# Data Mining Lab, Exercise 7

Team# 1: Poliakov Valerii, Holovnia Dmytro, Selvaraj Sinju

Dataset: vote1.csv

To work with apriori we need arules library. To use plot with rules generated by the apriori we need arulesViz library. Also, we need to read data set.

> library("arules")

> library("arulesViz")

> #Read the whole data set

> d<-read.csv("vote1.csv")

Generate association rules with no more than 5 antecedents, the confidence of 80% and the support of 30%.

> rules\_democrat<-apriori(d, parameter=list(supp=0.3, conf = 0.8, maxlen=5),

appearance = list(default="lhs", rhs="class=democrat"))

Apriori

Parameter specification:

confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext

0.8 0.1 1 none FALSE TRUE 5 0.3 1 5 rules TRUE

Algorithmic control:

filter tree heap memopt load sort verbose

0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 41

set item appearances ...[1 item(s)] done [0.00s].

set transactions ...[32 item(s), 138 transaction(s)] done [0.00s].

sorting and recoding items ... [31 item(s)] done [0.00s].

creating transaction tree ... done [0.00s].

checking subsets of size 1 2 3 4 5 done [0.00s].

writing ... [380 rule(s)] done [0.00s].

creating S4 object ... done [0.00s].

Let’s check first 10 rules ordered by confidence.

> inspect(sort(rules\_democrat[1:10], by="confidence", decreasing=TRUE))

lhs rhs support confidence coverage lift count

[1] {physician.fee.freeze=n} => {class=democrat} 0.6014493 1.0000000 0.6014493 1.568182 83

[2] {duty.free.exports=y} => {class=democrat} 0.4275362 0.9672131 0.4420290 1.516766 59

[3] {crime=n} => {class=democrat} 0.3913043 0.9642857 0.4057971 1.512175 54

[4] {superfund.right.to.sue=n} => {class=democrat} 0.4420290 0.9104478 0.4855072 1.427748 61

[5] {mx.missile=y} => {class=democrat} 0.4710145 0.9027778 0.5217391 1.415720 65

[6] {adoption.of.the.budget.re=y} => {class=democrat} 0.5144928 0.8987342 0.5724638 1.409379 71

[7] {religious.groups.in.schoo=n} => {class=democrat} 0.3043478 0.8936170 0.3405797 1.401354 42

[8] {aid.to.nicaraguan.contras=y} => {class=democrat} 0.5144928 0.8875000 0.5797101 1.391761 71

[9] {handicapped.infants=y} => {class=democrat} 0.3623188 0.8333333 0.4347826 1.306818 50

[10] {anti.satellite.test.ban=y} => {class=democrat} 0.4565217 0.8076923 0.5652174 1.266608 63

With help of arulesViz library we can use plot for the rules.

> plot(rules\_democrat)

Зображення, що містить текст, схема, Графік, ряд

Автоматично згенерований опис

Next step is to generate association rules with same settings except the confidence, it will be 95%:

> rules\_democrat95<-apriori(d, parameter=list(supp=0.3, conf = 0.95, maxlen=5),

appearance = list(default="lhs", rhs="class=democrat")) Apriori

Parameter specification:

confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext

0.95 0.1 1 none FALSE TRUE 5 0.3 1 5 rules TRUE

Algorithmic control:

filter tree heap memopt load sort verbose

0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 41

set item appearances ...[1 item(s)] done [0.00s].

set transactions ...[32 item(s), 138 transaction(s)] done [0.00s].

sorting and recoding items ... [31 item(s)] done [0.00s].

creating transaction tree ... done [0.00s].

checking subsets of size 1 2 3 4 5 done [0.00s].

writing ... [294 rule(s)] done [0.00s].

creating S4 object ... done [0.00s].

List part of the rules:

> inspect(sort(rules\_democrat95[1:10], by="confidence", decreasing=TRUE))

lhs rhs support confidence coverage lift count

[1]{physician.fee.freeze=n} => {class=democrat} 0.6014493 1.0000000 0.6014493 1.568182 83

[2]{physician.fee.freeze=n,

religious.groups.in.schoo=n} => {class=democrat} 0.3043478 1.0000000 0.3043478 1.568182 42

[3]{crime=n, duty.free.exports=y} => {class=democrat} 0.3260870 1.0000000 0.3260870 1.568182 45

[4]{superfund.right.to.sue=n, crime=n} => {class=democrat} 0.3405797 1.0000000 0.3405797 1.568182 47

[5]{mx.missile=y, crime=n} => {class=democrat} 0.3695652 1.0000000 0.3695652 1.568182 51

[6]{anti.satellite.test.ban=y, crime=n} => {class=democrat} 0.3623188 1.0000000 0.3623188 1.568182 50

[7]{adoption.of.the.budget.re=y, crime=n} => {class=democrat} 0.3840580 1.0000000 0.3840580 1.568182 53

[8]{aid.to.nicaraguan.contras=y, crime=n} => {class=democrat} 0.3913043 1.0000000 0.3913043 1.568182 54

[9]{duty.free.exports=y} => {class=democrat} 0.4275362 0.9672131 0.4420290 1.516766 59

[10]{crime=n} => {class=democrat} 0.3913043 0.9642857 0.4057971 1.512175 54

Now let’s check the plot for the confidence 95%.

> plot(rules\_democrat95)

Зображення, що містить текст, Графік, ряд, число

Автоматично згенерований опис

Comparing the 2 graphs we created based on the rules, we can see that the 95% confidence has eliminated all the noise we have with the 80% confidence. Also rules list shows us that with confidence 95% left hand side has up to 2 attributes.