Scalable Inference of Gene Regulatory Networks with the Spark Distributed Computing Platform

Cristóbal Barba-González and José García-Nieto and Antonio Benítez-Hidalgo and Antonio J. Nebro and José F. Aldana-Montes

Abstract Inference of Gene Regulatory Networks (GRNs) remains an important open challenge in computational biology. The goal of bio-model inference is to, based on time-series of gene expression data, obtain the sparse topological structure and the parameters that quantitatively understand and reproduce the dynamics of biological system. Nevertheless, the inference of a GRN is a complex optimization problem that involve processing S-System models, which include large amount of gene expression data from hundreds (even thousands) of genes in multiple time- series (essays). This complexity, along with the amount of data managed, make the inference of GRNs to be a computationally expensive task. Therefore, the genera- tion of parallel algorithmic proposals that operate efficiently on distributed process-ing platforms is a must in current reconstruction of GRNs. In this paper, a parallel multi-objective approach is proposed for the optimal inference of GRNs, since min- imizing the Mean Squared Error using S-System model and Topology Regulariza- tion value. A flexible and robust multi-objective cellular evolutionary algorithm is adapted to deploy parallel tasks, in form of Spark jobs. The proposed approach has been developed using the framework jMetal, so in order to perform parallel compu- tation, we use Spark on a cluster of distributed nodes to evaluate candidate solutions modeling the interactions of genes in biological networks.

Key words: Gene Regulatory Networks, Multi-objective, Metaheuristics, Distributed Computing,jMetal, Spark

Cristóbal Barba-González   José García-Nieto   Antonio Benítez-Hidalgo   Antonio J. Nebro   José F. Aldana-Montes

Dept. de Lenguajes y Ciencias de la Computación, Instituto de Investigación Biomédica de Málaga (IBIMA), University of Málaga, Campus de Teatinos, Málaga - 29071, Spain,

e-mail: `cbarba@lcc.uma.es`   e-mail: `jnieto@lcc.uma.es`   e-mail: `antonio.b@uma.es`

e-mail: `antonio@lcc.uma.es` e-mail: `jfam@lcc.uma.es`