Fine-grained parallelization of fitness functions in bioinformatics optimization problems: Gene selection for cancer classification and biclustering of gene expression data

Gomez-Pulido J.A.

Cerrada-Barrios J.L.

Trinidad-Amado S.

Lanza-Gutierrez J.M.

Fernandez-Diaz R.A.

Crawford B.

Soto R.

Background: Metaheuristics are widely used to solve large combinatorial optimization problems in bioinformatics because of the huge set of possible solutions. Two representative problems are gene selection for cancer classification and biclustering of gene expression data. In most cases, these metaheuristics, as well as other non-linear techniques, apply a fitness function to each possible solution with a size-limited population, and that step involves higher latencies than other parts of the algorithms, which is the reason why the execution time of the applications will mainly depend on the execution time of the fitness function. In addition, it is usual to find floating-point arithmetic formulations for the fitness functions. This way, a careful parallelization of these functions using the reconfigurable hardware technology will accelerate the computation, specially if they are applied in parallel to several solutions of the population. Results: A fine-grained parallelization of two floating-point fitness functions of different complexities and features involved in biclustering of gene expression data and gene selection for cancer classification allowed for obtaining higher speedups and power-reduced computation with regard to usual microprocessors. Conclusions: The results show better performances using reconfigurable hardware technology instead of usual microprocessors, in computing time and power consumption terms, not only because of the parallelization of the arithmetic operations, but also thanks to the concurrent fitness evaluation for

several individuals of the population in the metaheuristic. This is a good basis for building accelerated and low-energy solutions for intensive computing scenarios.

Biclustering

Cancer classification

Fitness function

Floating-point arithmetic

FPGA

Metaheuristics

Parallelism

Bioinformatics

Classification (of information)

Combinatorial optimization

Computer aided diagnosis

Computer hardware

Digital arithmetic

Diseases

Field programmable gate arrays (FPGA)

Genes

Hardware

Health

Heuristic algorithms

Heuristic programming

Optimization

Parallel processing systems

Reconfigurable hardware

Bi-clustering

Cancer classification

Fitness functions

Meta heuristics

Parallelism

Gene expression

algorithm

biology

classification

gene expression regulation

genetics

human

neoplasm

pathology

procedures

software

Algorithms

Computational Biology

Gene Expression Regulation, Neoplastic

Humans

Neoplasms

Software