

Title: Tanagra (machine learning)

URL: [https://en.wikipedia.org/wiki/Tanagra_\(machine_learning\)](https://en.wikipedia.org/wiki/Tanagra_(machine_learning))

PageID: 39478785

Categories: Category:Data mining and machine learning software, Category:Free data analysis software, Category:Free science software, Category:Java (programming language) software

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

Tanagra is a free suite of machine learning software for research and academic purposes developed by Ricco Rakotomalala at the Lumière University Lyon 2 , France. Tanagra supports several standard data mining tasks such as: Visualization, Descriptive statistics, Instance selection, feature selection , feature construction, regression , factor analysis , clustering , classification and association rule learning .

Tanagra is an academic project. It is widely used in French-speaking universities. Tanagra is frequently used in real studies and in software comparison papers.

History

The development of Tanagra was started in June 2003. The first version was distributed in December 2003. Tanagra is the successor of Sipina, another free data mining tool which is intended only for supervised learning tasks (classification), especially the interactive and visual construction of decision trees. Sipina is still available online and is maintained.

Tanagra is an "open source project" as every researcher can access the source code and add their own algorithms, as long as they agree and conform to the software distribution license.

The main purpose of the Tanagra project is to give researchers and students a user-friendly data mining software, conforming to the present norms of the software development in this domain (especially in the design of its GUI and the way to use it), and allowing the analyzation of either real or synthetic data.

From 2006, Ricco Rakotomalala made an important documentation effort. A large number of tutorials are published on a dedicated website. They describe the statistical and machine learning methods and their implementation with Tanagra on real case studies. The use of other free data mining tools on the same problems is also widely described. The comparison of the tools enables readers to understand the possible differences in the presentation of results.

Description

Tanagra works similarly to current data mining tools. The user can design visually a data mining process in a diagram. Each node is a statistical or machine learning technique, the connection between two nodes represents the data transfer. But unlike the majority of tools which are based on the workflow paradigm, Tanagra is very simplified. The treatments are represented in a tree diagram. The results are displayed in an HTML format. This makes it is easy to export the outputs in order to visualize the results in a browser. It is also possible to copy the result tables to a spreadsheet.

Tanagra makes a good compromise between statistical approaches (e.g. parametric and nonparametric statistical tests), multivariate analysis methods (e.g. factor analysis, correspondence analysis, cluster analysis, regression) and machine learning techniques (e.g. neural network, support vector machine, decision trees, random forest).

See also

Free and open-source software portal

Free statistical software

Data mining

List of numerical analysis software

References

External links

Tanagra Project home page

Sipina Project home page

Free Statistical Software on StatPages.net