A System to Find Genetic Networks Based on Weighted Majority Algorithm

Ayumi Shinohara[1] ayumi@i.kyushu-u.ac.jp
Masayuki Takeda[1] takeda@i.kyushu-u.ac.jp
Tomohiro Moriyama[1] moriyama@i.kyushu-u.ac.jp
Osamu Maruyama[2] maruyama@ims.u-tokyo.ac.jp
Takao Goto[2] takao@ims.u-tokyo.ac.jp
Satoru Miyano[2] miyano@ims.u-tokyo.ac.jp
Shigeru Muta[3] muta@grt.kyushu-u.ac.jp
Koushuke Tashiro[3] ktashiro@grt.kyushu-u.ac.jp
Satoru Kuhara[3] kuhara@grt.kyushu-u.ac.jp

[1] Department of Informatics, Kyushu University 33, Fukuoka 812-8581, Japan
[2] Human Genome Center, University of Tokyo, Tokyo 108-8639, Japan
[3] Graduate School of Genetic Resources Technology, Kyushu University, Fukuoka 812-8581, Japan

Keywords: genetic networks, weighted majority algorithm, network visualization, gene expression data

1 Introduction

Analyzing the interactions between genes by systematic gene disruptions and gene overexpressions is getting more important in Genome Science. DNA microarray technology enabled us to produce time series of gene expression patterns. Our research group have launched a project whose purpose is to reveal the gene regulatory networks among the 6,200 genes of *Saccharomyces cerevisiae* while many laboratories have also started similar project.

Some methods have been proposed to identify the gene regulatory networks from gene expression patterns [1, 2, 5, 6]. In our previous work [3, 4], we have introduced a *weighted network model* as an edge-weighted graph, where each weight reflects the strength of the interaction. We analyzed its computational complexity [3]. The simulation results showed that our algorithm to adjust weights incrementally predicts more accurate than the algorithm whose worst case performance is theoretically guaranteed [4]. In the poster, we show the overview of our system and experimental results using this system.

2 Method and Results

In Fig. 1, we illustrate our system. The core module of the system is to produce a genetic network as a weight matrix from given gene expression profiles. We have implemented our incremental weight adjusting algorithm, whose practical behaviors have verified in our previous work [4]. The problem here is that the size of the produced network is quite large. If we directly apply our system for gene expression profiles consisting of 6,200 genes, the output is a weight matrix of size $6,200 \times 6,200$. Unfortunately, there is no guarantee that the network is sparse. Thus it will be quite hard for the users to understand the connections among genes in the network. The standard graph drawing techniques will not work effectively.

We introduce two modules in our system in order to reduce the problem. The first one is the correlation analyzer of gene activations. It analyzes not only the standard correlation coefficient of gene activations, but also the correlation coefficient through some functions specified by the users. We can use it to condense the gene expression profile data, with respect to the users' interests.

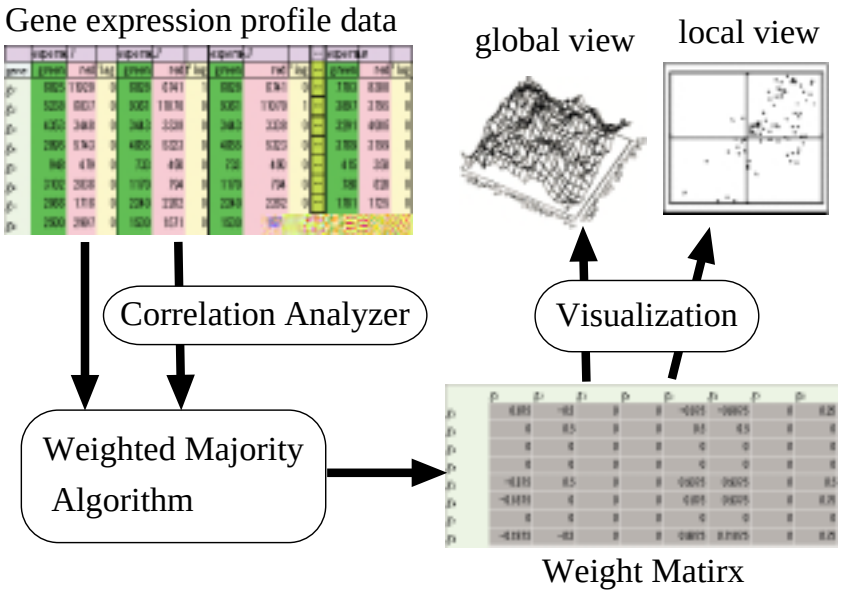

Figure 1: Overview of the system

The second module is to visualize the weighted network. We propose two views to capture the connections in the network. One is the global view, which rearranges a weight matrix by permuting columns and rows so that similar genes locate closely and dissimilar genes far from each other. The goal is related to other methods of visualization, such as the Self-Organizing Map (SOM) and the Multi-Dimensional Scaling (MDS). We analyze the computational complexity to rearrange the weight matrix optimally. We show that the problem can be reduced to the traveling salesman problem and vice versa. These results imply that the problem is computationally intractable to optimize, while some heuristic methods may be applicable.

The other view is called the local view, which shows how strongly a specified gene activates other genes, and how strongly it is activated by other genes on the Cartesian coordinates. Through the local view, the users can observe the behavior of each gene separately.

References

[1] Akutsu, M., Miyano, S. and Kuhara, S., Identification of genetic networks from a small number of gene expression patterns under the boolean network model, *Pacific Symposium on Biocomputing*, 4:17–28, 1999.

[2] Chen, T., He, H. and Church, G., Modeling gene expression with differential equations, *Pacific Symposium on Biocomputing*, 4:29–40, 1999.

[3] Noda, K., Shinohara, A., Takeda, M., Matsumoto, S., Miyano, S. and Kuhara, S., Finding genetic network from experiments by weighted network model, *Genome Informatics*, 9:141–150, 1998.

[4] Noda, K., Shinohara, A., Takeda, M., Matsumoto, S., Miyano, S. and Kuhara, S., Simulation results on finding genetic networks by weighted network model, *Pacific Symposium on Biocomputing (Poster abstracts)*, page 84, 1999.

[5] Szallasi, Z., Genetic network analysis in light of massively parallel biological data acquisition, *Pacific Symposium on Biocomputing*, 4:5–16, 1999.

[6] Weaver, D., Workman, C. and Stormo, G., Modeling regulatory networks with weight matrices, *Pacific Symposium on Biocomputing*, 4:112–123, 1999.