Balanced Neurofuzzy Models

Oleg Yu. Mytnyk

Institute of Applied System Analysis, Kyiv, Ukraine

`oleg@mytnyk.kiev.ua, oleg.mytnyk@gmail.com`

Abstract. *This paper is devoted to the problem of a high complexity of fuzzy knowledge bases which contain enormous number of compound fuzzy rules. In order to significantly decrease the number of fuzzy rules and increase their transparency we present balanced neurofuzzy models. These models use the idea of Gabor-Kolmogorov expansion for additive decomposition into univariate and bivariate neurofuzzy submodels as well as maximum entropy principle to ground independent use of these submodels. Each submodel generates simplified rules independently of other submodels and contributes to fuzzy knowledge base of reduced complexity. The last but not least advantage of balanced neurofuzzy models is that they can be regularized and learned by modern inductive methods. Although the present paper omits learning. We demonstrate the potential of balanced neurofuzzy approach on a toy example of wind-induced wave model.*

Keywords

fuzzy knowledge base (FKB), neurofuzzy model.

1 Introduction

The data mining competitions are an effective instrument to analyze the performance of a certain method from the growing variety of approaches. The last contests like DMC and CoIL Challenge[1] have proved the advantage of Bayesian approach, Gaussian process regression, group method of data handling and support vector methods. But their productivity is in contrast with their awkward integration in expert and decision support systems. On the other hand neurofuzzy modeling is an appealing resource for knowledge representation. Neurofuzzy models allow transforming prior experts' knowledge in the form of raw fuzzy rules into analytical models, learn these models on empirical data, and transform them back to empirically consistent fuzzy rules. This work tries to combine the precedence of modern inductive algorithms with transparency and interpretability of neurofuzzy modeling.

Fuzzy inference works with several kind of fuzzy rules [1]. In this paper we consider fuzzy rules of Mamdani's type only as the most transparent, human and understandable rules. Fuzzy modeling consists of two phases designated as structural and parametric identification. There are several approaches to parametric identification. The first one is to tune parameters of membership functions. Another one is to introduce a confidence for each rule. We follow second approach.

2 General Problem Statement

Definition 1. Let us define the canonical neurofuzzy model in the form of the following linear combination

$$f(\mathbf{x}) = \sum_i w_i \mu_{A^i}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}, \quad w_i \in R, \tag{1}$$

where $\{\mu_{A^i}(\mathbf{x})\}$ — real nonnegative fuzzy membership functions of input vector $\mathbf{x}$, which satisfy unity of support condition $\sum_i \mu_{A^i}(\mathbf{x}) = 1$; $\{A^i\}$ — fuzzy labels defined on input space $\mathcal{X}$.

---

[1] `http://www.data-mining-cup.com, http://www.liacs.nl/~putten/library/cc2000/`

Note that the definition of canonical neurofuzzy model is given irrelatively to fuzzy inference system (FIS). This makes possible neurofuzzy models to be used in different machine learning algorithms and be interpreted afterwards with help of FIS. The fundamental result with respect to neurofuzzy models was obtained in [2]. It has been shown how analytical model can be used in fuzzy inference chain. We state this result in the following redaction:

Theorem 1. Suppose that crisp input $\mathbf{x} \in \mathcal{X}$ and output $y \in \mathcal{Y} \subset R$ of fuzzy system are represented via real nonnegative fuzzy membership functions $\{\mu_{A^i}(\mathbf{x})\}$ and $\{\mu_{B^j}(y)\}$ correspondingly, where $\{A^i\}$, $\{B^j\}$ — fuzzy labels defined on input spaces $\mathcal{X}$ and $\mathcal{Y}$. In addition unity of support condition $\sum_i \mu_{A^i}(\mathbf{x}) = 1$ and $\forall j : \int \mu_{B^j}(y)dy = const$ condition hold true. If algebraic operators are used to implement the fuzzy logical functions, then defuzzified (using center of gravity method) result of Mamdani's fuzzy inference can be represented by neurofuzzy model:

$$y(\mathbf{x}) = \sum_i w_i \mu_{A^i}(\mathbf{x}). \tag{2}$$

Here the weight $w_i = \sum_j c_{ij} y_j^c$, where $y_j^c = \int \mu_{B^j}(y) y dy / \int \mu_{B^j}(y)dy$ — center of output fuzzy membership function μ_{B^j}. Coefficient c_{ij} is the confidence in the rule of Mamdani type: if $\mathbf{x} \in A^i$ then $y \in B^j$ ($\sum_j c_{ij} = 1$, $c_{ij} \in [0;1]$).

In order to generate fuzzy rules based on neurofuzzy model it is necessary to find rule confidences $\{c_{ij}\}$ for the given weights $\{w_i\}$. Confidences are a solution to the constrained linear systems for each i:

$$\left\{ \begin{array}{lcl} w_i & = & \sum_j c_{ij} y_j^c, \\ 1 & = & \sum_j c_{ij}, \\ 0 & \leq & c_{ij}. \end{array} \right. \tag{3}$$

Number of nonzero adjacent coefficients c_{ij} for given i depends on level of overlapping of output fuzzy membership functions. If only adjacent output fuzzy membership functions overlap then solution is unique.

In practice fuzzy membership functions are not defined for the whole input vector $\mathbf{x} = \text{col}\,(x^1, \ldots, x^n)$, but for each input variable x^k separately, because variables usually have different physical meaning. Let the definition domain $\mathcal{X}_k$ of variable x^k gets covered by finite number of fuzzy labels $A_k^i,\ i = 1, \ldots, m_k$ with fuzzy membership functions $\mu_{A_k^i}(x^k)$; let the definition domain $\mathcal{Y}$ of output y gets covered by finite number of fuzzy labels $B^i,\ i = 1, \ldots, m_0$ with fuzzy membership functions $\mu_{B^i}(y)$. Thus the full set of fuzzy rules contains $m_0 m_1 \cdots m_n$ rules. Each rule is defined as:

$$r\text{-rule}: \text{ if } x^1 \in A_1^{i_1} \text{ and if } x^2 \in A_2^{i_2} \text{ and } \ldots \text{ and if } x^n \in A_n^{i_n} \text{ then } y \in B^{i_0} \text{ with confidence } c_{ri_0}.$$

Index r corresponds to the sequence $\{i_0, i_1, \ldots, i_n\}$, $i_k = 1, \ldots, m_k$. This is co-called *the curse of dimensionality* expressed in exponential complexity $\mathcal{O}(m^{n+1})$ of fuzzy rules set. It prevents spreading FKBs in expert systems. Suppose that we have merely 4 inputs, 1 output and each domain gets covered by 3 fuzzy labels. Total number of fuzzy rules makes up to $3^{4+1} = 243$ rules. Obviously it is impossible to comprehend all relations between variables. It is very hard even to read rules because each rule is a compound statement. The curse of dimensionality abolishes the basic idea of fuzzy approach to make models transparent to human. As a result a number of investigations are devoted to the question of complexity reduction of fuzzy modeling. Among them numerous optimizations of fuzzy rules set [3, 4] but they do not remove exponential dependency in fact.

3 Solution

In order to significantly decrease the number of fuzzy rules and increase their transparency it is reasonably to make a decomposition of FKB. Here we address [5], where the idea of Gabor-Kolmogrov expansion is used for additive decomposition of multidimensional model into univariate and bivariate submodels. Fuzzy rules are generated for each submodel separately. Thereby exponential complexity of FKB can be reduced to quadratic. But unfortunately there is an unsolved problem, which prevents this approach from being used widely. We may not make a fuzzy inference based on a separate submodel unless independent use of this submodel (from other submodels) is proved.

Here we introduce a novel class of co-called balanced neurofuzzy models (BNFM) for which independent use of any submodel is well-grounded by maximum entropy principle.

Definition 2. Neurofuzzy model

$$f(\mathbf{x}) = b + \sum_{k=1}^{n} g_k(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} g_{pq}(x^p, x^q), \quad g_k(x^k) = \sum_i w_i^k \mu_{A_k^i}(x^k), \quad g_{pq}(x^p, x^q) = \sum_j w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q),$$

is called balanced if functions g_k, g_{pq} have zero expectation under the condition of multidimensional uniform distribution of random variate $\mathbf{x}$ on its definition domain. Neurofuzzy submodels are defined as $f_k = b + g_k$ and $f_{pq} = b + g_{pq}$. Value b is bias of BNFM and has a meaning of an average output value.

According to maximum entropy principle [6, 7] if we do have a lack of knowledge about variable x^k it is naturally to suppose that x^k can accept any value from its finite definition domain with the same probability. As a result this variable does not contribute to the model output since uncertainty disclosure using mathematical expectation gives zero. In turn, if we make a fuzzy inference based on a separate submodel then we imply that the variables which are absent in this submodel are uniformly distributed over their definition domains.

The definition 2 is correct and does not contradict with definition 1 since neurofuzzy submodels can be easily reduced to the canonical form (1):

$$f_k(x^k) = b + \sum_i w_i^k \mu_{A_k^i}(x^k) = \sum_i (w_i^k + b)\mu_{A_k^i}(x^k),$$

$$f_{pq}(x^p, x^q) = b + \sum_j w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q) = \sum_j (w_j^{pq} + b)\mu_{A_{pq}^j}(x^p, x^q),$$

by virtue of unity of support properties: $\sum_i \mu_{A_k^i}(x^k) = 1$, $\sum_j \mu_{A_{pq}^j}(x^p, x^q) = 1$.

Below we derive the sufficient condition of neurofuzzy model to be balanced.

Theorem 2. Let random variable $\mathbf{x}$ be uniformly distributed over its definition domain. If

$$\forall i : E\left[\mu_{A_k^i}(x^k)\right] = const_k, \quad \forall j : E\left[\mu_{A_{pq}^j}(x^p, x^q)\right] = const_{pq}, \tag{4}$$

$$\sum_i w_i^k = 0, \quad \sum_j w_j^{pq} = 0, \tag{5}$$

then corresponding neurofuzzy model is balanced (E — expectation operator).

Theorem 3. If only first condition (4) holds true, but second (5) does not hold:

$$\sum_i w_i^k = \Lambda_k \neq 0, \quad \sum_j w_j^{pq} = \Lambda_{pq} \neq 0,$$

then neurofuzzy model can be reduced to the balanced one.

This result allows all neurofuzzy models which are inherited from Gabor-Kolmogorov expansion to be balanced. And generate small FKB with simple rules.

4 Toy Example

We show the potential of presented approach on the following toy example. It is required to describe a dependency between the height of wind-induced wave and speed and direction of wind in shoaling waters of the given bay. The observational data is also given. The variable of direction (North, North-East, East, ...) is divided into two variables: sine (x^1) and cosine (x^2) of a clockwise angle from North to given direction. Model also includes wind speed (x^3).

Suppose we received the following BNFM using the inductive algorithm described in [8]:

$$f(\mathbf{x}) = 0.47 + 0.09\mu_{A_1^1}(x^1) - 0.09\mu_{A_1^2}(x^1) - 0.04\mu_{A_2^1}(x^2) + 0.04\mu_{A_2^2}(x^2) - 0.35\mu_{A_3^1}(x^3) + 0.35\mu_{A_3^2}(x^3), \tag{6}$$

where $\mu_{A_i^1}(x^i)$ and $\mu_{A_i^2}(x^i)$ are any valid fuzzy membership functions defined for input variables. Corresponding fuzzy labels: A_1^1 — West wind, A_1^2 — East wind, A_2^1 — South wind, A_2^2 — North wind, A_3^1 — weak wind, A_3^2 — strong wind. The model (6) is decomposed into three neurofuzzy submodels in the canonical form:

$$f_1(x^1) = 0.56\mu_{A_1^1}(x^1) + 0.38\mu_{A_1^2}(x^1), \; f_2(x^2) = 0.43\mu_{A_2^1}(x^2) + 0.51\mu_{A_2^2}(x^2), \; f_3(x^3) = 0.12\mu_{A_3^1}(x^3) + 0.82\mu_{A_3^2}(x^3).$$

The rules and their confidences are easy to find if output fuzzy membership functions are given as B-splines. In this case at most two adjacent coefficients are nonzero. Lets define fuzzy membership functions for wind-induced wave output variable normalized on $[0; 1]$ using five first order B-splines defined in nodes $\{-0.25;\ 0;\ 0.25;\ 0.5;\ 0.75;\ 1;\ 1.25\}$ as shown on figure 1. They correspond to fuzzy sets: B^1 — calm, B^2 — small wave, B^3 — average wave, B^4 — big wave, B^5 — storm with central points $\{0;\ 0.25;\ 0.5;\ 0.75;\ 1\}$. The rule confidences induced by neurofuzzy submodel f_1 are

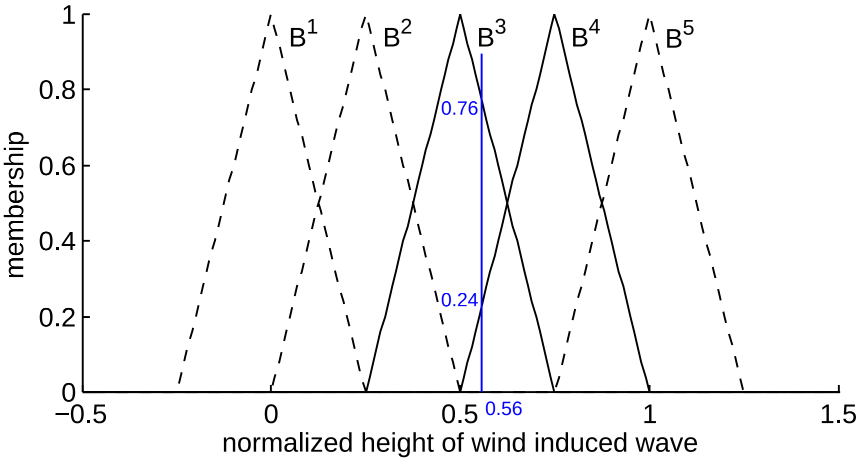

Fig. 1. Five first order B-splines defined in nodes $\{-0.25;\ 0;\ 0.25;\ 0.5;\ 0.75;\ 1;\ 1.25\}$ are shown. The blue vertical line shows graphical solution to the first system.

found by solving systems (3):

$$\left\{\begin{array}{lcl} 0.56 & = & 0\cdot c^1_{11} + 0.25\cdot c^1_{12} + 0.5\cdot c^1_{13} + 0.75\cdot c^1_{14} + 1\cdot c^1_{15} \\ 1 & = & c^1_{11} + c^1_{12} + c^1_{13} + c^1_{14} + c^1_{15} \\ 0 & \leq & c^1_{11},\ c^1_{12},\ c^1_{13},\ c^1_{14},\ c^1_{15} \\ \multicolumn{3}{l}{\text{at most two adjucent coeficients } c^1_{1j},\ c^1_{1j+1} \text{ are nonzero}} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} c^1_{11} = 0 \\ c^1_{12} = 0 \\ c^1_{13} = 0.76 \\ c^1_{14} = 0.24 \\ c^1_{15} = 0 \end{array}\right.$$

$$\left\{\begin{array}{lcl} 0.38 & = & 0\cdot c^1_{21} + 0.25\cdot c^1_{22} + 0.5\cdot c^1_{23} + 0.75\cdot c^1_{24} + 1\cdot c^1_{25} \\ 1 & = & c^1_{21} + c^1_{22} + c^1_{23} + c^1_{24} + c^1_{25} \\ 0 & \leq & c^1_{21},\ c^1_{22},\ c^1_{23},\ c^1_{24},\ c^1_{25} \\ \multicolumn{3}{l}{\text{at most two adjucent coeficients } c^1_{2j},\ c^1_{2j+1} \text{ are nonzero}} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} c^1_{21} = 0 \\ c^1_{22} = 0.48 \\ c^1_{23} = 0.52 \\ c^1_{24} = 0 \\ c^1_{25} = 0 \end{array}\right.$$

The confidences induced by submodels f_2 and f_3 are found in the same way:

$$f_2: \quad c^2_{1j} = \{0, 0.28, 0.72, 0, 0\} \quad c^2_{2j} = \{0, 0, 0.96, 0.04, 0\}$$

$$f_3: \quad c^3_{1j} = \{0.52, 0.48, 0, 0, 0\} \quad c^3_{2j} = \{0, 0, 0, 0.72, 0.28\}$$

Each submodel generates two fuzzy rules leading to six fuzzy rules in set:

rule 1: if wind direction is West then waves are average (0.76) or big (0.24);
rule 2: if wind direction is East then waves are small (0.48) or average (0.52);
rule 3: if wind direction is South then waves are small (0.28) or average (0.72);
rule 4: if wind direction is North then waves are average (0.96) or big (0.04);
rule 5: if wind is weak then calm (0.52) or waves are small (0.48);
rule 6: if wind is strong then waves are big (0.72) or storm (0.28);

where values in brackets describes rule confidences. Such transparent FKB allows experts in application domain (e.g. hydrometeorology) to understand and discuss this model. It is much simpler than for example similar ANFIS TS model, which has $2^3 = 8$ compound rules with linear function in the right side. First of all it is obviously that wind speed plays a key role even in shoaling waters. But wind direction also has some influence. For example West and North wind induces higher waves than East and South wind. This fact makes neurofuzzy model to gain experts' confidence because experts know that they consider Northwest coast and sea-breeze must produce higher waves than off-shore wind.

5 Discussion and Conclusion

In this paper we did not touch important question of finding an adequate structure and parameters at membership functions (w_i) of BNFM. We just outlined the interpretability performance of the new class of models. Presented approach reduces exponential complexity of fuzzy knowledge bases to quadratic as well as simplifies compound fuzzy rules. It allows experts in application domain to comprehend relations between variables and to participate in model fusion together with modelers. One more advantage of this modeling technique is that the fuzzy inference is possible under incompleteness of information about input variables. Even under complete uncertainty it is possible to estimate model output as its bias. There are also some open issues to discuss. For example how to treat a confidence rate for fuzzy rule. How to choose membership function for output variable.

References

[1] Castellano G.: A neurofuzzy methodology for predictive modeling. Ph.D. thesis. Department of Computer Science, University of Bari, Italy, 2001.

[2] Brown M., Harris C. J.: Neurofuzzy adaptive modelling and control. Hemel Hempstead, Prentice Hall, 1994.

[3] Hong X., Harris C. J., Chen S.: Robust neurofuzzy rule base knowledge extraction and estimation using subspace decomposition combined with regularization and D-optimality. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, SMC-34(1): 598-608, 2004.

[4] Roubos H., Setnes M.: Compact fuzzy models through complexity reduction and evolutionary optimization. *In Proc. of the Ninth IEEE International Conference on Fuzzy Systems*, vol. 2: 762-767, San Antonio, Texas, May 2000.

[5] Hong X., Harris C. J.: Generalized neurofuzzy network modeling algorithms using Bézier-Bernstein polynomial functions and additive decomposition. *IEEE Transactions On Neural Networks*, 11(4): 889–902, 2000.

[6] Jaynes E. T.: Prior probabilities. *IEEE Transactions On Systems Science and Cybernetics*, SSC-4(3): 227-241, 1968.

[7] Jaynes E. T.: Probability Theory: The Logic of Science. Cambridge University Press, 2003.

[8] Mytnik O. Yu.: Construction of Bayesian support vector regression in the feature space spanned by Bézier-Bernstein polynomial functions. *Cybernetics and Systems Analysis*, 43(4): 613-620, 2007.

Appendix

Proof of theorem 2.

$$E\left[g_k(x^k)\right] = \sum_i w_i^k E\left[\mu_{A_k^i}(x^k)\right] = 0 \cdot const_k = 0,$$

$$E\left[g_{pq}(x^p, x^q)\right] = \sum_j w_j^{pq} E\left[\mu_{A_{pq}^j}(x^p, x^q)\right] = 0 \cdot const_{pq} = 0.$$

Proof of theorem 3. For each submodel we do the following transformations:

$$g_k(x^k) = \sum_{i=1}^{L_k} w_i^k \mu_{A_k^i}(x^k) = \sum_{i=1}^{L_k} \left(w_i^k - \frac{\Lambda_k}{L_k} \right) \mu_{A_k^i}(x^k) + \frac{\Lambda_k}{L_k},$$

$$g_{pq}(x^p, x^q) = \sum_{j=1}^{L_{pq}} w_j^{pq} \mu_{A_{pq}^j}(x^p, x^q) = \sum_{j=1}^{L_{pq}} \left(w_j^{pq} - \frac{\Lambda_{pq}}{L_{pq}} \right) \mu_{A_{pq}^j}(x^p, x^q) + \frac{\Lambda_{pq}}{L_{pq}}.$$

Here we redefine bias and coefficients at membership functions in the following way:

$$\tilde{w}_i^k \longleftarrow w_i^k - \frac{\Lambda_k}{L_k}, \quad \tilde{w}_j^{pq} \longleftarrow w_j^{pq} - \frac{\Lambda_{pq}}{L_{pq}}, \quad \tilde{b} \longleftarrow b + \sum_{k=1}^{n} \frac{\Lambda_k}{L_k} + \sum_{p>q}^{n} \frac{\Lambda_{pq}}{L_{pq}}.$$

It is easy to check that newly created neurofuzzy model in the form

$$\tilde{f}(\mathbf{x}) = \tilde{b} + \sum_{k=1}^{n} \tilde{g}_k(x^k) + \sum_{p=1}^{n-1} \sum_{q=p+1}^{n} \tilde{g}_{pq}\left(x^p, x^q\right), \quad \tilde{g}_k(x^k) = \sum_i \tilde{w}_i^k \mu_{A_k^i}(x^k), \quad \tilde{g}_{pq}(x^p, x^q) = \sum_j \tilde{w}_j^{pq} \mu_{A_{pq}^j}(x^p, x^q),$$

satisfies both conditions (4) and (5). Thus it is balanced. Note that output of the new model is the same as output of initial model since $\tilde{f}(\mathbf{x}) = f(\mathbf{x})$.