

Association Analysis

Apriori Algorithm and Associate Rule generation

Members:

Jay Shah – 50205647(jaynaren)

Yash Jain – 50206851(yashnavi)

Akshay Shah – 50206543(akshaybh)

STAGE 1: Frequent item-sets generation using Apriori Algorithm.

1. Read the dataset file 'associationruletestdata.txt' in a multidimensional array and created candidate items and their calculated their frequency.
2. Pruned the candidate items according to the minimum support threshold, leaving us with the frequent candidate item-sets.
3. Perform self-join on the frequent item-sets and generate new candidate item-sets.
4. Repeated step 2 and step 3 till the list of generated frequent item-set is empty, i.e. No more frequent item-sets are generated. Stored all the frequent item-sets generated on all the stages into a single list/set.
5. Repeated from Step 2 to Step 5 for different value of support threshold and compare the results.
6. The result will be stored in two ways.
 - a. A Map which contains all the item-sets with their support count.
 - b. A queue, which contains all the Rule class object. This rule class object will contain two sets representing Right and Left side of the rule.

STAGE 2: Association rules generation from Frequent item-sets.

1. From the previous stage, we use the queue objects to generate rules.
2. De-queue a rule object.
3. Remove an item from left set and add it to right set.
4. Compute confidence for the rule and if it is above the threshold confidence add it to the Queue.
5. Repeat step 2 till the queue is empty.
6. Store all the rules into a Set or a list.

RESULT:

STAGE 1: Frequent Item- set generation

Support	Frequent Item-sets
30%	Length 1 Item-sets: 196
	Length 2 Item-sets: 5340
	Length 3 Item-sets: 5287
	Length 4 Item-sets: 1518
	Length 5 Item-sets: 438
	Length 6 Item-sets: 88
	Length 7 Item-sets: 11
	Length 8 Item-sets: 1
	Total: 12879
40%	Length 1 Item-sets: 167
	Length 2 Item-sets: 753
	Length 3 Item-sets: 149
	Length 4 Item-sets: 7
	Length 5 Item-sets: 1
	Total: 1077

50%	Length 1 Item-sets: 109
	Length 2 Item-sets: 63
	Length 3 Item-sets: 2
	Total: 174
60%	Length 1 Item-sets: 34
	Length 2 Item-sets: 2
	Total: 36
70%	Length 1 Item-sets: 7

STAGE 2: Sample template Query results.

Query	Result(No of rules)
RULE "ANY" ["G59_Up"]	26
RULE "NONE" ["G59_Up"]	91
RULE 1 ["G59_Up","G10_Down"]	39
BODY "ANY" ["G59_Up"]	9
BODY "NONE" ["G59_Up"]	108
BODY 1 ["G59_Up","G10_Down"]	17
HEAD "ANY" ["G59_Up"]	17
HEAD "NONE" ["G59_Up"]	100
HEAD 1 ["G59_Up","G10_Down"]	24
RULE 3	9
BODY 2	6
HEAD 1	117
1or1 "BODY" "ANY" ["G10_Down"],"HEAD" 1 ["G59_Up"]	24
1and1 "BODY" "ANY" ["G10_Down"], "HEAD" 1 ["G59_Up"]	1
1or2 "BODY" "ANY" ["G10_Down"],"HEAD" 2	11
1and2 "BODY" "ANY" ["G10_Down"],"HEAD" 2	0
2or2 "BODY" 1 "HEAD" 2	117
2and2 "BODY" 1 "HEAD" 2	3