Complete Round-Robin method for Data Mining

Clayton Todd, Dan Henderson, Boris Kovalerchuk
Dept. of Computer Science, Central Washington University, Ellensburg, WA, 98926,

Abstract.

The reliability of data mining applications heavily depends on testing of discovered patterns. The common testing approach is based on the Round-Robin method for available data. Usually training data are randomly chosen from the available data and the rest of the data are used for testing. Each time, the parameters of the system are identified using training data and then correctness of these parameters is tested on testing data. There are \(2^n\) pairs of training and testing data sets for \(n\) data objects. Actual Round-Robin implementations heuristically restrict this number of combinations. It decreases the amount of computations, but these smaller tests may be not valid to represent all \(2^n\) possible tests.

We developed a method to speed up a complete Round-Robin computational procedure using the concept of monotonicity and multithreaded programming for Windows NT. This method is applicable for both attribute-based learning and relational data mining methods. The method has been implemented. It has shown its effectiveness for Neural Networks based on backpropagation. Backpropagation using one processor with 1024 threads runs faster than without threads (experiment 1, ratio 3.0 with monotonicity; experiment 2, ratio 2.4 without monotonicity). Backpropagation using one processor with monotonicity runs faster than without monotonicity (experiment 3, ratio 4.0 with 1024 threads; experiment 4, ratio 3.4 without threads). A screen shot of the implementation for backpropagation Neural Network is presented in figure 1.

Figure 1. Main screen

Reference