
Accelerating GPU-based Evolutionary Induction of Decision Trees - Fitness Evaluation Reuse

Krzysztof Jurczuk, Marcin Czajkowski, Marek Kretowski
Faculty of Computer Science, Bialystok University of Technology
Bialystok, Poland
{k.jurczuk,m.czajkowski,m.kretowski}@pb.edu.pl

The rapid development of new technologies and parallel frameworks is a chance to overcome barriers of slow evolutionary induction of decision trees (DTs). This global approach, that search for the tree structure and tests simultaneously, is an emerging alternative to greedy top-down solutions. However, in order to be efficiently applied to big data mining both technological and algorithmic possibilities need to be fully exploit. This paper shows how by reusing information from previously evaluated individuals, we can accelerate GPU-based evolutionary induction of DTs on large-scale datasets even further. Noting that some of the trees or their parts may reappear during the evolutionary search, we have created a so-called repository of trees (splitted between GPU and CPU). Experimental evaluation is carried out on the existing Global Decision Tree system where the fitness calculations are delegated to the GPU, while the core evolution is run sequentially on the CPU. Results demonstrate that reusing information about trees from the repository (classification errors, objects' locations, etc.) can accelerate the original GPU-based solution. It is especially visible on large-scale data where the cost of the trees evaluation exceeds the cost for storing and exploring the repository.

Keywords: Evolutionary algorithms, Decision trees, Big data mining, Graphics processing unit (GPU), CUDA.