**Source Code output for Association Rules**

**Yu Deng**

library(plyr)

library(tidyverse)

library(arules)

library(lubridate)

library(arulesViz)

x=read.csv('./Desktop/test.csv')

x=x[complete.cases(x),]

#100 best sellers

tmp = x %>%

group\_by(sku, description) %>%

summarize(count = sum(quantity)) %>%

arrange(desc(count))

tmp %>%

ggplot(aes(x=reorder(description,count), y=count))+

geom\_bar(stat="identity",fill="indian red")+

coord\_flip() + labs(x='sku\_code')

head(tmp)

# A tibble: 6 x 2

sku count

<int> <int>

1 9861016 2389

2 9860713 1184

3 9860923 1004

4 9860515 337

5 9860696 290

6 9859890 283

#x\_sorted <- x[order(x$trans\_id),]

#get the dataset for association rule input

itemList <- ddply(x,c("trans\_id"),

function(df1)paste(df1$description,

collapse = ","))

colnames(itemList)[2] <- c("items")

write.csv(itemList,"market\_basket.csv", quote = FALSE, row.names = FALSE, col.names = FALSE)

#delete the ',' at the end of table

tr <- read.transactions('market\_basket.csv', format = 'basket', sep = ',')

summary(tr)

transactions as itemMatrix in sparse format with

232 rows (elements/itemsets/transactions) and

336 columns (items) and a density of 0.02864583

most frequent items:

9861016 9859975 9860180 9860515 9859902 (Other)

159 85 77 77 71 1764

element (itemset/transaction) length distribution:

sizes

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 27 28 29 30 32 33 34 35

94 28 11 8 6 5 5 3 4 3 3 3 3 2 2 2 2 3 2 2 3 4 1 2 3 2 2 3 1 3 4 2

36 37 41 42 44

3 4 2 1 1

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.000 2.000 3.000 9.625 14.000 44.000

includes extended item information - examples:

labels

1 100

2 1000

3 10000

rules <- apriori(tr, parameter = list(supp=0.001, conf=0.8))

rules <- sort(rules, by='confidence', decreasing = TRUE)

summary(rules)

set of 114987 rules

rule length distribution (lhs + rhs):sizes

2 3 4 5 6 7 8 9 10

91 1541 7144 17943 28100 28976 19888 8892 2412

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.00 6.00 7.00 6.59 8.00 10.00

summary of quality measures:

support confidence lift

Min. :0.1034 Min. :0.8000 Min. :1.235

1st Qu.:0.1034 1st Qu.:0.9062 1st Qu.:3.013

Median :0.1121 Median :0.9630 Median :3.364

Mean :0.1156 Mean :0.9483 Mean :3.337

3rd Qu.:0.1207 3rd Qu.:1.0000 3rd Qu.:3.932

Max. :0.3233 Max. :1.0000 Max. :5.043

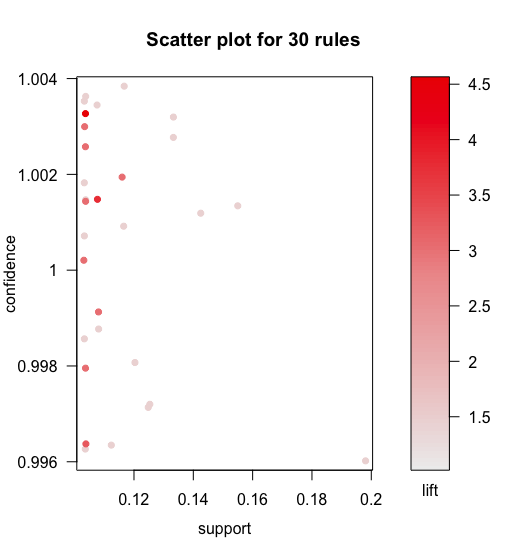
mining info:

data ntransactions support confidence

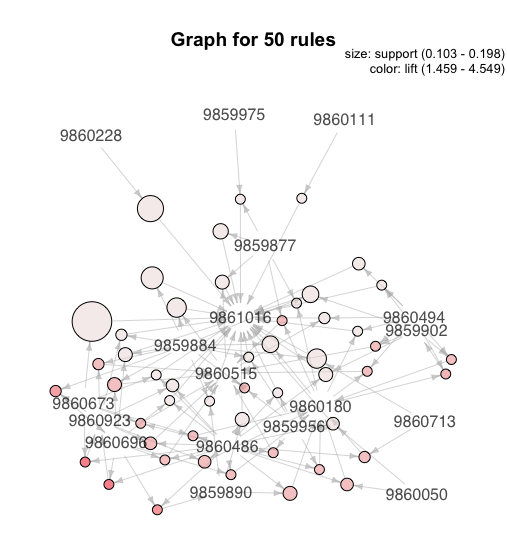
tr 232 0.1 0.8

topRules <- rules[1:30]

plot(topRules)



plot(topRules, method="graph")



plot(topRules, method = "grouped")

