

# COMP37332

## WEKA: Association Rule Mining tutorial<sup>1</sup>

In this tutorial, you will explore association rule mining. In the first part you will look at the contact-lenses.arff data. This small data set contains information about fitting contact lenses.

In the second part your specific task is to mine rules associating animal characteristics — such as hair, producing milk, etc. — to animal types, such as mammals.

You will use the implementation of the Apriori algorithm in Weka. Note that this implementation starts with a maximum support level, and reduces it until the desired number of rules is generated.

### **Task A**

A1. Download the contact lenses dataset from the Course web page (contact-lenses.arff) and load it into Weka. Check that all attributes are nominal (categorical).

A2. Change to the Associate Panel. Select 'Apriori' as associator. After pressing the start button, Apriori starts to build its model and writes its output into the output field. The first part of the output ('Run information') describes the options that have been set and the data set used. Make sure you understand all the data reported.

A3. The rules that have been generated are listed at the end of the output. By default, only the 10 most valuable rules according to their *confidence* level are shown. Each rule consists of some attribute values on a left hand side of the arrow, the arrow sign and the right hand side list of attribute values. Right of the arrow sign are the predicted attribute values. Rules have certain support and confidence values. The number before the arrow sign is the number of instances the rule applies to. The number after the arrow sign is the number of instances predicted correctly. The number in brackets after 'conf:' is the confidence of the rule. Analyse the rules mined from the data set. What are their confidence and support values? Examine the number of large itemsets – make sure you understand how this data has been calculated (check that the values you would get 'manually' are correct).

---

<sup>1</sup> Tailored after: *Labs 6&7:Association Rules* (MSCS 228: Data Mining) by Dr Craig A. Struble, Marquette University

## **Task B**

B1. Download the zoo dataset from the Course web page (zoo.arff) and load it into Weka. Examine the attributes and make sure you understand their meaning. Are all attributes nominal?

B2. In the preprocess area, deselect the animal and legs attributes. The animal attribute is the name of the animal, and is not useful for mining. The legs attribute is numeric and cannot be used directly with Apriori. Alternatively, you can try to use the Discretize Filter to discretize the legs attribute.

B3. After deselecting the attributes, use the *Apply Filters* button to generate a working relation that removes those attributes. Notice how the working relation changes, and has fewer attributes than the base relation.

B4. First, try using the Apriori algorithm with the default parameters. Record the generated rules.

B5. Vary the number of rules generated (click on the command that you are running). Try 20, 30, ... Record how many rules you have to generate before generating a rule containing *type=mammal*.

B6. Vary the maximum support until a rule containing *type=mammal* is the top rule generated. Record the maximum support needed.

B7. Select one generated rule that was interesting to you. Why was it interesting? What does it mean? Check its confidence and support – are they high enough?

B8. Suggest one improvement to the Apriori implementation in Weka that would have made this data mining lab easier to accomplish.