Monografías del Semin. Matem. García de Galdeano. **27:** 177–184, (2003).

Functional Network Models in Statistics

Enrique Castillo[1], Ali S. Hadi[2], Beatriz Lacruz[3] and Rosa E. Pruneda[4]

[1] Department of Applied Mathematics, University of Cantabria

[2] Department of Statistical Sciences, Cornell University, Ithaca (NY), USA and

Department of Mathematics, the American University in Cairo, Egypt

[3] Department of Statistical Methods, University of Zaragoza, Spain

[4] Department of Applied Mathematics, University of Castilla-La Mancha, Ciudad Real

Abstract

Functional networks are a general framework useful for solving a wide range of problems. In this paper we introduce the elements of functional networks and the steps involved in the modeling process, illustrated with several examples. Our main purpose is to show functional networks power as an unified approach for statistical applications.

Keywords: Functional networks, Conjugate families, Stability with respect to maxima operations, Optimal transformation, Regression

AMS Classification: 62J

1 Introduction

Functional networks are a very useful general framework for solving a wide range of problems. In probability and statistics they have been used for: characterization of univariate and bivariate distributions, finding conjugate families of distributions, obtaining reproductive families and stable families with respect to maxima operations, modeling fatigue problems, computing convenient posterior probability distributions, conditional specification of statistical models, time series and regression modeling, etc. Some of these applications have been developed in [2], [3], [4], [5] and [6].

The main purpose of this paper is to introduce functional networks and to show their power as an unified approach for statistical applications. In Section 2 the elements of a functional network are introduced. In Section 3 we describe the steps involved in model building. The process is illustrated and applied to solve the problem of stability with respect to maxima operations using exact learning. In Section 4, the technique is used to discover optimal transformations in the response and/or the explanatory variables in a linear regression model and it is applied to a real example in Section 5. Some concluding remarks are given in Section 6.

2 Elements of functional networks

A functional network consists of the following elements:

1. Several layers of storing units: These units are represented by small filled circles. They are used for storing input, output, and intermediate information.

2. One or more layers of functional units: These units are represented by open circles with the name of the unit inside the circle. They evaluate a set of input values and return a set of output values to the next layer of storing units. Thus, each of these units represents a function.

3. A set of directed links: The computing units are connected to the storing units by directed arrows. The arrows indicate the direction of information flow. Converging arrows to an intermediate or output unit indicate that the functions from which they emanate must produce identical outputs and represent constraints which arise from the characteristics of the problem at hand.

The elements are illustrated by the following example:

Example 1: Conjugate Families. Let X be a random variable which belongs to a parametric family of distributions with likelihood function $L(\mathbf{x}, \theta)$, where $\mathbf{x}$ stands for a sample value and $\theta \in \Theta$ is a possibly vector-valued parameter. Let $F(\theta; \eta)$ be the prior probability density function. A classical problem in Bayesian statistics is to find a parametric family of probability density functions such that both the prior and the posterior probability density function, $F(\theta; G(\mathbf{x}; \eta))$, belong to the family. Using Bayes' theorem, this problem can be represented by the functional equation

$$F(\theta; G(\mathbf{x}; \eta)) = H(\mathbf{x}; \theta) F(\theta; \eta), \tag{1}$$

where $H(\mathbf{x}; \theta) = h(\mathbf{x}) L(\mathbf{x}; \theta)$ and G specifies the value of the new parameter. This functional equation can be represented by the functional network depicted in Figure 1.

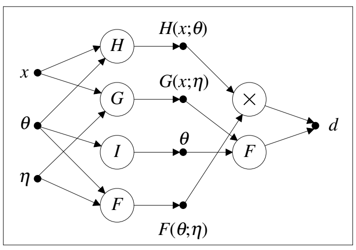

Figure 1: Functional network associated with conjugate families problem.

1. This functional network has three layers of storing units: the first (left) layer contains three input units ($\mathbf{x}, \theta$, and η), the second layer consists of four intermediate units, ($H(\mathbf{x};\theta)$, $G(\mathbf{x};\eta)$, θ and $F(\theta;\eta)$) and the third (right) consists of one output unit (d).

2. The first layer of functional units consists of four functions H, G, I (the identity function) and F. The second layer has two functions: the product operator "$\times$" and F.

3. The product operator, "$\times$", and F must give identical output, which is represented by the output unit d.

3 Model building

The steps required in functional networks modeling will be introduced using the following example:

Example 2: Stability with respect to Maxima Operations. Let X and Y be two independent random variables which cumulative probability distribution functions (CDF) belong to the parametric family $\{F(z;\theta), \theta \in \Theta\}$. Then the CDF of the random variable $Z = \max(X, Y)$ is $T(z; a, b) = F(z; a)F(z; b)$. If we wish X, Y and Z to be stable with respect to maxima operations (i.e., to belong to the same parametric family), we have the following functional equation

$$F(z; G(a; b)) = F(z; a)F(z; b), \tag{2}$$

represented by the functional network in Figure 2(a).

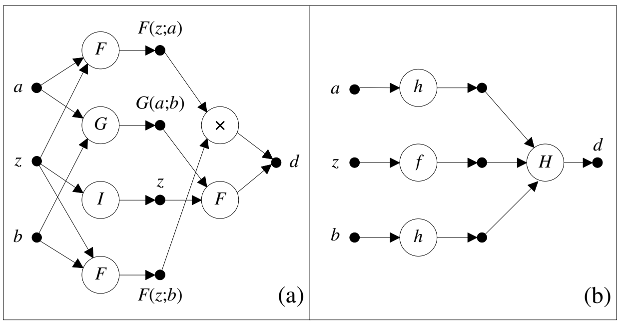

Figure 2: (a) Functional network associated with the stability with respect to maxima operations, and (b) the corresponding simplified functional network.

Modeling by functional networks consists of:

1. Specifying the initial topology: The elements and structure of a functional network is based on the characteristics of the problem at hand. In Example 2, the functional network is obtained from the functional equation which represents the stability with respect to maxima operations.

2. Simplifying functional networks: Sometimes, solving the functional equation which describes the problem at hand allows us to simplify the initial topology. The result is an equivalent functional network with a simpler structure. Methods for solving functional networks can be found in [1] and [7].

Example 2 (continued): A solution of the functional equation (2) is

$$F(x;y) = f(x)^{h(y)}, \quad G(x;y) = h^{-1}[h(x) + h(y)], \tag{3}$$

where f is a CDF and h is a positive and invertible function, both arbitrary. Replacing (3) in (2), we get $f(z)^{h(a)+h(b)}$. Figure 2(b) represents the simplified functional network where $H(h(a), f(z), h(b)) = f(z)^{h(a)+h(b)}$.

3. Checking uniqueness. For a given functional network, several functional units can lead to the same output for any input. In order to solve estimation problems we need to know what conditions must hold for uniqueness.

Example 2 (continued): Assuming that we have two sets of functions $\{f_1, h_1\}$ and $\{f_2, h_2\}$ such that $f_1(z)^{h_1(a)+h_1(b)} = f_2(z)^{h_2(a)+h_2(b)}$, then

$$\frac{h_2(a) + h_2(b)}{h_1(a) + h_1(b)} = \frac{\log f_1(z)}{\log f_2(z)} = k, \tag{4}$$

where k is an arbitrary constant. To obtain uniqueness, we just need to fix the value of some of the functions f_1, f_2, h_1, or h_2 at a point.

4. Learning functional units: To complete the model building process, it is necessary to learn the functional units. There are two types of methods: exact and approximate. Exact learning is illustrated with an example based on the problem of stability with respect to maxima operations. Approximate learning is discussed in Sections 4 and 5.

Example 2 (continued): The solution of the functional equation (2) given by the equation (3) can be applied to the problem of fatigue of longitudinal elements problem. In this case, $F(x, y)$ is the survival function of an element with length y and $G(x, y) = x + y$, then $h(x) = cx$. Therefore, $F(x, y) = f(x)^{cy}$, where c is a positive constant and f is a survival function, both arbitrary.

4 Optimal transformations in a regression model

In this section we illustrate the approximate learning method using an example based on the problem of finding optimal transformations of the response and/or the explanatory variables in a linear regression model.

4.1 Selecting and simplifying the initial topology

We consider the additive model

$$f(y) = h_1(x_1) + h_2(x_2) + \ldots + h_k(x_k), \tag{5}$$

where y is the response variable, $x_1, x_2, \ldots, x_k$ are the explanatory variables and f, $h_1, h_2, \ldots, h_k$ are parametric functions known or unknown, which define the required transformations of the variables in order to obtain a linear model in the parameters. Furthermore, we assume that f is an invertible function.

The functional network associated with the equation is shown in Figure 3. Since the estructure has no converging arrows in the output node, this functional network can not be simplified.

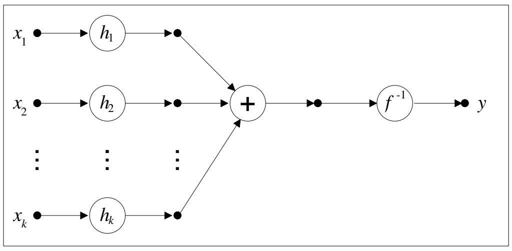

Figure 3: Functional network associated with the additive model.

4.2 Uniqueness problem

Given two sets of functions $\{h_1, h_2, \ldots, h_k, f\}$ and $\{h_1^\star, h_2^\star, \ldots, h_k^\star, f^\star\}$ such that,

$$f^{-1}[h_1(x_1) + h_2(x_2) + \ldots + h_k(x_k)] = f^{\star -1}[h_1^\star(x_1) + h_2^\star(x_2) + \ldots + h_k^\star(x_k)]; \ \ \forall x_1, \ldots, x_k,$$

we obtain a functional equation whose general solution is:

$$\begin{array}{rcl} h_1^\star(x) & = & a h_1(x) + b_1, \\ h_2^\star(x) & = & a h_2(x) + b_2, \\ \ldots & \ldots & \ldots \\ h_k^\star(x) & = & a h_k(x) + b_k, \\ f^\star(x) & = & a f(x) + b_1 + b_2 + \ldots + b_k, \end{array} \tag{6}$$

where a and $b_1, b_2, \ldots, b_k$ are arbitrary constants (see [2], page 98, for details). Then to obtain uniqueness we need to fix the functions f and $h_1, h_2, \ldots, h_k$ at a point.

4.3 Approximate learning

The approximate learning consists of estimating the functional units using a linear combination of linearly independent functions and an estimation method to obtain the coefficients. Some possible families are:

1. Polinomial family: $\Phi = \{1, x, x^2, \ldots, x^q\}$,
2. Exponential family: $\Phi = \{1, e^x, e^{-x}, e^{2x}, e^{-2x} \ldots, e^{qx}, e^{-qx}\}$, and
3. Fourier family: $\Phi = \{1, \sin x, \cos x, \sin(2x), \cos(2x), \ldots, \sin(qx), \cos(qx)\}$.

Given a set of observations $D = \{y_i, x_{1i}, \ldots, x_{ki}; i = 1, 2, \ldots, n\}$, the approximation of the additive model can be written as:

$$\sum_{j=1}^{q_0} \alpha_j \phi_{0j}(y_i) = \sum_{j=1}^{q_1} \beta_{1j} \phi_{1j}(x_{1i}) + \cdots + \sum_{j=1}^{q_k} \beta_{kj} \phi_{kj}(x_{ki}) + \varepsilon_i, \; i = 1, \ldots, n, \tag{7}$$

where ϕ_{sj}, $j = 1, \ldots, q_s$, $s = 0, 1, \ldots, k$, are the elements of the chosen family, and ε_i is the error.

In matrix form we can write:

$$\mathbf{Y}\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8}$$

where, using the polynomial family, $\mathbf{Y}$ and $\mathbf{X}$ contain q_0 and $1 + \sum_{s=1}^{k} q_s$ columns, respectively. For instance, if $k = 2$ and $q_0 = q_1 = q_2 = 2$, the rows of $\mathbf{Y}$ and $\mathbf{X}$ are $\{y, y^2\}$ and $\{1, x_1, x_1^2, x_2, x_2^2\}$, respectively. Notice that the constant is included only in one of the matrices as a condition to obtain uniqueness.

To obtain the parameters, we use the least squares method subject to one more constraint necessary for uniqueness, that is, we minimize $(\mathbf{Y}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y}\boldsymbol{\alpha} - \mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{w}_0\boldsymbol{\beta} = c$.

To select the best model we apply a search procedure based on a goodness-of-fit measure. We choose the adjusted correlation coefficient

$$R_a^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2/(n-p)}{\sum_{i=1}^{n} \left(\hat{f}(y_i) - \overline{\hat{f}(y_i)}\right)^2 /(n-1)}, \tag{9}$$

where p is the number of parameters and $\overline{\hat{f}(y)} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(y_i)$. This measure allows us to compare models when different families of polynomials are used.

5 An example

To illustrate the use of the approximate learning method we present an example based on a set of real data analyzed in [8] and obtained from [9]. The response variable is the average January minimum temperature (in Fahrenheit degrees) and the predictors are the latitude and longitude of 56 U.S. cities.

Using the additive model $f(y) = g_1(x_1) + g_2(x_2)$, [8] reports a study in which no transformation is made to temperature and latitude (f and g_1 are the identity function) and a cubic polynomial (g_2) is used for longitude. This model obtains $R_a^2 = 0.8971$ and

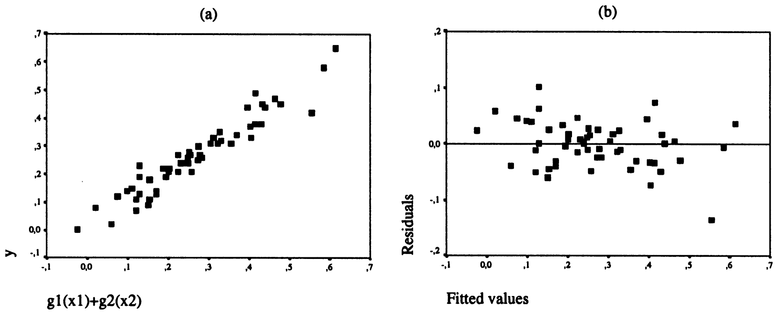

Figure 4: Scatter plot of (a) y versus $\hat{g}_1(x_1)+\hat{g}_2(x_2)$, and (b) residuals versus fitted values.

leads to a good linear approximation, as we can see in Figure 4 (a). However, Figure 4 (b) shows some specification problems.

Using the approximate learning method proposed in this paper, with polynomial family $\Phi = \{1, t, t^2, t^3\}$, and the exhaustive search method, the obtained model is

$$y^2 = 113.10 - 13.74\,x_1 + 25.64\,x_1^2 - 16.08\,x_1^3 + 15.05\,x_2 - 16.88\,x_2^2 + 6.19\,x_2^3, \qquad (10)$$

with $R_a^2 = 0.9771$. The obtained model is more complex than that suggested in [8], but Figures 5 (a) and (b) show a very good linear approximation and no pattern in the residuals plot.

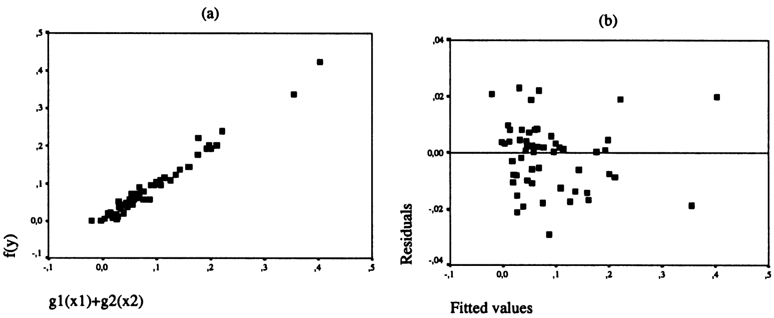

Figure 5: Scatter plot of (a) $\hat{f}(y)$ versus $\hat{g}_1(x_1) + \hat{g}_2(x_2)$, and (b) residuals versus fitted values.

6 Conclusions

In this work we have shown an unified approach based on functional networks to solve statistical problems. The characteristics of the problem at hand allow us to select the structure of the functional network. It can sometimes be simplified solving the associate functional equation. Once we have solved the uniqueness problem, functional units should be learned using exact or approximate learning.

A specially interesting problem is to select regression models when the response or the explanatory variables should be transformed. This problem can be solved via approximate learning in functional networks. In this work we have chosen the additive model to represent the regression problem and the least squares method for parameter estimation. Other functional equations that, for example, allow for including interactions, and other estimation methods have been used to solve the problem of transformations in linear regression. They can be found in [5] and [6].

Acknowledgements

We thank the University of Zaragoza (Project 228-34), DGICYT (Project PB98-0421) and Iberdrola for partial support of this work.

References

[1] Aczél, J. *Lectures on Functional Equations and Their Applications*, **Vol.** 19, Mathematics in Science and Engineering, Academic Press, New York, (1966).

[2] Castillo, E., Cobo, A., Gutiérrez, J. M., and Pruneda, R. E. *An Introduction to Functional Networks with Applications*, Kluwer Academic Publishers: New York, (1998).

[3] Castillo, E. and Gutiérrez, J. M. Nonlinear Time Series Modeling and Prediction Using Functional Networks. Extracting Information Masked by Chaos, *Phisics Letters A*, **Vol.** 244, 71-84, (1998).

[4] Castillo, E., Gutiérrez, J. M., Hadi, A. S. and Lacruz, B. Some Applications of Functional Networks in Statistics and Engineering, *Technometrics*, **Vol.** 43, Issue 1, 10-24, (2001).

[5] Castillo, E., Hadi, A. S. and Lacruz, B. Optimal Transformations in Multiple Linear Regression Using Functional Networks, *Proceedings of the International WorkConference on Artificial and Natural Neural Networks. IWANN 2001. Lecture Notes in Computer Science*, 2084, Part I, 316-324, (2001).

[6] Castillo, E., Hadi, A. S., Lacruz, B. and Pruneda, R. E. Semi-Parametric Nonlinear Regression and Transformation Using Functional Networks, Sent for publication, (2001).

[7] Castillo, E. and Ruiz-Cobo, R. *Functional Equations in Science and Engineering*, Marcel Dekker: New York, (1992).

[8] Peixoto, J. L. A Property of Well-Formulated Polynomial Regression Models, *The American Statistician*, 44, 26-30, (1990).

[9] *http://lib.stat.cmu.edu/DASL/Stories/USTemperatures.html*