Chapter 13 – Association Rules关联规则分析

Instructor: Zach Zhizhong ZHOU,
Shanghai Jiao Tong University
主讲教师：周志中，上海交通大学

Data Mining for Business Intelligence

Shmueli, Patel & Bruce

Titanic - 1

```
library(arules) ## needed for Association Rules Mining
setwd("C:/BA/AR")

Titanic <- read.csv("Titanic.csv",header = TRUE)

rules <- apriori(Titanic)
inspect(rules)
```

Titanic - 2

```
# 下面找出后项为“Survived”的规则。

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

# minlen - an integer value for the minimal number of
items per item set (default: 1)

# supp - a numeric value for the minimal support of an
item set (default: 0.1)

# conf - a numeric value for the minimal confidence of
rules/association hyperedges (default: 0.8)

rules.sorted <- sort(rules, by="lift")

inspect(rules.sorted)
```

Titanic - 3

```
# 找到冗余规则
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)

# 移除冗余规则
rules.pruned <- rules.sorted[!redundant]
inspect(rules.pruned)
```