Genomics & Informatics Vol. 1(1) 32-39, September 2003

Gene Expression Pattern Analysis via Latent Variable Models Coupled with Topographic Clustering

Jeong-Ho Chang, Sung Wook Chi, and Byoung-Tak Zhang*

Biointelligence Laboratory, School of Computer Science and Engineering, Seoul National University, Seoul, Korea

Abstract

We present a latent variable model-based approach to the analysis of gene expression patterns, coupled with topographic clustering. Aspect model, a latent variable model for dyadic data, is applied to extract latent patterns underlying complex variations of gene expression levels. Then a topographic clustering is performed to find coherent groups of genes, based on the extracted latent patterns as well as individual gene expression behaviors. Applied to cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae*, the proposed method could discover biologically meaningful patterns related with characteristic expression behavior in particular cell cycle phases. In addition, the display of the variation in the composition of these latent patterns on the cluster map provided more facilitated interpretation of the resulting cluster structure. From this, we argue that latent variable models, coupled with topographic clustering, are a promising tool for explorative analysis of gene expression data.

Keywords: gene expression pattern, latent variable model, Fisher kernel, topographic clustering

Introduction

DNA microarrays from cDNA chips or oligonucleotide chips provide a global, parallel view on the expression patterns of hundreds or thousands of genes in a cell at a specific time, under a specific experimental conditions or processes (Baldi and Hatfield, 2002; Shamir and Sharan, 2002). They are among the most powerful and versatile tools for functional genomics, and advances in technologies for these high-density arrays are enabling to produce enormous amount of expression data. For the efficient exploration and analysis of these massive microarray data, appropriate analysis tools are needed which are different from conventional methods for the traditional one-gene-in-one-experiment paradigm.

Clustering is a key step to analyzing gene expression data, where genes are systematically grouped together according to their similarity in expression patterns. Among the earliest and extensively used algorithms are hierarchical clustering (Eisen *et al.*, 1998; Spellman *et al.*, 1998; Scherf *et al.*, 2000) and k-means (Herwig *et al.*, 1999; Tavazoie *et al.*, 1999). Eisen *et al.* (Eisen *et al.*, 1998) analyzed the gene expression data of budding yeast *Saccharomyces cerevisiae* using the hierarchical clustering algorithm and showed that genes of known similar functions were grouped together with the clustering. Tavazoie *et al.* (Tavazoie *et al.*, 1999) used the k-means algorithm to identify transcriptional regulatory sub-networks in yeast. Despite their successful applications in many other biological tasks as well as these, however, the algorithms suffer from some shortcomings, such as lack of robustness, incompetence of proper handling of global structure (hierarchical clustering) and non-structure in resulting clusters (k-means clustering) (Tamayo *et al.*, 1999; Fowlkes *et al.*, 2002).

In this paper, we present an effective approach for gene expression pattern analysis, based on latent variable models and topographic clustering. Latent variable models are a powerful tool for discovering latent structure underlying data objects. They are well suited to extract correlational patterns of variables and provide a compact, useful representation of data. Also, a natural similarity metric between data can be derived from latent variable models (Jaakkola and Haussler, 1999; Tsuda *et al.*, 2002). Topographic clustering algorithms such as SOM (self-organizing maps) (Kohonen, 1997) are an attractive method for providing not only good clustering performance but also easy visualization and interpretation of the results (Tamayo *et al.*, 1999; Törönen *et al.*, 1999).

We use an aspect model that is a latent variable model for co-occurrence data to extract significant patterns underlying gene expression data. Using these patterns and original expression data, two cluster maps are produced from a topographic clustering. In the first map, each cluster is represented by the average expression levels of genes

*Corresponding author: E-mail btzhang@bi.snu.ac.kr

Accepted 4 April 2003

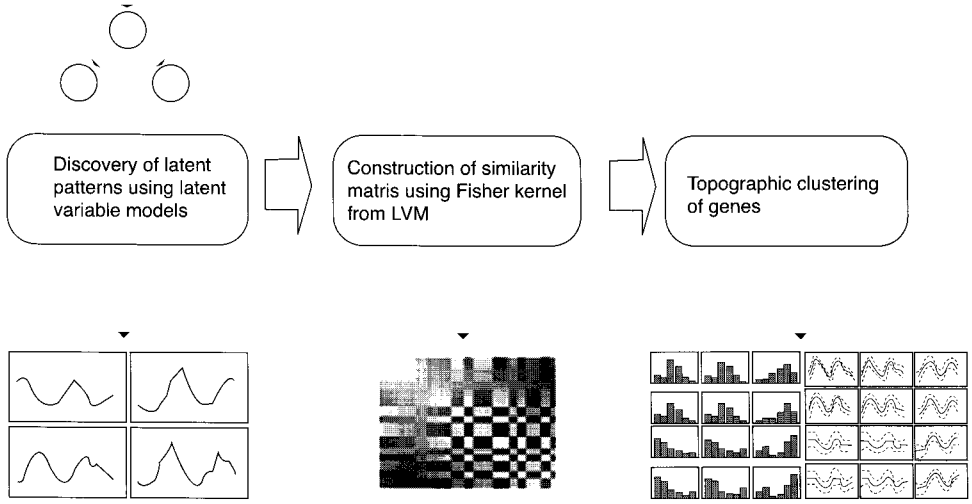

Fig. 1. The overall procedure for gene expression profiling based on latent variable models coupled with topographic clustering.

in the cluster. The second map summarizes the clusters in terms of its characteristics in convex combinations of latent patterns.

We applied our method to the expression data of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae*. Our results show that the latent patterns discovered by the aspect model are well reflecting the characteristic expression behaviors across various cell cycle phases. Two independent cluster maps provided by topographic clustering are shown to be useful in their ability to provide an intuitive, integrated understanding of complex variations of hundreds of genes during the progress of cell cycle.

Materials and Methods

Our approach to gene expression analysis consists of three steps: (1) identification of meaningful patterns inherent in gene expression data using a latent variable model; (2) construction of a similarity matrix containing similarity values of all gene pairs; (3) clustering of genes based on the similarity matrix and a topographic clustering algorithm. Fig. 1 summarizes the overall procedure of our approach to gene expression data analysis.

Latent pattern analysis based on aspect model

As the first step to the analysis of gene expression data, we use an aspect model (Hofmann, 2001) which is a latent variable model for co-occurrence data. Latent variable models are a powerful approach to probabilistic modeling where a set of observed variables are supplemented with additional latent variables (Bishop, 1999). Besides their efficiency for density estimation, the latent variable models are very useful in discovering latent structure underlying a set of data.

Let D be an $N \times M$ matrix where M is the number of data and N is the number of attributes and each element is indexed by (a_j, x_i), $(1 \leq i \leq M,\ 1 \leq j \leq N)$. In the aspect model, the joint probability of $P(a_j, x_i)$ is decomposed by introducing a latent class variable $z \in Z = \{z_1, z_2, \cdots, z_K\}$,

$$\begin{aligned} P(a_j, x_i) &= P(x_i)P(a_j \mid x_i) \\ &= P(x_i)\sum_{k=1}^{K} P(a_j \mid z_k)P(z_k \mid x_i). \end{aligned} \tag{1}$$

This indicates that the conditional probability or sample specific attribute distribution $P(a_j \mid x_i)$is approximated by a convex combination of K aspects $P(a_j \mid z_k)$. The goal of the modeling by the aspect model is to estimate class-specific attribute distribution $P(a_j \mid z_k)$ and data specific class distribution $P(z_k \mid x_i)$ from data set.

Formally, the aspect model is learned by maximizing the log data-likelihood L=log $p(D)$:

$$L = \sum_{i=1}^{M}\sum_{j=1}^{N} r(a_j, x_i) \log P(a_j, x_i), \tag{2}$$

where $r(a_j, x_i)$ is the value of a_j in the data x_i. The Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) is used to estimate the parameters maximizing L. In the E-step, the posterior probabilities $P(z_k \mid a_j, x_i)$ are computed and in the M-step, the parameters $P(a_j \mid z_k)$ and $P(z_k \mid x_i)$ are estimated. These two steps are iteratively alternated until convergence.

Estimating similarity by the fisher kernel

One of the most important things in clustering algorithms is how to define (dis)similarity measure between data, and the choice is quite subjective. In kernel methods like support vector machines (Cristianini and Shawe-Taylor, 2000; Schölkopf and Smola, 2001), this corresponds to selecting an appropriate *kernel function*.

Given two data vectors x_i and x_j in the input space X, a valid kernel function $k(x_i, x_j)$ can be represented as the inner product between two feature vectors in a (high dimensional) feature space F, that is,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle, \tag{3}$$

where ϕ is a fixed mapping from X to F, that is, $\phi: X \rightarrow F$. The mapping ϕ can be viewed as a preprocessing step to draw out meaningful features from data, and the technique of kernel function provides an efficient way of inner-product calculation in the feature space F (Cristianini and Shawe-Taylor, 2000; Graepel *et al.*, 1998).

In this paper, we utilize the kernel function presented in (Hofmann, 2000), where an intrinsic kernel, named the *Fisher kernel*, was derived from the aspect model. The Fisher kernel (Jaakkola and Haussler, 1999) is an approach to constructing kernels from generative probabilistic models and defines the similarity between data based on information-geometric principles using the gradient space of the generative model.

Let $l(x_i)$ be the expected log-probability of a sample x_i under the aspect model parameterized by $\theta=\{\theta_1, \theta_2\}$ where $\theta_1=\{P(z_k)$ and $\theta_2=\{P(a_l|z_k)\}$. The kernel is derived by computing the Fisher scores $\nabla_\theta l(x_i)$ and the Fisher information matrix $I(\theta)=E\{\nabla_\theta l(x) \nabla_\theta^T l(x)\}$. With the square-root transformation of parameters $\theta=\{\theta_1, \theta_2\}$ and the approximation of $I(\theta)$ as the identity matrix (Hofmann, 2000), the kernel from the aspect model is given by

$$\begin{aligned} k(x_i, x_j) &= \nabla_\theta^T l(x_i) I(\theta)^{-1} \nabla_\theta l(x_j) \\ &\approx k_1(x_i, x_j) + k_2(x_i, x_j), \end{aligned} \tag{4}$$

where

$$\begin{aligned} k_1(x_i, x_j) &= \nabla_{\theta 1}^T l(x_i) \nabla_{\theta 1} l(x_j) \\ &= \sum_k \frac{P(z_k \mid x_i) P(z_k \mid x_j)}{P(z_k)}, \end{aligned} \tag{5}$$

$$\begin{aligned} k_2(x_i, x_j) &= \nabla_{\theta 2}^T l(x_i) \nabla_{\theta 2} l(x_j) \\ &= \sum_l \tilde{P}(a_l \mid x_i) \tilde{P}(a_l \mid x_j) \sum_k \frac{P(z_k \mid a_l, x_i) P(z_k \mid a_l, x_j)}{P(a_l \mid z_k)}. \end{aligned} \tag{6}$$

Here, $\tilde{p}(a_l, x_i)$ in the Equation (6) is an empirical distribution estimated by $r(a_l, x_i) / \sum_m r(a_m, x_i)$. Eventually, this similarity metric is the linear summation of k_1 and k_2 with equal weights: $k_1(x_i, x_j)$ is the similarity in the representation on the latent space. $k_2(x_i, x_j)$ calculates the inner product of the original two data x_i and x_j, where products of the corresponding attributes are weighted by the degree of overlap of respective posterior probabilities (the second line in Equation (6)) (Hofmann, 2000).

Clustering by kernel-based soft topographic mapping

The kernel-based soft topographic mapping (STMK) (Graepel *et al.* 1998) is a topographic clustering algorithm based on kernel-based distance measures and principles from statistical physics. It can provide not only a stable and good clustering algorithm, but also a topographic map of the clustered data. By using a specific kernel function and the kernel trick, the algorithm can also perform clustering efficiently in a feature space related to the original data space in a nonlinear way, compared to the basic SOM.

In the SMTK, clustering is defined in terms of an optimization problem. The cost function E to be optimized is given as

$$E = \sum_{i=1}^{M} \sum_{j=1}^{C} m_{ij} e_{ij} \tag{7}$$

where M is the number of data and C is the number of clusters. The binary variable m_{ij} indicates whether the ith sample belongs to the jth cluster, and e_{ij} is the error occurred by assigning the ith sample to the jth cluster. Given a mapping ϕ from input space X to a feature space F as explained above, the partial assignment cost e_{ij} is defined as

$$e_{ij} = \frac{1}{2} \sum_{l=1}^{C} h_{jl} \left\| \phi(\mathbf{x}_i) - \mathbf{c}_l \right\|^2, \qquad \sum_{l=1}^{C} h_{jl} = 1 \quad (\forall j), \tag{8}$$

where $\phi(x_i)$ and c_l are a sample and a cluster center in a feature space F, respectively. h_{jl} is a neighborhood function between jth and lth clusters, and determines the coupling between the two clusters as in SOM.

The STMK algorithm provides an efficient procedure to find a good solution to the minimization of the cost function in Equation (7), based on deterministic annealing (Rose *et al.*, 1990) and the EM algorithm. Starting with a random initialization of e_{ij} for all data and clusters, the algorithm iteratively updates the parameters using the EM algorithm with some temperature-annealing schedule. In the E-step, expectation values of m_{ij}, the probability that the ith sample is assigned to the jth cluster, is estimated for each pair of data and clusters. Then all the parameters related with the calculation of cluster centers in a feature space F are updated in the M-step.

Dataset

We applied our method to the yeast *Saccharomyces*

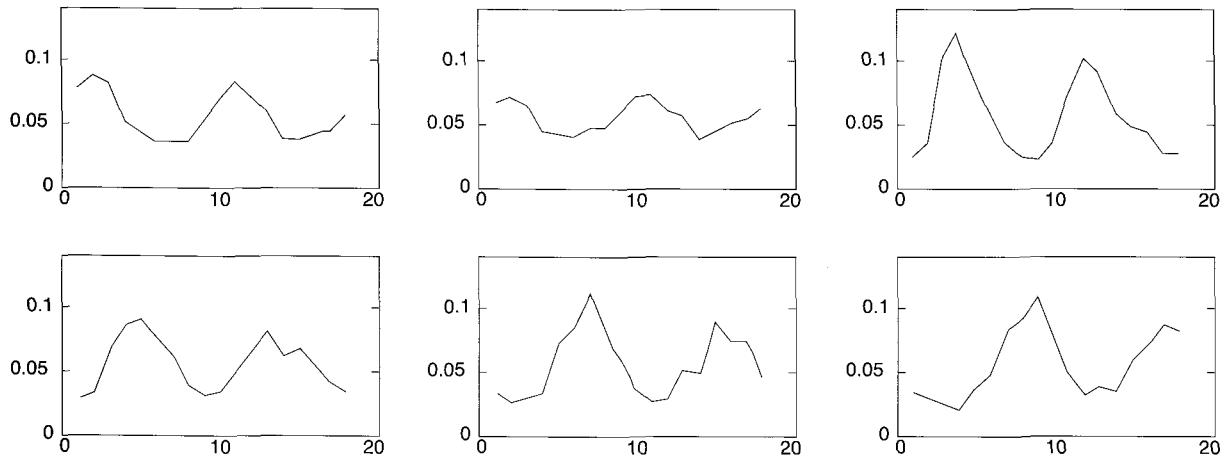

Fig. 2. The six inherent patterns for the yeast cell cycle data, extracted by an aspect model with K=6. Each pattern is characterized by its conditional distribution $P(a_j \mid z_k)/\sum_{i=1}^{18} P(a_i \mid z_k$ Among the 73 conditions, patterns over 18 time points from α factor-based synchronization are shown. For a clear presentation, patterns were reordered in ascending order according to their first time point of the peak expression, and probability values were rescaled. The peak time indices are (2, 11), (2, 11), (4, 12), (5, 13), (7, 15), and (9, 17), from pattern 1 to pattern 6.

cerevisiae cell cycle data (Spellman *et al.*, 1998). This dataset (available at *http://cellcycle-www.stanford.edu*) contains the time series of relative expression measurements of more than 6,000 genes from cell cultures synchronized by three independent methods. In the first set, the α mating factor pheromone was used to arrest cells in G1 and samples were taken in every 7 minutes throughout 140 minutes. In the second set, centrifugal elutriation was used to collect small G1 cells in every 30 minutes during 6.5 hours. In the third, cdc15 temperature-sensitive mutant was used to arrest cells in late mitosis and samples were taken in every 10 minutes during 300 minutes. Additionally, samples from arrest of a cdc28 temperature-sensitive mutant (Cho *et al.*, 1998) were included in the dataset. The analysis of these data sets enables researchers to identify cycles or waves of expressions that are meaningful in biological processes.

Spellman and his colleagues (Spellman, 1998) have identified 800 genes whose expressions are *cell cycle-regulated* using their periodicity and correlation analysis methods. Among the 800 yeast genes, we selected 700 genes with at most 4 missing entries over 73 conditions: 18 time points over α arrest, 24 over cdc15 arrest, 17 over cdc28 arrest, and 14 over elutriation. For each gene, the missing values were filled by the average values of the either side of the time points, assuming the smooth variation of gene expression levels during the cell cycle. Finally, the dataset is represented by a 73×700 matrix where each gene is represented by a column vector containing 73 attributes of gene expression levels.

Results

Initially, we applied the aspect model to extract latent patterns inherent in the gene expression variation during the progress of cell cycle. In the aspect model, observed values are assumed to be count-valued. Prior to the analysis, therefore, all the expression values v_{ji} ($1 \leq i \leq 700$, $1 \leq j \leq 73$) were first transformed to positive integer values by $r(a_j, a_i) = [100 \times (1/(1+\exp(-v_{ji})))]$[1].

The yeast cell cycle data was analyzed with varying number of latent factors $z=\{z_1, z_2, \cdots, z_K\}$, from K=5 to K=10. The result with K=6 is shown in Fig. 2, where the expression patterns are plotted across the 18 time points from the α-factor experiments. It can be seen that different periodicities of gene expression are captured in the six patterns. Compared with the cell cycle phase described in (Spellman *et al.*, 1998), these patterns relatively well reflect peak expression behaviors in respective cell cycle phases: M/G1-phase (patterns 1 and 2), G1-phase (pattern 3), G1/S-phase (pattern 4), S/G2-phase (pattern 5), and M-phase (pattern 6).

Additionally, we clustered the 700 genes using the learned aspect model in Fig. 2 and the STMK algorithm.

[1] By multiplying a constant 100, we actually consider to the second decimal point of the values (in the range of 0 to 1) obtained by $1/(1+\exp(-v_{ji}))$.

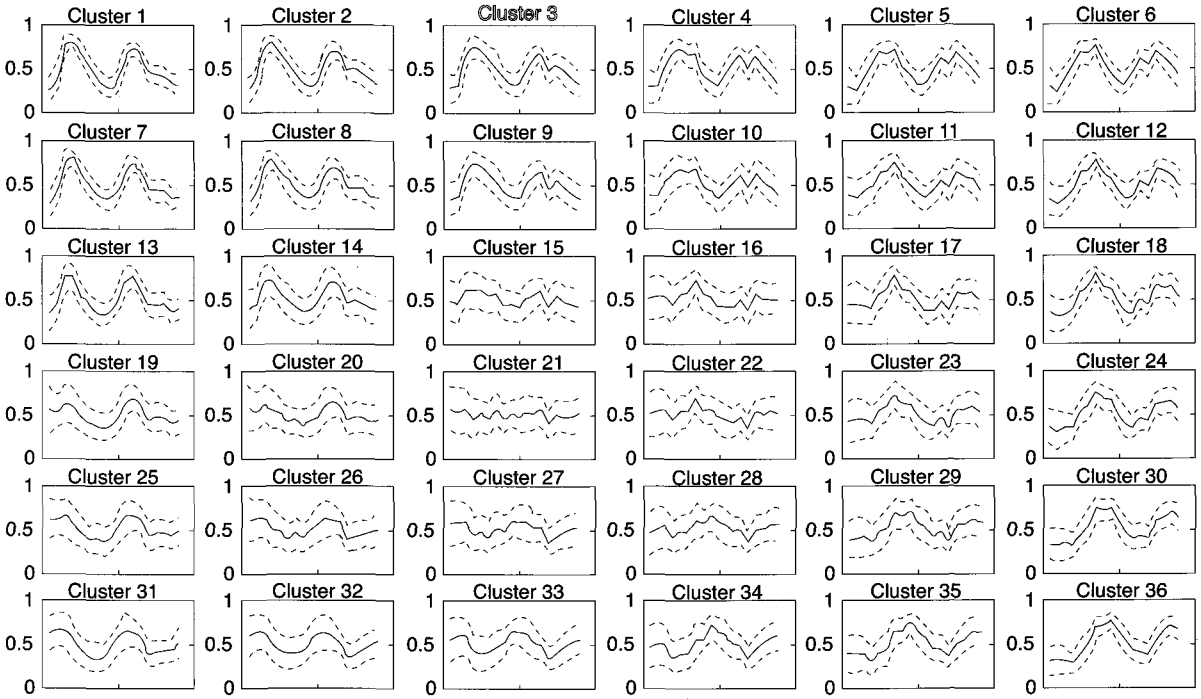

Fig. 3. The STMK clustering map of genes from yeast cell cycle data. 700 genes with less than or equal to 4 missing values, among Spellman's 800 putative cell cycle-regulated genes, were clustered into 36 (6×6) groups. Expression levels (y-axis) over the 18 time points (x-axis) from the experiment of α-factor synchronization are shown. In each cluster, the solid line is the mean expression values of the genes in the cluster and the dotted line is the standard deviation in expression levels of those.

The kernel function in Equation (4) was used to calculate similarities between two genes. Fig. 3 shows the resulting cluster map with 36 clusters on the 6×6 grid. Each cluster is represented by the average expression pattern of genes in the cluster. Different periodicities in gene expression are shown across the clusters on the grid and the adjacent clusters have similar patterns.

Cluster structure is also presented in terms of latent patterns (Fig. 4). A gene x_i is mapped into the latent space produced by the aspect model and is represented by a vector of probabilities, $\mathbf{y}_i = (P(z_1|x_i), P(z_2,x_i), \cdots P(z_6|x_i)), \ \sum_l P(z_l|x_i) = 1$ Then cluster centers C_k ($1 \leq k \leq 36$) in the latent space are expressed as linear combinations of y_i by

$$\mathbf{c}_k = \sum_{i=1}^{700} a_{ik}\mathbf{y}_i, \qquad a_{ik} = \frac{\sum_{l=1}^{36} h_{kl} P(x_j \in cluster l)}{\sum_{j=1}^{700}\sum_{l=1}^{36} h_{kl} P(x_j \in cluster l)},$$

where a_{ik} is the contribution of the ith sample to the cluster k.

The genes in a cluster are now distinguished from those in other clusters by the combination characteristics of the latent patterns. The cluster 1 of the top-left-most, for example, shows the highest probabilities in the pattern 3. The third pattern in Fig. 2 characterizes genes of which the first peaks of their expression are observed at the measurement points around 21 minutes from the start of the α-factor experiments. This is a typical expression behavior of the genes in the G1 phase group identified in (Spellman *et al.*, 1998), and all 45 genes (*CDC45, CLB5, MSH2, MSH6,* etc) in the cluster 1 are found in the G1 phase group. Nearby clusters of cluster 1 in Fig. 4 shows the similar combination patterns, and most of the genes in those clusters are found in the G1 phase group. Similar analysis is possible for clusters 6, 31, 36: patterns 4 and 5 are dominant in cluster 6, patterns 1 and 2 in cluster 31, and pattern 6 in cluster 36. Genes in those clusters are found in the phase groups of which typical expression characteristics are similar to the dominant patterns. 35 genes (*HHF1, HTA1, HTA2, HTB2,* etc) among 39 in the cluster 6, 21 (*ACE2, CLB1, CLB2, MOB1,* etc) genes among 24 in cluster 31, and all 35 genes (*ASH1, CDC46, MFA2, SIC1,* etc) in cluster 36 belong to S phase group, M/G1 phase group, and G2/M phase group, respectively. The constitution of all the clusters in terms of phase groups identified by Spellman *et al.* are described in Table 1.

A closer examination of Fig. 4 provides another

Table 1. The assignment table of 700 genes to 36 clusters. Cluster enumeration is the same as those in Figures 3 and 4. In each cluster, genes assigned to the cluster are divided according to their phase group specified by Spellman et al. Note that these classification of genes in a cluster are not rigid, since the classification at the boundary between two phases are rather ambiguous (Spellman et al., 1998). For example, six of the eight genes (PSA1, CBF2, YLL012W, YBL009W, SVL3, KAR3, SPC98, SRL1) in cluster 5, classified as being peak-expressed in G1 phase, are actually among the first 14 genes from the boundary of G1 and S phase.

G1 (45)	G1 (27)	G1 (31)	G1 (20) S (2)	G1 (8) S (13)	S (21) S/G2 (3)
G1 (23)	G1 (13)	G1 (12)	G1 (3)	S (3)	S (5) S/G2 (12)
G1 (27) M/G1 (8)	G1 (11) M/G1 (1)	G1 (7) M/G1 (1)	S (8), S/G2 (2) G1 (2), G2/M (2)	S (5) S/G2 (10)	S (1) S/G2 (30)
G1 (6) M/G1 (8)	G1 (4) M/G1 (8)	G1 (4), G2/M (1), S (2), S/G2 (2)	S (4), S/G2 (5), G2/M (5)	S/G2 (15) G2/M (2)	S/G2 (16) G2/M (2)
G1 (10) M/G1 (16)	G2/M (2) M/G1 (6) S (1)	G1 (1), M/G1 (3), G2/M (8), S/G2(1)	S/G2 (3) G2/M (10)	S/G2 (3) G2/M (14)	S/G2 (11) G2/M (4)
M/G1 (35), G1 (3), G2/M (1)	G2/M (8), G1 (1), M/G1 (5)	G2/M (29)	G2/M (26)	G2/M (23) S/G2 (1)	G2/M (35)

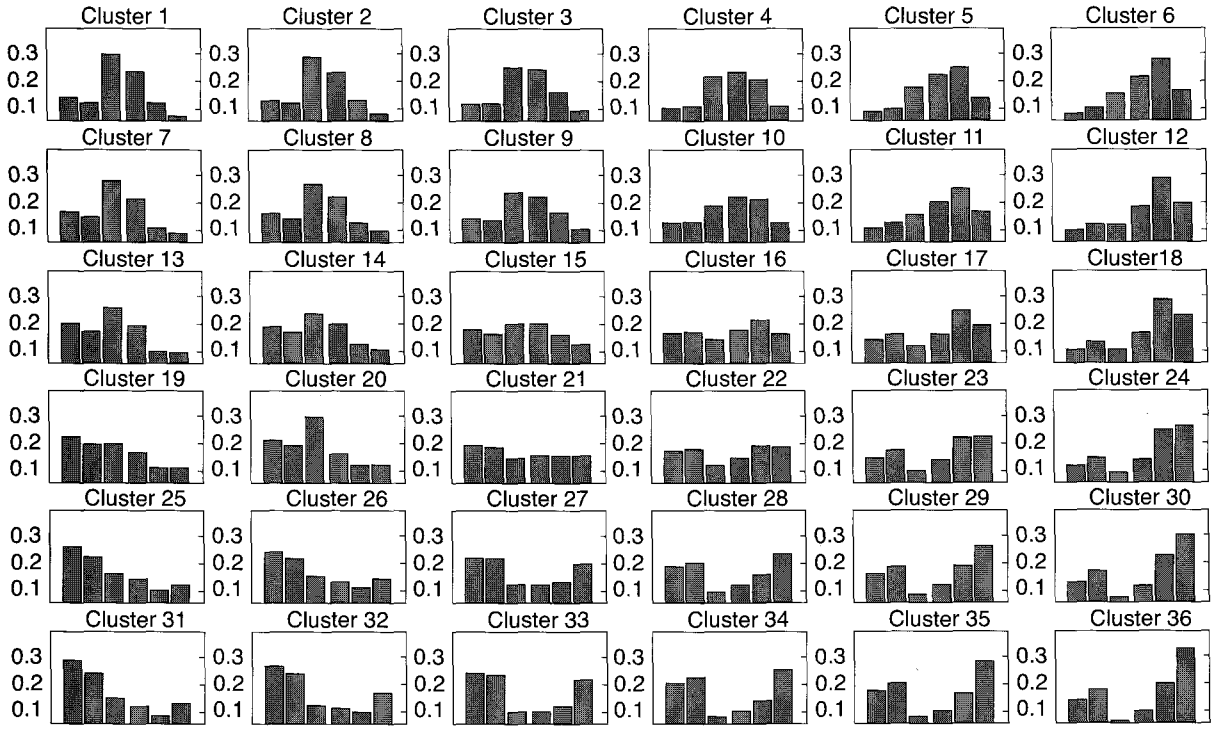

Fig. 4. The STMK clustering map with representation according to the composition of latent patterns. When each gene x_i is represented in the latent space by $y_i=\{P(z_1|x_i),\cdots, P(z_5|x_i)\}$ based on six patterns (x-axis) in Figure 2, each cluster shows the average responsibilities (y-axis) of the latent patterns for expression behaviors of genes in the cluster.

interesting interpretation. When we go from cluster 1 to cluster 6 along the top margin, for example, we can see that pattern 3 declines steadily and pattern 5 rises instead. This shows the distinction among genes in different clusters in terms of the peak time in their expression, where the first peak time of the genes with higher probabilities of pattern 3 is earlier than that of the genes with the opposite combination. Referred to the G1 phase group ordered by their peak times (Spellman, 1998), genes in cluster 1 and cluster 2 are found in the earlier part than

those in cluster 3 and 4. Most of the genes in cluster 5 are located in the boundary of G1 phase group and S phase group, and most of those in cluster 6 are found in the S phase group. Similar tendencies are observed in subsequent routes: from cluster 6 to cluster 36 along the right margin, from cluster 36 to cluster 31 along the bottom margin (Table 1). In this way, a proper discovery and exploitation of the inherent structures in gene expression behaviors helps us for further understanding of expression time-series data.

Conclusions

We have presented an effective approach for gene expression pattern analysis, based on latent variable models and topographic clustering. Applied to the expression analysis of the cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae*, various distinctive patterns underlying the data set were extracted by an efficient latent variable model, that is, the aspect model. The identified patterns well correspond to the typical expression behaviors of genes regulated in specific cell cycle phases. In topographical clustering, based on the STMK algorithm and a similarity measure derived from the aspect model, the discovered gene clusters are well reflecting their characteristic expression patterns during the progress of cell cycle.

In addition, our apprach has provided two informative maps on the cluster structure. The first map depicts each cluster by the expression variation of genes in the cluster, across the measured time points. In the second map, the clusters are depicted by the combination patterns of *latent properties* inherent in the data. A meta-level analysis and visualization based on these coupled maps enables a more principled interpretation of gene expression patterns.

As future works, we are considering to incorporate another type of information for gene expression data analysis. The learning in latent variable models is basically unsupervised. In some cases, however, existing labeled data can be used during training to improve performance. For the yeast data, for example, more than 100 genes have been already identified as being cell cycle-regulated. Hierarchical latent variable models and semi-supervised learning in the model could be applied to utilizing this information. Additionally, we will study on the selection of appropriate number of latent factors of aspect models in the analysis of gene expression data. In probabilistic generative models like aspect models, this problem can be reduced to a model selection problem. Then, standard techniques from statistics, based on various criteria such as BIC and MDL, could be used in our works.

Acknowledgments

This research was supported by the Korean Ministry of Science and Technology under the NRL, IMT-2000, and BrainTech programs.

References

Baldi, P. and Hatfield, G.W. (2002). DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modelings. (Cambrige, UK, Cambridge University Press).

Bishop, C.M. (1999). Latent variable models. In Learning in Graphical Models, Jordan, M.I.,ed. (Cambrige; The MIT Press), pp.371-403.

Cho, R.J. et al. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell 2, 65-73.

Cristianini, N. and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (Cambridge; Cambridge University Press).

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1-38.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863-14868.

Fowlkes, C., Shan, Q., Belongie, S., and Malik, J. (2002). Extracting global structure from gene expression profiles. In edso, Methods of Microarray Data Analysis II: Papers from CAMDA 01. Lin, S.M. and Johnson, K.F., (Norwell, MA: Kluwer Academic Publishers), pp. 81-90.

Graepel, T., Burger, M., and Obermayer, K. (1998). Self-organizing maps: Generalizations and new optimization techniques. Neurocomputing 21, 173-190.

Herwig, R., Poutska, A. J., Muller, C., Bull, C., Lehrach, H., and O' Brien, J. (1999). Large-scale clustering of cDNA-fingerprinting data. Genome Research 9, 1093-1105.

Hofmann, T. (2000). Learning the similarity of documents: an information-geometric approach to document retrieval and categorization. In Advances in Neural Information Processing Systems 12, 914-920.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42, 177-196.

Jaakkola, T. and Haussler, D. (1999). Exploiting generative models in discriminative classifiers. In Advances in Neural Information Processing Systems 11, 487-493.

Kohonen, T. (1997). Self-Organizing Maps (New York: Springer-Verlag).

Rose, K., Gurewitz, E., and Fox, G. (1990). A deterministic annealing approach to clustering. Pattern Recognition Letters 11, 589-594.

Scherf, U. et al. (2000). A gene expression database for the molecular pharmacology of cancer. Nature Genetics 24, 236-244.

Schölkopf, B. and Sm Nola, A.J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. (Cambridge, MA: MIT Press).

Shamir, R. and Sharan, R. (2002). Algorithmic approaches to clustering gene expression data. In Jiang, T., Smith, T., Xu, Y., and Zhang, M., edso, Current Topics in Computational Biology, Jiang, T., Smith, T., Xu, Y., and Ahang, M., (Cambridge, MA: MIT Press), pp 269-299.

Spellman, P.T. et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9, 3273-3297.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S. Dmitrovsky, E., Lander, E. S., and Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96, 2907-2912.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nature Genetics 22, 281-285.

Törönen, P., Kolehmainen, M., Wong, G., and Castren, E. (1999). Analysis of gene expression data using self-organizing maps. FEBS Letters 451(2), 142-146.

Tsuda, K., Kin, T., and Asai, K. (2002). Marginalized kernels for biological sequences. Bioinformatics 18(Suppl 1), S268-S275.