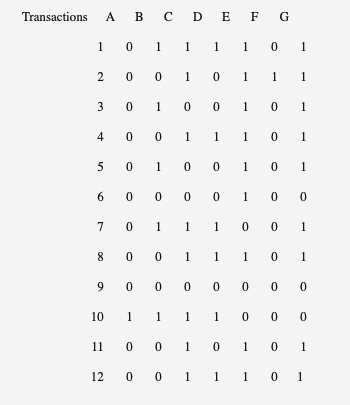
Assign6-1 Association Analysis

Jonathan De Los Santos

2/28/2021

# Calculate Support, Confidence, & Lift

The following data shows the numbers of transactions of a retail store. The manager of the store wants to analyze associations among the transactions shown below:



Assignment Table

The manager found some association rules and wants to figure out their levels of “Support,” “Confidence,” and “Lift.” Specifically, she asks you to:

## a. Rule 1

Compute “Support,” “Confidence,” and “Lift” for the rule of {C&G} –>{B}

**Support**

**Confidence**

**Lift**

## b. Rule 2

Compute “Support,” “Confidence,” and “Lift” for the rule of {C&E} –>{B}

**Support**

**Confidence**

**Lift**

# Association Analysis in R

## Data Prep and Inspection

The below tasks require you to analyze data using R. R provides a dataset named “Epub.” Help provide some explanation about the following data: The data contains 15,729 rows and 936 columns. Inspect the first 10 items, and demonstrate item frequency of the items.

### Inspection

The Epub data contained in stock R is loaded and inspected below.

library(arules)  
data("Epub")  
#summary(Epub)  
inspect(Epub[1:10])

## items transactionID TimeStamp   
## [1] {doc\_154} session\_4795 2003-01-01 19:59:00  
## [2] {doc\_3d6} session\_4797 2003-01-02 06:46:01  
## [3] {doc\_16f} session\_479a 2003-01-02 09:50:38  
## [4] {doc\_11d,doc\_1a7,doc\_f4} session\_47b7 2003-01-02 17:55:50  
## [5] {doc\_83} session\_47bb 2003-01-02 20:27:44  
## [6] {doc\_11d} session\_47c2 2003-01-03 09:18:04  
## [7] {doc\_368} session\_47cb 2003-01-03 13:40:57  
## [8] {doc\_11d,doc\_192} session\_47d8 2003-01-03 18:00:01  
## [9] {doc\_364} session\_47e2 2003-01-04 11:48:36  
## [10] {doc\_ec} session\_47e7 2003-01-04 14:58:48

### Frequency

We can find the frequency of these items as well:

itemFrequency(Epub[,1:10])

## doc\_11d doc\_13d doc\_14c doc\_14e doc\_150 doc\_151   
## 0.0226333524 0.0009536525 0.0024794965 0.0017801513 0.0015894208 0.0007629220   
## doc\_153 doc\_154 doc\_155 doc\_156   
## 0.0006357683 0.0013351135 0.0010808062 0.0031152648

## Build Analysis

Investigate association rules using the data. You need to set the criteria for “Support” and “Confidence,” and provide your rationale to set the levels of “Support” and “Confidence.”

First let’s try the default parameters of Support = 0.1 and Confidence = 0.8 for apriori()

apriori(Epub)

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.8 0.1 1 none FALSE TRUE 5 0.1 1  
## maxlen target ext  
## 10 rules TRUE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 1572   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[936 item(s), 15729 transaction(s)] done [0.00s].  
## sorting and recoding items ... [0 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 done [0.00s].  
## writing ... [0 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

## set of 0 rules

This data does not lend itself to any logical rules about the amount of support would make it interesting. The total number of Epubs is = 6% of the number of total transactions, so that should be the absolute maximum.

From here I was left to whittle down a better support number with testing, and I landed on 0.001 or th of a percent to get 120 rules. Let’s set minimum length to 2 since we’d like to see at least pairs of relationships.

epubRules <- apriori(Epub, parameter = list(support = 0.001,  
 confidence = 0.1,  
 minlen = 2))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.1 0.1 1 none FALSE TRUE 5 0.001 2  
## maxlen target ext  
## 10 rules TRUE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 15   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[936 item(s), 15729 transaction(s)] done [0.00s].  
## sorting and recoding items ... [481 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 done [0.00s].  
## writing ... [120 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

## Visualize Analysis

When you investigate the rules generated by your algorithms, the rules are displayed with some criteria. Sort the association rules by the “Lift,” and visualize the rules using the “arulesViz” package. Provide your interpretation of the output of your visualization.

# Rule Prep and Inspection

Now that we have our rules, we can save them into a dataframe for sorting.

inspect(epubRules[1:10])

## lhs rhs support confidence coverage lift count  
## [1] {doc\_506} => {doc\_507} 0.001207960 0.6551724 0.001843728 303.09432 19   
## [2] {doc\_507} => {doc\_506} 0.001207960 0.5588235 0.002161612 303.09432 19   
## [3] {doc\_470} => {doc\_4c6} 0.001080806 0.2048193 0.005276877 11.18612 17   
## [4] {doc\_714} => {doc\_574} 0.001080806 0.3695652 0.002924534 113.97826 17   
## [5] {doc\_574} => {doc\_714} 0.001080806 0.3333333 0.003242418 113.97826 17   
## [6] {doc\_4bf} => {doc\_4ac} 0.001080806 0.5000000 0.002161612 77.10294 17   
## [7] {doc\_4ac} => {doc\_4bf} 0.001080806 0.1666667 0.006484837 77.10294 17   
## [8] {doc\_6e9} => {doc\_6e8} 0.001207960 0.6785714 0.001780151 333.53906 19   
## [9] {doc\_6e8} => {doc\_6e9} 0.001207960 0.5937500 0.002034459 333.53906 19   
## [10] {doc\_6e9} => {doc\_6e7} 0.001271537 0.7142857 0.001780151 321.00000 20

erules <- DATAFRAME(epubRules)  
lift\_erules <- erules[order(erules$lift, decreasing = TRUE),]  
lift\_erules[1:10,]

## LHS RHS support confidence coverage lift  
## 120 {doc\_6e7,doc\_6e8} {doc\_6e9} 0.001080806 0.8095238 0.001335113 454.7500  
## 119 {doc\_6e7,doc\_6e9} {doc\_6e8} 0.001080806 0.8500000 0.001271537 417.8016  
## 118 {doc\_6e8,doc\_6e9} {doc\_6e7} 0.001080806 0.8947368 0.001207960 402.0947  
## 8 {doc\_6e9} {doc\_6e8} 0.001207960 0.6785714 0.001780151 333.5391  
## 9 {doc\_6e8} {doc\_6e9} 0.001207960 0.5937500 0.002034459 333.5391  
## 10 {doc\_6e9} {doc\_6e7} 0.001271537 0.7142857 0.001780151 321.0000  
## 11 {doc\_6e7} {doc\_6e9} 0.001271537 0.5714286 0.002225189 321.0000  
## 1 {doc\_506} {doc\_507} 0.001207960 0.6551724 0.001843728 303.0943  
## 2 {doc\_507} {doc\_506} 0.001207960 0.5588235 0.002161612 303.0943  
## 14 {doc\_6e8} {doc\_6e7} 0.001335113 0.6562500 0.002034459 294.9187  
## count  
## 120 17  
## 119 17  
## 118 17  
## 8 19  
## 9 19  
## 10 20  
## 11 20  
## 1 19  
## 2 19  
## 14 21

### Rules Plot

Below the rules are plotted by lift and reduced to 10 rules for useful visibility. Unsurprisingly, several of these associations are between documents that appear sequential. One extremely interesting grouping of rules are the links betwee doc\_6e7,8,9 which also have the greatest overall lift values.

library(arulesViz)

## Warning: package 'arulesViz' was built under R version 4.0.2

## Loading required package: grid

plot(epubRules[1:10,], method = "graph", measure = "lift", shading = "confidence")

