PEC 5 - Reglas de Asociación

Fernando Antonio Barbeiro Campos - [fbarbeiro@uoc.edu](mailto:fbarbeiro@uoc.edu)

3 de May, 2018

Table of Contents

my\_file <- "lastfm.csv"  
cat("\n", file = my\_file, append = TRUE)  
tdata <- read.transactions(file = my\_file, rm.duplicates = FALSE, skip = 1, sep = ",")  
head(tdata)

## transactions in sparse format with  
## 6 transactions (rows) and  
## 16163 items (columns)

class(tdata)

## [1] "transactions"  
## attr(,"package")  
## [1] "arules"

inspect(head(tdata))

## items   
## [1] {1,f,Germany,red hot chili peppers}   
## [2] {1,f,Germany,the black dahlia murder}  
## [3] {1,f,Germany,goldfrapp}   
## [4] {1,dropkick murphys,f,Germany}   
## [5] {1,f,Germany,le tigre}   
## [6] {1,f,Germany,schandmaul}

size(head(tdata))

## [1] 4 4 4 4 4 4

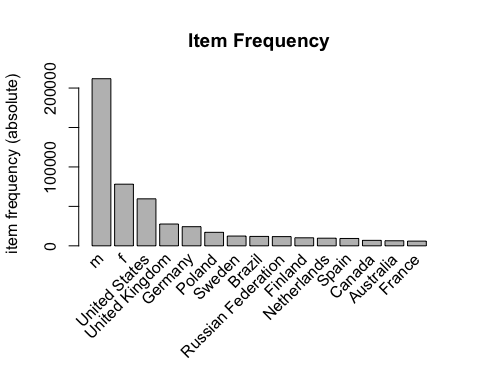
frequentItems <- eclat (tdata, parameter = list(supp = 0.07, maxlen = 15))

## Eclat  
##   
## parameter specification:  
## tidLists support minlen maxlen target ext  
## FALSE 0.07 1 15 frequent itemsets FALSE  
##   
## algorithmic control:  
## sparse sort verbose  
## 7 -2 TRUE  
##   
## Absolute minimum support count: 20297   
##   
## create itemset ...   
## set transactions ...[16163 item(s), 289958 transaction(s)] done [0.09s].  
## sorting and recoding items ... [5 item(s)] done [0.00s].  
## creating bit matrix ... [5 row(s), 289958 column(s)] done [0.02s].  
## writing ... [7 set(s)] done [0.00s].  
## Creating S4 object ... done [0.00s].

inspect(frequentItems)

## items support count   
## [1] {m,United Kingdom} 0.07730430 22415  
## [2] {m,United States} 0.13852696 40167  
## [3] {m} 0.73052994 211823  
## [4] {f} 0.26945971 78132  
## [5] {United States} 0.20540216 59558  
## [6] {United Kingdom} 0.09531725 27638  
## [7] {Germany} 0.08363625 24251

itemFrequencyPlot(tdata, topN=15, type="absolute", main="Item Frequency")



rules <- apriori (tdata, parameter = list(supp = 0.001, conf = 0.5))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.5 0.1 1 none FALSE TRUE 5 0.001 1  
## 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: 289   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[16163 item(s), 289958 transaction(s)] done [0.09s].  
## sorting and recoding items ... [360 item(s)] done [0.01s].  
## creating transaction tree ... done [0.11s].  
## checking subsets of size 1 2 3 done [0.00s].  
## writing ... [236 rule(s)] done [0.00s].  
## creating S4 object ... done [0.04s].

rules\_conf <- sort (rules, by="confidence", decreasing=TRUE)  
inspect(head(rules\_conf))

## lhs rhs support confidence lift   
## [1] {India} => {m} 0.002196870 0.9578947 1.311233  
## [2] {nas} => {m} 0.001203623 0.9356568 1.280792  
## [3] {slayer} => {m} 0.001596783 0.9186508 1.257513  
## [4] {2pac} => {m} 0.001058774 0.9002933 1.232384  
## [5] {dire straits} => {m} 0.001272598 0.9000000 1.231982  
## [6] {rage against the machine} => {m} 0.002445182 0.8963338 1.226964  
## count  
## [1] 637   
## [2] 349   
## [3] 463   
## [4] 307   
## [5] 369   
## [6] 709

rules\_lift <- sort (rules, by="lift", decreasing=TRUE)   
inspect(head(rules\_lift))

## lhs rhs support confidence lift   
## [1] {modest mouse} => {United States} 0.001427793 0.5328185 2.594026  
## [2] {regina spektor} => {f} 0.001096711 0.5417376 2.010459  
## [3] {paramore} => {f} 0.001189827 0.5051245 1.874582  
## [4] {britney spears} => {f} 0.001196725 0.5036284 1.869031  
## [5] {India} => {m} 0.002196870 0.9578947 1.311233  
## [6] {nas} => {m} 0.001203623 0.9356568 1.280792  
## count  
## [1] 414   
## [2] 318   
## [3] 345   
## [4] 347   
## [5] 637   
## [6] 349