論 文 の 内 容 の 要 旨

論文題目　Statisitical inference of gene networks from time course gene expression profiles using state space models
（状態空間モデルを利用した時系列遺伝子発現プロファイルからの統計的推測）

氏　　名　広瀬 修

By ingeniously interacting enormous number of biochemical molecules, cells grow, divide, and differentiate themselves, and maintain their homeostasis. Even the same biochemical molecule can vary its function according to the situation such that when and where it exists in a cell, in what level of differentiation, or cells are in what condition such as heat, cold, and starvation. Thus, for understanding living organisms, it is insufficient to understand a function of each biochemical molecule, and it is essential to study relationships or networks between biochemical molecules. A gene regulatory network, one of such networks between biochemical molecules, describes the control system of gene expressions. Though many researchers have studied reconstruction of gene regulatory networks from time course gene expression data, existing methods for gene network estimation have been limited, mainly because the length of the time course data is fairly short, e.g. typically less than 10, whereas the number of genes involved ranges from 10^2 to 10^4. Obviously, the length of time course gene expression profiles is not sufficient to infer such a large gene network. For example, the maximum likelihood estimator of the vector autoregressive model does not exist if the number of genes is greater than the length of the time course. In this thesis, we challenge the imbalance between dimensionality and the length of time course and propose a novel method to infer large scale gene networks and an extension of the method.

A key idea to overcome the imbalance is to explore temporal networks of the transcriptional modules which are sets of genes sharing a common function or are involved in the same pathway rather than the use of gene-level networks. In the context of gene expression analysis, the transcriptional modules may be defined by the groups of transcriptionally co-expressed genes. In this thesis, we provide an approach to identify the potential transcriptional modules and map them onto the gene-level networks, i.e. the module-based gene networks. The proposed method is based on the state space model which has a potential to construct large gene networks from time course gene expression profiles. We applied the proposed method to the time course gene expression profiles of human umbilical vein endothelial cells during growth factor deprivation-induced apoptosis. The estimated gene network suggested that the TRAF1 gene played an important role for the cell cyclearrest and inflammatory response during apoptosis.

The extension of the proposed method is the mixture of state space models.Recently, novel kinds of time course gene expression profiles that existing methods may fail to analyze have appeared. Baranzini et al. investigated the longitudinal gene expression change of multiple sclerosis (MS) patients with treatments of recombinant interferon B (rIFN-B). In this data set, each MS patient is characterized by a gene expression matrix whose column vectors are gene expression vectors for corresponding observed time points. They aimed at classifying 53 MS patients, composed of 33 good responders and 20 poor responders for the therapy of rIFN-B.Hence, the problem is to classify samples where each sample is characterized by matrix data. In this thesis, we present a novel clustering method based on a mixture model that make use of time series of datasets effectively. State space models are used as component models of the mixture in order to handle high dimensional time series and to avoid the over-parameterization by considering dimension reduction. The proposed method addreses the following tasks: (1) clustering samples according to temporal patterns of gene expressions, (2) automatic detection of genes that discriminate identified clusters, (3) estimation of a gene network for each cluster. We applied the proposed method to the dataset of MS patients.
As a result, we succeeded to separate MS patients according to the response level of the therapy of rIFN-B, and to select genes that discriminate groups of patients.