Topic : Data mining – R - association rules and apriori algorithm
Author : Ming-Chang Lee
Date : 2009.03.30

```
# Topic  : Data mining - association rules and apriori algorithm
# Author : Ming-Chang Lee
# Date   : 2009.03.30
# Note   : 資料來源, 參考 northwind.mdb 北風資料庫

## step 1. load arules package
library(arules)

## step 2. Import and transform data

workpath <- "C:/temp"
setwd(workpath)     # Set working directory. Make c:\temp directory and
copy csv file to the directory
getwd()            # Get working directory

# import raw data (csv file)
nw_data <- read.table(file="northwind_trans.csv", header = TRUE, sep
= "," )
class(nw_data)      # "data.frame"
mode(nw_data)       # "list"

# group data by orderID. using column 1(OrderID) and 2(ProductName) only
nw_temp <- tapply(nw_data[,2], nw_data[,1], function(x) paste(x))
nw_temp[1]          # get the first row data
names(nw_temp[1])   # get the name of first row data
nw_temp[[1]]        # get the element of first row data
class(nw_temp)      # "array"
mode(nw_temp)       # "list"

nw2 <- vector("list", length(nw_temp)) # length(nw_temp)=830
for (i in seq(nw_temp)) names(nw2)[i] <- names(nw_temp[i])
for (i in seq(nw_temp)) nw2[[i]] <- nw_temp[[i]]
class(nw2)          # "list"
mode(nw2)           # "list"
```

```
# force data into transactions
nw <- as(nw2, "transactions")
class(nw)          # "transactions"
mode(nw)           # "S4"

# step 3. analyze data
# generate level plots to visually inspect binary incidence matrices
image(nw) # result - Figure 1 Level plot
summary(nw)

# step 4. find 1-items (L1)
# provides the generic function itemFrequency and the frequency/support
for all single items in an objects based on itemMatrix.
itemFrequency(nw, type = "relative") # default: "relative"
itemFrequency(nw, type = "absolute")

# step 5.
# create an item frequency bar plot for inspecting the item frequency
distribution for objects based on itemMatrix
itemFrequencyPlot(nw) # result- Figure 2 Item frequency bar plot

# step 6.
# mine association rules
# rules <- apriori(nw) # Mine association rules using default Apriori
algorithm
rules1 <- apriori(nw, parameter = list(supp = 0.05, target = "maximally
frequent itemsets")) # set parameters
rules2 <- apriori(nw, parameter = list(supp = 0.001, conf = 0.8, target
= "rules")) # set parameters

# step7.
# display results
inspect(nw)       # display transactions
inspect(rules1)    # display maximally frequent itemset
inspect(rules2[1])  # display association

# reference:
```

```
# Data Mining: Han, J. and Kamber, M. (2006) Concepts and Techniques,
Second Edition, Morgan Kaufmann.
# http://r-forge.r-project.org/projects/arules
# end
```

分析結果:

```
> inspect(rules1)    # display maximally frequent itemset
   items       support
1  {item062} 0.05783133
2  {item056} 0.06024096
3  {item075} 0.05542169
4  {item059} 0.06506024
5  {item041} 0.05662651
6  {item016} 0.05180723
7  {item031} 0.06144578
8  {item024} 0.06265060
9  {item002} 0.05421687
10 {item060} 0.06144578
> inspect(rules2[1])  # display association
  lhs         rhs          support confidence     lift
1 {item048,
   item061} => {item043} 0.001204819         1 29.64286
>
```

item048, item061 => item043 {support=0.12%, confidence=1}

Note: 僅列出第一個 rule

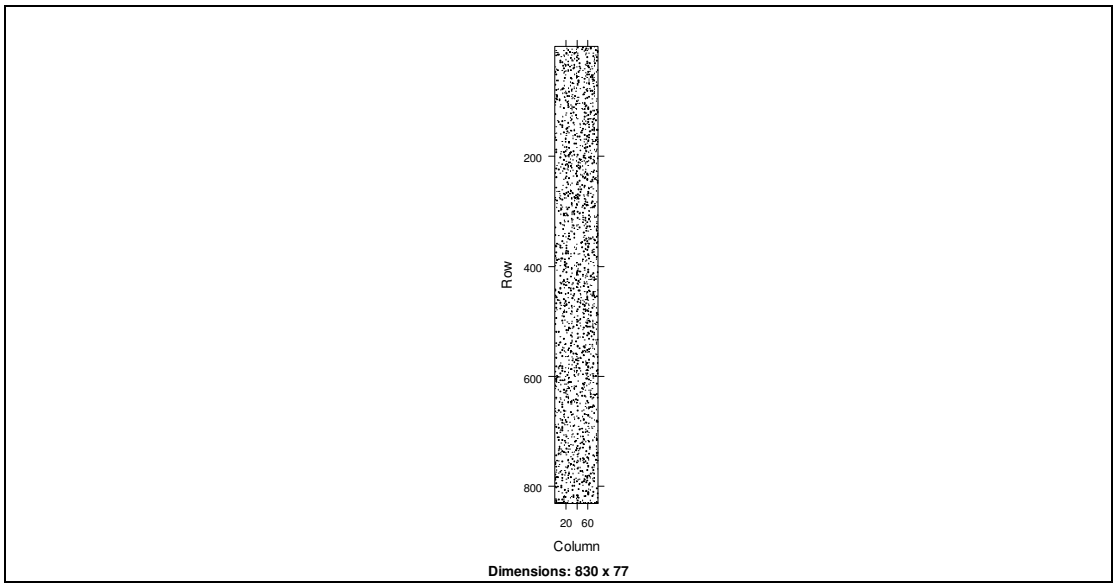

Figure 1 Level plot

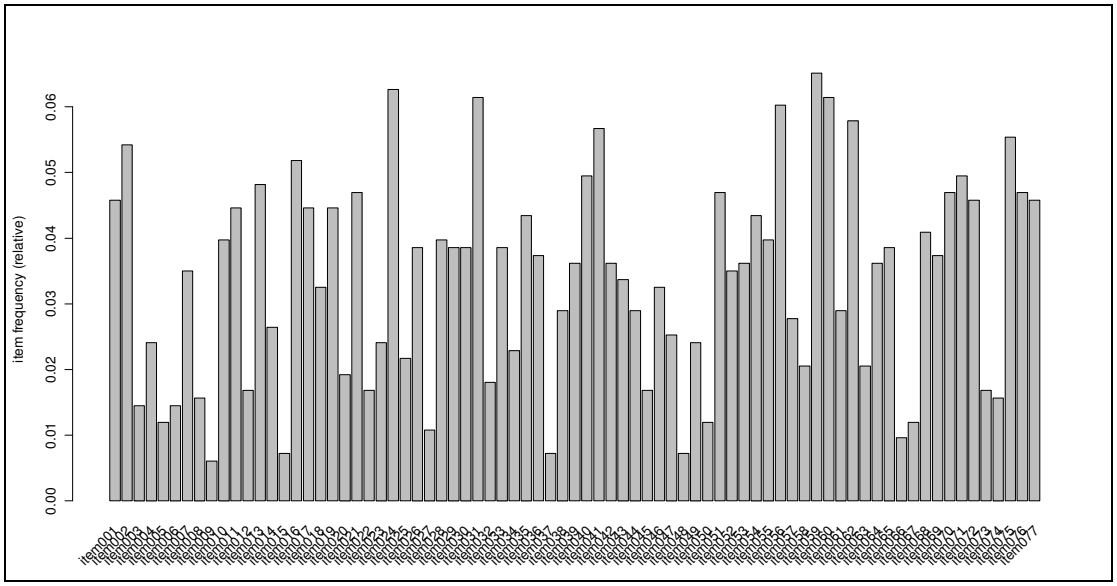

Figure 2 Item frequency bar plot

Reference

1. Data Mining: Han, J. and Kamber, M. (2006) Concepts and Techniques, Second Edition, Morgan Kaufmann.
2. R – arules Package, http://r-forge.r-project.org/projects/arules