reglas de asociacion

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ # Paso 1: Obtener y procesar la base de datos #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

library(arules)

## Loading required package: Matrix

##   
## Attaching package: 'arules'

## The following objects are masked from 'package:base':  
##   
## abbreviate, write

dvds <- read.transactions("dvdtrans-modificado.csv",   
 sep = ",",  
 format = "basket", header = TRUE)

dvds

## transactions in sparse format with  
## 9 transactions (rows) and  
## 10 items (columns)

summary(dvds)

## transactions as itemMatrix in sparse format with  
## 9 rows (elements/itemsets/transactions) and  
## 10 columns (items) and a density of 0.2777778   
##   
## most frequent items:  
## Gladiator Patriot Sixth Sense Braveheart Green Mile (Other)   
## 7 6 5 1 1 5   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 2 3 4   
## 3 5 1   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 2.000 3.000 2.778 3.000 4.000   
##   
## includes extended item information - examples:  
## labels  
## 1 Braveheart  
## 2 Gladiator  
## 3 Green Mile

dim(dvds)

## [1] 9 10

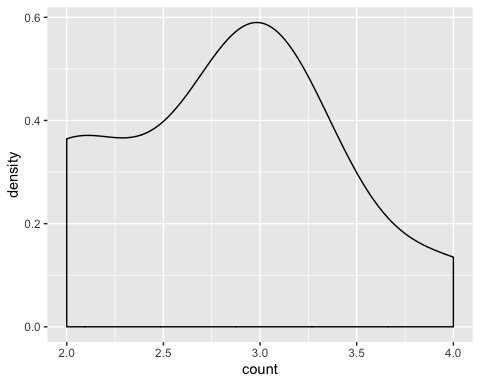
basketSizes <- size(dvds)  
basketSizes

## [1] 3 2 3 3 3 2 2 3 4

summary(basketSizes)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 2.000 3.000 2.778 3.000 4.000

library(ggplot2)  
ggplot(data.frame(count = basketSizes)) +  
 geom\_density(aes(x=count))



# Mostrar las transacciones  
labels(dvds) # me das las etiqutas de las transacciones

## [1] "{Braveheart,Gladiator,Patriot}"   
## [2] "{LOTR1,LOTR2}"   
## [3] "{Gladiator,Patriot,Sixth Sense}"   
## [4] "{Gladiator,Patriot,Sixth Sense}"   
## [5] "{Gladiator,Patriot,Sixth Sense}"   
## [6] "{Harry Potter1,Harry Potter2}"   
## [7] "{Gladiator,Patriot}"   
## [8] "{Gladiator,Patriot,Sixth Sense}"   
## [9] "{Gladiator,Green Mile,LOTR,Sixth Sense}"

colnames(dvds)[1:5] # me devuelve el nombre de los items

## [1] "Braveheart" "Gladiator" "Green Mile" "Harry Potter1"  
## [5] "Harry Potter2"

# Mostrar un subconjunto de transacciones (p. ej. las cinco primeras)  
inspect(dvds[1:5])

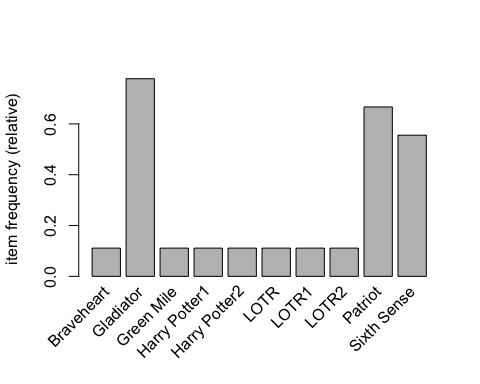
## items   
## [1] {Braveheart,Gladiator,Patriot}   
## [2] {LOTR1,LOTR2}   
## [3] {Gladiator,Patriot,Sixth Sense}  
## [4] {Gladiator,Patriot,Sixth Sense}  
## [5] {Gladiator,Patriot,Sixth Sense}

El soporte es la proporcion de los itemsets

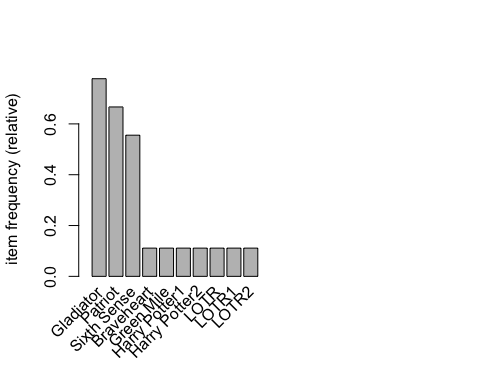
# Mostrar el soporte (proporción de transacciones) de un item (p. ej. de los tres primeros)  
itemFrequency(dvds)

## Braveheart Gladiator Green Mile Harry Potter1 Harry Potter2   
## 0.1111111 0.7777778 0.1111111 0.1111111 0.1111111   
## LOTR LOTR1 LOTR2 Patriot Sixth Sense   
## 0.1111111 0.1111111 0.1111111 0.6666667 0.5555556

# Visualizar el soporte de los items (p. ej. de aquellos items con una proporcion mayor a 0.15)  
itemFrequencyPlot(dvds, support = 0.1)



# Visualizar el soporte de los items (p. ej.de los 20 ?tems con mayor soporte)  
itemFrequencyPlot(dvds, topN = 20)

 #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ # Paso 2: Entrenar el modelo con los datos #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dvdsrules <- apriori(dvds, parameter = list(support = 0.40,  
 confidence = 0.25, minlen = 2))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.25 0.1 1 none FALSE TRUE 5 0.4 2  
## maxlen target ext  
## 10 rules FALSE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 3   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[10 item(s), 9 transaction(s)] done [0.00s].  
## sorting and recoding items ... [3 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 done [0.00s].  
## writing ... [9 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

dvdsrules

## set of 9 rules

ay que ver si los parametros son adecuados

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
# Paso 3: Evaluar el modelo  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
summary(dvdsrules)

## set of 9 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 2 3   
## 6 3   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 2.000 2.000 2.333 3.000 3.000   
##   
## summary of quality measures:  
## support confidence lift count   
## Min. :0.4444 Min. :0.6667 Min. :1.200 Min. :4.000   
## 1st Qu.:0.4444 1st Qu.:0.7143 1st Qu.:1.200 1st Qu.:4.000   
## Median :0.4444 Median :0.8000 Median :1.286 Median :4.000   
## Mean :0.5185 Mean :0.8339 Mean :1.248 Mean :4.667   
## 3rd Qu.:0.5556 3rd Qu.:1.0000 3rd Qu.:1.286 3rd Qu.:5.000   
## Max. :0.6667 Max. :1.0000 Max. :1.286 Max. :6.000   
##   
## mining info:  
## data ntransactions support confidence  
## dvds 9 0.4 0.25

# Mostrar las tres primeras reglas de asociacion  
inspect(dvdsrules[1:9])

## lhs rhs support confidence lift   
## [1] {Sixth Sense} => {Patriot} 0.4444444 0.8000000 1.200000  
## [2] {Patriot} => {Sixth Sense} 0.4444444 0.6666667 1.200000  
## [3] {Sixth Sense} => {Gladiator} 0.5555556 1.0000000 1.285714  
## [4] {Gladiator} => {Sixth Sense} 0.5555556 0.7142857 1.285714  
## [5] {Patriot} => {Gladiator} 0.6666667 1.0000000 1.285714  
## [6] {Gladiator} => {Patriot} 0.6666667 0.8571429 1.285714  
## [7] {Patriot,Sixth Sense} => {Gladiator} 0.4444444 1.0000000 1.285714  
## [8] {Gladiator,Sixth Sense} => {Patriot} 0.4444444 0.8000000 1.200000  
## [9] {Gladiator,Patriot} => {Sixth Sense} 0.4444444 0.6666667 1.200000  
## count  
## [1] 4   
## [2] 4   
## [3] 5   
## [4] 5   
## [5] 6   
## [6] 6   
## [7] 4   
## [8] 4   
## [9] 4

# Subconjuntos de reglas  
rules <- subset(dvdsrules, items %in% "Sixth Sense")  
inspect(rules)

## lhs rhs support confidence lift   
## [1] {Sixth Sense} => {Patriot} 0.4444444 0.8000000 1.200000  
## [2] {Patriot} => {Sixth Sense} 0.4444444 0.6666667 1.200000  
## [3] {Sixth Sense} => {Gladiator} 0.5555556 1.0000000 1.285714  
## [4] {Gladiator} => {Sixth Sense} 0.5555556 0.7142857 1.285714  
## [5] {Patriot,Sixth Sense} => {Gladiator} 0.4444444 1.0000000 1.285714  
## [6] {Gladiator,Sixth Sense} => {Patriot} 0.4444444 0.8000000 1.200000  
## [7] {Gladiator,Patriot} => {Sixth Sense} 0.4444444 0.6666667 1.200000  
## count  
## [1] 4   
## [2] 4   
## [3] 5   
## [4] 5   
## [5] 4   
## [6] 4   
## [7] 4

# Subconjuntos de reglas  
rules <- subset(dvdsrules, lhs %in% "Sixth Sense")  
inspect(rules)

## lhs rhs support confidence lift   
## [1] {Sixth Sense} => {Patriot} 0.4444444 0.8 1.200000  
## [2] {Sixth Sense} => {Gladiator} 0.5555556 1.0 1.285714  
## [3] {Patriot,Sixth Sense} => {Gladiator} 0.4444444 1.0 1.285714  
## [4] {Gladiator,Sixth Sense} => {Patriot} 0.4444444 0.8 1.200000  
## count  
## [1] 4   
## [2] 5   
## [3] 4   
## [4] 4

## Se observa que a las personas que han visto “Sixth Sense” , se les recomienda ver las peliculas “Patriot” y “Gladiator”