

## Efficient Fuzzy Logic-Based Algorithm for Microarray Network Identification and Prediction in Bioinformatics

P. Carbonell

Department of Systems Engineering and Control, Polytechnic University of Valencia

Pza Ferrándiz i Carbonell, 2, 03801 Alcoi, Spain, phone: +34 96 6528458; e-mail: pjcarbon@isa.upv.es

### Introduction

High-throughput experiments in Bioinformatics are characterised by huge amounts of data which need to be processed in order to extract as much meaningful information as possible. Typically, data is analysed for pattern recognition, classification, and network modelling. For instance, the complete genome of certain specimen is composed of big collections of nucleotides or amino acids sequences. These sequences have to be analysed in order to detect candidates to coding and/or regulating regions. Adding even more complexity to this scenario, gene products (proteins) interact in complex regulation networks of activators and repressors [1].

An excellent experimental source of analysis and detection of these potential protein pathways are microarrays [2]. In microarrays, levels of gene expression are automatically measured under different conditions, tissues, etc, resulting in a high dimension data array which has to be explored.

Fuzzy logic has been shown to be a useful and reliable technique for characterising experimental data in microarrays, as well as in other areas of Bioinformatics [3]. In [4], a simple microarray problem was successfully addressed by using fuzzy logic, although the given example is not easily extensible to a more complex framework, due to the fact of what is called the curse of dimensionality: a combinatorial explosion, even for a small number of genes and experiments. In [5], this problem was avoided by using the so-called union rule configuration (URC) method [6]. This method is based on the assumption of additive separability of the fuzzy rule base [7], requirement which is verified in these authors by considering only some particular interrelations among genes.

Based on these previous works, it is apparent the need for a more general methodology, who takes into account all the possible combinations among several genes, performs an exhaustive search, providing at the same time a computationally efficient algorithm for the computation of higher order interactions which avoids the curse of dimensionality.

### Methods

Our study will be limited to a special type of fuzzy systems which will be called triangular fuzzy system, characterized by some internal symmetry and simple structure. Our objective will be to take advantage of these properties in order to make the system additively separable, and therefore, computational simple.

**Definition 1.** A standard fuzzy system  $\Phi(\mathbf{x}, \boldsymbol{\mu}, \mathbf{z})$  verifies the following properties:

1. The input space has been pre-processed and normalised so that:  $(x_1, x_2) \in \left[-\frac{N}{2}, \frac{N}{2}\right] \times \left[-\frac{N}{2}, \frac{N}{2}\right]$ .

2. The input space for each variable is partitioned in  $N$  segments defined by the centres:  $[x_{k0}, x_{k1}, \dots, x_{kN}]$ ,  $k = 1, 2$ .

3. There are  $N+1$  fuzzy input sets with membership functions  $\mu_{ki}(x_k)$   $i = 0 \dots N$ .

4. The membership functions verify  $\mu_{ki}(x_{ki}) = 1$  and  $\mu_{ki}(x_k) = 0$  iff  $x_k \leq x_{ki-1}$  or  $x_k \geq x_{ki+1}$ .

5. The membership functions are normalized, i.e.:

$$\mu_{ki}(x_k) = \begin{cases} \mu_{ki-1}^*(x_k) = 1 - \mu_{ki-1}(x_k), & x_{ki-1} \leq x_k \leq x_{ki}, \\ \mu_{ki+1}^*(x_k) = 1 - \mu_{ki+1}(x_k), & x_{ki} \leq x_k \leq x_{ki+1}. \end{cases}$$

6. The output of the fuzzy rule-based system is calculated using the centre average defuzzification of the rule table  $z_{ij}$ :

$$7. z(x_1, x_2) = \sum_{i=0}^N \sum_{j=0}^N z_{ij} \prod_{k=1}^2 \mu_{ki}(x_k). \quad (1)$$

**Proposition 1.** A standard fuzzy system  $\Phi(\mathbf{x}, \boldsymbol{\mu}, \mathbf{z})$  with triangular-shaped membership functions can be transformed into an additively separable standard fuzzy

system  $\Delta(\hat{\mathbf{x}}, \boldsymbol{\omega}, \boldsymbol{\alpha})$  by rotating the input variables by 45 degrees, leading to the following defuzzification:

$$z(x_1, x_2) = \hat{z}(\hat{x}_1, \hat{x}_2) = \sum_{i=0}^N \sum_{k=0}^2 \alpha_{ki} \omega_{ki}(\hat{x}_k). \quad (2)$$

The intuitive idea is that by rotating the input variables by 45 degrees, a new reference system across the diagonals is established where the grid formed by the centres of the membership functions is disentangled. First, we give the input variables in the new coordinates:

$$\hat{\mathbf{x}} = \mathbf{T}\mathbf{x} \quad \mathbf{T} = 2^{-1/2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \quad (3)$$

**Lemma 1.** Given a point in the plane  $(x, y)$ , the product of its coordinates equals half the difference between the squares of the second and first diagonal coordinates:

$$xy = \frac{1}{2}(\hat{x} + \hat{y})(-\hat{x} + \hat{y}) = \frac{\hat{y}^2 - \hat{x}^2}{2}. \quad (4)$$

Therefore, the product of the two variables is transformed in an additively separable function of each new coordinate. This property will be very useful since, as we will see, every fuzzy output of a fuzzy triangular system can be decomposed into products of the input variables.

In a triangular standard fuzzy system, each fuzzy membership function is defined as a linear function of the variable  $\mu_{ki}(x_k) = a_{ki}x_k + b_{ki}$ ,  $x_{ki} \leq x_k \leq x_{ki+1}$ . At a particular point  $(x_1, x_2)$ , where  $x_{1i} \leq x_1 \leq x_{1i+1}$  and  $x_{2j} \leq x_2 \leq x_{2j+1}$ , the output of equation (1) is reduced to:

$$z(x_1, x_2) = \sum_i^{i+1} \sum_j^j z_{ij} \mu_{1i}(x_1) \mu_{2j}(x_2). \quad (4)$$

If we rotate the plane by 45 degrees and walk across the diagonals, assuming that the membership functions are arranged symmetrically:

$$x_1 = x_2 \rightarrow \begin{cases} \hat{x}_1 = 0, \\ \hat{x}_2 = 2^{1/2} x_2, \end{cases} \quad (6)$$

where  $\mu_{1i}(x_1) = \mu_{2j}(x_2)$ .

$$x_1 = -x_2 \rightarrow \begin{cases} \hat{x}_1 = 2^{1/2} x_1, \\ \hat{x}_2 = 0, \end{cases} \quad (5)$$

where  $\mu_{2j}(x_2) = \mu_{1N-i}(-x_1) = \mu_{1i}(x_1)$ .

This last expression, combined with the conditions given in Definition 1 of the triangular fuzzy system, allows the separation of variables along the diagonals, leading to the following scalar functions, where indexes  $i$  and  $j$  are kept for referring to  $x_1$  and  $x_2$  in order to improve readability:

$$\begin{cases} z(x_2, x_2) = \sum_{k=0}^2 \beta_{kj} \mu^k_{2j}(x_2), \\ \beta_{2j} = z_{jj} + z_{j+1j+1} - z_{jj+1} - z_{j+1j}, \\ \beta_{1j} = z_{jj+1} + z_{j+1j} - 2z_{j+1j+1}, \\ \beta_{0j} = z_{j+1j+1}, \end{cases} \quad (8)$$

$$\begin{cases} z(x_1, x_1) = \sum_{k=0}^2 \beta_{ki} \mu^k_{1i}(x_1), \\ \beta_{2i} = z_{iN-i} + z_{i+1N-i-1} - z_{iN-i+1} - z_{i+1N-i}, \\ \beta_i = z_{iN-1} + z_{i+1N-i} - 2z_{i+1N-i-1}, \\ \beta_{0i} = z_{i+1N-i-1}. \end{cases} \quad (9)$$

In the new coordinates framework, the input space can be divided again into segments:

$$\hat{z}(\hat{x}_1, 0) = \sum_{k=0}^2 \beta_{ki} \mu^k_{1i}(2^{-1/2} \hat{x}_1), \quad (10)$$

$$\hat{z}(0, \hat{x}_2) = \sum_{k=0}^2 \beta_{kj} \mu^k_{2j}(2^{-1/2} \hat{x}_2). \quad (11)$$

Expressions (10) and (11) represent the projection of surface  $\hat{z}(\hat{x}, \hat{y})$  along the new axis. In order to determine the structure of the sought additively separable fuzzy system, output equation (2) is particularized to the projection on the axis:

$$\hat{z}(\hat{x}_1, 0) = \sum_{i=0}^N \alpha_{1i} \omega_{1i}(\hat{x}_1) + S_{\hat{x}_2}(0), \quad (12)$$

$$\hat{z}(0, \hat{x}_2) = S_{\hat{x}_1}(0) + \sum_{j=0}^N \alpha_{2j} \omega_{2j}(\hat{x}_2). \quad (13)$$

$S_{\hat{x}_1}(0)$  and  $S_{\hat{x}_2}(0)$  represent the value of the contribution of each variable to the output, at the origin of coordinates  $\hat{z}(0,0)$ . At a particular point  $(\hat{x}_1, \hat{x}_2)$  expressed in the new coordinates, where  $\hat{x}_{1i} \leq \hat{x}_1 \leq \hat{x}_{1i+1}$  and  $\hat{x}_{2j} \leq \hat{x}_2 \leq \hat{x}_{2j+1}$ , the output will be given by:

$$\begin{aligned} \hat{z}(\hat{x}_1, \hat{x}_2) &= \alpha_{1i} \omega_{1i}(\hat{x}_1) + \alpha_{1i+1} \omega^*_{1i+1}(\hat{x}_1) + \\ &+ \alpha_{2j} \omega_{2j}(\hat{x}_2) + \alpha_{2j+1} \omega^*_{2j+1}(\hat{x}_2). \end{aligned} \quad (14)$$

Taking into account property 4 of the system and expressions (12) and (13), a system of linear equations can be established corresponding to the reticule:

$$\hat{z}(\hat{x}_{1i}, 0) = \alpha_{1i} + S_{\hat{x}_2}(0) \quad i = 0 \dots N, \quad (15)$$

$$\hat{z}(0, \hat{x}_{2j}) = S_{\hat{x}_1}(0) + \alpha_{2j} \quad j = 0 \dots N, \quad (16)$$

An additional equation came from the origin of coordinates:

$$\hat{z}(0,0) = S_{\hat{x}_1}(0) + S_{\hat{x}_2}(0). \quad (17)$$

The following last equation of the linear system is used in order to give some symmetry to the system:

$$S_{\hat{x}_1}(0) = S_{\hat{x}_2}(0). \quad (18)$$

Expressions (15) – (18) contain  $2N + 2$  parameters and form  $2N + 2$  non-singular linear equations. Therefore, the system is solvable. On the other hand, the membership functions are obtained by considering again the different segments along the diagonals. i.e., if  $(\hat{x}_1 \leq \hat{x}_1 \leq \hat{x}_{1i+1}, \hat{x}_2 = 0)$  and  $(\hat{x}_1 = 0, \hat{x}_2 \leq \hat{x}_2 \leq \hat{x}_{2j+1})$ :

$$\hat{z}(\hat{x}_1, 0) = \alpha_{1i} \omega_{1i}(\hat{x}_1) + \alpha_{1i+1} \omega_{1i}^*(\hat{x}_1) + S_{\hat{x}_2}; \quad (19)$$

$$\hat{z}(0, \hat{x}_2) = \alpha_{2j} \omega_{2j}(\hat{x}_2) + \alpha_{2j+1} \omega_{2j}^*(\hat{x}_2) + S_{\hat{x}_1}. \quad (20)$$

Arriving finally at the general expressions:

$$\omega_{1i}(\hat{x}_1) = \frac{\sum_{k=0}^2 \beta_{ki} \mu^k_{1i}(2^{-1/2} \hat{x}_1) - S_{\hat{x}_2} - \alpha_{1i+1}}{\alpha_{1i} - \alpha_{1i+1}}, \quad (21)$$

$$\omega_{2j}(\hat{x}_2) = \frac{\sum_{k=0}^2 \beta_{kj} \mu^k_{2j}(2^{-1/2} \hat{x}_2) - S_{\hat{x}_1} - \alpha_{2j+1}}{\alpha_{2j} - \alpha_{2j+1}}. \quad (22)$$

In case some solution of the system of equations verifies  $\alpha_{1i} = \alpha_{1i+1}$ , it implies that  $\hat{z}(\hat{x}_1, 0) = \hat{z}(\hat{x}_{1i+1}, 0)$  and therefore, it suffices to make  $\alpha_{1i+1} = 0$ . By doing that, we will obtain denormalized membership functions around  $\hat{x}_{1i+1}$ . Nevertheless, in order to verify properties 4 and 5 in Definition 1, these functions can be normalized to the well-known fuzzy basis functions [8].

## Results

A simple fuzzy rule system  $\Phi(\mathbf{x}, \boldsymbol{\mu}, \mathbf{z})$  is considered with 2 inputs  $\mathbf{x} = (x_1, x_2)$  and 2 triangular membership  $\boldsymbol{\mu} = [\mu_{11}(x_1), \mu_{12}(x_1), \mu_{21}(x_2), \mu_{22}(x_2)]$  functions that divide the input space into 2 segments:  $x_{1i} = x_{2j} = [-1, 0, 1]$ . These two inputs represent the levels of expression of two different genes from a microarray experiment at a given conditions, tissue, etc.

The output of the fuzzy system  $z(x_1, x_2)$  models the interdependence of a third gene respect to the two input genes. In order to determine which fuzzy rule base is the best fit for a set of experiments, the  $4^2$  possible output vectors corresponding to each of the combinations must be computed. Tables 1-4 show some examples of these combinations, along with their interpretations. The left part of each table is the original fuzzy rule table  $z_{ij}$ , which is

not additively separable. The right part show the separation performed after the rotation, giving two fuzzy rule tables  $\alpha_{ij}$ , corresponding each one to one input. For this simple example, where  $n = 2$ , both tables have the same dimensions, since  $2n = 2^2$ , although the difference will be apparent for higher ones.

**Table 1.** Rules for  $x_1, x_2$  cross-activators

$x_1 / x_2$	L	H		$\omega_1$	$\omega_2$
L	0	0	$\hat{x}_1$	-0.125	0.875
H	0	1	$\hat{x}_2$	1	0

**Table 2.** Rules for  $x_1$  activator,  $x_2$  repressor

$x_1 / x_2$	L	H		$\omega_1$	$\omega_2$
L	0	0	$\hat{x}_1$	-0.125	0
H	1	0	$\hat{x}_2$	0.875	-0.125

**Table 3.** Rules for  $x_1$  activator,  $x_2$  indifferent

$x_1 / x_2$	L	H		$\omega_1$	$\omega_2$
L	0	0	$\hat{x}_1$	-0.25	0.75
H	1	1	$\hat{x}_2$	0.75	-0.25

**Table 4.** Rules for  $x_1, x_2$  activators with mutual repression

$x_1 / x_2$	L	H		$\omega_1$	$\omega_2$
L	0	1	$\hat{x}_1$	-0.25	0
H	1	0	$\hat{x}_2$	1	0

Figures 1-3 plot the fuzzy system obtained after rotation, corresponding to the system shown in Table 1 (cross-activation of the inputs). Figure 1-2 plot the new fuzzy membership functions for input  $\hat{x}_1$  and  $\hat{x}_2$  respectively. In the second case, the membership functions are denormalized, due to the facts exposed at the end of the previous section. If desired, the membership functions can be normalized by using the fuzzy basis function transformation, although from a practical point of view, there is no advantage from doing that operation.

Figure 3 is the output surface plot for the rotated system of Table 1, brought back to the original system of coordinates. The computed output coincides at any point with the original output, as expected since the transformation method is an exact expression.

## Discussion

The method presented in this paper allows the computation of the output of a particular type of fuzzy systems in polynomial time by means of an additive separation, and therefore avoiding the exponential time associated with the curse of dimensionality. The algorithm presented here will be very useful, since this algorithm allows to perform an exhaustive and deeper search within the data network, and therefore to increase the possibilities of finding hidden complex interactions among genes. Furthermore, this method can also be used in other data mining problems.

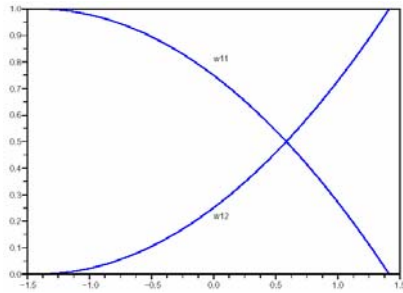


Fig. 1. Membership functions for  $\hat{x}_1$  of the cross-activator

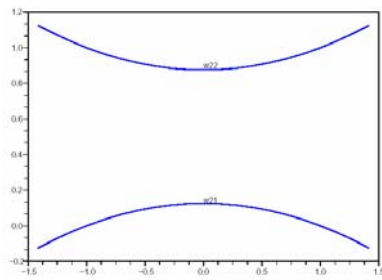


Fig. 2. Membership functions for  $\hat{x}_2$  of the cross-activator

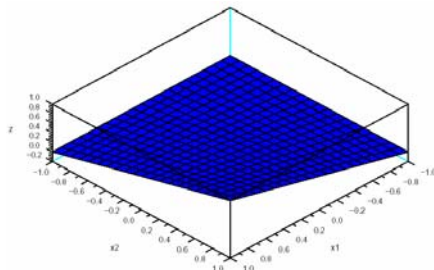


Fig. 3. Output surface for the crossactivator

The proposed method makes the computation of high dimension complex fuzzy systems more feasible. Furthermore, this method can also be used in other data mining problems in Bioinformatics applications, where it

is usually necessary to compute a big number of input combinations.

This study has presented the basic idea and has shown its reliability by means of a simple example. The next step in this research study will be to formalize the results for  $n$  dimensions, by using the superposition property of additive systems. Additionally, the development of an open source code library which implements the algorithm, and the study of its hardware implementation will be another research path to be explored.

## References

1. Sontag, E. Molecular systems biology and control // IEEE Conference on Decision and Control CDC05, Seville, 2005.
2. Liu, E.T. Gene array technologies in biological investigations // Proc. of the IEEE. – 2005. Vol. 93.- No. 4. - P. 737-749.
3. Resson, H., Natarajan, P., Varghese, R.S., Musavi, M.T. Applications of fuzzy logic in genomics // Fuzzy Sets and Systems. – 2005. - No. 152, No. 1. - P. 125-138.
4. Woolf, P.J., Wang, Y. A fuzzy logic approach to analyzing gene expression data // Physiol. Genomics. – 2000. - No. 3., - P. 9-15.
5. Sokhansanj, B.A., Fitch, J.P., Quong, J.N., Quong, A.A. Linear fuzzy gene network models obtained from microarray data by exhaustive search // BMC Bioinformatics. – 2004.- Vol. 5.- No. 108.
6. Combs, W.E., Andrews, J.E. Combinatorial rule explosion eliminated by a fuzzy rule configuration // IEEE Trans. On Fuzzy Systems. – 1998. - Vol. 6. - No. 1. - P. 1-11.
7. Weinschenk, J.J., Marks, R.J., Combs, W.E. Layered URC fuzzy systems: a novel link between fuzzy systems and neural networks // Int. Joint Conference on Neural Networks. – 2003. - Vol. 4. - P. 2995-3000.
8. Wang, L.X., Mendel, J.M. Fuzzy basis functions, universal approximation, and orthogonal least-square learning // IEEE Trans. On Neural Networks. – 1992 - Vol. 3, No. 5. - P. 807-814.

Presented for publication 2006 01 27

## P. Carbonell. Efficient Fuzzy Logic-Based Algorithm for Microarray Network Identification and Prediction in Bioinformatics // Electronics and Electrical Engineering. – Kaunas Technologija, 2006. – No. 3(67). – P. 37–40.

Fuzzy logic-based systems have been proposed as a tool for analyzing experimental data in microarrays in order to detect and predict gene regulation networks. The main drawback of this approach is the so-called curse of dimensionality associated with high dimension exhaustive searches. In this paper, it is shown that simple standard fuzzy systems can be transformed into additively separable systems by using an appropriate rotation of the input variables. By doing this, the algorithm for computing the output of the fuzzy system takes polynomial time and, therefore, it is computationally efficient for finding higher dimension interaction among genes. Ill. 3, bibl. 8 (in English; summaries in English, Russian and Lithuanian).

## П. Карбонелл. Применение алгоритма логики „Fuzzy“ в идентификации микромассивов сетей в биоинформатике // Электроника и электротехника. – Каунас: Технология, 2006. – № 3(67).- С. 37–40.

Описывается способ применения экспериментальных данных микромассивов для анализа и прогнозирования методов регулирования генов сетей. Система создана на основе алгоритмов логики „Fuzzy“. Главной проблемой данного метода является большой объем нахождения информации. Указано, что стандартные системы „Fuzzy“ могут быть трансформированы в отдельные аддитивные системы, используя ротации входных перемен. Выходной сигнал системы „Fuzzy“ рассчитывается алгоритмом, названным полиномиальным временем, и является эффективным относительно затрат расчета высшего порядка связи между генами. Ил. 3, библи. 8 (на английском языке; рефераты на английском, русском и литовском, яз.).

## P. Carbonell. Efektyvus „Fuzzy“ logikos algoritmas mikromasyvų tinklui identifikuoti ir prognozėms daryti bioinformatikoje // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2006. – Nr. 3(67). – P. 37–40.

„Fuzzy“ logikos pagrindu veikiančios sistemos buvo pasiūlytos kaip įrankis eksperimentinių duomenų mikromasyvams analizuoti siekiant detektuoti ir prognozuoti genų reguliavimo tinklus. Didžiausias šio metodo trūkumas yra vadinamoji erdviškumo problema, susijusi su didelės apimties išsamiais paieškomis. Straipsnyje parodyta, kad paprastos standartinės „fuzzy“ sistemos gali būti transformuojamos į adityviai atskiriamas sistemas, panaudojant atitinkamą įėjimo kintamųjų rotaciją. Taigi „fuzzy“ sistemos išėjimo signalo skaičiavimo algoritmas apibūdinamas „polinominiu laiku“ ir dėl šios priežasties jis yra efektyvus skaičiavimo sąnaudų požiūriu ieškant aukštesnės eilės sąveikos tarp genų. Ill. 3, bibl. 8 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).