

Report for Data Mining Assignment 1

Language used : C++

Preprocessing on data:

Stored vote.arff file in txt format

y is treated as 1 and n treated as 0 republic treated as 0 , democrat treated as 1

Binary attributes stored in 2d array named a

For each attribute a negation attribute is also considered when attribute value is 0 then negation of that attribute

has value 1 and vice versa.

? treated as no for that attribute

Hash tree generation

To count support hash tree is used to solve problem with least complexity. Hash tree is generated each time

when k size frequent itemsets are calculated. To calculate k+1 size frequent itemset k size itemset are used

and hash tree implementation is used to get all possible frequent itemsets.

Rule generation

After getting all frequent itemsets rules are generated from those. For each frequent dataset

lattice of rules is generated and is pruned whenever it gets below min_conf. And then achieved rules are stored in a vector and then it is sorted in descending order of confidence for that rule.

Output rules

Special rules with high confidence are included in output where rhs includes only republic or democrat attribute.

Compilation steps

Compile rule_mining.cpp along with file input.txt in same directory

Analysis:

min_support = 230

min_conf = 0.9

Number of rules generated = 258

min_support = 230

min_conf = 0.8

Number of rules generated= 325

min_support = 200

min_conf = 0.9

Number of rules generated = 567

min_support = 200

min_conf = 0.8

Number of rules generated = 735

Optimized rules generated :

NO export-administration-act-south-africa support = 350 ==> democrat support = 427

NO duty-free-exports support = 317 ==> democrat support = 427

NO el-salvador-aid support = 354 ==> republic support = 398

duty-free-exports support = 414 ==> democrat support = 427

NO handicapped-infants && NO duty-free-exports support = 385 ==> democrat support = 427

physician-fee-freeze && NO duty-free-exports support = 378 ==> democrat support = 407

crime && NO education-spending support = 104 ==> republic support = 117

NO handicapped-infants && NO superfund-right-to-sue && NO duty-free-exports support = 255
==> democrat support = 427

Team Members:

Ayush Sharma 2014A7PS039H

Uddhav Mishra 2014A7PS048H

Chitresh Sharma 2014A7PS121H

Varun Singh 2014A7PS184H