1 Theme

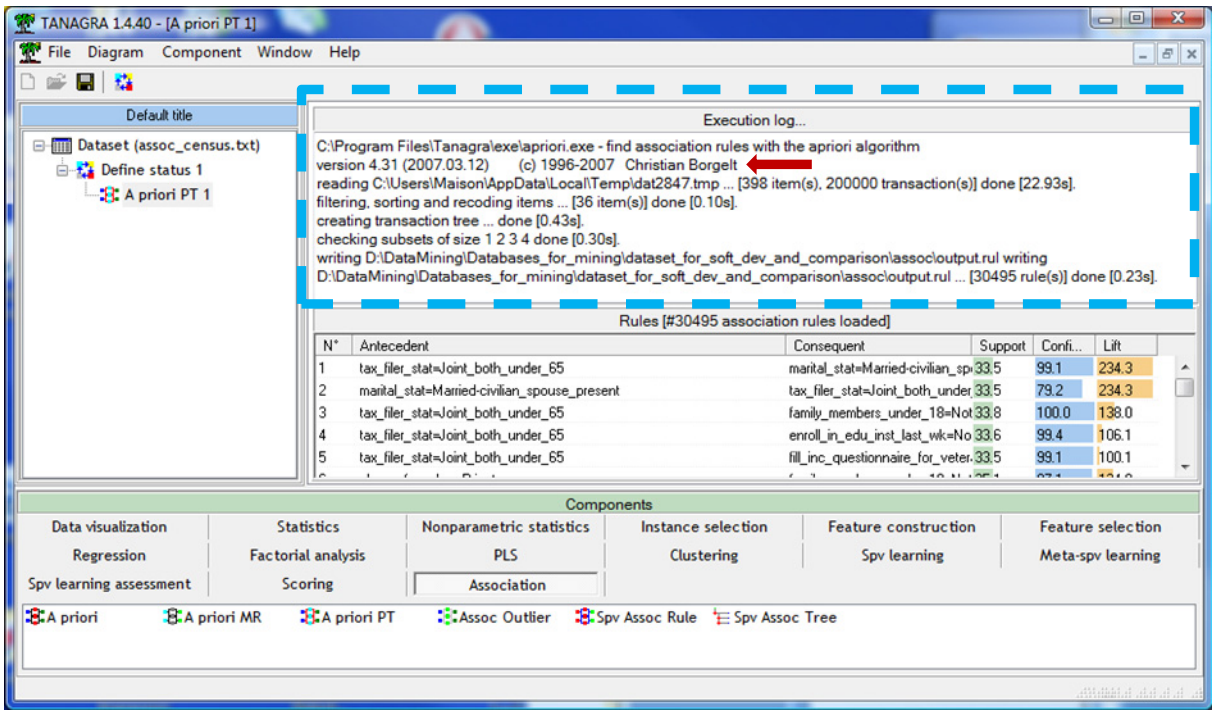
Updating of the APRIORI PT component, based on the Borgelt's apriori.exe 5.57 program.

A PRIORI PT is a tool dedicated for the extraction of association rules. This is one of the few components of Tanagra based on external library. We use the Borgelt's "apriori.exe" program. Until the version 1.4.40 of Tanagra, we used the 4.31 version of "apriori.exe". From the Tanagra 1.4.41, we introduce the latest update 5.57 (2011/09/02). Even if the settings of the tool are slightly modified, we observe that the extracted rules and the readings of the results are identical.

We take again a former tutorial to describe the behavior of this component (Association Rule Learning using A PRIORI PT). Thus, we do not detail the construction of the diagram here. We try above all to highlight the improvement of the library, especially about the computation time. We observe that this improvement is really impressive.

2 APRIORI.EXE 4.31 (via la version 1.4.40 de Tanagra)

We import the assoc_census.txt data file. We define the diagram. We use the default settings for A PRIORI PT (min support = 0.33, min confidence = 0.75, max itemset length = 4). We launch the calculations, we obtain the following results.

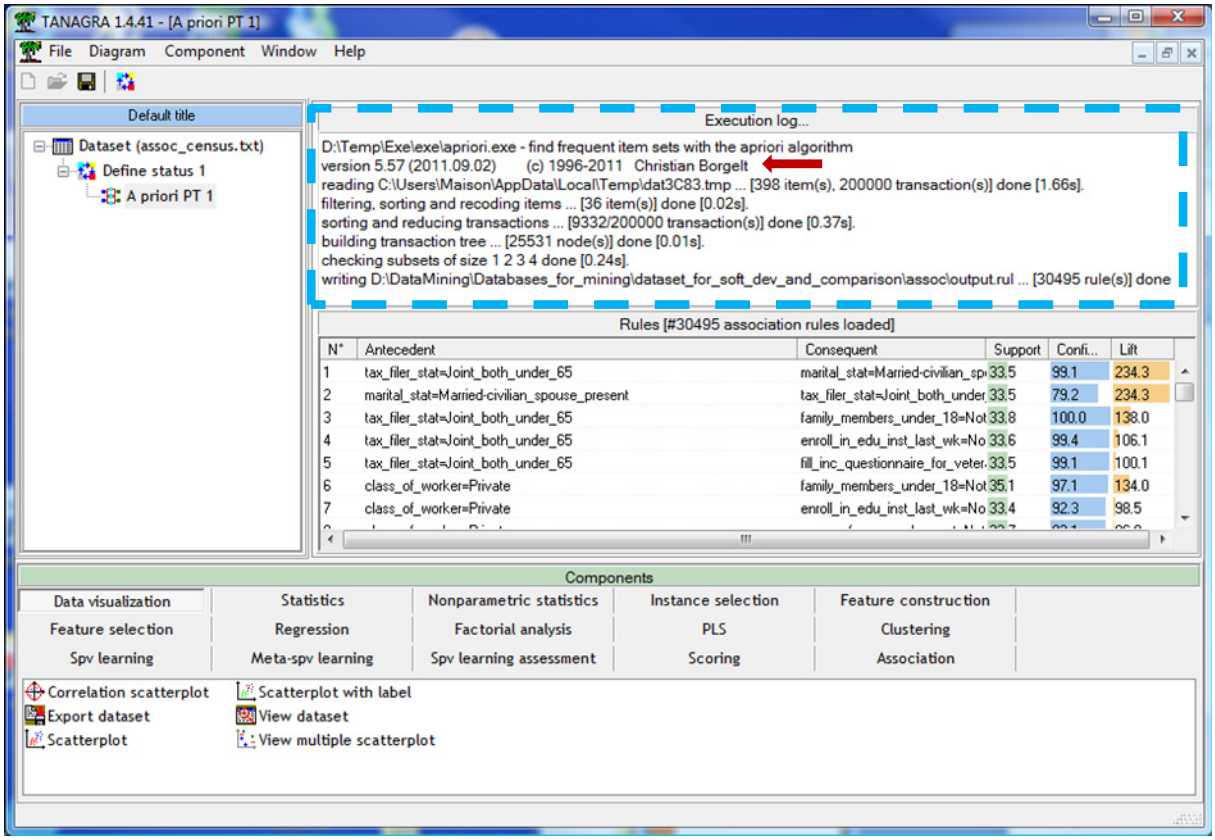


30.495 rules are extracted.

In a hidden manner, Tanagra creates a temporary file (transaction file format) that is sent to the "apriori.exe" library. The calculation time for this preparation is about 4 seconds. Then, the external library takes over for the processing. We observe that the rule extraction duration is rather fast. The calculation time is mainly penalized by the data file importation (*reading*). We see below that the improvement concerns this step.

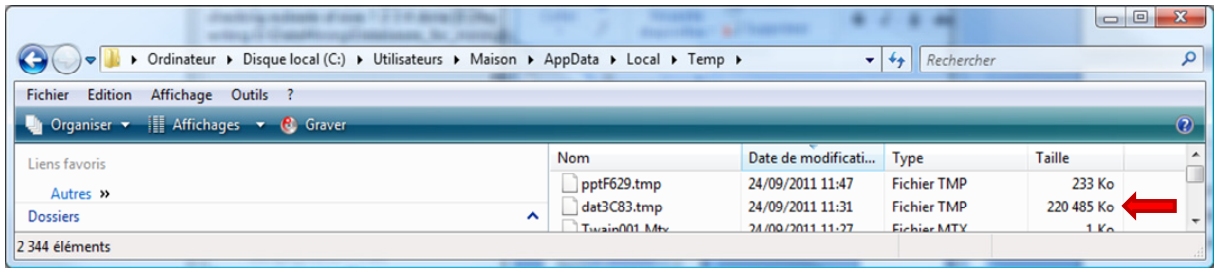
3 APRIORI.EXE 5.57 (via la version 1.4.41 de Tanagra)

With the latest version of Tanagra (1.4.41), we obtain the following results.



With the same settings, we obtain the same set of rules.

We observe that an additional step is detailed during the process (*sorting and reducing transactions*). We observe above all that the importation (*reading*) duration is drastically reduced. **On the same computer, the reading time fell from 22.93 seconds to 1.66 seconds!** We note that the temporary transaction file size is about 220 MB. It would appear difficult to make better.



4 Conclusion

Usually, I do not like to running after the versions of the libraries. Because, when we deal with a new version, we must check if the old treatments are still viable, if we obtain the same results than with the older versions. In the case of "apriori.exe", the checks that I made confirm the good behavior of the tool. And we have found that the calculation time is improved in spectacular proportions.