Practical 02

WEKA Tool

We use WEKA (www.cs.waikato.ac.nz/ml/weka/), an open source data mining tool. WEKA is developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. WEKA is a tool for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data pre-processing, feature reduction, classification, regression, clustering, and association rules. It also includes visualization tools. The new machine learning algorithms can be used with it and existing algorithms can also be extended with this tool.

Association Rule Mining.

Association rule mining is considered as a Major technique in data mining applications. It reveals all interesting relationships, called associations, in a potentially large database. However, how interesting a rule is depends on the problem a user wants to solve. Existing approaches employ different parameters to guide the search for interesting rules.

The key strength of association rule mining is that all interesting rules are found.

EXERCISE: Mining Association Rule with WEKA Explorer

Apriori Algorithm

1. To get a feel for how to apply Apriori algorithm, start by mining association rules from the **weather.nominal.arff** dataset of Lab One. Note that Apriori algorithm expects data that is purely nominal: if there are any numeric attributes they must be discretized first.

2. Like in the previous example choose Associate and Click Start button on the left of the window, the algorithm begins to run. The output is showing in the right window.

3. You could re-run Apriori algorithm by selecting different parameters, such as lowerBoundMinSupport, minMetric (min. confidence level), and different evaluation metric (confidence vs. lift), and so on.

FP- Growth Algorithm

Find all the frequent item sets of the contact-lenses.arff dataset using the FP growth algorithm with default parameters. And try with changing the parameters as well.