Untitled

Akhil Gupta

May 13, 2017

library(shiny)  
library(dummies)

## dummies-1.5.6 provided by Decision Patterns

#library(ggplot2)   
library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

#library(MASS)   
#library(tree)   
library(ISLR)   
#library(cluster)   
#library(fpc)   
library(arules)

## Loading required package: Matrix

##   
## Attaching package: 'arules'

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

library(arulesViz)

## Loading required package: grid

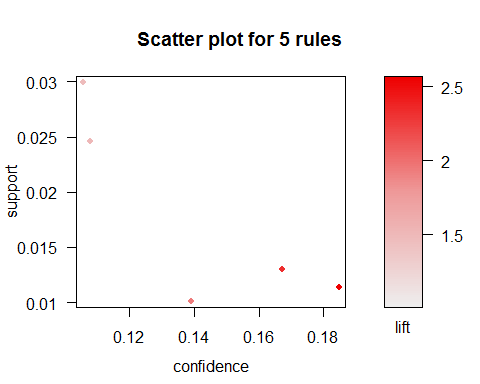
#library(GGally)  
#library(mice) #helps to fill values which are blank  
  
getwd();

## [1] "C:/Users/gupta/DataMiningProject/Akhil"

setwd("C:/Users/gupta/DataMiningProject")  
data<-read.csv(file="app\_metadata\_cleaned\_removed\_min\_downloads\_above\_5m.csv", header=TRUE) #Reading the dataset  
  
num.vars <- sapply(data, is.numeric)  
  
Data <- data  
  
  
#setting the variables to null so that they do not come in our analysis. These variables are useless for our computations.  
Data$APP\_ID <- NULL  
Data$APP\_NAME <- NULL  
Data$DOWNLOADS <- NULL  
Data$CURRENT\_VERSION <- NULL  
Data$LASTUPDATED <- NULL  
Data$DEVELOPER\_SITE <- NULL  
Data$DEVELOPER\_CONTACT <- NULL  
Data$DEVELOPER\_NAME <- NULL  
Data$MIN\_REQUIRED\_ANDROID <- NULL  
  
Data$MIN\_REQ\_ANDROID\_FIRST <- as.factor(Data$MIN\_REQ\_ANDROID\_FIRST) #declare variable as a factor  
Data1<-Data[c("CATEGORY","CONTENT\_RATING","MIN\_REQ\_ANDROID\_FIRST","spam")] #for the association rules we will consider the following  
Data2<-dummy.data.frame(Data1, names=c("")) #Create dummy variables  
Data3<-data.frame(sapply(Data1, as.numeric)) #use the variables as numericals  
Data1$MIN\_REQ\_ANDROID\_FIRST <- as.factor(Data$MIN\_REQ\_ANDROID\_FIRST)  
Data1$CATEGORY <- as.factor(Data$CATEGORY)  
Data1$CONTENT\_RATING <- as.factor(Data$CONTENT\_RATING)  
Data1$spam <- as.factor(Data$spam)  
###association rules applied ###  
rules<-apriori(Data1, parameter=list(supp=.01, conf=.1),appearance =list(default="lhs",rhs="spam=1"), control=list(verbose=F))  
  
rules<-sort(rules ,decreasing=TRUE,by="confidence")   
inspect(rules[1:5]) #inspect the top 5 rules which were sorted by confidence

## lhs rhs support confidence lift  
## [1] {CATEGORY=Entertainment,   
## MIN\_REQ\_ANDROID\_FIRST=2} => {spam=1} 0.01138743 0.1847134 2.561180  
## [2] {CATEGORY=Entertainment} => {spam=1} 0.01295812 0.1672297 2.318757  
## [3] {CATEGORY=Personalisation,   
## CONTENT\_RATING=LowMaturity} => {spam=1} 0.01007853 0.1392405 1.930667  
## [4] {CONTENT\_RATING=LowMaturity,   
## MIN\_REQ\_ANDROID\_FIRST=2} => {spam=1} 0.02460733 0.1081081 1.498994  
## [5] {CONTENT\_RATING=LowMaturity} => {spam=1} 0.02997382 0.1057248 1.465949

library(arulesViz)  
plot(rules,measure=c("confidence","support"),shading="lift") #Plot the cofidence vs support and shade the life to get a good idea of where the rules range is



plot(rules, method="matrix3D" ,measure="lift") #this is a 3d representtion of all the antecedents and consequents with lift shown

## Itemsets in Antecedent (LHS)  
## [1] "{CATEGORY=Entertainment,MIN\_REQ\_ANDROID\_FIRST=2}"   
## [2] "{CATEGORY=Entertainment}"   
## [3] "{CATEGORY=Personalisation,CONTENT\_RATING=LowMaturity}"  
## [4] "{CONTENT\_RATING=LowMaturity,MIN\_REQ\_ANDROID\_FIRST=2}"   
## [5] "{CONTENT\_RATING=LowMaturity}"   
## Itemsets in Consequent (RHS)  
## [1] "{spam=1}"

