hw2.R

asahikuroki222

2021-11-11

#Use package arules  
library(arules)

## Loading required package: Matrix

##   
## Attaching package: 'arules'

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

### Import dataset  
### a)  
movie\_df <- read.csv("movies\_binary.csv")  
movie\_matrix <- as.matrix(movie\_df)  
?as  
movie\_transactions <- as(movie\_matrix, Class = "transactions")  
inspect(head(movie\_transactions))

## items   
## [1] {Fracture.2007.,   
## JohnWick.2014.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,  
## BigHero6.2014.,   
## VeryBadThings.1998.,   
## DownwithLove.2003.,   
## ForbiddenKingdom.The.2008.,   
## MeettheRobinsons.2007.,   
## BourneUltimatum.The.2007.,   
## HappyGilmore.1996.,   
## BillabongOdyssey.2003.,   
## TheDUFF.2015.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## Up.2009.,   
## LionKing.The.1994.,   
## Hobbit.AnUnexpectedJourney.The.2012.}   
## [2] {Warrior.2011.,   
## Departed.The.2006.,   
## Fracture.2007.,   
## MyBestFriend.sWedding.1997.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,  
## BigHero6.2014.,   
## VeryBadThings.1998.,   
## MeettheRobinsons.2007.,   
## BourneUltimatum.The.2007.,   
## HappyGilmore.1996.,   
## BillabongOdyssey.2003.,   
## TheDUFF.2015.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## Proof.1991.,   
## TopGun.1986.,   
## Up.2009.,   
## LionKing.The.1994.,   
## AlmostFamous.2000.,   
## Hobbit.AnUnexpectedJourney.The.2012.}   
## [3] {Departed.The.2006.,   
## Fracture.2007.,   
## JohnWick.2014.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,  
## BigHero6.2014.,   
## ForbiddenKingdom.The.2008.,   
## MeettheRobinsons.2007.,   
## BourneUltimatum.The.2007.,   
## HappyGilmore.1996.,   
## TheDUFF.2015.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## TopGun.1986.,   
## Anchorman.TheLegendofRonBurgundy.2004.,   
## Up.2009.,   
## LionKing.The.1994.}   
## [4] {Warrior.2011.,   
## Departed.The.2006.,   
## Fracture.2007.,   
## MyBestFriend.sWedding.1997.,   
## Interstellar.2014.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,  
## ForbiddenKingdom.The.2008.,   
## MeettheRobinsons.2007.,   
## BourneUltimatum.The.2007.,   
## HappyGilmore.1996.,   
## EternalSunshineoftheSpotlessMind.2004.,   
## TheDUFF.2015.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## Proof.1991.,   
## TopGun.1986.,   
## Up.2009.,   
## LionKing.The.1994.,   
## AlmostFamous.2000.}   
## [5] {Fracture.2007.,   
## MyBestFriend.sWedding.1997.,   
## JohnWick.2014.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## PulpFiction.1994.,   
## DownwithLove.2003.,   
## ForbiddenKingdom.The.2008.,   
## MeettheRobinsons.2007.,   
## BourneUltimatum.The.2007.,   
## HappyGilmore.1996.,   
## BillabongOdyssey.2003.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## Proof.1991.,   
## TopGun.1986.,   
## InvincibleIronMan.The.2007.,   
## AlmostFamous.2000.,   
## Hobbit.AnUnexpectedJourney.The.2012.}   
## [6] {Warrior.2011.,   
## Departed.The.2006.,   
## Fracture.2007.,   
## MyBestFriend.sWedding.1997.,   
## Interstellar.2014.,   
## JohnWick.2014.,   
## StarWars.EpisodeIV.ANewHope.1977.,   
## DownwithLove.2003.,   
## ForbiddenKingdom.The.2008.,   
## MeettheRobinsons.2007.,   
## HappyGilmore.1996.,   
## TheDUFF.2015.,   
## DarkKnightRises.The.2012.,   
## X.Men.2000.,   
## X.MenOrigins.Wolverine.2009.,   
## TopGun.1986.,   
## InvincibleIronMan.The.2007.,   
## Anchorman.TheLegendofRonBurgundy.2004.,   
## AlmostFamous.2000.,   
## Hobbit.AnUnexpectedJourney.The.2012.}

### b) Total number and names of unique movies  
# 32 unique movies  
itemInfo(movie\_transactions)

## labels  
## 1 Warrior.2011.  
## 2 Departed.The.2006.  
## 3 Fracture.2007.  
## 4 MyBestFriend.sWedding.1997.  
## 5 Interstellar.2014.  
## 6 JohnWick.2014.  
## 7 StarWars.EpisodeIV.ANewHope.1977.  
## 8 StarWars.EpisodeV.TheEmpireStrikesBack.1980.  
## 9 PulpFiction.1994.  
## 10 HarryPotterandtheSorcererStone.2001.  
## 11 HarryPotterandtheChamberofSecrets.2002.  
## 12 BigHero6.2014.  
## 13 VeryBadThings.1998.  
## 14 DownwithLove.2003.  
## 15 ForbiddenKingdom.The.2008.  
## 16 MeettheRobinsons.2007.  
## 17 BourneUltimatum.The.2007.  
## 18 HappyGilmore.1996.  
## 19 EternalSunshineoftheSpotlessMind.2004.  
## 20 BillabongOdyssey.2003.  
## 21 TheDUFF.2015.  
## 22 DarkKnightRises.The.2012.  
## 23 X.Men.2000.  
## 24 X.MenOrigins.Wolverine.2009.  
## 25 Proof.1991.  
## 26 TopGun.1986.  
## 27 InvincibleIronMan.The.2007.  
## 28 Anchorman.TheLegendofRonBurgundy.2004.  
## 29 Up.2009.  
## 30 LionKing.The.1994.  
## 31 AlmostFamous.2000.  
## 32 Hobbit.AnUnexpectedJourney.The.2012.

### c)  
itemFrequencyPlot(movie\_transactions, ylim = c(0, 1), topN = 10,   
 main = "Movies by support %", col = "#053259" )

Chart, bar chart

Description automatically generated

# Top2 movies   
# No1. StarWars.EpisodeIV.ANewHope.1977.: 0.8998999  
# No2. StarWars.EpisodeV.TheEmpireStrikesBack.1980. : 0.8588589   
# Within all the people in this dataset, 89.9% of the people have watched StarWars.EpisodeIV  
# and 85.8% of the people have watched StarWars.EpisodeV.  
# This shows how popular these two movies are.   
  
### d) Transactions that contain both up.2009 and lion king  
up2009\_and\_lionking <- subset(movie\_transactions,   
 items %ain% c("Up.2009.", "LionKing.The.1994."))  
summary(up2009\_and\_lionking)

## transactions as itemMatrix in sparse format with  
## 776 rows (elements/itemsets/transactions) and  
## 32 columns (items) and a density of 0.6022068   
##   
## most frequent items:  
## Up.2009.   
## 776   
## LionKing.The.1994.   
## 776   
## StarWars.EpisodeIV.ANewHope.1977.   
## 725   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.   
## 691   
## X.Men.2000.   
## 685   
## (Other)   
## 11301   
##   
## element (itemset/transaction) length distribution:  
## sizes  
## 10 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27   
## 1 3 9 13 36 49 85 101 110 123 90 73 43 23 9 6 2   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.00 17.00 19.00 19.27 21.00 27.00   
##   
## includes extended item information - examples:  
## labels  
## 1 Warrior.2011.  
## 2 Departed.The.2006.  
## 3 Fracture.2007.

# Answer: 776 transactions  
  
### E)   
?apriori  
movie\_rules <- apriori(movie\_transactions, list(supp = 0.7, conf = 0.7, minlen = 2, maxlen = 2))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.7 0.1 1 none FALSE TRUE 5 0.7 2  
## maxlen target ext  
## 2 rules TRUE  
##   
## Algorithmic control:  
## filter tree heap memopt load sort verbose  
## 0.1 TRUE TRUE FALSE TRUE 2 TRUE  
##   
## Absolute minimum support count: 699   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[32 item(s), 999 transaction(s)] done [0.00s].  
## sorting and recoding items ... [9 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2

## Warning in apriori(movie\_transactions, list(supp = 0.7, conf = 0.7, minlen  
## = 2, : Mining stopped (maxlen reached). Only patterns up to a length of 2  
## returned!

## done [0.00s].  
## writing ... [22 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

inspect(head(movie\_rules))

## lhs rhs support confidence coverage lift count  
## [1] {HarryPotterandtheChamberofSecrets.2002.} => {HarryPotterandtheSorcererStone.2001.} 0.7217217 1.0000000 0.7217217 1.216809 721  
## [2] {HarryPotterandtheSorcererStone.2001.} => {HarryPotterandtheChamberofSecrets.2002.} 0.7217217 0.8781973 0.8218218 1.216809 721  
## [3] {Up.2009.} => {LionKing.The.1994.} 0.7767768 1.0000000 0.7767768 1.237918 776  
## [4] {LionKing.The.1994.} => {Up.2009.} 0.7767768 0.9615861 0.8078078 1.237918 776  
## [5] {Up.2009.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7257257 0.9342784 0.7767768 1.038203 725  
## [6] {StarWars.EpisodeIV.ANewHope.1977.} => {Up.2009.} 0.7257257 0.8064516 0.8998999 1.038203 725

summary(movie\_rules)

## set of 22 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 2   
## 22   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2 2 2 2 2 2   
##   
## summary of quality measures:  
## support confidence coverage lift   
## Min. :0.7057 Min. :0.8065 Min. :0.7217 Min. :0.9827   
## 1st Qu.:0.7227 1st Qu.:0.8474 1st Qu.:0.8078 1st Qu.:1.0173   
## Median :0.7397 Median :0.8809 Median :0.8478 Median :1.0382   
## Mean :0.7485 Mean :0.8945 Mean :0.8387 Mean :1.0689   
## 3rd Qu.:0.7598 3rd Qu.:0.9327 3rd Qu.:0.8589 3rd Qu.:1.0993   
## Max. :0.8589 Max. :1.0000 Max. :0.8999 Max. :1.2379   
## count   
## Min. :705.0   
## 1st Qu.:722.0   
## Median :739.0   
## Mean :747.7   
## 3rd Qu.:759.0   
## Max. :858.0   
##   
## mining info:  
## data ntransactions support confidence  
## movie\_transactions 999 0.7 0.7

# X -> Y  
# mean value of support: 0.7485 -> 74.8% of the people in the dataset   
# watch these two movies combination of X and Y. High support value,   
# which shows that these movies are popular  
# mean value of confidence: 0.8945 -> When people watch Movie X, 89.45%   
# of the time they watch Movie Y   
# mean value of lift: 1.06 -> Low lift value. People who watch movies X,   
# watch movie Y only 1.06 times more than all the other people.   
  
### f)  
inspect(movie\_rules)

## lhs rhs support confidence coverage lift count  
## [1] {HarryPotterandtheChamberofSecrets.2002.} => {HarryPotterandtheSorcererStone.2001.} 0.7217217 1.0000000 0.7217217 1.2168088 721  
## [2] {HarryPotterandtheSorcererStone.2001.} => {HarryPotterandtheChamberofSecrets.2002.} 0.7217217 0.8781973 0.8218218 1.2168088 721  
## [3] {Up.2009.} => {LionKing.The.1994.} 0.7767768 1.0000000 0.7767768 1.2379182 776  
## [4] {LionKing.The.1994.} => {Up.2009.} 0.7767768 0.9615861 0.8078078 1.2379182 776  
## [5] {Up.2009.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7257257 0.9342784 0.7767768 1.0382025 725  
## [6] {StarWars.EpisodeIV.ANewHope.1977.} => {Up.2009.} 0.7257257 0.8064516 0.8998999 1.0382025 725  
## [7] {LionKing.The.1994.} => {X.Men.2000.} 0.7137137 0.8835192 0.8078078 1.0420728 713  
## [8] {X.Men.2000.} => {LionKing.The.1994.} 0.7137137 0.8417946 0.8478478 1.0420728 713  
## [9] {LionKing.The.1994.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7057057 0.8736059 0.8078078 1.0171706 705  
## [10] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {LionKing.The.1994.} 0.7057057 0.8216783 0.8588589 1.0171706 705  
## [11] {LionKing.The.1994.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7397397 0.9157373 0.8078078 1.0175991 739  
## [12] {StarWars.EpisodeIV.ANewHope.1977.} => {LionKing.The.1994.} 0.7397397 0.8220245 0.8998999 1.0175991 739  
## [13] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7507508 0.9135201 0.8218218 1.0636440 750  
## [14] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {HarryPotterandtheSorcererStone.2001.} 0.7507508 0.8741259 0.8588589 1.0636440 750  
## [15] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7627628 0.9281364 0.8218218 1.0313774 762  
## [16] {StarWars.EpisodeIV.ANewHope.1977.} => {HarryPotterandtheSorcererStone.2001.} 0.7627628 0.8476085 0.8998999 1.0313774 762  
## [17] {X.Men.2000.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7277277 0.8583235 0.8478478 0.9993767 727  
## [18] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {X.Men.2000.} 0.7277277 0.8473193 0.8588589 0.9993767 727  
## [19] {X.Men.2000.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7497497 0.8842975 0.8478478 0.9826621 749  
## [20] {StarWars.EpisodeIV.ANewHope.1977.} => {X.Men.2000.} 0.7497497 0.8331479 0.8998999 0.9826621 749  
## [21] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.8588589 1.0000000 0.8588589 1.1112347 858  
## [22] {StarWars.EpisodeIV.ANewHope.1977.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.8588589 0.9543938 0.8998999 1.1112347 858

# 22 rules found  
# Obvious rule: X -> Y and Y -> X.   
# We could see that from first 2 rules. HarryPotter.2002 -> HarryPotter.2001   
# and HarryPotter.2001 -> HarryPotter2002  
# Less obvious rule: People who watch X.men watch Lion King.   
# It is less obvious because X men and lion king are in two different genre   
# (Action and Musical)  
  
### g) Order the rules in increasing order by confidence  
inspect(sort(movie\_rules, by = "confidence", decreasing = FALSE ))

## lhs rhs support confidence coverage lift count  
## [1] {StarWars.EpisodeIV.ANewHope.1977.} => {Up.2009.} 0.7257257 0.8064516 0.8998999 1.0382025 725  
## [2] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {LionKing.The.1994.} 0.7057057 0.8216783 0.8588589 1.0171706 705  
## [3] {StarWars.EpisodeIV.ANewHope.1977.} => {LionKing.The.1994.} 0.7397397 0.8220245 0.8998999 1.0175991 739  
## [4] {StarWars.EpisodeIV.ANewHope.1977.} => {X.Men.2000.} 0.7497497 0.8331479 0.8998999 0.9826621 749  
## [5] {X.Men.2000.} => {LionKing.The.1994.} 0.7137137 0.8417946 0.8478478 1.0420728 713  
## [6] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {X.Men.2000.} 0.7277277 0.8473193 0.8588589 0.9993767 727  
## [7] {StarWars.EpisodeIV.ANewHope.1977.} => {HarryPotterandtheSorcererStone.2001.} 0.7627628 0.8476085 0.8998999 1.0313774 762  
## [8] {X.Men.2000.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7277277 0.8583235 0.8478478 0.9993767 727  
## [9] {LionKing.The.1994.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7057057 0.8736059 0.8078078 1.0171706 705  
## [10] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {HarryPotterandtheSorcererStone.2001.} 0.7507508 0.8741259 0.8588589 1.0636440 750  
## [11] {HarryPotterandtheSorcererStone.2001.} => {HarryPotterandtheChamberofSecrets.2002.} 0.7217217 0.8781973 0.8218218 1.2168088 721  
## [12] {LionKing.The.1994.} => {X.Men.2000.} 0.7137137 0.8835192 0.8078078 1.0420728 713  
## [13] {X.Men.2000.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7497497 0.8842975 0.8478478 0.9826621 749  
## [14] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7507508 0.9135201 0.8218218 1.0636440 750  
## [15] {LionKing.The.1994.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7397397 0.9157373 0.8078078 1.0175991 739  
## [16] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7627628 0.9281364 0.8218218 1.0313774 762  
## [17] {Up.2009.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7257257 0.9342784 0.7767768 1.0382025 725  
## [18] {StarWars.EpisodeIV.ANewHope.1977.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.8588589 0.9543938 0.8998999 1.1112347 858  
## [19] {LionKing.The.1994.} => {Up.2009.} 0.7767768 0.9615861 0.8078078 1.2379182 776  
## [20] {HarryPotterandtheChamberofSecrets.2002.} => {HarryPotterandtheSorcererStone.2001.} 0.7217217 1.0000000 0.7217217 1.2168088 721  
## [21] {Up.2009.} => {LionKing.The.1994.} 0.7767768 1.0000000 0.7767768 1.2379182 776  
## [22] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.8588589 1.0000000 0.8588589 1.1112347 858

# Highest Confidence  
# HarryPotter.2002 -> HarryPotter,,2001 and Up.2009 -> Lionking.1994   
# and Starwars.1980 -> StarWars1977. These three all have a confidence of 1.0   
# which means, if people watch Harry Potter.2002 100% of the time,   
# they watch HarryPotter.2001. This applies to all three rules we found  
# Lowest Confidence   
# StarWars.1977 -> Up.2009  
# Confidence of 80%. When people watch StarWars.1977, 80% of the time   
# they watch Up.2009  
  
### h) Order rules in increasing order by lift  
inspect(sort(movie\_rules, by = "lift", decreasing = FALSE))

## lhs rhs support confidence coverage lift count  
## [1] {X.Men.2000.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7497497 0.8842975 0.8478478 0.9826621 749  
## [2] {StarWars.EpisodeIV.ANewHope.1977.} => {X.Men.2000.} 0.7497497 0.8331479 0.8998999 0.9826621 749  
## [3] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {X.Men.2000.} 0.7277277 0.8473193 0.8588589 0.9993767 727  
## [4] {X.Men.2000.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7277277 0.8583235 0.8478478 0.9993767 727  
## [5] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {LionKing.The.1994.} 0.7057057 0.8216783 0.8588589 1.0171706 705  
## [6] {LionKing.The.1994.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7057057 0.8736059 0.8078078 1.0171706 705  
## [7] {LionKing.The.1994.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7397397 0.9157373 0.8078078 1.0175991 739  
## [8] {StarWars.EpisodeIV.ANewHope.1977.} => {LionKing.The.1994.} 0.7397397 0.8220245 0.8998999 1.0175991 739  
## [9] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7627628 0.9281364 0.8218218 1.0313774 762  
## [10] {StarWars.EpisodeIV.ANewHope.1977.} => {HarryPotterandtheSorcererStone.2001.} 0.7627628 0.8476085 0.8998999 1.0313774 762  
## [11] {Up.2009.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7257257 0.9342784 0.7767768 1.0382025 725  
## [12] {StarWars.EpisodeIV.ANewHope.1977.} => {Up.2009.} 0.7257257 0.8064516 0.8998999 1.0382025 725  
## [13] {X.Men.2000.} => {LionKing.The.1994.} 0.7137137 0.8417946 0.8478478 1.0420728 713  
## [14] {LionKing.The.1994.} => {X.Men.2000.} 0.7137137 0.8835192 0.8078078 1.0420728 713  
## [15] {HarryPotterandtheSorcererStone.2001.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7507508 0.9135201 0.8218218 1.0636440 750  
## [16] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {HarryPotterandtheSorcererStone.2001.} 0.7507508 0.8741259 0.8588589 1.0636440 750  
## [17] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.8588589 1.0000000 0.8588589 1.1112347 858  
## [18] {StarWars.EpisodeIV.ANewHope.1977.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.8588589 0.9543938 0.8998999 1.1112347 858  
## [19] {HarryPotterandtheChamberofSecrets.2002.} => {HarryPotterandtheSorcererStone.2001.} 0.7217217 1.0000000 0.7217217 1.2168088 721  
## [20] {HarryPotterandtheSorcererStone.2001.} => {HarryPotterandtheChamberofSecrets.2002.} 0.7217217 0.8781973 0.8218218 1.2168088 721  
## [21] {LionKing.The.1994.} => {Up.2009.} 0.7767768 0.9615861 0.8078078 1.2379182 776  
## [22] {Up.2009.} => {LionKing.The.1994.} 0.7767768 1.0000000 0.7767768 1.2379182 776

# Highest lift  
# Up.2009 -> LionKing.1994 and LionKing.1994 and Up.2009  
# Lift value of 1.24 shows that people who watch up.2009 are 1.24 times   
# more likely to watch Lionking and vice versa  
# Lowest lift  
# X.Men -> StarWars.1977 and StarWars.1977 -> X.men  
# Lift value of 0.98 shows that people who watch X.men are less likely   
# to watch StarWars and vice versa  
  
### i)   
# 2 rules I trust   
# 1st: HarryPotter.2002 -> HarryPotter.2001  
# 2nd: StarWars.1980 -> StarWars1977]  
# These two rules have Confidence of 1 and lift of over 1 which means   
# when people watch the left hand side of the movie, 100% of the time   
# they watch the right hand side too. And they are more likely to watch   
# them than all the other people. Also, these rules make sense because   
# when people watch a new episode of a series they will also watch   
# the previous one too. Otherwise, they will get confused.   
# 2 rules I don't trust  
# 1st: StarWars.1980 -> X.men.2000  
# 2nd: StarWars.1977 -> X.men.2000  
# They have Confidence of about 0.847 and 0.833 which means when people   
# watch StarWars movies, they only watch X.men about 83 - 85 % of the time.   
# Also, they have Lift value of less than one, which means people   
# who watch Star.Wars are less likely to watch X.men than others.   
  
### j)  
# Recommendations to the company:   
# I would suggest the streaming company to give recommendation to customers based   
# on the rules we found. We found out that people tend to watch movies that are   
# in the same genre or series. That is supported by high confidence and lift values.   
# For instance {Up.2009 -> LionKing.1994} have a confidence of 1 and lift of   
# 1.23. This makes sense because they belong in the same genre. This applies to   
# harry potter and StarWars too.   
# Therefore, if a customer watches an action movie then the company should recommend  
# action movies to a customer. However, we cannot just recommend any action movie   
# because those two movies might have low confidence or lift value. If two movies  
# have low confidence or lift value then that means they are generally not watched   
# together. Thus, I would suggest the company to recommend movies that have higher   
# confidence and lift value.   
  
  
### k)  
bad\_rules <- apriori(movie\_transactions, list(supp = 0.5, conf = 0.6, minlen = 2))

## Apriori  
##   
## Parameter specification:  
## confidence minval smax arem aval originalSupport maxtime support minlen  
## 0.6 0.1 1 none FALSE TRUE 5 0.5 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: 499   
##   
## set item appearances ...[0 item(s)] done [0.00s].  
## set transactions ...[32 item(s), 999 transaction(s)] done [0.00s].  
## sorting and recoding items ... [23 item(s)] done [0.00s].  
## creating transaction tree ... done [0.00s].  
## checking subsets of size 1 2 3 4 5 6 done [0.00s].  
## writing ... [620 rule(s)] done [0.00s].  
## creating S4 object ... done [0.00s].

inspect(head(bad\_rules))

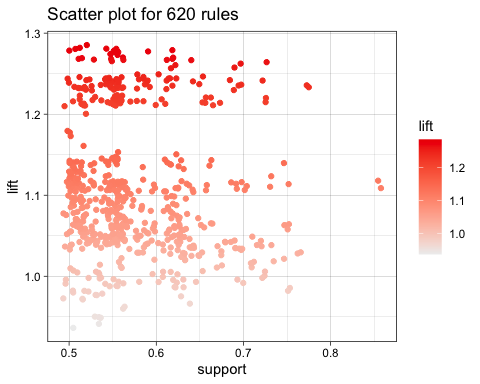
## lhs rhs support confidence coverage lift count  
## [1] {X.MenOrigins.Wolverine.2009.} => {X.Men.2000.} 0.5035035 1.0000000 0.5035035 1.1794569 503  
## [2] {PulpFiction.1994.} => {HarryPotterandtheChamberofSecrets.2002.} 0.5135135 0.7056396 0.7277277 0.9777170 513  
## [3] {HarryPotterandtheChamberofSecrets.2002.} => {PulpFiction.1994.} 0.5135135 0.7115118 0.7217217 0.9777170 513  
## [4] {PulpFiction.1994.} => {Up.2009.} 0.5355355 0.7359010 0.7277277 0.9473777 535  
## [5] {Up.2009.} => {PulpFiction.1994.} 0.5355355 0.6894330 0.7767768 0.9473777 535  
## [6] {PulpFiction.1994.} => {LionKing.The.1994.} 0.5625626 0.7730399 0.7277277 0.9569602 562

summary(bad\_rules)

## set of 620 rules  
##   
## rule length distribution (lhs + rhs):sizes  
## 2 3 4 5 6   
## 71 189 216 120 24   
##   
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.000 3.000 4.000 3.737 4.000 6.000   
##   
## summary of quality measures:  
## support confidence coverage lift   
## Min. :0.5005 Min. :0.6766 Min. :0.5005 Min. :0.9297   
## 1st Qu.:0.5285 1st Qu.:0.8460 1st Qu.:0.5756 1st Qu.:1.0467   
## Median :0.5566 Median :0.8927 Median :0.6296 Median :1.0881   
## Mean :0.5763 Mean :0.8976 Mean :0.6471 Mean :1.1050   
## 3rd Qu.:0.6176 3rd Qu.:0.9814 3rd Qu.:0.7137 3rd Qu.:1.1435   
## Max. :0.8589 Max. :1.0000 Max. :0.8999 Max. :1.2824   
## count   
## Min. :500.0   
## 1st Qu.:528.0   
## Median :556.0   
## Mean :575.7   
## 3rd Qu.:617.0   
## Max. :858.0   
##   
## mining info:  
## data ntransactions support confidence  
## movie\_transactions 999 0.5 0.6

# Answer: 620 rules found  
  
### L)  
library(arulesViz)  
plot(bad\_rules, measure = c("support", "lift"), shading = "lift")

## To reduce overplotting, jitter is added! Use jitter = 0 to prevent jitter.



### m)  
# First of all rules that have lift of <=1.0 should be excluded.   
# Lift value of less than or equal to one means, they either do not have   
# any association or they negatively affect each other.  
badrules\_badlift <- subset(bad\_rules, subset = lift <= 1)  
inspect(sort(badrules\_badlift, by = "confidence"))

## lhs rhs support confidence coverage lift count  
## [1] {PulpFiction.1994.,   
## LionKing.The.1994.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.5005005 0.8896797 0.5625626 0.9886430 500  
## [2] {X.Men.2000.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.7497497 0.8842975 0.8478478 0.9826621 749  
## [3] {PulpFiction.1994.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.6346346 0.8720770 0.7277277 0.9690823 634  
## [4] {X.Men.2000.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.7277277 0.8583235 0.8478478 0.9993767 727  
## [5] {PulpFiction.1994.,   
## X.Men.2000.} => {StarWars.EpisodeIV.ANewHope.1977.} 0.5385385 0.8526149 0.6316316 0.9474553 538  
## [6] {PulpFiction.1994.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.6176176 0.8486933 0.7277277 0.9881638 617  
## [7] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {X.Men.2000.} 0.7277277 0.8473193 0.8588589 0.9993767 727  
## [8] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {X.Men.2000.} 0.7277277 0.8473193 0.8588589 0.9993767 727  
## [9] {HarryPotterandtheSorcererStone.2001.} => {X.Men.2000.} 0.6956957 0.8465286 0.8218218 0.9984440 695  
## [10] {StarWars.EpisodeIV.ANewHope.1977.,   
## HarryPotterandtheSorcererStone.2001.,   
## DarkKnightRises.The.2012.} => {X.Men.2000.} 0.5005005 0.8460237 0.5915916 0.9978485 500  
## [11] {StarWars.EpisodeIV.ANewHope.1977.,   
## DarkKnightRises.The.2012.} => {X.Men.2000.} 0.5755756 0.8418741 0.6836837 0.9929542 575  
## [12] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5535536 0.8417047 0.6576577 0.9927544 553  
## [13] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5535536 0.8417047 0.6576577 0.9927544 553  
## [14] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5535536 0.8417047 0.6576577 0.9927544 553  
## [15] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5535536 0.8417047 0.6576577 0.9927544 553  
## [16] {StarWars.EpisodeIV.ANewHope.1977.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5625626 0.8400598 0.6696697 0.9908143 562  
## [17] {StarWars.EpisodeIV.ANewHope.1977.,   
## HarryPotterandtheSorcererStone.2001.,   
## HarryPotterandtheChamberofSecrets.2002.} => {X.Men.2000.} 0.5625626 0.8400598 0.6696697 0.9908143 562  
## [18] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.} => {X.Men.2000.} 0.6296296 0.8386667 0.7507508 0.9891712 629  
## [19] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.} => {X.Men.2000.} 0.6296296 0.8386667 0.7507508 0.9891712 629  
## [20] {StarWars.EpisodeIV.ANewHope.1977.,   
## HarryPotterandtheSorcererStone.2001.} => {X.Men.2000.} 0.6386386 0.8372703 0.7627628 0.9875243 638  
## [21] {PulpFiction.1994.,   
## X.Men.2000.} => {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} 0.5285285 0.8367670 0.6316316 0.9742777 528  
## [22] {StarWars.EpisodeIV.ANewHope.1977.} => {X.Men.2000.} 0.7497497 0.8331479 0.8998999 0.9826621 749  
## [23] {X.Men.2000.} => {HarryPotterandtheSorcererStone.2001.} 0.6956957 0.8205431 0.8478478 0.9984440 695  
## [24] {PulpFiction.1994.} => {HarryPotterandtheSorcererStone.2001.} 0.5955956 0.8184319 0.7277277 0.9958751 595  
## [25] {PulpFiction.1994.,   
## X.Men.2000.} => {HarryPotterandtheSorcererStone.2001.} 0.5085085 0.8050713 0.6316316 0.9796178 508  
## [26] {StarWars.EpisodeIV.ANewHope.1977.,   
## PulpFiction.1994.} => {LionKing.The.1994.} 0.5005005 0.7886435 0.6346346 0.9762762 500  
## [27] {PulpFiction.1994.} => {LionKing.The.1994.} 0.5625626 0.7730399 0.7277277 0.9569602 562  
## [28] {PulpFiction.1994.} => {Up.2009.} 0.5355355 0.7359010 0.7277277 0.9473777 535  
## [29] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## X.Men.2000.} => {PulpFiction.1994.} 0.5285285 0.7262724 0.7277277 0.9980001 528  
## [30] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## X.Men.2000.} => {PulpFiction.1994.} 0.5285285 0.7262724 0.7277277 0.9980001 528  
## [31] {HarryPotterandtheSorcererStone.2001.} => {PulpFiction.1994.} 0.5955956 0.7247259 0.8218218 0.9958751 595  
## [32] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {PulpFiction.1994.} 0.6176176 0.7191142 0.8588589 0.9881638 617  
## [33] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.} => {PulpFiction.1994.} 0.6176176 0.7191142 0.8588589 0.9881638 617  
## [34] {StarWars.EpisodeIV.ANewHope.1977.,   
## X.Men.2000.} => {PulpFiction.1994.} 0.5385385 0.7182911 0.7497497 0.9870327 538  
## [35] {X.Men.2000.,   
## LionKing.The.1994.} => {PulpFiction.1994.} 0.5115115 0.7166900 0.7137137 0.9848327 511  
## [36] {StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.} => {PulpFiction.1994.} 0.5345345 0.7120000 0.7507508 0.9783879 534  
## [37] {StarWars.EpisodeIV.ANewHope.1977.,   
## StarWars.EpisodeV.TheEmpireStrikesBack.1980.,   
## HarryPotterandtheSorcererStone.2001.} => {PulpFiction.1994.} 0.5345345 0.7120000 0.7507508 0.9783879 534  
## [38] {HarryPotterandtheChamberofSecrets.2002.} => {PulpFiction.1994.} 0.5135135 0.7115118 0.7217217 0.9777170 513  
## [39] {HarryPotterandtheSorcererStone.2001.,   
## HarryPotterandtheChamberofSecrets.2002.} => {PulpFiction.1994.} 0.5135135 0.7115118 0.7217217 0.9777170 513  
## [40] {StarWars.EpisodeIV.ANewHope.1977.,   
## HarryPotterandtheSorcererStone.2001.} => {PulpFiction.1994.} 0.5415415 0.7099738 0.7627628 0.9756035 541  
## [41] {PulpFiction.1994.} => {HarryPotterandtheChamberofSecrets.2002.} 0.5135135 0.7056396 0.7277277 0.9777170 513  
## [42] {StarWars.EpisodeIV.ANewHope.1977.} => {PulpFiction.1994.} 0.6346346 0.7052280 0.8998999 0.9690823 634  
## [43] {LionKing.The.1994.} => {PulpFiction.1994.} 0.5625626 0.6964064 0.8078078 0.9569602 562  
## [44] {Up.2009.} => {PulpFiction.1994.} 0.5355355 0.6894330 0.7767768 0.9473777 535  
## [45] {Up.2009.,   
## LionKing.The.1994.} => {PulpFiction.1994.} 0.5355355 0.6894330 0.7767768 0.9473777 535  
## [46] {StarWars.EpisodeIV.ANewHope.1977.,   
## LionKing.The.1994.} => {PulpFiction.1994.} 0.5005005 0.6765900 0.7397397 0.9297296 500

# badrules\_badlift is a subset of badrules where lift <= 1, and   
# I sorted the rules by Confidence in descending order.   
# By looking at the subset we could clearly see that these rules contain  
# movies that are not in the same genre.   
# For instance the first rule is PulpFiction.1994, LionKing -> StarWars.1977   
# and it has confidence of 0.889 and lift of 0.988.   
# It means when people watch PulpFiction and LionKing, 88.9 % of the time   
# they watch StarWars. And people who watch PulpFiction and Lionking are   
# less likely to watch StarWars.   
# This makes sense because they don't belong in the same genre.   
# This applies to most of the rules in this subset.   
# Most of the movies in the left hand side does not belong to   
# the movie genre on the right hand side.   
  
  
### n)   
# Music streaming services. We could apply association rules to music streaming  
# service because people often have similarity in their music choices.

# They could mine association between genre, artists, and songs. If they see # a lot of association between some artists, they could use this association as a recommendation for customers.

# The data we need would be a list of artists, songs  
# and its genre for each of the customers. For instance they might be able to   
# find association of {Drake -> Travis Scott}, by mining which artists  
# people listen to. Then, if this association have a high confidence and a lift  
# value they could recommend Travis Scott's song to people who listen to Drake