Lab 2

Association Rule Mining

The Titanic Dataset

- downloaded “titanic.raw.rdata” from http://www.rdatamining.com/data.

```
> str(Titanic)

 table [1:4, 1:2, 1:2, 1:2] 0 0 35 0 0 0 17 0 118 154 ...
 - attr(*, "dimnames")=List of 4
  ..$ Class   : chr [1:4] "1st" "2nd" "3rd" "Crew"
  ..$ Sex     : chr [1:2] "Male" "Female"
  ..$ Age     : chr [1:2] "Child" "Adult"
  ..$ Survived: chr [1:2] "No" "Yes"
```

The Titanic Dataset

```
> str(titanic.raw)

data.frame:       2201 obs. of  4 variables:
 $ Class   : Factor w/ 4 levels "1st","2nd","3rd",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ Sex     : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 2 ...
 $ Age     : Factor w/ 2 levels "Adult","Child": 2 2 2 2 2 2 2 2 2 2 ...
 $ Survived: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```
> head(titanic.raw)
  Class  Sex   Age Survived
1   3rd Male Child       No
2   3rd Male Child       No
3   3rd Male Child       No
4   3rd Male Child       No
5   3rd Male Child       No
6   3rd Male Child       No
```

Basic `arules()` function

```
> install.packages("arules")
> library(arules)
> rules.all <- apriori(titanic.raw)
> rules.all
> inspect(rules.all)
```

arules() parameters

```
> rules<-apriori(titanic.raw, control=list(verbose=F),
                 parameter=list(minlen=2,supp=0.005,conf=0.8),
                 appearance = list(rhs=c("Survived=No",
                                         "Survived=Yes"),
                 default="lhs"))

> quality(rules)<-round(quality(rules),digits=3)
> rules.sorted <- sort(rules,by="lift")
> inspect(rules.sorted)
```

`arules()` parameters

Some key parameters:

- rhs containing survival information only, the rest not interested: `rhs=c("Survived=No", "Survived=Yes")`
- `Default ="lhs"` => All other items can appear in the lhs.
- To suppress the details of progress use `verbose=F`.
- `minlen=2` means `lhs` consists of at least two items.

Removing Redundancy

- Rule 2 is a redundant rule for Rule 1. Adding `sex=Female` will not affect the rule.

```
    lhs                rhs             support confidence  lift
1   {Class=2nd,
     Age=Child}     => {Survived=Yes}    0.011      1.000 3.096
2   {Class=2nd,
     Sex=Female,
     Age=Child}     => {Survived=Yes}    0.006      1.000 3.096
```

Removing Redundancy

- To find redundant rules:

```
> # find redundant rules
> subset.matrix <- is.subset(rules.sorted, rules.sorted)
> subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
> redundant <- colSums(subset.matrix, na.rm=T) >= 1
> which(redundant)

[1] 2 4 7 8

> # remove redundant rules
> rules.pruned <- rules.sorted[!redundant]
> inspect(rules.pruned)
```

Matrix: `lower.tri()`

```
> m2 <- matrix(1:16, 4, 4)
> lower.tri(m2)
> m2[lower.tri(m2)] <- NA
> m2
```

```
> m2
     [,1] [,2] [,3] [,4]
[1,]    1    5    9   13
[2,]   NA    6   10   14
[3,]   NA   NA   11   15
[4,]   NA   NA   NA   16
```

Interpreting Rules

```
> rules <- apriori(titanic.raw,
+                  parameter = list(minlen=3, supp=0.002, conf=0.2),
+                  appearance = list(rhs=c("Survived=Yes"),
+                                    lhs=c("Class=1st", "Class=2nd", "Class=3rd",
+                                          "Age=Child", "Age=Adult"),
+                                    default="none"),
+                  control = list(verbose=F))
> rules.sorted <- sort(rules, by="confidence")
> inspect(rules.sorted)
```

Interpreting Rules

	lhs	rhs	support	confidence	lift
1	{Class=2nd, Age=Child}	=> {Survived=Yes}	0.010904134	1.0000000	3.0956399
2	{Class=1st, Age=Child}	=> {Survived=Yes}	0.002726034	1.0000000	3.0956399
3	{Class=1st, Age=Adult}	=> {Survived=Yes}	0.089504771	0.6175549	1.9117275
4	{Class=2nd, Age=Adult}	=> {Survived=Yes}	0.042707860	0.3601533	1.1149048
5	{Class=3rd, Age=Child}	=> {Survived=Yes}	0.012267151	0.3417722	1.0580035
6	{Class=3rd, Age=Adult}	=> {Survived=Yes}	0.068605179	0.2408293	0.7455209

Visualization Rules

```
> install.packages("arulesViz")
> library(arulesViz)
> plot(rules.all)
> plot(rules.all, method="grouped")
> plot(rules.all, method="graph")
```

Visualization Rules