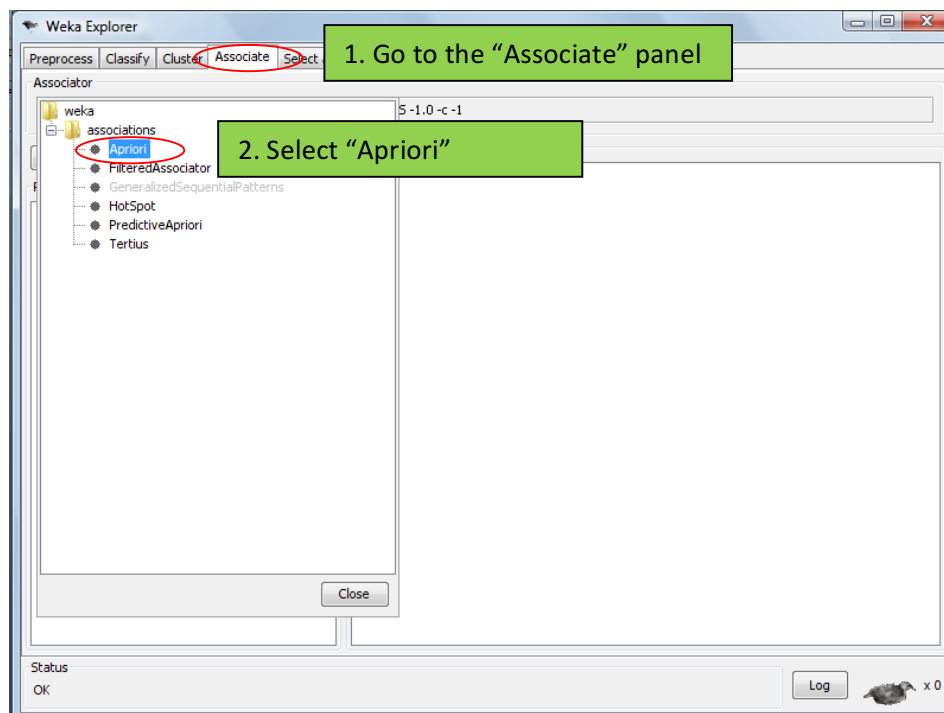


Weka Association Rules

Start Weka Explorer. Import the nominal weather data (weather_nominal_arff.arff) on the “Preprocess” panel.

1. Using Apriori

Go to the “Associate” panel. Click the “Choose” button on the top left, and then choose “weka->associations->Apriori”. Apriori needs nominal attributes, so you will need to discretize numeric values if there are numeric attributes in your data set. But the weather dataset is all nominal.



Click on the “Start” button, and you will see the result shortly.

Tune parameters

Click on the textbox on the right of the “Choose” button on the “Associate” panel. In the pop-up menu “weka.gui.GenericObjectEditor” you can edit the parameters. Here are some parameters you can play with:

delta: decrease rate of minimum support; the minimum support values get reduced by this rate and the algorithm stops either when the lowerBoundMinSupport is reached or when numRules is reached.

lowerBoundMinSupport: lower bound for minimum support.

metricType: the approach to ranking rules.

minMetric: the threshold for selecting rules.

numRules: number of best rules to be displayed; it's the confidence threshold for Aprior.

outputItemSets: whether to output item sets. You will see item sets and their frequencies in data set if it's set to true.

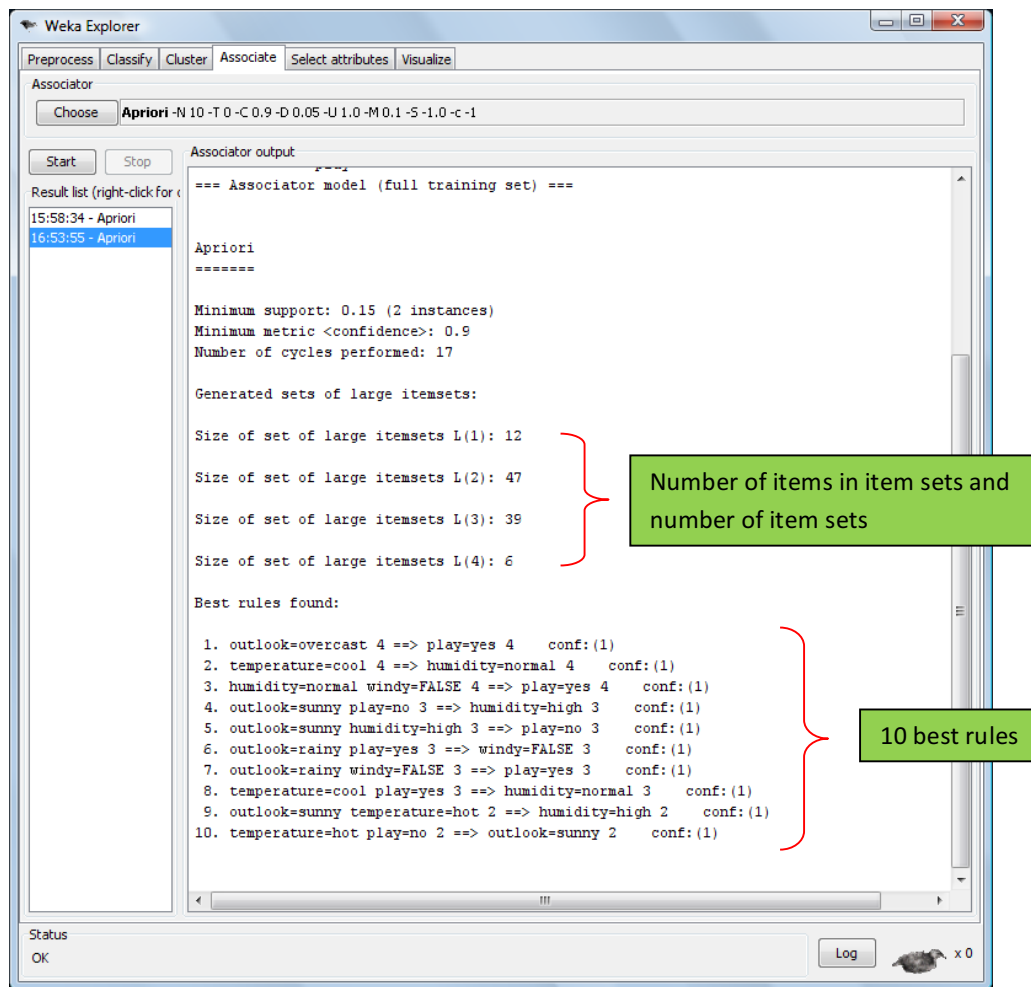
upperBoundMinSupport: upper bound for minimum support.

The lowerBoundMinSupport decreases as the algorithm iterates. The table below shows how it'd iterate based on the default setting. Note that lowerBoundMinSupport decreases based on value of delta, meaning it is 0.05 less than its value of previous iteration.

iteration	lowerBoundMinSupport	delta
1	1	0.05
2	0.95	0.05
3	0.9	0.05
...	...	0.05
19	0.1	0.05

Explaining Results

The chart below shows the result (the part after run information section). The first part of the chart lists minimum support, minimum metric, and number of cycles performed. We can find that the algorithm stopped before lowerBoundMinSupport is reached, since minimum support=0.15 in result. Correspondingly, the number of cycles performed is 17, which means there are 17 iterations of applying Aprior before the program ceased. The minimum metric (confidence threshold here) is 0.9 as specified in parameter settings.



The second part lists size of item sets and the corresponding number of item sets at this size. For example, the first line means there are 12 item sets when size of item set is 1.

The third part shows rules. Let's look at the first line for example. The antecedent (left side) is "outlook=overcast", and its frequency in data set is 4; the consequent (right side) is "play=yes" and its frequency is 4 when the antecedent is true; the confidence is 1 (right 4 divided by left 4).

Note that Support for the rule is not explicitly listed, just the support for the LHS and the RHS.

Also, when the rules are sorted by the Confidence measure, the results don't report the other measures. So open the ObjectEditor and under the MetricType, select Lift, save and Start the algorithm. Now note that the results report all 4 measures when Lift is used to sort the rules.

2. Other algorithms

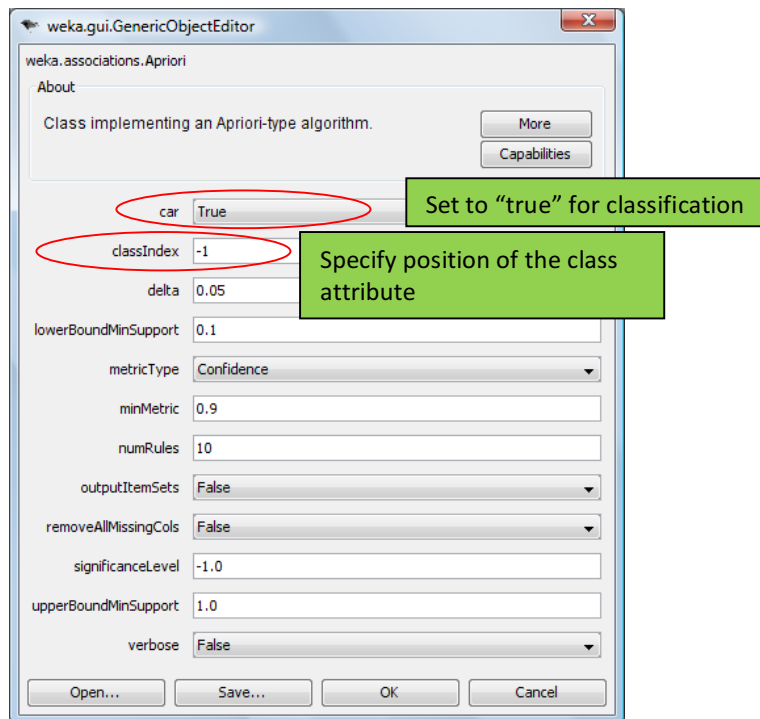
Besides Apriori algorithm, note that there are other algorithms like FPGrowth.

3. Using association rules for classification

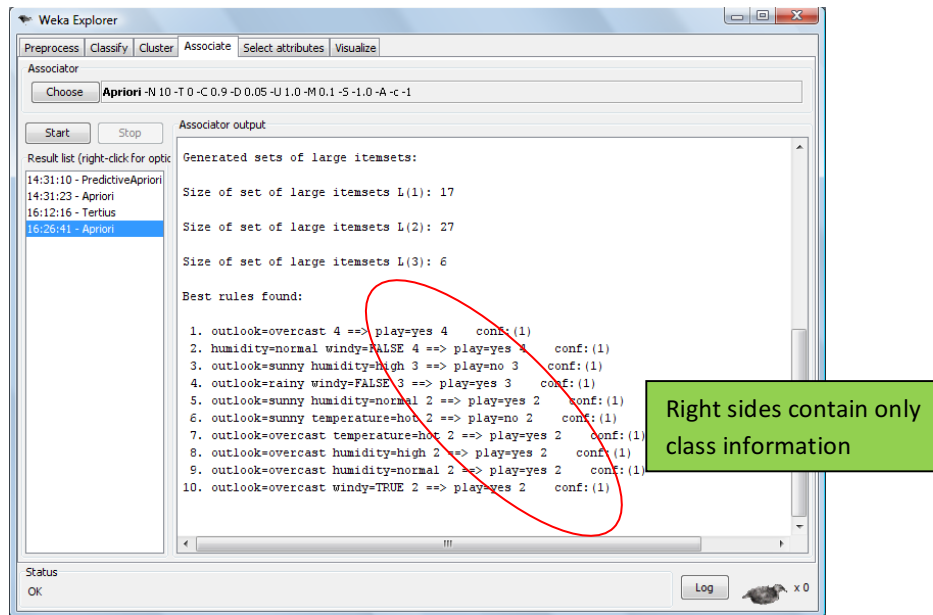
If we constrain the consequent (right side) of rules to be the class attribute, then we get association rules for classification. This restricts the right hand side of the rules to be the class label.

Still choose the “Apriori” algorithm as in section 1. Then click on the parameter textbox besides the “Choose” button to invoke the “weka.gui.GenericObjectEditor”.

Set two parameters: 1) set “car” to true; this tells Weka that the right side of rules will be class attribute; 2) set classIndex by specifying index of the class attribute; the default value “-1” means the class value is the last column, and you can keep it as default if that’s the case in your data; otherwise you need to specify the position of your class attribute. Set the MetricType to be Confidence.



Then click on the “Start” button. In the result section, you will see rules whose right sides contain only class attribute. For example, according to the first rule, if outlook is overcast, then the class value (play) is “yes”. When a new instance comes in, you will need match its attribute values with the mined rules and decide which class it belongs to.



4. Convert transaction data to record data for Weka

Weka cannot directly read in transaction data to run association rules. This wiki entries showed how to use Weka filters to pre-process the transaction data for association rule mining.

<https://weka.wikispaces.com/How+can+I+use+transactional+data+in+Weka%3F>

Exercise: Titanic Data Set

Since the Weather dataset is small, we will also look at the Titanic dataset. Load this dataset and in Preprocessing, use Discretize to convert Numeric attributes into Nominal.

Run the Apriori algorithm first with the Confidence measure. Try lowering the confidence to .8 in order to get more rules.

Now run the Apriori algorithm with the Lift measure. Look at rules with Lift greater than 1 and as high Confidence as possible. Also estimate the Support of these rules.

Post three interesting rules.