Lab 9

Association rule learning

Alex Becheru
irlab.becheru.net
irlab@becheru.net

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases.

Example

database:

alpha beta epsilon

alpha beta theta

alpha beta epsilon

alpha beta theta

rules:

- 100% alpha & beta
- 50% alpha, beta & epsilon
- 50% alpha, beta & theta

Association rules are intended to identify strong rules discovered in databases using different measures of interestingness

Support measure

supp(X), where X is an item in the data set

supp(alpha,beta,epsilon)=2/4=0.5

Confidence of a rule

conf(X=>Y)=supp(X,Y)/supp(X)

conf(X=>Y)= nr of items where where both X and Y appear / nr of items where X appears

conf(alpha,beta => epsilon) = 2/4 = 0.5

Lift of a rule

lift(X => Y) = supp(X,Y)/ supp(X) * supp(Y)

The ratio of the observed support to that expected if X and Y were independent.

Conviction of a rule

$$conv(X \Rightarrow Y) = (1 - supp(Y))/(1-conf(X \Rightarrow Y))$$

The frequency that the rule makes an incorrect prediction if X and Y were independent

Apriori algorithm

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time.

Apriori treshold

A frequent item is an item that appears in the database more frequent than a threshold.

(given by us)

Example Apriori

Itemsets:

Threshold =3

1,2,3,4

1,2,4

1,2

2,3,4

2,3

3,4

2,4

Items of 1 element

Itemsets:

1,2,3,4

1,2,4

1,2

2,3,4

2,3

3,4

2,4

Item	Support
1	3
2	6
3	4
4	5

Items of 2 elements

Itemsets:
1,2,3,4
1,2,4
1,2
2,3,4
2,3
3,4
2,4

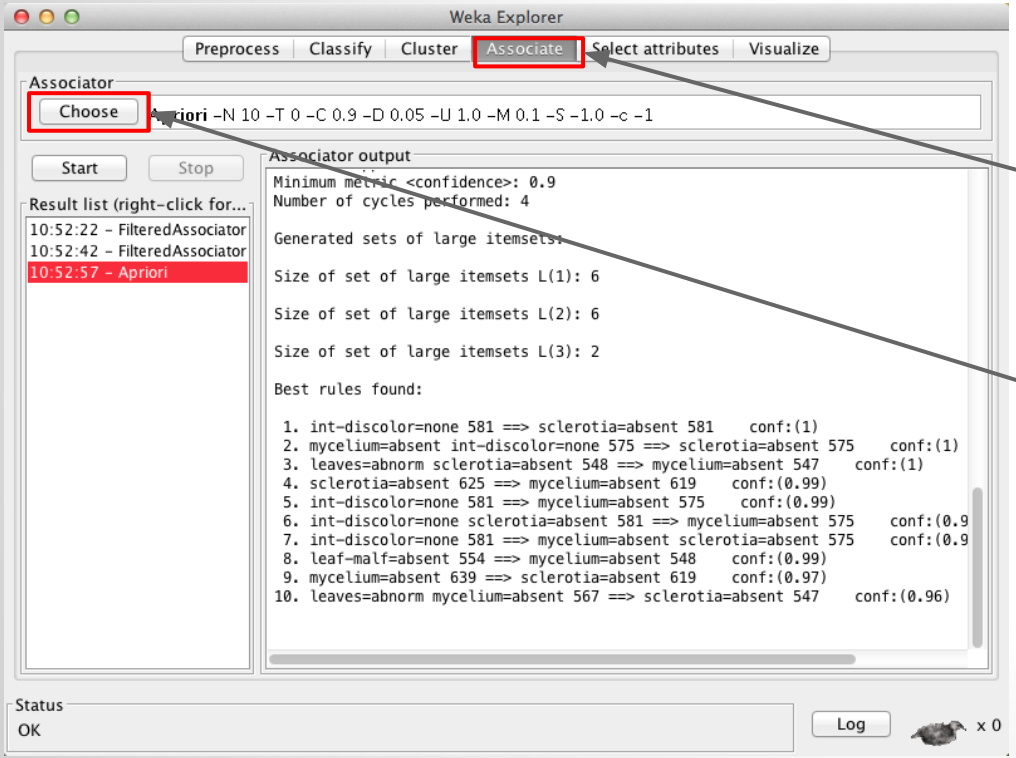
Item	Support
1,2	3
1,3	1
1,4	2
2,3	3
2,4	4
3,4	3

Items of 3 elements

Itemsets:
1,2,3,4
1,2,4
1,2
2,3,4
2,3
3,4
2,4

Item	
2,3,4	2

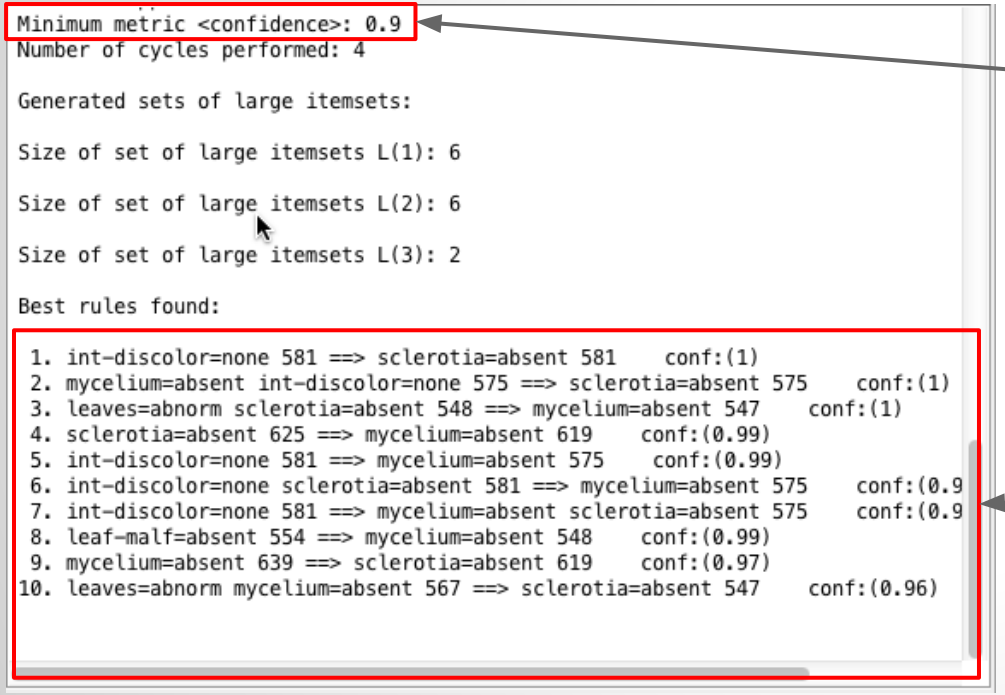
Association rules in Weka



Associate panel in Explorer

Choose algorithm

Apriori output



Minimum confidence

Best rules discovered

Ex.1

- open soybean.arff
- use the apriori algorithm to discover new rules

Questions?


THANK
YOU