Titanic survival analysis using Association Rules

Surabhi Chouhan

December 20, 2016

```
library("sqldf")
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Loading required package: RSQLite
```

```
library("plyr")
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
```

```
load("C:/Users/Surabhi/Downloads/Surabhi docs/titanic.raw(1).rdata")
t <- titanic.raw
View(t)

#Descriptive Stats
#percentage people survived
surv <- count(sqldf("select Survived from t where Survived=='Yes' "))
```

```
## Loading required package: tcltk
```

```
## Warning: Quoted identifiers should have class SQL, use DBI::SQL() if the
## caller performs the quoting.
```

```
surv
```

```
##   Survived freq
## 1      Yes  711
```

```
percentage <- (surv$freq/2201)*100
percentage
```

```
## [1] 32.3035
```

```
#percentage of people that were children
child <- tapply(t$Age,t$Age,count)
child
```

```
## $Adult
##       x freq
## 1 Adult 2092
##
## $Child
##       x freq
## 1 Child  109
```

```
percent <- (child$Child$freq/2201)*100
percent
```

```
## [1] 4.952294
```

```
#percentage of people that were female
female <- tapply(t$Sex,t$Sex,count)
female
```

```
## $Female
##        x freq
## 1 Female  470
##
## $Male
##      x freq
## 1 Male 1731
```

```
percent1 <- (female$Female$freq/2201)*100
percent1
```

```
## [1] 21.35393
```

```
#percentage of people that were in first class
first <- tapply(t$Class,t$Class,count)
first
```

```
## $`1st`
##     x freq
## 1 1st  325
##
## $`2nd`
##     x freq
## 1 2nd  285
##
## $`3rd`
##     x freq
## 1 3rd  706
##
```

```
## $Crew
##      x freq
## 1 Crew  885

p_first <- (first$`1st`$freq/2201)*100
p_first

## [1] 14.76602

#More Descriptive Stats

#percentage of children survived
child_surv <- sqldf("select count(*) from t where Age='Child' AND
Survived='Yes' ")
child_surv

##   count(*)
## 1       57

p1 <- (child_surv/2201)*100
p1

##   count(*)
## 1 2.589732

#percentage of female survived
female_surv <- sqldf("select count(*) from t where Sex='Female' AND
Survived='Yes' ")
female_surv

##   count(*)
## 1      344

p2 <- (female_surv/2201)*100
p2

##   count(*)
## 1 15.62926

#percentage of first class people survived
first_surv <- sqldf("select count(*) from t where Class='1st' AND
Survived='Yes' ")
first_surv

##   count(*)
## 1      203

p3 <- (first_surv/2201)*100
p3

##   count(*)
## 1  9.22308
```

```r
#percentage of third class people survived
third_surv <- sqldf("select count(*) from t where Class='3rd' AND
Survived='Yes' ")
third_surv
```

```
##   count(*)
## 1      178
```

```r
p4 <- (third_surv/2201)*100
p4
```

```
##   count(*)
## 1 8.087233
```

```r
#Writing a Function

myFunction <- function(class,sex,age,surv)
{
  i <- 0
  for (i in length(t)){
      if(class=='1st' && sex=='Female' && age=='Adult' && surv=='Yes'){
        df <- data.frame(class,sex,age,surv)
        return(df)
      }
  }
}

#Function calling with different arguments to check the functionality of it
myFunction('1st','Female','Adult','Yes')
```

```
##   class    sex   age surv
## 1   1st Female Adult  Yes
```

```r
myFunction('1st','Male','Child','Yes')
myFunction('3rd','Female','Adult','No')


percentFunction <- function()
{
  num <- length(df)
 return((num/2201)*100)
}

percentFunction()
```

```
## [1] 0.04543389
```

```r
ruleset <- apriori(t,parameter=list(support=0.07,confidence=0.4)) #76 rules
generated
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.4    0.1    1 none FALSE            TRUE       5    0.07      1
##  maxlen target   ext
##      10  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 154
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[10 item(s), 2201 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 3 4 done [0.00s].
## writing ... [76 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
summary(ruleset)
```

```
## set of 76 rules
##
## rule length distribution (lhs + rhs):sizes
##  1  2  3  4
##  4 27 35 10
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   2.671   3.000   4.000
##
## summary of quality measures:
##     support          confidence          lift
##  Min.   :0.07497   Min.   :0.4021   Min.   :0.6563
##  1st Qu.:0.09530   1st Qu.:0.6229   1st Qu.:0.9687
##  Median :0.21308   Median :0.7970   Median :1.0521
##  Mean   :0.27988   Mean   :0.7733   Mean   :1.1392
##  3rd Qu.:0.39164   3rd Qu.:0.9278   3rd Qu.:1.1887
##  Max.   :0.95048   Max.   :1.0000   Max.   :2.3017
##
## mining info:
##  data ntransactions support confidence
##     t          2201    0.07        0.4
```

```
inspect(ruleset)
```

```
##      lhs                                   rhs            support
## [1]  {}                                 => {Class=Crew}   0.40208996
## [2]  {}                                 => {Survived=No}  0.67696502
```

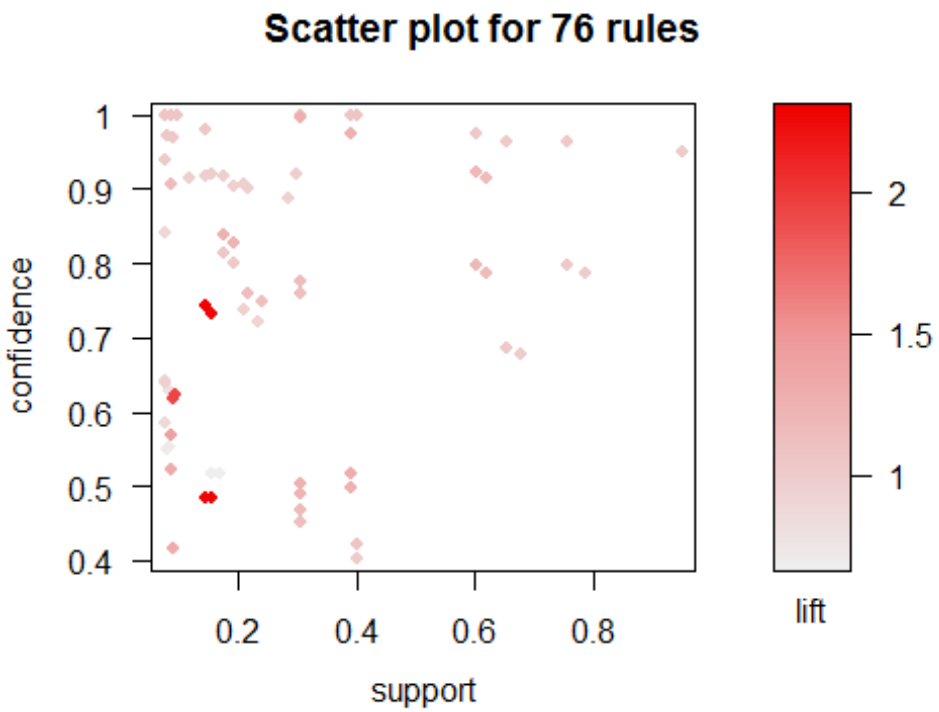
```
## [3]  {}                                => {Sex=Male}     0.78646070
## [4]  {}                                => {Age=Adult}    0.95047706
## [5]  {Class=2nd}                       => {Survived=No}  0.07587460
## [6]  {Class=2nd}                       => {Sex=Male}     0.08132667
## [7]  {Class=2nd}                       => {Age=Adult}    0.11858246
## [8]  {Class=1st}                       => {Survived=Yes} 0.09223080
## [9]  {Class=1st}                       => {Sex=Male}     0.08178101
## [10] {Class=1st}                       => {Age=Adult}    0.14493412
## [11] {Sex=Female}                      => {Class=3rd}    0.08905043
## [12] {Sex=Female}                      => {Survived=Yes} 0.15629259
## [13] {Survived=Yes}                    => {Sex=Female}   0.15629259
## [14] {Sex=Female}                      => {Age=Adult}    0.19309405
## [15] {Class=3rd}                       => {Survived=No}  0.23989096
## [16] {Class=3rd}                       => {Sex=Male}     0.23171286
## [17] {Class=3rd}                       => {Age=Adult}    0.28487051
## [18] {Survived=Yes}                    => {Sex=Male}     0.16674239
## [19] {Survived=Yes}                    => {Age=Adult}    0.29713766
## [20] {Class=Crew}                      => {Survived=No}  0.30577010
## [21] {Survived=No}                     => {Class=Crew}   0.30577010
## [22] {Class=Crew}                      => {Sex=Male}     0.39164016
## [23] {Sex=Male}                        => {Class=Crew}   0.39164016
## [24] {Class=Crew}                      => {Age=Adult}    0.40208996
## [25] {Age=Adult}                       => {Class=Crew}   0.40208996
## [26] {Survived=No}                     => {Sex=Male}     0.61971831
## [27] {Sex=Male}                        => {Survived=No}  0.61971831
## [28] {Survived=No}                     => {Age=Adult}    0.65333939
## [29] {Age=Adult}                       => {Survived=No}  0.65333939
## [30] {Sex=Male}                        => {Age=Adult}    0.75738301
## [31] {Age=Adult}                       => {Sex=Male}     0.75738301
## [32] {Class=2nd,Survived=No}           => {Age=Adult}    0.07587460
## [33] {Class=2nd,Age=Adult}             => {Survived=No}  0.07587460
## [34] {Class=2nd,Sex=Male}              => {Age=Adult}    0.07632894
## [35] {Class=2nd,Age=Adult}             => {Sex=Male}     0.07632894
## [36] {Class=1st,Survived=Yes}          => {Age=Adult}    0.08950477
## [37] {Class=1st,Age=Adult}             => {Survived=Yes} 0.08950477
## [38] {Class=1st,Sex=Male}              => {Age=Adult}    0.07950931
## [39] {Class=1st,Age=Adult}             => {Sex=Male}     0.07950931
## [40] {Class=3rd,Sex=Female}            => {Age=Adult}    0.07496592
## [41] {Sex=Female,Survived=Yes}         => {Age=Adult}    0.14357110
## [42] {Sex=Female,Age=Adult}            => {Survived=Yes} 0.14357110
## [43] {Age=Adult,Survived=Yes}          => {Sex=Female}   0.14357110
## [44] {Class=3rd,Survived=No}           => {Sex=Male}     0.19173103
## [45] {Class=3rd,Sex=Male}              => {Survived=No}  0.19173103
## [46] {Class=3rd,Survived=No}           => {Age=Adult}    0.21626533
## [47] {Class=3rd,Age=Adult}             => {Survived=No}  0.21626533
## [48] {Class=3rd,Sex=Male}              => {Age=Adult}    0.20990459
## [49] {Class=3rd,Age=Adult}             => {Sex=Male}     0.20990459
## [50] {Class=Crew,Survived=Yes}         => {Sex=Male}     0.08723308
## [51] {Sex=Male,Survived=Yes}           => {Class=Crew}   0.08723308
## [52] {Class=Crew,Survived=Yes}         => {Age=Adult}    0.09631985
```

```
## [53] {Sex=Male,Survived=Yes}              => {Age=Adult}    0.15356656
## [54] {Age=Adult,Survived=Yes}             => {Sex=Male}     0.15356656
## [55] {Class=Crew,Survived=No}             => {Sex=Male}     0.30440709
## [56] {Class=Crew,Sex=Male}                => {Survived=No}  0.30440709
## [57] {Sex=Male,Survived=No}               => {Class=Crew}   0.30440709
## [58] {Class=Crew,Survived=No}             => {Age=Adult}    0.30577010
## [59] {Class=Crew,Age=Adult}               => {Survived=No}  0.30577010
## [60] {Age=Adult,Survived=No}              => {Class=Crew}   0.30577010
## [61] {Class=Crew,Sex=Male}                => {Age=Adult}    0.39164016
## [62] {Class=Crew,Age=Adult}               => {Sex=Male}     0.39164016
## [63] {Sex=Male,Age=Adult}                 => {Class=Crew}   0.39164016
## [64] {Sex=Male,Survived=No}               => {Age=Adult}    0.60381645
## [65] {Age=Adult,Survived=No}              => {Sex=Male}     0.60381645
## [66] {Sex=Male,Age=Adult}                 => {Survived=No}  0.60381645
## [67] {Class=3rd,Sex=Male,Survived=No}     => {Age=Adult}    0.17582917
## [68] {Class=3rd,Age=Adult,Survived=No}    => {Sex=Male}     0.17582917
## [69] {Class=3rd,Sex=Male,Age=Adult}       => {Survived=No}  0.17582917
## [70] {Class=Crew,Sex=Male,Survived=Yes}   => {Age=Adult}    0.08723308
## [71] {Class=Crew,Age=Adult,Survived=Yes}  => {Sex=Male}     0.08723308
## [72] {Sex=Male,Age=Adult,Survived=Yes}    => {Class=Crew}   0.08723308
## [73] {Class=Crew,Sex=Male,Survived=No}    => {Age=Adult}    0.30440709
## [74] {Class=Crew,Age=Adult,Survived=No}   => {Sex=Male}     0.30440709
## [75] {Class=Crew,Sex=Male,Age=Adult}      => {Survived=No}  0.30440709
## [76] {Sex=Male,Age=Adult,Survived=No}     => {Class=Crew}   0.30440709
##      confidence lift
## [1]  0.4020900  1.0000000
## [2]  0.6769650  1.0000000
## [3]  0.7864607  1.0000000
## [4]  0.9504771  1.0000000
## [5]  0.5859649  0.8655764
## [6]  0.6280702  0.7986034
## [7]  0.9157895  0.9635051
## [8]  0.6246154  1.9335843
## [9]  0.5538462  0.7042261
## [10] 0.9815385  1.0326798
## [11] 0.4170213  1.3000904
## [12] 0.7319149  2.2657450
## [13] 0.4838256  2.2657450
## [14] 0.9042553  0.9513700
## [15] 0.7478754  1.1047474
## [16] 0.7223796  0.9185196
## [17] 0.8881020  0.9343750
## [18] 0.5161744  0.6563257
## [19] 0.9198312  0.9677574
## [20] 0.7604520  1.1233254
## [21] 0.4516779  1.1233254
## [22] 0.9740113  1.2384742
## [23] 0.4979780  1.2384742
## [24] 1.0000000  1.0521033
## [25] 0.4230402  1.0521033
```

```
## [26] 0.9154362  1.1639949
## [27] 0.7879838  1.1639949
## [28] 0.9651007  1.0153856
## [29] 0.6873805  1.0153856
## [30] 0.9630272  1.0132040
## [31] 0.7968451  1.0132040
## [32] 1.0000000  1.0521033
## [33] 0.6398467  0.9451696
## [34] 0.9385475  0.9874489
## [35] 0.6436782  0.8184492
## [36] 0.9704433  1.0210066
## [37] 0.6175549  1.9117275
## [38] 0.9722222  1.0228782
## [39] 0.5485893  0.6975420
## [40] 0.8418367  0.8856992
## [41] 0.9186047  0.9664669
## [42] 0.7435294  2.3016993
## [43] 0.4831804  2.2627237
## [44] 0.7992424  1.0162522
## [45] 0.8274510  1.2222950
## [46] 0.9015152  0.9484870
## [47] 0.7591707  1.1214326
## [48] 0.9058824  0.9530818
## [49] 0.7368421  0.9369090
## [50] 0.9056604  1.1515647
## [51] 0.5231608  1.3011038
## [52] 1.0000000  1.0521033
## [53] 0.9209809  0.9689670
## [54] 0.5168196  0.6571461
## [55] 0.9955423  1.2658514
## [56] 0.7772622  1.1481571
## [57] 0.4912023  1.2216230
## [58] 1.0000000  1.0521033
## [59] 0.7604520  1.1233254
## [60] 0.4680111  1.1639463
## [61] 1.0000000  1.0521033
## [62] 0.9740113  1.2384742
## [63] 0.5170966  1.2860221
## [64] 0.9743402  1.0251065
## [65] 0.9242003  1.1751385
## [66] 0.7972406  1.1776688
## [67] 0.9170616  0.9648435
## [68] 0.8130252  1.0337773
## [69] 0.8376623  1.2373791
## [70] 1.0000000  1.0521033
## [71] 0.9056604  1.1515647
## [72] 0.5680473  1.4127369
## [73] 1.0000000  1.0521033
## [74] 0.9955423  1.2658514
```

```
## [75] 0.7772622  1.1481571
## [76] 0.5041384  1.2537952

plot(ruleset)
```




```
#most interesting and useful rules

#Contains a value for lift which is large enough to make this a good rule,
and conditions which will be a valid condition to check
#[36] {Class=1st,Survived=Yes}  => {Age=Adult}    0.08950477 0.9704433
1.0210066

#Contains large value for lift, and conditions which will be a valid
condition to check
#[46] {Class=3rd,Survived=No}   => {Age=Adult}    0.21626533 0.9015152
0.9484870

#Contains maximum value for lift, and conditions which will be a valid
condition to check
#[42] {Sex=Female,Age=Adult}    => {Survived=Yes} 0.14357110 0.7435294
2.3016993
```