IFSA 알고리즘을 이용한 유전자 상호 관계 분석
(Analysis of Interactions in Multiple Genes using IFSA (Independent Feature Subspace Analysis))

김 혜 진 † 최 승 진 †† 방 승 양 †††
(Hye Jin Kim) (Seungjin Choi) (Sung Yang Bang)

요 약 세포는 환경 변화 및 자극으로부터 자신을 보호하기 위해 유전자가 발현하여 생명을 유지 시스템을 갖고 있다. 유전자의 발현은 비정상적인 상태의 세포를 환경을 조절, 변화시켜 정상으로 바꾸기 위한 기능, 발달단계에 필요한 기능 등 생명현상에 필요한 특수 역할을 수행한다. 따라서 각 유전자의 기능을 아는 것은 생물학적으로 상당히 의미 있는 일이다. 본 논문에서는 유전자 기능을 알아보기 위해 발현 패턴을 통해 같을 때, 유사한 형태 혹은 시차를 갖고 동일한 형태로 발현하는 유전자들은 같은 기능을 한다는 가정을 하였다. 이 가정에 기반하여 각 유전자들을 기능에 따라 분류하였다. (1) IFSA선형 모델을 적용하여 데이타를 잘 나타내 줄 수 있는 특징 패턴을 찾았으며 (2) 이 특징 패턴으로부터 본 논문에서 제안한 Membership Scoring Function을 이용하여 유전자를 필터링(filtering) 하였다. 이 유전자들은 기존의 ICA(Independent Component Analysis) 방법에서 보다 IFSA 방법이 더 효과적으로 각 기능에 따른 유전자 그룹을 찾아내줌을 GO(Gene Ontology)에서 확인할 수 있었다. 이는 시차 혹은 위상 변화에 상관없이 데이타를 잘 나타낼 수 있는 IFSA의 특성이, ICA보다 생물학적인 변수를 더 고려해 줄 수 있기 때문이라고 생각된다[1]. 이 논문의 또 다른 주요 작업은 유전자의 상호작용 관계로부터 유전자 네트웍을 얻어내는 것이다. 유전자 네트웍은 같은 그룹 내에서 유전자간의 상관 계수를 구하고 가장 높은 상관도를 보이는 유전자쌍을 연결시켜 얻게되었다. 이 네트웍 역시 GO 해석에서 그 유효성을 확인하였다.

키워드 : 마이크로어레이, 유전자 네트웍, 독립공간분석방법

Abstract The change of external/internal factors of the cell rquires specific biological functions to maintain life. Such functions encourage particular genes to interact/regulate each other in multiple ways. Accordingly, we applied a linear decomposition model IFSA, which derives hidden variables, called the 'expression mode' that corresponds to the functions. To interpret gene interaction/regulation, we used a cross-correlation method given an expression mode. Linear decomposition models such as principal component analysis (PCA) and independent component analysis (ICA) were shown to be useful in analyzing high dimensional DNA microarray data, compared to clustering methods. These methods assume that gene expression is controlled by a linear combination of uncorrelated/indepdendent latent variables. However, these methods have some difficulty in grouping similar patterns which are slightly time-delayed or asymmetric since only exactly matched patterns are considered. In order to overcome this, we employ the (IFSA) method of [1] to locate phase- and shift-invariant features. Membership scoring functions play an important role to classify genes since linear decomposition models basically aim at data reduction not but at grouping data. We address a new function essential to the IFSA method. In this paper we stress that IFSA is useful in grouping functionally-related genes in the presence of time-shift and expression phase variance. Ultimately, we propose a new approach to investigate the multiple interaction information of genes.

Key words : microarray, genetic network, independent feature subspace analysis

· This work was supported by National Core Research Center for Systems Bio-Dynamics and POSTECH Basic Research Fund.

† 정 회 원 : 포항공과대학교 컴퓨터공학과
marisan@postech.ac.kr
†† 정 회 원 : 포항공과대학교 컴퓨터공학과 교수
seungjin@postech.ac.kr
††† 종신회원 : 포항공과대학교 컴퓨터공학과 교수
sybang@postech.ac.kr

논문접수 : 2004년 6월 10일
심사완료 : 2005년 11월 16일

1. INTRODUCTION

Cells maintain their lives by appropriate chemical reactions to external stimuli and to internal needs. A cellular function is represented by the expression

of genes bound by a set of proteins, such as signaling substances, transcription factors, and their binding sites in the promoter regions of genes. The regulatory relationship between genes is associated with many variables, such as transcriptional modification factors, signaling substances, and modification of the chromatin structure. That is, gene expression cannot completely explain the regulatory network but this data still contains useful information concerning cellular function. Genes control the induction or repression of other genes in a regulatory network. Various pathways in the biological network consist of a systematic structure of biological function, insight into which can provide hints for solving biological problems.

Genomic-scale gene expression data are provided by high-throughput methods such as microarray. Gene expression data measured at different time points represent biologically functional behaviors of genes at different expression levels, so that one can obtain new insights about regulatory machinery. For example, [2,3] analyzed extensive time-series data on gene expression in yeast, and clustered coexpressed genes, each of which activates another gene.

Current microarray technologies cannot perform experiments repeatedly, so that they may not be analyzed for more accuracy. In addition, to perform a cellular function, a set of genes involved in the function cooperate with each other and work within a limited time interval. However, because several functions are carried out simultaneously for cells to cope with external or internal stimuli, one cannot be certain that all genes expressed contemporaneously are functionally related to each other. On the other hand, a gene can perform multi-functions, but a gene only acts under specific circumstances; therefore, in order to identify a whole function of a gene, we need time series data measured under different cell or tissue conditions. To date, various algorithms have been employed to microarray time series data. Theses include: (1) Bayesian Network; (2) Hierarchical Clustering; (3) Differential equations; (4) Edge detection; (5)Linear decomposition models such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), etc. These methods attempt to investigate gene functions by clustering or detecting the gene-gene relationship. Assuming that gene expression levels are continuous, Bayesian networks provide gene interactions in terms of conditional probabilities [4,5]. Hierarchical clustering method finds groups of genes based on the similarity measurement between their gene expression profiles [6]. Differential Equations and Edge detection are a technique catching the difference of expression levels between two or more points for a local area. All methods already listed concern the direct relationship between gene pairs. On the other hand, Linear modelbased methods explicitly describe dominant functions in terms of expression modes by effective genes: Singular value Decomposition (SVD) [7], PCA [8], ICA [9, 10, 11] and Bayes Decomposition [12].

In this paper, we address the problem of determining a group consisting of genes having a functional relation, as well as having multiple interactions between genes using a membership scoring function.

(a) A virtual biological function may be represented by a feature profile.

By a linear decomposition model, we can extract feature patterns. Alter [4] and Libermeister [10] already showed that a feature pattern explains one type of cell cycle behavior in terms of an 'expression mode.' These papers showed that a few feature patterns related to the cell-cycle can predict regulated genes effectively.

(b) Analysis of phase-shift, directionality, and periodicity analysis for time-series data integration.

To complete a cellular behavior, expression profiles of interactive genes tend to be time-lagged because it takes some time for a gene to affect other genes. On the other hand, a gene can activate or inhibit other genes. Using their directionality of expression patterns, we can distinguish which gene is an activator or inhibitor of other genes. Using periodicity, [2,3] showed the example of clustering cell cycle data. They integrated genes into several cell cycle phases such as G1, S, S/G2,

M/G1. Detecting directionality and phase shift of expression profiles, [13,14] identified the direct regulatory relationship of genes.

(c) Select genes with functional interactions.

Linear decomposition models can explain cellular behavior as an expression mode but are inadequate to interpret which type of relationship between genes is placed on. The techniques that consider periodicity, phase- shift- invariance, or directionality of expression profiles are useful to recognize the gene-gene relationship in a direct way, but are limited to explain the functional relationship as well as multiple interactions among genes. Therefore, the problem addressed in this paper is to select functionally interactive genes based on a *Membership Soring Function* we devised. Firstly, feature patterns are extracted from the microarray by linear decomposition models. Secondly, by statistical analysis are feature patterns selected that assume that the patterns correspond to biological functions. Thirdly, genes can be clustered by their dominant role to engender feature profiles using the membership scoring function. Finally, we explain the relationship between genes by detecting their correlation in a group, expecting all genes in a group to be all connected each other.

2. METHODS

2.1 Independent Feature Subspace Analysis

A linear decomposition model assumes that the data matrix X = (X_{nt}) where the elements Xnt represents the expression level of gene n in the tth time points of samples, with respect to n = {1,...,N}, t = {1,..., T}. Classical ICA model decomposes X into S and A matrices, i.e., $X_{nt} = \sum_k S_{nk}A_{kt}$ subject to minimizing statistical dependency among S_t. In [10], row vectors of A $\in R^{KT}$ correspond to expression modes. Mutidemensional ICA [2] generalizes the ICA by allowing the components in a K-tuple to be dependent but different K-tuples to be independent. The IFSA [1] embeds the invariant feature subspaces in mutidemensional ICA by considering the probability distributions for the K-tuples of latent variables that are spherically symmetric, i.e., depend only on the norm.

The model of IFSA

$$X_{nt} = \sum_k (S_{nk}A_{kl} \; k \in j) \qquad (1)$$

where $k \in j$ implies k-th member of the j-th group, k = {1,...,K} and j = {1,...,J} (see Figure 1. In contrast to ICA, the IFSA aims at finding a linear transformation W (which corresponds to A^{-1}) such that feature subspaces (obtained by taking the square root of the sum of K-tuples, naming 'energy of responses,') become independent but components in feature subspace is allowed to be dependent.

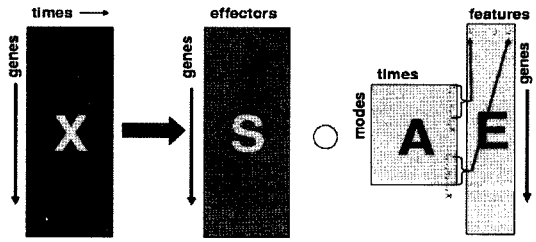


Figure 1 The microarray data matrix **X** is decomposed into a product of two matrices, i.e., **S** and **A**, gene effects and expression modes, respectively. Invariant feature patterns $\{E_j\}^{J_{j=1}}$ (see Eqn. (2)) play a critical role in grouping similar patterns as well as preserving shift- and phase-invariance

We assume that the data matrix X is already whitened by the V matrix. In other words, the row vectors of A are confined to be orthogonal to each other and are normalized to have a unit norm. Alternatively, $\overline{\omega_i}$ equal to the rows of the new separating matrix

$(VA)^{-1}$. The non-orthogonal factor is reflected in a whitening transform. To avoid abuse of notations, we use the notation X for the whitened data matrix. We consider the case where latent variables are divided into J number of K-tuples (where K represents the dimension of subspace). For the sake of simplicity, we assume an identical dimension K for every feature subspace. The j-th feature subspace is denoted by F_j. The value $E_j(\vec{x})$ in Fj with data vector $\vec{x}$ is given by

$$E_j(\vec{x}) = \sum_{k \in F_j} < \vec{w_k}, \vec{x} >^2 \quad (2)$$

where $< \cdot, \cdot >$ is the inner product.

IFSA intends to minimize information loss and extract invariant features. The invariant property of IFSA refers to the complex cell model of V1 of the brain. The response of complex cell can be modelled by the norm of the projection of the input vector (gene expression profiles) onto a linear subspace, according to the classical energy models. IFSA maximizes the independence between the norms of projections, or energies, generating a subspace, which corresponds to K-tuple patterns as if a complex cell characterizes a representative feature through K simple cells. The log-likelihood L of the data can be written by,

$$L = \sum_{n=1}^{N} \sum_{j=1}^{J} \log p(\sum_{k \in F_j} s_k^2) + N \log |\det W| \quad (3)$$

where $p(\sum_{k \in F_j} s_k^2 = p_j(s_k, k \in F_j)(s_k = < \vec{w_k}, \vec{x} >)$ represents the probability density inside the j th K-tuple of s_k.

We follow the notation in [1] in order to archive W matrix. The IFSA finds a linear transformW which maximizes the log-likelihood Eq.(3). Learning independent subspaces is carried out by a stochastic gradient ascent method, whose updating rule has the form

$$\Delta \vec{w_k} \propto \vec{x} < \vec{w_k}, \vec{x} > \varphi(\sum_{r \in F_{j(k)}} < \vec{w_k}, \vec{x} >^2 \quad (4)$$

where j (k) is the index of the feature subspace which $\vec{w_k}$ belongs to and φ is the score function, i.e., $\varphi = \frac{p'}{p}$. More details on IFSA (including the description of hypothesize density p) can be found in [1].

2.2 Membership Scoring Function

For the linear decomposition model, X = SA, we interpret gene expression data based on the following assumption previously suggested by [10].

(a) A Expression mode, the row of A, can be regarded as a latent variable which represents the feature of expression profiles of samples.

(b) A row vector of S, a effector of expression modes, shows how much a gene influences the occurrence of each expression mode.

The membership scoring function is a function to decide which cluster a gene belongs to. By virtue of the assigned scores, genes can be clustered in a specific expression mode in a group, on the behalf of a expression mode. Three types of membership scoring function are listed as follows.

2.2.1 Membership Scoring Function for PCA

PCA assumes that its components are sorted by their eigenvalues in a descending order; the first component seizes a maximal variance of data. Therefore, the membership scoring function for PCA is simply a function of expression modes as shown below. Denoting $J_s(n) = \sum_k |S_{nk}|$, we let the function ψ_n

$$\psi_n = \max_k \frac{|S_{nk}|}{J_s(n)} \quad (5)$$

2.2.2 Membership Scoring Function for ICA

A paper [10] took the variances of row vectors of A into account for the membership scoring function. Since the paper employed ICA as a linear decomposition model, informative components were determined based on the component variances. Since FastICA, used in [10], maximizes the statistical independence between columns of S by mutual information, the scoring function also considers mutual information. In this paper, we address simpler scoring function for ICA for convention. Let $J_A^k = \sum_k A_{kt}^2$, then the membership scoring function for ICA is as follows.

$$\varphi_n = \max_k |S_{nk}| / J_s(n) + J_A^k / < J_A > \quad (6)$$

2.2.3 Membership Scoring Function for IFSA

IFSA, similar to ICA, maximizes sparseness (or nongaussianity) of data components. However, IFSA regards the dependence among its components as important information. Therefore, IFSA acquires the dependence information by generating J groups which contain K components, simultaneously allowing K-tuples to be dependent, whereas k components belonging to different groups to be independent. To obtain independent feature space, IFSA defines an energy function, represented by the norm of projection of the input vectors. IFSA algorithm estimates the statistical independence between feature subspaces using energy function,

E. To take the amount of information of each subspace into account, we devised the scoring function for the purpose of involving the effects of $J_E^{k(j)}$ on each member in j-th group. Defining the energy function of IFSA as Eq.(2), we obtain the membership scoring function for IFSA:

$$\varphi_{n,k(j)} = \max_{k \in j} S_{nk} / J_s(n) \times E_j / < E > \quad (7)$$

3. RESULTS

We applied membership scoring functions to time-series microarray data with linear decomposition models. The data consists of 6178 ORFs and 77 tissue samples during the cell replication cycle in the budding yeast *Saccaromyces cerevisiae*[16]. We evaluated the amount of genes' contribution to expression modes in PCA/ICA case or a feature subspace in IFSA case and clustered genes by the estimated scores. Afterwards, we employed a correlation method in order to divide the genes of a cluster into two groups on the basis of the sign of genes' correlation values with an expression mode (beyond 99% significant level); one group corresponds to an activation related group and another an inhibition related one. Finally, we measured the cross-correlation between expression profiles in the same group, for the purpose of revealing the pairwise relations of genes. Cross-correlation provides time-lagged information, thus we can distinguish patterns into three types: concurrently expressed pairs, a gene prior to another (activators, or inhibitors), and a gene following after (activatee or inhibitee).

To analyze the potential for determining regulatory pairs from Spellman data sets, we began by constructing a database of genes, categorized by phenotypes of mRNA regulated with the cell cycle. Genes have 5 phases of cell cycle: Early G1, late G1, S, G2,and M phase. Their phenotypes are categorized by cell cycle regulators, chromosome, nuclear segregation, budding, directional growth, DNA replication, DNA repair and recombination, transcription, unknown/complex phenotype, mating pathway, glycolysis and respiration, biosynthesis, miscellaneous, and unknown function. Among these genes, we found well-studied 386 genes, which helps to compare clustering results by IFSA to ICA and PCA, and to figure out what the clustered genes imply.

3.1 Threshold and Membership Scoring Function

The number of genes we archived by linear methods relies on a membership scoring function and its threshold value. We defined threshold values in order to filter out unrelated genes in a cluster according to correlation coefficients between a gene profile and a expression mode with 99% significant level. To a maximal extent, we obtained 307 genes for PCA, 240 for ICA, and 143 for IFSA in total(see Figure 2). Different from PCA, ICA and IFSA algorithms depends on the number of group and the number of members and hence, the result can be flexible. The higher threshold, the tighter relationship to the expression mode a gene has. Therefore, the expert could draw his/her knowledge to decide how much close each other and how many genes he wanted.

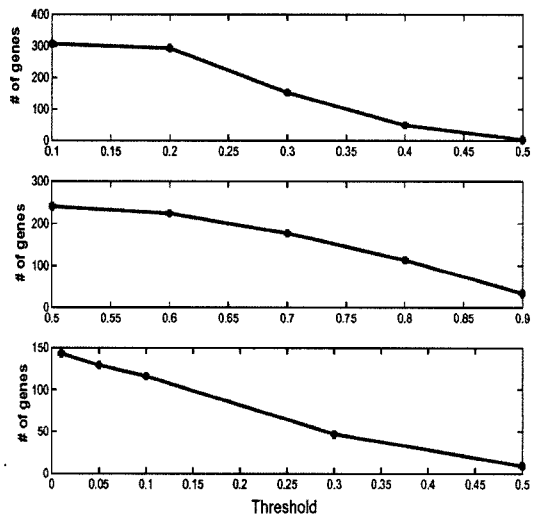


Figure 2 Threshold versus the number of genes in a cluster. For illustration of the threshold effect, we randomly choose three clusters

3.2 Multiple Interaction Among Genes

Multiple interaction among genes have been issued for a long time. However, wet-experimental technologies have not allowed us to investigate multiple interaction simultaneously. Even if technically possible, it is difficult to interpret their

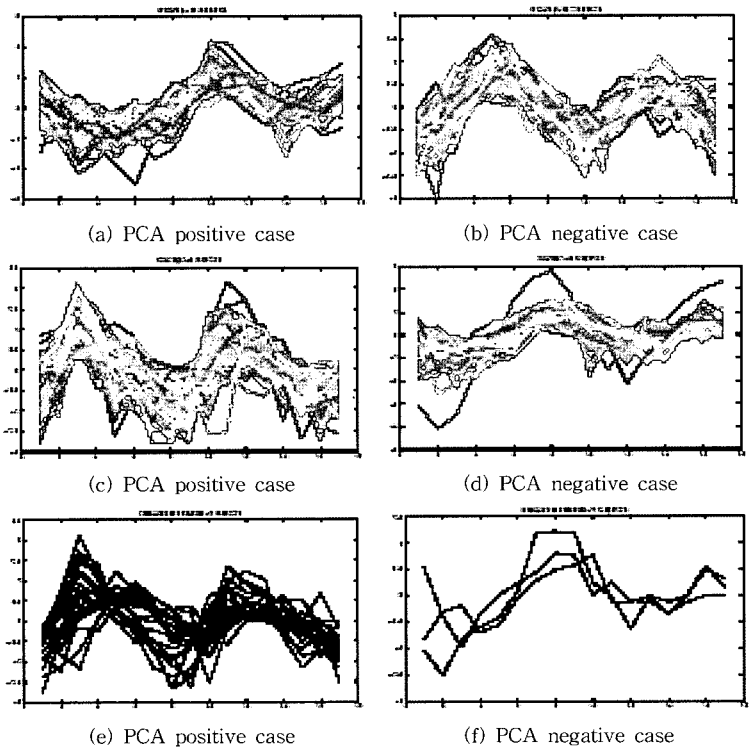
relationship. In dry-experiments, the problem is also considered to be one of NP-complete problems, whose proof and more details can be found in [15]. By now, gene studies in dry-experiments are in two models: clustering, for instance, linear models, and pairwise interpretation, for example, bayesian network and edge detection. The bayesian network reguires high computation complexity and its continuity assumption is not satisfied when the number of time points in the time-series data is small. The edge detection method considers time delays and asymmetry between rising and falling edges of expression profiles, hence pairwise regulatory genes were revealed. In PCA or ICA case of linear models, genes in a cluster can be regarded as having multiple interaction between them, but it does not take into account time delay and asymmetry of expression profile, thus this misses information, such as a mode and its lagged mode can identify a regulation relationship.

The IFSA model has a subspace, interpreting an energy function, E_j with K-tuple expression modes. Based on phase-shift invariant features of IFSA, we propose a model of gene expression based on the following assumption.

(1) For an expression mode, if the sign of correlation coefficients with gene profiles implies directionality; the positive sign means an activation-related movement, meanwhile, the negative sign means an inhibition-related one (for a general linear model case).

(2) For different expression modes in the same group, high cross-correlation value may imply that genes play a different role but are related hierarchically, such as regulatory relation (for only a IFSA specific case).

In Figure 3, we draw expression profiles in a cluster by a membership scoring function for PCA, ICA and IFSA. As shown, the membership scoring function cluster genes well. Moreover, the sign of correlation is a good basis for directionality of expression profiles.



(a) PCA positive case (b) PCA negative case

(c) PCA positive case (d) PCA negative case

(e) PCA positive case (f) PCA negative case

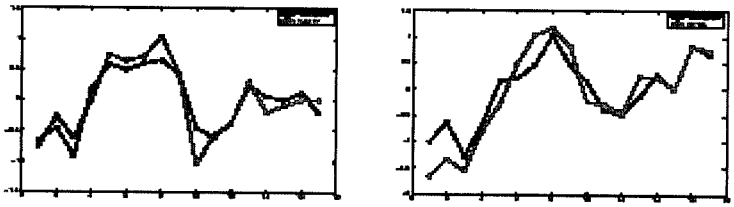
Figure 3 Analysis Result Comparisons Among PCA, ICA, and IFSA

On the other hand, Figure 4 represents pairwise interactions by measuring cross-correlation. Cross-correlation considers the time-shift correlation between two genes as well as direct similarity. Figure 4(a) and (b) display that those model and cross-correlation well matched the most similar patterns; IFSA model is good at direct similar pattern matching, as well as seen in Figure 4. Note that we identified that IFSA model was useful for detecting the expression pattern having time-lagged or asymmetric closeness Figure 4(c).

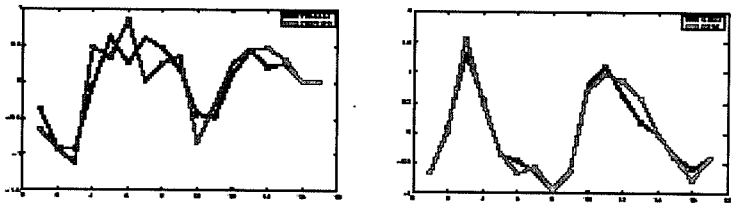
To take a close look at multiple interaction, we demonstrated the gene-gene interaction usign graph. The nodes in a graph implies genes in a cluster by membership scoring function. The edges were lied only if there exists biological relationship between two genes. To illustrate how effective our method and IFSA, we chose a category, *DNA replication*. It was found that the proposed membership scoring function successfully clustered genes belonging to *DNA replication* group by PCA, ICA and IFSA algorithm. The threshold was adjusted to maximize the number of genes of *DNA replication*. However, the effectiveness depended on each algorithm. In the PCA case, it contained about 17%, In the case of ICA, the cluster had about 33%. IFSA, however, contained 83%. In the cluster of *DNA replication*, there were 53 genes by setting the IFSA model with the group Size, J = 3 and the number of group components, K = 2. The connection between 53 genes in the first group is represented in Figure 5. In the group, for instance, CDC54, MCM2, and CDC46, corresponding to the first mode with the positive sign of a correlation value, belong to the Cdc46p/Mcm2p/Mcm3p family. The colored nodes in Figure 5 represented that the samed-colored nodes are biologically related by GO[17] and MIPS[18].

4. DISCUSSION

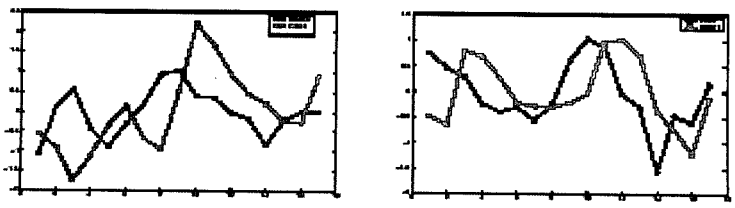
By constructing genetic network or clustering genes, many researchers tried to investigate the roles of genes in the biological system. Microarray technology and other biotechnologies made it



(a) Pair-wise Interaction by PCA model



(b) Pair-wise Interaction by ICA model



(c) Pair-wise Interaction by IFSA model

Figure 4 Pair-wise Interaction Comparison Among PCA, ICA and IFSA

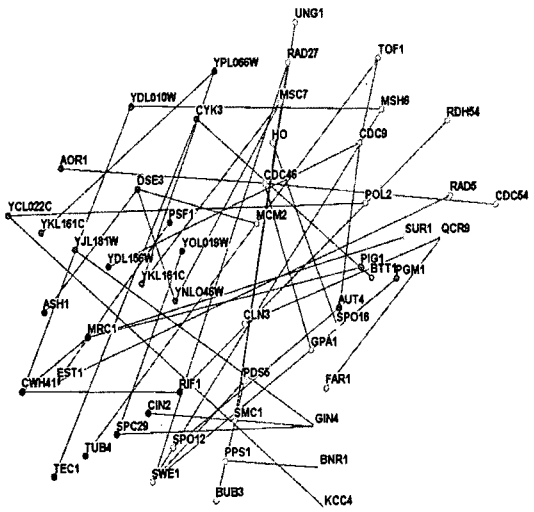


Figure 5 Mutiple Interactions by IFSA: The same colored circles represent that their characteristics is identical (based on GO and MIPS database). Each edge shows interaction between two genes. For instance, CDC54, MCM2, and CDC46 linked in the picture work for DNA unwinding, pre-replicative complex formation and maintenance and DNA replication initiation with 4.49×10^{-9}. Those present only 12 among 7270 in total of yeast genes in SGD database

possible to obtain thousands of gene expressions under the same condition. To track the expression profiles of genes and detect edge changes could lead to the discovery of pairwise regulatory relations. Meanwhile, clustering methods identified where genes are active in the cell cycle. However, both cannot satisfy our goals: the former cannot recognize multiple relations and it is weak to annotate their functional relation at the algorithm level; the latter loses information of the gene-gene connection.

In this paper, we show a new approach to identify gene information from time series data. To overcome previous limitations, we introduced two methods: membership scoring function and IFSA. As mentioned previously, membership scoring function properly cluster genes with linear decomposition models. Possessing phase and shift invariant characteristics, IFSA makes it possible to observe time delays and asymmetry between gene profiles, thus finding a biologically more reliable connection.

In light of the limitations of complexity, directional multiple interaction of genes remains a desirable but difficult problem. We regard the way to construct the genetic network by considering each gene relation as a bottom-up approach. On the contrary, we regard the top-down approach. Therefore, we attempt to catch the functional information first. Afterwards, we calculate the correlation with members in the same cluster, and we may be able to find the entire gene relation.

Meanwhile, linear models may only identify dominant pattern related functions. That is, they have originally created for the purpose of data visualization or reduction method. Hence, if some genes play a critical role in biological processes but have a statistically small portion, IFSA (like PCA and ICA) cannot recognize the pattern as an expression mode. To overcome this limitation, a linear decomposition model that is more sensitive to biological data is required. A linear decomposition method allowing prior knowledge may help in this context, as it would result in more weights to such genes and success to identify minor but critical genes.

References

[1] A. Hyvärinen at al, "Emergence of Phase , Shift Invarianct Features by Decomposition of Natural Images into Independent Feature Subspaces," Neuro Computing, Vol.12, No.7, pp.1705-1720, 2000.

[2] J. -F. Cardoso, "Multidimensional Independent Component Analysis," ICASS 1998.

[3] P. T. Spellman at al, "Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization," Molecular Biology of the Cell, Vol.9, pp. 3273-3297, 1998.

[4] O. Alter at al, "Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling," Pacific Symposium Biocomputing, Vol.97, No.18, pp.10101-10106, 2000f.

[5] N. Friedman at al, "Using Bayesian Networks to Analyze Expression Data. Journal of Molecular Biology," Vol.7, pp.601-620, 2000.

[6] M.B. Eisen at al, "Cluster Analysis, Display of Genome-wide Expression Pattern," Pacific Symposium Biocomputing, Vol.95, pp.14863-14868, 1998.

[7] N. Holter at al, "Fundamental Patterns Underlying Gene Expression Profiles: Simplicity from Complexity," Pacific Symposium Biocomputing, Vol.97, pp.8409-8414, 2000.
[8] S. Raychaudhuri at al, "Principal Components Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series," Proc. Pacific Symp. Biocomputing, pp.452-463, 2000.
[9] G. Hori , at al, "Blind Gene Classification Based on Independent Component Analysis of Microarray Data," Proc. ICA, 2001.
[10] W. Liebermeister "Linear Modes of Gene Expression Determined by Independent Component Analysis," Bioinformatics, Vol.18, No.1, pp.51-60, 2002.
[11] T. D. Moloshok at al, "Application of Bayesian Decomposition for Analysing Microarray Data," Bioinformatics, Vol.18, No.4, pp.566-575, 2002.
[12] A. M. Martoglio at al, "A Decomposition Model to Track Gene Expression Signatures: Preview on Observer-Independent Classification of Ovarian Cancer," Bioinformatics, Vol.18, No.12, pp.1617-1624, 2002.
[13] Vladimir Filkov at al, "Identifying gene regulartory networks from experimental data," Journal of Parallel Computing, Vol.27, pp.141-162, 2001.
[14] Andrew T. Kwon at al, "Inference of transcriptional regulation relationships from gene expression data," Bioinformatics, Vol.19, No.8, pp. 905-912, 2003.
[15] T. Chen,H.L at al, "MModeling Gene Expression with Differential Equations," Pacific Symposium Biocomputing, pp.29-40, 1999.
[16] P.T.Spellman, at al., "Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," Molecular Biology of the Cell, vol.9, pp.3273-3279, 1998.
[17] http://geneontology.org
[18] http://mips.gsf.de/genre/proj/yeast/index.jsp

김 혜 진
2001년 포항공과대학교 화학공학과(학사). 2003년 포항공과대학교 컴퓨터공학과(석사). 2004년~현재 한국전자통신연구원 연구원. 관심분야는 Machine learning, Bayesian learning, ICA, Robot vision, Speaker recognition

최 승 진
1987년 서울대학교 전기공학과(학사). 1989년 서울대학교 전기공학과(석사). 1996년 University of Notre Dame, EE(박사) 1997년 일본 이화학연구소, Frontier 연구원. 1997년 9월~2001년 2월 충북대학교 전기전자공학부 조교수. 2001년 2월~현재 포항공과대학교 컴퓨터공학과 부교수. 관심분야는 Machine learning, Probabilistic graphical models, Bayesian learning, and ICA

방 승 양
1966년 일본 Kyoto대학 전기공학(학사) 1969년 서울대학교 전기공학(석사). 1974년 미국 University of Texas 전산학(박사). 1974년~1981년 미국 Bell 연구소 근무. 1981년~1985년 한국전자기술연구소 시스템부 실장, 부장 역임, (주)유니온 시스템 전무. 1986년~현재 포항공대 컴퓨터공학과 교수. 관심분야는 패턴인식, 신경회로망