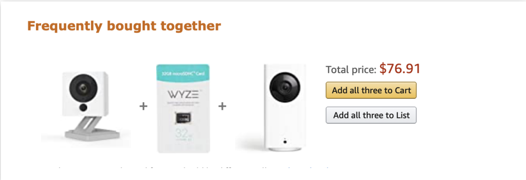
associaltion rules

note:plant-based food data sets as example

1. Introduction

> "Customers who bought this item also bought....."
> after you finish payment, often you can see "you might also like items....."
> after you watched a movie online, "you might also like movies...."



"Recommendation" is ubiquitous in daily life, have you ever thought of why these "recommendations" come to you? what behind "recommendations" is "association rules". websites applied various association rules on users history data to predict purchase combinations or hobby associations, etc.
Suppose you are operating a brand, how to build star "gift set"? how to decide which combinations in the set? to sell a lipstick seperately or put it in a gift set with BB cushion?
Association rules can help solve above questions.

2. "Apriori"

2.1 how to understand it in manageable few data sets

The commonly used one is "apriori" algorithm, see as below:

user_id	food	society	entertnmt	science	art	music
1	1	1		1	1	
2	1		1	1		
3	1		1			
4			1			1
5	1	1	1			1
support	P(f)=4/5	P(s)=2/5	P(e)=4/5	P(s)=2/5	P(a)=1/5	P(m)=2/5
confidence						
lift						

note: {loseweight} -> {food}

Left-hand-side (LHS or Antecedent) -> Right-hand-side (RHS or Consequent)

support: 支持度，frequency= item frequency/transaction records，can calculate single frequency or joint frequency, but denominator can only be transaction records amount,

from which we can identify the **star items or popular features.** P(A), P(A and B) .
confidence: 条件概率, 置信度 conditional probablity, calculate probablity of post/purchase science given post/purchase food; P(A and B)/ P(A)
{food}->{science} = support (food and science)/support(food) = (2/5)/(4/5)= 1/2
{science}->{food} = support (food and science)/support(science) =(2/5)/(2/5)= 1
lift: 提升度， to evaluate indenpence of two items/features/attributes
lift{science->food}
= confidence (s->f)/support(f)
=support(f and s)/support(science)/ support(food) =5/4 >1

> *lift >1, positive results, recommend.*
> *lift =1, independent, no relation between 2 items.*
> *lift <1, negative results.*
> ***lift better above 3, which means excellent association.***
> *lift {A->B} = lift {B->A}*

Another example is about "game player" and "game cards", lift{gameplayer->game cards}<1; purchasing game cards can be indenpendent.

2.2 when it comes to large data set software R is necessary

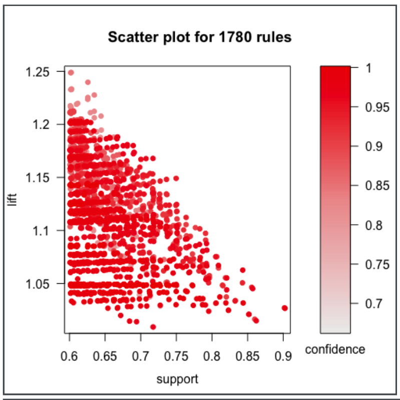
```
##transactions or sparse matrix(need data manipulation)
library(arules)
library(arulesViz)
####transaction import data
data <-read.csv("~/Desktop/test.csv",header=T)
dlist <- apply(data,1,function(x) colnames(data)[unlist(x,use.names=F)]) ##c
trans <- as(dlist,"transactions")
inspect(trans) ##check the data after manipulation

#Apriori
rules=apriori(trans,parameter=list(support=0.1,confidence=0.3,minlen=2))
# to check food relation rules
# rules=apriori(trans, parameter=list(support=0.01,confidence=0.1,minlen=2),
rules
rules<-sort(rules,by='support') ##rank by support
inspect(rules[1:10]) #top 10 rules by support
# transform rules and export
R1<-as(rules,'data.frame') #transform the data to data.frame
###export and check

#data visualization on the rules
install.packages(arulesViz)
library (arulesViz )
#scatter plot
plot(rules, measure = c("support", "lift"), shading = "confidence")
plot(rules, measure = c("support", "lift"), shading ="confidence", interact
```

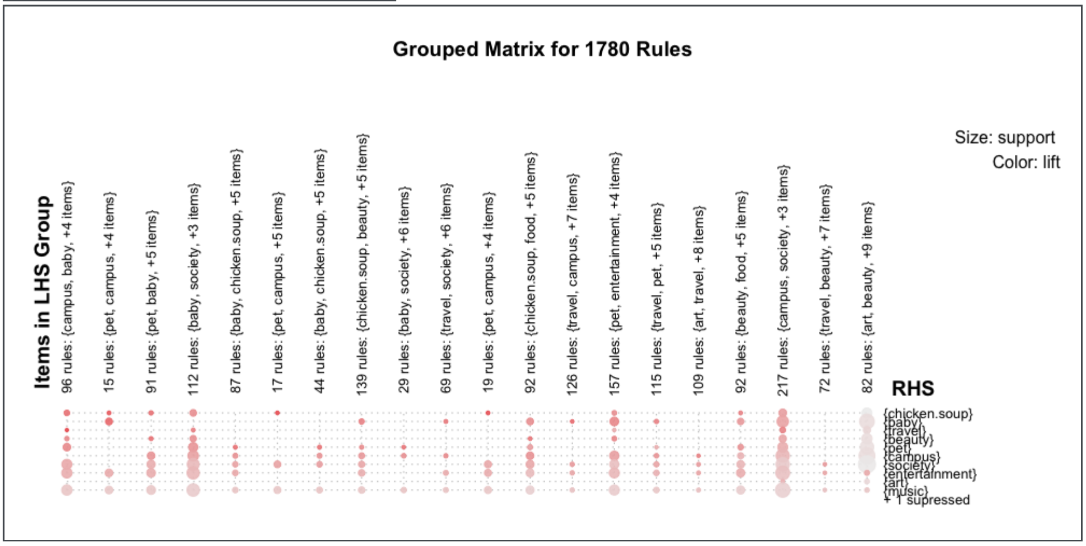
```
##grouped/graph
plot(rules,method="grouped")
plot(rules[1:50],method="graph")
#two key plot, shades of dot means amount of items containing in the rules
plot(rules,shading="order", control=list(main="Two_key plot"))
```

running the codes, rules result was generated:

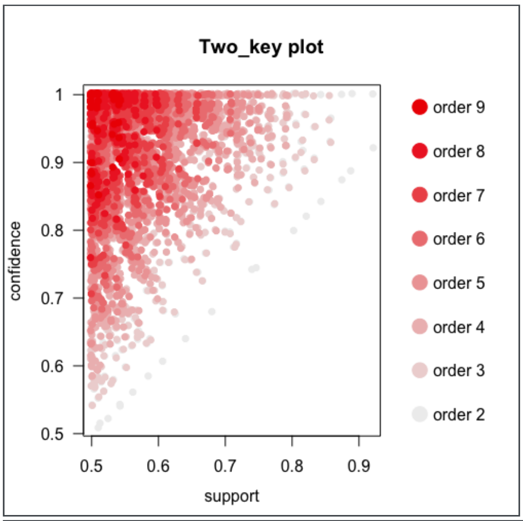
	rules	support	confidence	coverage	lift	count
95	{music} => {food}	0.9027778	0.9798995	0.9212963	1.027467	195
96	{food} => {music}	0.9027778	0.9466019	0.9537037	1.027467	195
93	{entertainment} => {food}	0.8611111	0.9687500	0.8888889	1.015777	186
94	{food} => {entertainment}	0.8611111	0.9029126	0.9537037	1.015777	186
91	{entertainment} => {music}	0.8564815	0.9635417	0.8888889	1.045854	185
92	{music} => {entertainment}	0.8564815	0.9296482	0.9212963	1.045854	185
89	{society} => {food}	0.8518519	0.9735450	0.8750000	1.020804	184
90	{food} => {society}	0.8518519	0.8932039	0.9537037	1.020804	184
394	{entertainment,music} => {food}	0.8425926	0.9837838	0.8564815	1.031540	182
395	{entertainment,food} => {music}	0.8425926	0.9784946	0.8611111	1.062085	182
396	{food,music} => {entertainment}	0.8425926	0.9333333	0.9027778	1.050000	182
87	{society} => {music}	0.8333333	0.9523810	0.8750000	1.033740	180
88	{music} => {society}	0.8333333	0.9045226	0.9212963	1.033740	180
83	{campus} => {food}	0.8240741	0.9780220	0.8425926	1.025499	178
84	{food} => {campus}	0.8240741	0.8640777	0.9537037	1.025499	178

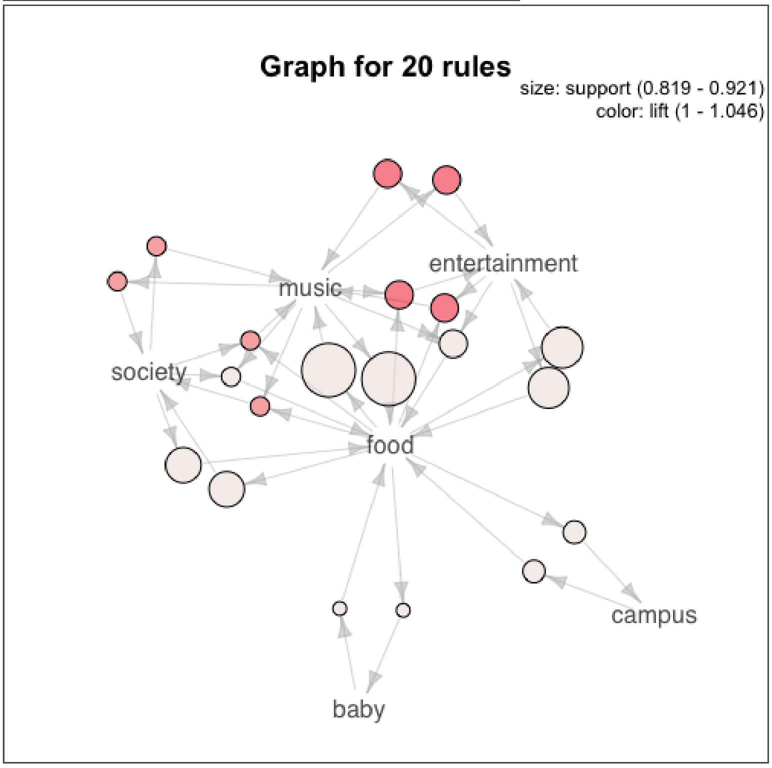
2.3 plot rules

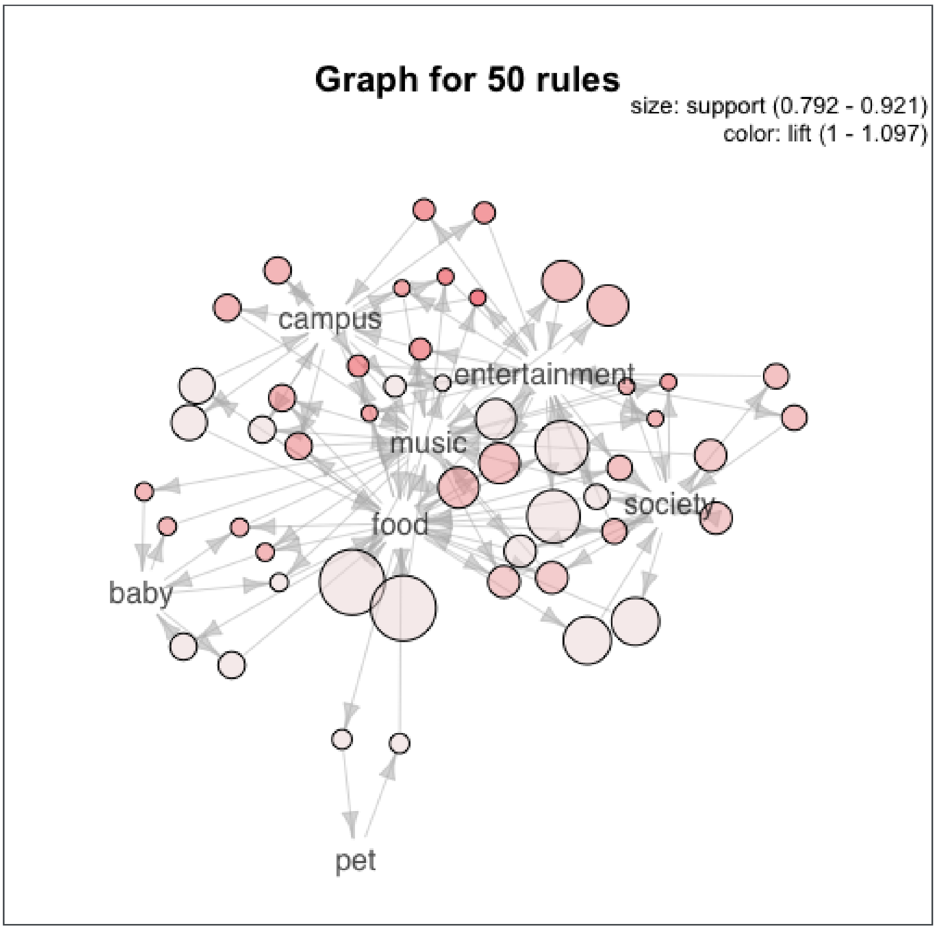




two_key plot


Two_key plot
1
0.9
0.8
0.7
0.6
0.5
confidence
0.5
0.6
0.7
0.8
0.9
support
order 9
order 8
order 7
order 6
order 5
order 4
order 3
order 2


Graph for 20 rules
size: support (0.819 - 0.921)
color: lift (1 - 1.046)
entertainment
music
society
food
campus
baby



3. explanations on graph

–scatter plot

–grouped graph

–two key plot, order=color shades, the darker the shades, the more itemes contain in the rule.

–graph plot, bubble size present support=P(lhs and rhs), items frequency.
bubbule color refers to "lift", the higher the "lift" is, the closer relations lhs&rhs are, which means they could be recommended together.

–interactive graph

4. Conculsion

"Shop basket analysis" is a well-known example to apply association rules on **items**, while we can also investigate relations among **people**, its network analysis (another story), most website recommendations consider **items attributes as well as personalities**.

It can be applied on business cases like……

–create gift sets or promotion sets on e-commerce channel or in physical stores

–joint marketing or advertising, to identify related brands or products

–preference prediction, to know what other hobbies your target audience have when you know they prefer food/music….

–……. (depend on the real requirements)

2020.07.21