Armando Navarro
John Sanchez

User Manual for Tour of Data Mining Algorithms

Implementation Project by Armando Navarro and John Sanchez

Note that Java 8 is required to run this program successfully. If an exception occurs, make sure you have the latest version of Java installed.

Before you run the program, ensure the folowing:

- Latest version of the Java Runtime Environment is installed
- All test data set files are in the same folder/directory as the dm-proj.jar file
- Open a terminal in the folder containing the .jar and testing files

Each option shown has a short-hand single-character specifier and a long-hand specifier. All options can be shown by running the program without options, as described below.

Since it is a Java jar, no compilation is needed. You can simply type the following to get the help message:

```
java -jar dm-proj.jar
```

APRIORI

To run Apriori on the simple data set, type:

```
java -jar dm-proj.jar -a apriori -i apriori_simple-test.txt --min-sup 2 -d comma
```

Note that **-d** specifies the input file attribute delimiter, which is space by default. However, the simple data set uses a comma delimiter, hence the option is shown above.

To run Apriori on the Belgium retail market data set, type:

```
java -jar dm-proj.jar -a apriori -i apriori_belgium-retail-market.dat --min-sup 1500
```

To see how to specify children per node or max bucket size, pull up the help message as described above. These options are not required.

Armando Navarro
John Sanchez

ID3

To run ID3 on the simple data set, type:

```
java -jar dm-proj.jar -a id3 -i id3_simple-training.txt --label-index
              4 --testing-file id3_simple-testing.txt
```

In addition, training file “id3_simple-training-2.txt” and “id3_simple-testing-2.txt” are also provided. Simply modify the line above.

XMeans

To run Xmeans on the simple data set, type:

```
java -jar dm-proj.jar -a xmeans -i xmeansMidTest --min-k 1 --max-k 5
```

This will only show the centroids of the clusters to reduce output clutter. For more verbose output, use -v (this option only applies to XMeans). Since data sets ca

In addition, training file “xmeansSmallTest” is also provided. Simply modify the line above.