Winston: A Data Mining Assistant

Štefan Pero, Tomáš Horváth

Institute of Computer Science
Faculty of Science
Pavol Jozef Šafárik University in Košice

We will present the software Winston in our presentation. It is intended for small data mining applications and for students of data mining. The idea is to have a modular, easily extensible system. The modules can belong to one of three packages: The DATA package responds to input the data from several types of sources (i.e. text files, databases, Excel, Access). The TASK package contains data manipulation methods, such as selection and projection from a table, join of tables, etc. It will contains preprocessing methods, too, e.g. handling missing values, outliers detections, etc. The third is the ALGORITM package, which consists of the data mining algorithms. The system contains a wizard to easily add a new module. The main part of the system is the graphical interface (the SCHEMA panel), where the user can build a so-called schema of a data mining process, i.e. reading the data, preprocessing them and launch a data mining algorithm. Each module have its own properties, defined in an XML file. The user can see the data in each step of the process by switching to the VIEW panel. Since, the system is easy-to-use and modular, it can be used in teaching as well as in small data mining applications by researchers in other fields of science, e.g. physics, biology, ...