

JClose User Guide

Nicolas Pasquier

Contact : I3S Laboratory – University of Nice / CNRS UMR-6070,
2000 route des Lucioles, 06903 Valbonne, France

E-mail: Nicolas.Pasquier@unice.fr

1 Installation

To run JClose, you need the Java Runtime Environment (JRE). The latest version of the Java Runtime Environment can be downloaded from the page: <http://www.java.com/en/download/>.

All distributions include a graphical installer which covers all the installation procedure.

Note: to check whether Java is installed on you machine, visit this page

<http://www.java.com/en/download/help/testvm.xml>

2 Usage

Just download the `JClose-1.0.jar` file and run it in command-line mod. To display the in-line help, just type in the terminal the command:

```
java -jar JClose-1.0.jar
```

The format of the command to launch JClose is:

```
java Close dataFile -g -i -k -t -s=minSupport -c=minConfidence
```

Description of the command-line parameters:

Parameter	Issue
<code>dataFile</code>	Name of the <code>.csv</code> or <code>.arff</code> data file. If the file is not located in the same directory as the jar file, you must specify the complete path on disk in the name.
<code>-g</code>	Generate <code>.data</code> and <code>.alias</code> files. These files are necessary if they are not already present in data file folder, as a result of a previous execution of JClose on the same dataset.
<code>-k</code>	Do not delete all temporary files generated after the execution. These files are the <code>.data</code> , <code>.alias</code> , <code>.fci</code> , <code>.mme</code> , <code>.mma</code> files used to improve the efficiency of the application on large datasets.
<code>-i</code>	Ignore first attribute (line ID). This option is required if the first column of the data contains a line identifier, such as the Gene ID for instance. If this option is not used, the first column of the data will be used as an item and appear in association rules.
<code>-t</code>	Use only tabulation as cell separator (spaces, commas, etc.). This option is required if data columns contain special characters.
<code>-s</code>	Minimum support value in $[0.0, 1.0]$ (percentage)
<code>-c</code>	Minimum confidence value in $[0.0, 1.0]$ (percentage)

3 Results

The JClose application generates three files in text format:

- `fileName.fci.nom` : contains the description of *equivalence classes* extracted from the dataset. Each line describes the *generator*, *closure* and *support_count* of an equivalence class
- `fileName.mme.nom` : contains all *exact minimal non-redundant association rules* extracted. Each line contains a minimal non-redundant association rules with confidence = 100% (exact association rule).

- `fileName.mma.nom` : contains all *partial minimal non-redundant association rules* extracted. Each line contains a minimal non-redundant association rules with confidence $< 100\%$ (partial, or approximate, association rule).

The generation of these files depends on the `minsupport` and `minconfidence` thresholds.

If no equivalence class has a support greater or equal than the one defined by the `minsupport` parameter, no file will be generated.

If no partial association rule has a confidence greater or equal than the one defined by the `minconfidence` parameter, the `fileName.mma.nom` file will not be generated.

4 References

Details on equivalence classes of itemsets and minimal non-redundant association rules can be found in the following publication:

[Nicolas Pasquier](#), [Rafik Taouil](#), [Yves Bastide](#), [Gerd Stumme](#) and [Lotfi Lakhal](#), [Generating a Condensed Representation for Association Rules](#) (2005), in: Journal of Intelligent Information Systems, 24:1(29-60).