Comparing Evolutionary Algorithms on the Problem of Network Inference

Christian Spieth
Centre for Bioinformatics
Tübingen (ZBIT)
72076 Tübingen, Germany

spieth@informatik.uni-tuebingen.de

Rene Worzischek
Centre for Bioinformatics
Tübingen (ZBIT)
72076 Tübingen, Germany

renewor@web.de

Felix Streichert
Centre for Bioinformatics
Tübingen (ZBIT)
72076 Tübingen, Germany

streiche@informatik.uni-tuebingen.de

ABSTRACT

In this paper, we address the problem of finding gene regulatory networks from experimental DNA microarray data. We focus on the evaluation of the performance of different evolutionary algorithms on the inference problem. These algorithms are used to evolve an underlying quantitative mathematical model. The dynamics of the regulatory system are modeled with two commonly used approaches, namely linear weight matrices and S-systems and a novel formulation, namely H-systems. Due to the complexity of the inference problem, some researchers suggested evolutionary algorithms for this purpose. However, in many publications only one algorithm is used without any comparison to other optimization methods. Thus, we introduce a framework to systematically apply evolutionary algorithms and different types of mutation and crossover operators to the inference problem for further comparative analysis.

Categories and Subject Descriptors

I.2 [**Computing Methodologies**]: ARTIFICIAL INTELLIGENCE—*Miscellaneous*; J.3 [**Computer Applications**]: LIFE AND MEDICAL SCIENCES—*Biology and genetics*

General Terms

ALGORITHMS, PERFORMANCE

Keywords

Evolutionary Computation, Inference, Systems Biology

1. INTRODUCTION

Systems biology has become one of the major research areas in biology over the past few years. Due to tremendous progress in experimental methods like DNA microarrays, several thousand expression levels of genes in an organism can be measured in parallel under specific environmental conditions. This enables researchers to examine intracellular processes on a systemic level. The inference of gene regulatory networks from experimental data is one of the main unsolved problems in the post-genomic area. A gene regulatory network (GRN) is an abstract model representing

dependencies between genes using a directed graph. In this graph, each node is a gene or component of the regulatory system and each edge represents a regulatory impact from one component to the other (e.g. activation or suppression of the transcription and translation of the dependent gene).

Several publications addressing the problem of inferring gene regulatory networks can be found in the literature. De Jong gives a good overview about related work in [1]. A major part of the work done in this field is using deterministic mathematical models to simulate regulatory networks. One kind of those deterministic models are linear models like the weighted matrix model [8, 9]. These models have only a small number of system parameters compared to S-systems but are often not flexible enough to model biological systems in detail, since they model the dependencies linearly. S-systems, on the other hand, model dynamic systems in a nonlinear manner. They consist of a set of differential equations describing the changes in expression over time. However, they show a significant higher number of system parameters. S-systems have been recently examined in [3, 4]. Most applications of deterministic models use evolutionary algorithms (EA) to determine the correct parameters of the mathematical model. EAs have proven to be successful in finding parameters of mathematical models representing GRNs.

So far, the authors commonly describe only the algorithm of their choice on the problem of identifying the unknown parameters of the model. Some publications show the results of the proposed method only in comparison to one standard algorithm, thus lacking the possibility of impartially evaluating the performance of other approaches. Therefore, we designed a framework to systematically compare optimization algorithms on a set of artificial benchmark problems, which is accessible through the JCell project website[1] [6].

2. MATHEMATICAL MODELING

The genetic dependencies of a cell can be abstracted by a directed graph with N nodes representing N genes. Each gene g_i produces a certain amount of mRNA x_i when expressed and changes the concentration of the mRNA level over time: $\vec{x}(t + 1) = h(\vec{x}(t))$, $\vec{x}(t) = (x_1, \cdots, x_n)$. Here, function h represents the changes of expression levels from one state to the next. To model this function, several approaches can be found in the literature. We decided to use

---

[1]http://www.jcell.de

the two most popular deterministic models, namely linear weight matrices [8], H-systems [2], and S-systems [5].

3. CONCLUSIONS

In this paper, we introduced the framework JCell that was developed to allow users to evaluate different algorithms on a set of well-defined benchmark systems to obtain comparable results. Several optimization algorithms together with a variety of mathematical models are implemented to study the performance on the inference problem.

Further on, we systematically examined the performance of standard evolutionary algorithms and the impact of mutation and crossover on the problem of inferring gene regulatory networks from microarray data. The comprehensive study was performed on an Opteron cluster with 16 dualcore CPUs with 2,2GHz and 2GB RAM per node. The overall computation time amounted approximately 400h.

Additionally, we fine-tuned the three most promising algorithms, namely an evolution strategy with CMA, a real-valued genetic algorithm with UNDX, and differential evolutions, to examine their best performance. The experiments in this paper showed that although some evolutionary mechanisms like the real-valued GA with UNDX crossover and differential evolution are able to cope with the complex and highly multi-modal search space, only the CMA mutation operator solved this type of optimization efficiently and at the same time reliable. Even with additional fine-tuning of the corresponding algorithm settings for the realGA and the DE, no alternative approach was able to equal the performance of the ES with CMA with respect to the RSE. A conclusion of this paper is that to choose a suitable model and an appropriate optimization method is crucial for the inference, and the results illustrated that evolutionary algorithms are well suited for the problem of network inference.

In a related publication [7], we present a comprehensive study of mathematical formulations, namely weight matrices, H-systems, and S-systems for modeling dynamic systems of different mathematical properties and size.

4. ACKNOWLEDGMENTS

5. ADDITIONAL AUTHORS

Additional authors: Jochen Supper (Centre for Bioinformatics Tübingen, Nora Speer (Centre for Bioinformatics Tübingen, and Andreas Zell (Centre for Bioinformatics Tübingen).

6. REFERENCES

[1] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103, January 2002.

[2] K. Hadeler. Gedanken zur Parameteridentifikation. Personal Communication, 2003.

[3] S. Kikuchi, D. Tominaga, M. Arita, K. Takahashi, and M. Tomita. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, 19(5):643–650, 2003.

[4] S. Kimura, K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya. Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, 21(7):1154–1163, 2005.

[5] M. A. Savageau. 20 years of S-systems. In *Canonical Nonlinear Modeling. S-systems Approach to Understand Complexity*, pages 1–44, 1991.

[6] C. Spieth. JCell : A java framework for inferring genetic networks. Technical Report WSI-2005-07, Centre for Bioinformatics Tübingen, University of Tübingen, 2005.

[7] C. Spieth, N. Hassis, F. Streichert, J. Supper, K. B. N. Speer and, and A. Zell. Comparing mathematical models on the problem of network inference. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 2006.

[8] D. Weaver, C. Workman, and G. Stormo. Modeling regulatory networks with weight matrices. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 4, pages 112–123, 1999.

[9] M. K. S. Yeung, J. Tegner, and J. J. Collins. Reverse engineering gene networks using singular value decomposition and robust regression. In *Proceedings of the National Academy of Science*, volume 99, pages 6163–6168, 2002.