5. Workshop Automatisierungstechnische Verfahren für die Medizin vom 15.-16. Oktober 2004 in Saarbrücken

„An Approach to Structural Modelling of Eucaryotic Gene Dynamics inferring Reed-Muller Forms from Microarray Time Series Data“

S. Faisal, G. Lichtenberg, H. Werner
Arbeitsbereich Regelungstechnik, TU Hamburg-Harburg, Hamburg, Deutschland
E-Mail: s.faisal@ tu-harburg.de, lichtenberg@ tu-harburg.de, h.werner@ tu-harburg.de

Band: „Tagungsband, Automed 2004“
Editors: W. I. Steudel
ISBN: 3-00-013509-X
Pages: 85-86

An Approach to Structural Modelling of Eucaryotic Gene Dynamics inferring Reed-Muller Forms from Microarray Time Series Data

S. FAISAL, G. LICHTENBERG, H. WERNER

Arbeitsbereich Regelungstechnik, Technische Universität Hamburg-Harburg
Eissendorfer Straße 40, 21073 Hamburg

E-mail: {s.faisal, lichtenberg, h.werner}@ tu-harburg.de

INTRODUCTION

It is well known that each activity of a living cell is controlled by the expression levels of the genes present in it. Moreover, pattern of genes expressed in a cell is characteristic of its current state. Recently developed microarray technology has made it possible to measure simultaneously the expression levels of thousands of genes at a certain sampling time [see Webpage http://genome-www5.stanford.edu].

With the emergence of these technologies, it is possible to model genetic systems in a much more detailed way than before. This can help in identifying complex genetic diseases as well as in drug discovery etc.

Generally, all microarray data in time series for an organism can formally be represented by a set

$$\aleph = \left\{ (x(0),...,x(T)) \mid x(k) \in \square^{n} \right\}$$

of sequences of n measured gene expression levels T+1 discrete sampling time points. The modelling problem is then to find a transition function f such that

$$f(x(k)) = x(k+1).$$

There have been many approaches used so far to describe this phenomenon. In general, all efforts can be divided into two categories: discrete models and continuous models. Discrete modelling efforts include mainly Boolean Network [Akutsu1999] or Bayesian Network models [Aburatani2003]. On the other hand, continuous modelling approaches include differential equations models [Dehoon2003] or linear models [D'haeseleer1999].

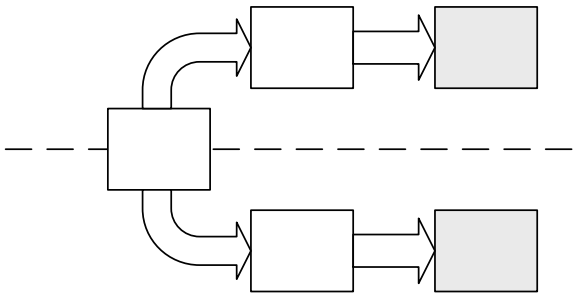

Fig. 1: Two ways of Gene Dynamics Modelling

The actual situation in gene dynamics modelling can be visualized by figure 1: Discrete models like Boolean Networks have the advantage that they share many properties with the Genetic Networks such as global complex behavior, periodicity etc.

However, these techniques start by quantizing the data, (I) and finally identify (II) a discrete model (D) that can be either deterministic or nondeterministic or stochastic (Dynamical Bayesian Networks), that reflects the qualitative behavior of the measured data [Akutsu1999]. Biological information (formulated as rules) can be included in the identification step, for instance the canalizing property of Boolean models that has been postulated.

The other type of modelling methods uses quantitative approaches, which do not quantize, but filter and cluster the data (III). Through identification techniques (IV), continuous models such as differential equation models are derived from the filtered data. The resulting models are able to approximate the continuous levels of gene expression.

Both approaches have severe draw backs. With the discrete approach it is not possible to model the intermediate expression levels occurring in microarray data. On the other hand, with quantitative approaches there is no possibility of defining biological constraints as well as they have difficulties in handling extreme values occurring in gene expression data.

MATERIALS AND METHODS

It is clear from figure 1, that the two modelling approaches lead to entirely two different paths which are neither compatible nor comparable.

This paper claims that there exists such a link which can join the two modelling paths as described in Figure 2. The missing link between the qualitative and quantitative approach can be obtained by using well-known methods of continuous representations of Boolean functions. The connecting link is the block (F), which includes a set of discrete models, represented by continuous polynomials known as Reed-Muller-Forms [Franke1994].

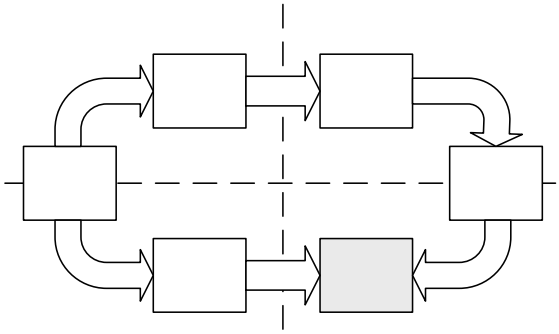

Fig. 2: Proposed Modelling Cycle

A Reed-Muller-Form of n variables has the form

$$f(x) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{j=2}^{n} \sum_{i=1}^{j-1} a_{ij} x_i x_j + \sum_{k=3}^{n} \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} a_{ijk} x_i x_j x_k + \ldots + a_{123\ldots n} x_1 x_2 \ldots x_n.$$

These polynomials have the ability to describe both, continuous as well as discrete functions. These representations coincide with Boolean functions at Boolean values of the variables x_i but otherwise are continuous functions.

RESULTS

For the sake of showing applicability, the results of applying modelling cycle on yeast time series data, [http://genome-www5.stanford.edu] of four randomly chosen genes (TFC3, EFB1, SSA1, FUN14) are presented here. Linearly normalized expression levels of the above mentioned four genes are shown in Figure 3.

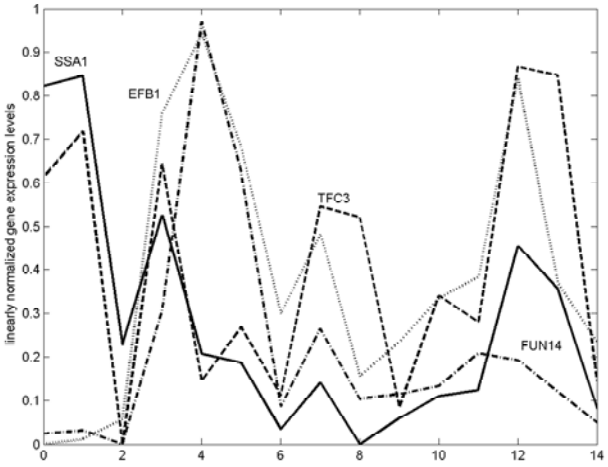

Fig. 3: Normalized Time Series Data of Yeast Genes

For this particular data set it is assumed that the expression level of each gene at time k+1 is effected by the expression level of all four genes at time k. Following the steps of modelling cycle 30 canalizing Boolean functions consistent with the measurements, corresponding to gene TFC3 were obtained. In figure 4, all 30 canalizing polynomials results along with the time series measurement of gene TFC3 are shown. Optimization of the least square error yielded an optimal Reed-Muller-Form, that is given in figure 4 by the thick line.

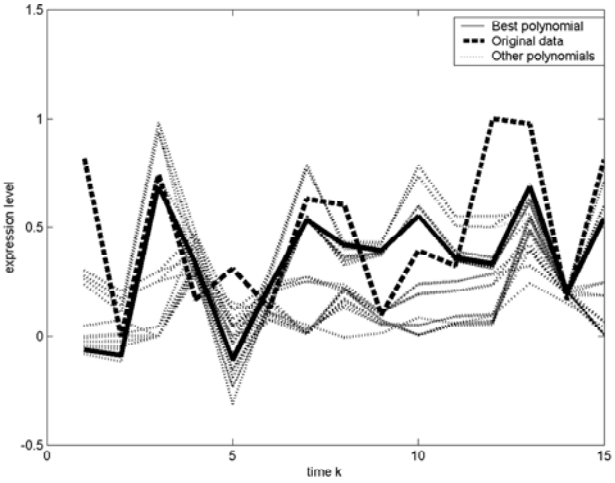

Fig. 4: Time Series Results of Structured Model

DISCUSSION

A new approach to modelling gene dynamics has been presented which has the capability of both, discrete as well as continuous model restrictions. The continuous restrictions arise from continuous levels of measured gene expression data while the discrete ones come from structural restrictions caused by principles of biology.

REFERENCES

[Aburatani2003]
S. Aburatani, T. Goto, S. Imoto, S. Kim, S. Kuhara, S. Miyano, K. Tashiro, "Bayesian Network and Nonparametric heteroscedastic regression for nonlinear modeling of genetic networks", *J Bioinf Comp biology*, No 2, p 231, 2003

[Akutsu1999]
T. Akutsu, S. Kuhara, S. Miyano "Identification of Genetic Networks from a Small Number of Gene Expression Patterns under the Boolean Network Model" in *Proc. Pacific. Symp. Biocomp.*, Hawai, 1999

[D'haeseleer1999]
P. D'haeseleer, S. Fuhrman, R. Somogyi, X. Wen, "Linear Modelling of mRNA Expression Levels during CNS Development and Injury" in *Proc. Pacific. Symp. Biocomp.*, Hawai, 1999

[Dehoon2003]
J. Dehoon, S. Imoto, S. Miyano, N. Ogasawara, "Inferring Gene Regulatory Networks from Time Ordered Gene Expression Data of Bacillus Subtilis using Differential Equation", in *Proc. Pacific. Symp. Biocomp.*, Hawai, 2003

[Franke1994]
D. Franke, "Sequentielle Systeme", Vieweg, 1994.