Computational Discovery of Cis-Regulatory Sites in Eukaryotic Genomes by Collective Inference

Wouter Van Delm  Yves Moreau  Bart De Moor

Dept. of Electrical Engineering (ESAT) - SCD - SISTA
Katholieke Universiteit Leuven, Belgium
wvandelm@esat.kuleuven.be

1 Problem Setting

Abstract: Motif discovery in eukaryotic genomes is an abstraction process that describes regulatory DNA sequences as collections of cis-regulatory sites. The result enables researchers to decipher complex gene regulatory networks. In silico methods prioritize cis-regulatory sites and aim to increase the efficiency of arduous wet lab research. Computational discovery of cis-regulatory sites is however often an ill-posed problem, resulting in many false positive detections.

2 Approach

1.1 Motif Discovery in Eukaryotes

A CBIP is a tuple (X, Y, E, r) [3]. The general approach enhances knowledge integration and cross-fertilization of solution algorithms.

Example. A HMM in CBIP terms:

\[ x' = (O', H', \theta), \quad E = (D_{i}, D_{j}, D_{k}) \]

where \( x_{j} \) is the state of the system. \( O' \) is the observation from the system.

The (Bayesian-like) inference task becomes:

\[ \phi(x, y) = \sum_{\phi} \phi \sum_{x} \phi(x) \]

Global dynamics are similar to that of bounded rational noncooperative games.

2.1 Constraint-Based Inference in Games

A CBIP is an abstraction process that describes regulatory DNA sequences as collections of cis-regulatory sites. The result enables researchers to decipher complex gene regulatory networks. In silico methods prioritize cis-regulatory sites and aim to increase the efficiency of arduous wet lab research. Computational discovery of cis-regulatory sites is however often an ill-posed problem, resulting in many false positive detections.

We searched for a heart specific cis-regulatory module. \( m_{c ntr} \) used data from Transfac and MotifSampler. Validation of results with Ern-0081 was not convincing so far.

2.2 Parallel Computation in a MAS

Characteristics:

- Heterogeneous multi-agent system
- No central control, synchronous updates
- Non-active / emergent cooperation
- Agents have alternative utility functions

3 Preliminary Results

Abstract: In a pilot study, we implemented a 3-agent system (module structure, location, motif matrices). We observed the performance on a synthetic dataset and searched for heart-specific cis-regulatory modules in a small case study. Preliminary results suggest that the proposed framework can deal consistently with heterogeneous data and allows for efficient distributed processing. The observations encourage us to further investigate the design of the coordination scheme.

3.1 Performance Analysis

Synthetic data with realistic SNR.

The outcome of interest is the output of \( A_{c m e t} \) : a sample with combinations made out of 20 matrices.

Performance was measured on the average combination (PD over the 20 matrices).

The effect of the sample size (variance) seemed to be important.

Sample size (X) versus performance (Y).

Since the game may have more than one Nash equilibrium, convergence to suboptimal solutions needs to be investigated more in the future.

3.2 Case Study

We searched for a heart specific cis-regulatory module. \( m_{c ntr} \) used data from Transfac and MotifSampler. Validation of results with Ern-0081 was not convincing so far.

References

Acknowledgements

The authors are thankful to the Canadian Centre in bioinformatics for providing the motif database. Research funded by PRIN grants for the Italian Ministry of University and Research (Ministero dell’Università e della Ricerca Scientifica e Tecnologica, Rome) for the study of vision systems (grant 03017824) and for the study of computational linguistics (grant 03015037). Research supported by the European Union’s seventh Framework Programme (FP7-KA2-CT-2007-205290: MULTILOG, “Multi-Agent Systems: a cooperative computational framework for complex systems”). This work was supported by the European Training Network “Complexity, Emergence and Control in Social, Economic and Biological Systems”, EU grant 501987 (HTRA).