiles desde la perspectiva del data mining y es por eso que buscamos valores de Lift bien diferentes de 1.



Vamos a realizar un análisis de la medida Lift con el dataset *Groceries* existente en el paquete **arules** de R.

```
> Groceries
transactions in sparse format with
 9835 transactions (rows) and
 169 items (columns)
```

```
itemFrequencyPlot(Groceries, support = 0.01, cex.names=0.8, main="MinSup=0.01")
```



```
rules = apriori(Groceries, parameter = list(support = 0.01, confidence = 0.1))
```

Apriori



**Bases de Datos Masivas (11088)**  
**Departamento de Ciencias Básicas**  
Noviembre de 2017

```
Parameter specification:
confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext
0.1 0.1 1 none FALSE TRUE 5 0.01 1 10 rules FALSE

Algorithmic control:
filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [435 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
```

```
Rules <- sort(rules, by="confidence", decreasing=TRUE)

inspect(head(rules, n = 4, by = "lift"))
```

	lhs	rhs	support	confidence	lift
[1]	{citrus fruit,root vegetables}	=> {other vegetables}	0.010	0.59	3.0
[2]	{tropical fruit,root vegetables}	=> {other vegetables}	0.012	0.58	3.0
[3]	{curd,yogurt}	=> {whole milk}	0.010	0.58	2.3
[4]	{other vegetables,butter}	=> {whole milk}	0.011	0.57	2.2

La primera regla *{citrus fruit,root vegetables} => {other vegetables}* tiene un soporte de 1% (98 itemsets) y la confianza de la regla es aceptable 0.59. El lift en este caso nos dice que las personas que compran cítricos, papas, batatas o zanahorias son tres veces más propensos a comprar otros vegetales.

Por otro lado, en el caso de la regla 3 *{curd,yogurt} => {whole milk}* los que queso de cabra y yogur son dos veces más propensos a comprar leche entera.

## Referencias

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons. (pág. 259)

Hasher, M., Hornik, K., Grun, B., & Buchta, C. Introduction to arules—A computational environment for mining association rules and frequent item sets, 2007.