

# ML – Association Rules - Exercise01

Daniel Ferreira Zanchetta y Lais Silva Almeida Zanchetta

19/02/2020

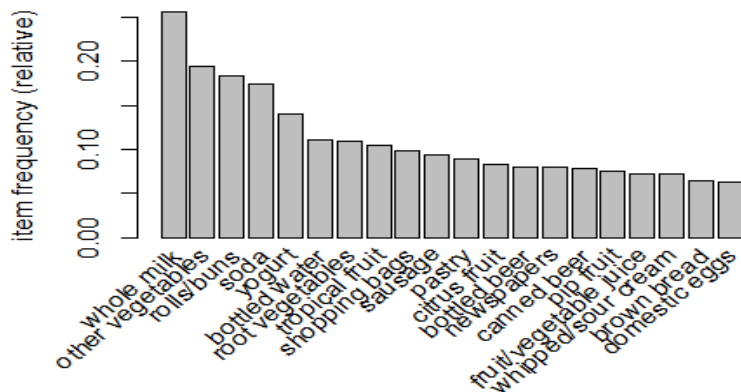
## Resumen de las Reglas de Asociación interesantes encontradas

- 1) {bottled beer,red/blush wine} => {liquor}
- 2) {tropical fruit,whole milk,yogurt} => {root vegetables}
- 3) {onions,whole milk} => {root vegetables}
- 4) {beef,rolls/buns} => {root vegetables}
- 5) {hamburger meat,yogurt,whipped/sour cream} => {butter}
- 6) {pip fruit,whole milk,yogurt,frozen meals} => {tropical fruit}
- 7) {root vegetables,whole milk,flour} => {whipped/sour cream}
- 8) {sausage} => {bread and backed goods}
- 9) {bread and backed goods,fruit} => {dairy produce}
- 10) {sausage} => {bread and backed goods}

## Versión más detallada del paso a paso

Lo primero que hemos hecho después de cargar la transacción de Groceries, ha sido aplicar un Summary y allí descubrimos que los ítems que estaban siendo presentados con más frecuencia eran: whole milk, other vegetables, rolls/buns, soda e yogurt.

Para demostrar esta frecuencia de determinados ítems, hemos hecho un plot de frecuencia de los 20 ítems, que sale en orden descendente



```
groceries.rules <- apriori(Groceries, parameter = list(support=0.005,
confidence=0.3,minlen=2),control=list(verbose=F))
groceries.rules.sortedlift <- sort(groceries.rules,by="lift")
groceries.rules.sortedlift <-
groceries.rules.sortedlift[!is.redundant(groceries.rules.sortedlift)]
inspect(groceries.rules.sortedlift[1:10])
```

En un primer intento para encontrar las reglas, hemos llegado a algunas reglas con un valor de lift bastante significativo, una que nos llamó la atención por su lift de 35.71 ha sido de {bottled beer,red/blush wine} => {liquor}, con la cual añadimos a nuestra de reglas interesantes. Para llegar hasta este valor, utilizamos support de 0.001 y confidence de 0.3. Sin embargo, en general, que no nos han interesado demasiado pues han parecido reglas un poco obvias. Mismo cuando cambiamos el support a

0.005. Un ejemplo en concreto es de las reglas [2] {herbs} => {root vegetables}. También nos ha dado la impresión que hay valores comunes que podrían ser quitados del algoritmo apriori (“other vegetables” y “shopping bags”), con la intención de intentar obtener reglas mas interesantes.

```
groceries.rules.filtered <- apriori(Groceries,  
parameter=list(support=0.001,confidence =  
0.55,minlen=4,maxlen=5),appearance = list(none=c("other  
vegetables","shopping bags")),control=list(verbose=F))  
groceries.rules.filtered.sorted <-  
sort(groceries.rules.filtered,by="lift")  
groceries.rules.filtered.sorted <-  
groceries.rules.filtered.sorted[!is.redundant(groceries.rules.filtered.so  
rted)]  
inspect(groceries.rules.filtered.sorted[1:10])
```

Por lo tanto, ha sido esto que hicimos. Primero intentamos jugar con un support de 0.001, confidence a 0.55, minimo de LHS a 4 y maximo a 5. Nos ha sorprendido la regla {hamburger meat,yogurt,whipped/sour cream} => {butter}, principalmente por su lift de 11.27.

Enseguida, cambiamos support a 0.005, confidence a 0.3, y minimo de LHS a 2. En el resultado ya empezaran a venir en repetidas combinaciones “whole milk” en los LHS, por lo tanto lo que nos parece un ítem más general. También hemos percibido RHS de “root vegetables” con un lift bastante elevado. Entonces, lo siguiente que hicimos ha sido centrarnos en los “root vegetables”:

```
root.vegetables.rule <- sort(subset(groceries.rules.filtered.sorted, rhs  
%in% "root vegetables"), by="lift")  
inspect(root.vegetables.rule)
```

Con el resultado seleccionamos algunas reglas que han parecido interesantes: {tropical fruit,whole milk,yogurt} => {root vegetables} -> esta parece involucrar muchos items de quien va para hacer compras para una familia {onions,whole milk} => {root vegetables} {beef,rolls/buns} => {root vegetables}

Lo siguiente que hicimos ha sido de continuar en el focalizar en algunos RHS que parecían interesantes, como “tropical fruit”:

```
tropical.fruit.rule <- sort(subset(groceries.rules.filtered.sorted, rhs  
%in% "tropical fruit"), by="lift")  
inspect(head(tropical.fruit.rule,5))
```

Del resultado consideramos: {pip fruit,whole milk,yogurt,frozen meals} => {tropical fruit} Ha sido ahora la vez de mirar el RHS yogurt, y vemos por primera vez en los LHS “pork”, no obstante no es una regla que nos ha interesado de todo: {pork,tropical fruit,fruit/vegetable juice} => {yogurt}

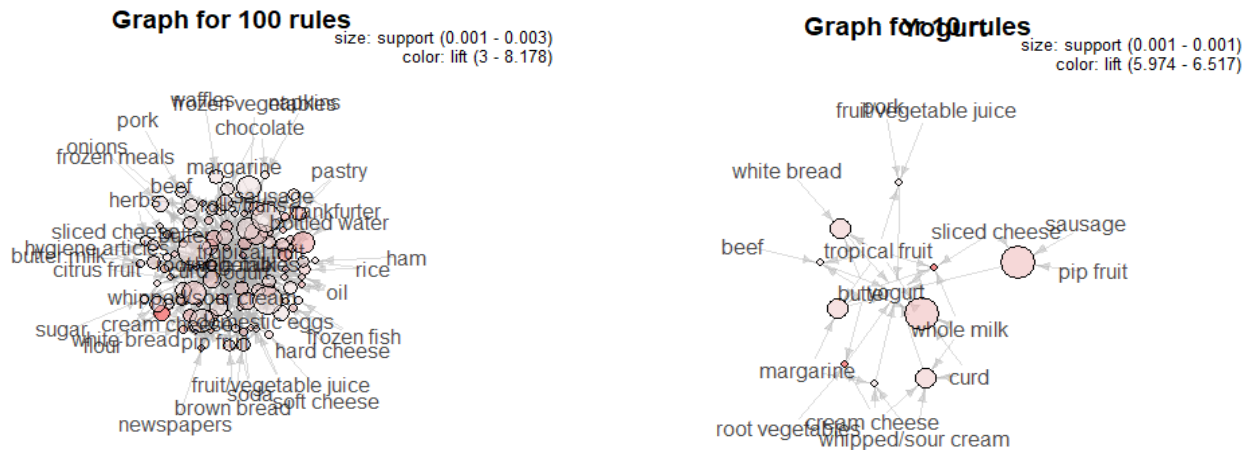
```
yogurt.rule <- sort(subset(groceries.rules.filtered.sorted, rhs %in%  
"yogurt"), by="lift")  
inspect(head(yogurt.rule,10))  
  
root.vegetables.rule <- sort(subset(groceries.rules.filtered.sorted, lhs  
%in% "root vegetables" & lift > 3), by="lift")  
inspect(root.vegetables.rule[1:10])
```

El siguiente que hicimos ha sido de regenerar reglas, pero ahora considerando un LHS que había salido en otras reglas: “root vegetables”. Quedamos con esta regla de interesante: {root vegetables,whole milk,flour} => {whipped/sour cream}.

Abajo analizamos algunos gráficos con algunos ítems que vimos en la análisis. Con los parámetros utilizados, el gráfico de las reglas de Root Vegetables no ha sido muy eficiente de analizar. No obstante, el análisis en las reglas de Yogurt queda evidente la implicación de Pip Fruit y Whole Milk.

```
#plot for root vegetables
plot(root.vegetables.rule,method="graph",interactive=FALSE,shading="lift"
)

plot(yogurt.rule[1:10],method="graph",interactive=FALSE,shading="lift")
```



También hemos agregado los ítems de groceries por level2, que nos es dado con grupos como por ejemplo: sausage, vegetables, poultry, beef, fruit, fish, cheese dairy produce.

```
#Aggregating by Level2, found in itemInfo
#itemInfo(Groceries)
groceries_level_2 <- aggregate(Groceries, by="level2")
groceries.rules.level2 <-
apriori(groceries_level_2,parameter=list(support=0.07,
confidence=0.5,minlen=2))

inspect(groceries.rules.level2)
```

La regla {sausage} => {bread and backed goods} nos ha parecido interesante. Hemos notado que con un support de 0.07 y confidence de 0.5 los valores de lift van más próximos a 1 en relación a los valores que habíamos encontrados en las ejecuciones anteriores. Aun sobre los mismos parámetros, la siguiente regla pueden ser de personas que les gusten a desayunos variados: {bread and backed goods,fruit} => {dairy produce}

```
groceries_level_3 <- addAggregate(Groceries, by="level2")
groceries.rules.level3 <-
apriori(groceries_level_3,parameter=list(support=0.1,
confidence=0.3,minlen=2))

multirules <- filterAggregate(groceries.rules.level3)
inspect(head(multirules, by = "lift"))
```

Por fin, intentamos a través de la análisis Multi-level, para analizar la relación de ítems individuales con ítems de grupo, y considerando un support de 0.1 y confidence de 0.3. La regla que nos ha parecido interesado ha sido la {sausage} => {bread and backed goods}.