Proceedings of the 7th International Conference on Applied Informatics
Eger, Hungary, January 28–31, 2007. Vol. 2. pp. 93–100.

Banded approximation with diffusion-neural-network*

Zoltán Makó, Júlia Salamon

Department of Mathematics and Informatics
Sapientia University, Romania
e-mail: makozoltan@sapientia.siculorum.ro
salamonjulia@sapientia.siculorum.ro

Abstract

In this paper we show the principle of information diffusion by using the properties of quasi-triangular fuzzy numbers and we apply this principle to construct a diffusion-neural-network for the banded approximation of the exchange rates.

Keywords: artificial neural network, information diffusion, fuzzy number, prognosis of the exchange rates

MSC: 62M20, 03E72

1. Introduction

Generally, the data are facts characterizing the phenomenon of the real world and information is such a structured sample of these data that helps the exploration of the phenomenon. In many cases the data are only a part of the facts, so the information deduced from them is uncertain. For example: if there are only few observations to the examination of a phenomenon then the information concluded will be uncertain.

If there are only few data available in the examination of a phenomenon we can assign these to some already existing statistical distribution (the Bayes method). The structured sample will have an informational value. The question arises: what to do in the case when we do not know a priory statistical distribution? Many successful solutions of the practical problems show that in such a case the theory of fuzzy sets can be applied with a very good efficiency (L. A. Zadeh, 1975).

The present paper deals with the diffusion principle known from the theory of the fuzzy sets and the application of the principle. In the first part we explain the

*This work was supported by the Research Programs Institute of Foundation Sapientia, Grant No. 1046/2006.

basic concepts of quasi-triangular fuzzy numbers with the help of the triangular norm. In the second part we show the principle of information diffusion with the help of the quasi-triangular fuzzy numbers. In the third part we apply the principle of information diffusion to construct an artificial neural networks for the banded approximation of the exchange rates.

2. Quasi-triangular fuzzy numbers

The fuzzy set concept was introduced in mathematics by K. Menger in 1942, and reintroduced in the system theory by L. A. Zadeh in 1965.

Definition 2.1. Let X be a set. A mapping $\mu\colon X \to [0,1]$ is called *membership function*, and the set $\bar{A} = \{(x, \mu(x)) : x \in X\}$ is called fuzzy set on X. The membership function of $\bar{A}$ is denoted by $\mu_{\bar{A}}$. The collection of all fuzzy subsets of X we will denote by $\mathcal{F}(X)$.

The construction of membership function of fuzzy numbers is an important problem in vagueness modeling. Theoretically, the shape of fuzzy numbers must depend on the applied triangular space. The membership function must be defined in such a way that the change of the triangular norm modifies the shape of fuzzy number, but the calculus with them remain valid.

We noticed that, if the model constructed on the computer does not comply the requests, then we choose another norm. If in the model we use the quasi-triangular fuzzy numbers introduced by M. Kovács in 1992, then the model will not have to be reconstructed, only the new norm gets a new model.

Let $p \in [1, +\infty]$ and $g\colon [0,1] \to [0, \infty]$ be a continuous, strictly decreasing function with the boundary properties $g(1) = 0$ and $\lim\limits_{t \to 0} g(t) = g_0 \leqslant \infty$. The quasi-triangular fuzzy number we define in the fuzzy triangular space $(\mathcal{F}(\mathbb{R}), T_{gp}, N)$, where

$$T_{gp}(x,y) = g^{[-1]}\left(\left(g^p(x) + g^p(y)\right)^{\frac{1}{p}}\right)$$

is an Archimedean triangular norm generated by g and

$$N(x) = \begin{cases} 1 - x & \text{if} \quad g_0 = +\infty, \\ g^{-1}(g_0 - g(x)) & \text{if} \quad g_0 \in \mathbb{R}. \end{cases}$$

is a negation operation, where

$$g^{[-1]}(t) = \begin{cases} g^{-1}(t) & \text{if} \quad 0 \leqslant t < g_0, \\ 0 & \text{if} \quad t \geqslant g_0. \end{cases}$$

Definition 2.2. The *set of quasi-triangular fuzzy numbers* is

$$\mathcal{N}_g = \Big\{\bar{A} \in \mathcal{F}(\mathbb{R}) : \text{there is } a \in \mathbb{R}, d > 0 \text{ such that}$$
$$\mu_{\bar{A}}(x) = g^{[-1]}(|x - a|/d) \text{ for all } x \in \mathbb{R}\Big\} \bigcup$$
$$\left\{\bar{A} \in \mathcal{F}(\mathbb{R}) : \text{ there is } a \in \mathbb{R} \text{ such that } \mu_{\bar{A}}(x) = \chi_{\{a\}}(x) \text{ for all } x \in \mathbb{R}\right\},$$

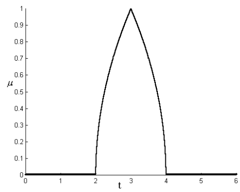

Figure 1: The quasi-triangular fuzzy number $\langle 3, 1 \rangle$

where χ_A is characteristic function of the set A. The elements of $\mathcal{N}_g$ will be called *quasi-triangular fuzzy numbers* generated by g with center λ and spread d and we will denote them with $\langle \lambda, d \rangle$.

Remark 2.3. If $\langle \lambda, d \rangle \in \mathcal{N}_g$ and $d > 0$, then α-levels of $\langle \lambda, d \rangle$ is $[\langle \lambda, d \rangle]^\alpha = [\lambda - dg(\alpha), \lambda + dg(\alpha)]$ and if $d = 0$, then $[\langle \lambda, d \rangle]^\alpha = \{\lambda\}$, for all $\alpha \in [0, 1]$.

Definition 2.4. The T-*Cartesian product*'s membership function of fuzzy sets $\bar{A}_i \in \mathcal{F}(X_i)$, $i = 1, \ldots, n$ is defined as

$$\mu_{\bar{A}}(x_1, x_2, \ldots, x_n) =$$
$$T\left(\mu_{\bar{A}_1}(x_1), T\left(\mu_{\bar{A}_2}(x_2), T\left(\ldots T\left(\mu_{\bar{A}_{n-1}}(x_{n-1}), \mu_{\bar{A}_n}(x_n)\right) \ldots\right)\right)\right),$$

for all $(x_1, x_2, \ldots, x_n) \in X_1 \times X_2 \times \cdots \times X_n$.

Example 2.5. Let $g\colon [0, 1] \to [0, \infty]$ be a function given by $g(t) = 1 - t^2$ for all $t \in [0, 1]$. Then the membership functions of quasi-triangular fuzzy numbers $\langle a, d \rangle$ is

$$\mu(t) = \begin{cases} 0 & \text{if} \quad t \leqslant a - d, \\ \sqrt{1 - \frac{a}{d} + \frac{t}{d}} & \text{if} \quad a - d < t \leqslant a, \\ \sqrt{1 + \frac{a}{d} - \frac{t}{d}} & \text{if} \quad a < t \leqslant a + d, \\ 0 & \text{if} \quad t > a + d \end{cases} \qquad \text{if } d > 0.$$

The graph of quasi-triangular fuzzy number $\langle 3, 1 \rangle$ we can see on the Figure 1.

3. The diffusion of information

Let A be a sample of data in a given normed space X that comes from the observation of some phenomenon. Let us denote a real relation with R. The method that defines the R from sample A is called an *operator*. We denote the set of all operators by Γ. Examples for operators: data series analysis, correlation examination, hypothesis examination, the method of artificial neural networks, etc.

Definition 3.1. Let R be a relation in X. The sample A is a *correct-data* set to R on universe $U \subseteq X$ if there exists an operator γ such that we can obtain a relation $R(\gamma, A)$ equal to the restriction of R at U.

Definition 3.2. Let R be a relation in X. The sample A is an *incomplete-data* set to R on universe $U \subseteq X$ if there does not exist an operator such that we can obtain the restriction of R at U from A.

Definition 3.3. Let $A = \{\mathbf{x}_k : k = 1, 2, \ldots, n\}$ be a deterministic sample in universe $U \subseteq X$. The *characteristic function* of A is $\chi_A : A \times U \to \{0, 1\}$, where

$$\chi_A(\mathbf{x}_k, \mathbf{u}) = \begin{cases} 1 & \text{if} \quad \mathbf{u} = \mathbf{x}_k, \\ 0 & \text{if} \quad \mathbf{u} \neq \mathbf{x}_k. \end{cases}$$

Definition 3.4. We consider a division U_j, $j = 1, \ldots, m$ of universe U, i.e.

$$U = \bigcup_{j=1}^{m} U_j, \;\; U_j \cap U_k = \emptyset \text{ if } j \neq k.$$

The *characteristic function* of the division U_j is $\chi_m\colon A \times U \to \{0, 1\}$, where

$$\chi_m(\mathbf{x}_k, \mathbf{u}) = \begin{cases} 1 & \text{if} \quad \mathbf{x}_k \in U_j, \\ 0 & \text{if} \quad \mathbf{x}_k \notin U_j, \end{cases} \quad \text{for all } \mathbf{u} \in U_j.$$

The characteristic function is replaceable with membership function $\mu\colon A \times U \to [0, 1]$. In this case, the value $\mu(x_k, u)$ shows how far the sample's element x_k is in set U_j. For example, if $X = \mathbb{R}$ then the membership function of quasi-triangular fuzzy number

$$\mu(x_k, u) = g^{[-1]}\left(\frac{|x_k - u|}{d}\right)$$

is a membership function of division to interval with centre x_k and length $2d$. Therefore, $\mu(:, u_j)$ is the membership function of U_j, for all $u_j \in U_j$.

Definition 3.5. The family of membership functions $\mu(:, u_j)\colon U \to [0, 1]$, $j = 1, \ldots, m$ is a *fuzzy division* of the U.

Definition 3.6. Let A be a sample of universe U. The function $\mu\colon A \times U \to [0, 1]$ is a *scattering function of the information*, if i) $\mu(\mathbf{x}_k, \mathbf{x}_k) = 1$, for all $\mathbf{x}_k \in A \cap U$; ii) for all $\mathbf{x}_k \in A$ and for all $\mathbf{u}, \mathbf{v} \in U$, if $\|\mathbf{x}_k - \mathbf{u}\| \leqslant \|\mathbf{x}_k - \mathbf{v}\|$, then $\mu(\mathbf{x}_k, \mathbf{u}) \geqslant \mu(\mathbf{x}_k, \mathbf{v})$.

For all elements $\mathbf{x}_k$ of the sample A the scattering function defines a fuzzy number with centre in $\mathbf{x}_k$ and membership function $\mu(\mathbf{x}_k, :)\colon U \to [0, 1]$. The simplest scattering function is $\mu_t = \chi$. This function will be called *trivial scattering function of the information*.

The scattering function of the information shows how far the data $\mathbf{u}$ can be the correct-data of a phenomenon. For example, if $\mathbf{u}$ is in sample A then $\mathbf{u}$ is totally correct-data of the phenomenon. Using the scattering function μ the sample A can be expand with new elements and so we get a sample notated by $A(\mu, U)$ with elements $(\mathbf{x}_k, \mathbf{u}_j, \mu(\mathbf{x}_k, \mathbf{u}_j)) \in A \times U \times [0, 1]$, where $\mathbf{u}_j \in U$, $j = 1, \ldots, p$ and $k = 1, \ldots, n$.

If $X = \mathbb{R}^n$, then it possible to define scattering function with help of quasi-triangular fuzzy numbers. We fuzzified all elements of sample A, i.e. for all components x_{ki} of vector $\mathbf{x}_k \in A$ we assign a quasi-triangular fuzzy number $\langle x_{ki}, \lambda(x_{ki})\rangle$ with spread $\lambda(x_{ki}) > 0$, $i = 1, \ldots, n$.

As follows from the definition of T_{gp}-Cartesian product the scattering function of information is given by

$$\mu\left((x_{k1}, x_{k2}, \ldots, x_{kn}), (u_1, u_2, \ldots, u_n)\right)$$
$$= \begin{cases} g^{-1}\left(\left[\left(\frac{|x_{k1}-u_1|}{\lambda(x_{k1})}\right)^p + \cdots + \left(\frac{|x_{kn}-u_n|}{\lambda(x_{k1})}\right)^p\right]^{\frac{1}{p}}\right) & \text{if} \quad |x_{ki} - u_i| \leqslant \lambda(x_{ki})\, g_0 \\ & \quad \text{for all } i = 1, \ldots, n, \\ 0 & \quad \text{otherwise.} \end{cases}$$

Let R be a relation on universe $U \subset X$ and γ be an operator. If we are using the sample $A = \{\mathbf{x}_k : k = 1, 2, \ldots, n\}$ to estimate the relation R, then our method is a *nondiffusion estimator*, and if we are using the sample $A(\mu, U)$, where μ is a nontrivial information scattering function, then our method is a *diffusion estimator.* The trivial information scattering function yields a nondiffusion estimator.

Theorem 3.7 (Principle of information diffusion). *Let R be a relation on universe $U \subset X = \mathbb{R}^n$, where U is a convex set. Let $A = \{\mathbf{x}_k : k = 1, 2, \ldots, n\}$ be a deterministic sample for estimation of R on universe U. We assume that γ is the best operator of relation R for some measurement of the error. The sample A is incomplete-data set of the relation R on U if and only if there exists a nontrivial information scattering function μ such that if we apply the operator γ to fuzzified sample $A(\mu, U)$, then we get a better estimation of R.*

The proof of this theorem see in Z. Makó (2006) and C. F. Huang (2006).

4. The approximation property of artificial neural network

The neural network can be understood as a mapping $f\colon \mathbb{R}^n \to \mathbb{R}^m$, defined by $\mathbf{y} = f(\mathbf{x}) = g(W\mathbf{x})$, where $\mathbf{x}$ is the input vector, $\mathbf{y}$ is the output vector, W is the weight matrix and g is the activation function. The mapping f can be decomposed into a chaining of mappings; the result is a multi-layer network $\mathbb{R}^n \to \mathbb{R}^p \to \mathbb{R}^q \to \ldots \to \mathbb{R}^m$. The algorithm for computing W is often called the training algorithm. The most popular neural network are the multi-layer back-propagation networks whose training algorithm is the well-known gradient descendent method. Such networks are called *back-propagation (BP) networks.*

An artificial neural network is a learning machine whose function depends on the training examples. So, the machine does not recognize the real relation but it determines a numerical relation among the state parameters. According to the principle of information diffusion we can increase the certainty of the determined

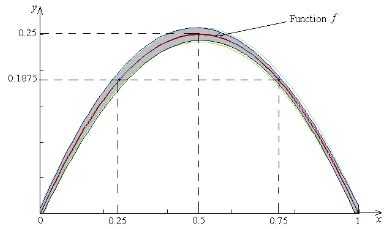

Figure 2: The banded approximation of the function f

relation if we multiply the number of the training examples with the help of an appropriate information scattering function or if we apply a banded approach. Neural networks trained in this manner are called *diffusion-neural-networks* (C. F. Huang and C. Moraga, 2004).

A number of authors have discussed the universal approximation property of BP networks. For example, in 1989 K. Hornik et al. proved that the multi-layer networks can approximate the continuous function to any degree of accuracy, i.e. *multi-layer networks have the universal approximation property.* After that, in 1995 J. Wray and G. G. R. Green showed that the universal approximation property does not hold in practice for networks implemented on computers.

4.1. Banded approximation

Let $f\colon [a,b] \to \mathbb{R}$ be a given continuous function and $A = \{((x_k, f(x_k)) \in [a,b] \times \mathbb{R} : k = 1, 2, \ldots, n\}$ be a given sample. We diffuse the information derived from this sample with the generator function $g : [0,1] \to [0, \infty]$, where $g(1) = 0$ and $g_0 = \lim_{t \to 0} g\,(t) \leqslant +\infty$. Thus, we obtain the fuzzified

$$\bar{A} = \{(\langle x_k, \alpha_k \rangle, \langle f\,(x_k)\,, \beta_k \rangle) \ : \ k = 1, 2, \ldots, n\}\,,$$

sample, where $\alpha_k, \beta_k \geqslant 0$ are the spread of quasi-triangular fuzzy numbers $\langle x_k, \alpha_k \rangle$ and $\langle f\,(x_k)\,, \beta_k \rangle$. Above derivative sample can be used to train a conventional BP-network with two input values x_k and α_k, and two output values $o_1(x_k, \alpha_k)$ and $o_2(x_k, \alpha_k)$. After the training we get a weight system where

$$H = \sum_{k=1}^{n} \left[(f\,(x_k) - o_1(x_k, \alpha_k))^2 + (\beta_k - o_2(x_k, \alpha_k))^2 \right]$$

the sum of square errors is less than a given number $\delta > 0$. The trained network for any input values x and 0 return the output values y and β. Using the generator function g, we can construct a band $[y - g(\gamma)\beta, y + g(\gamma)\beta]$ around to function f for any level value $\gamma \in [0,1]$. The approximation has precision ε on the level γ, if the distance between f and γ-level set $[y - g(\gamma)\beta, y + g(\gamma)\beta]$ is less than ε.

For illustration of this method, we consider the function $f\colon [0,1] \to \mathbb{R},\ \ f(x) = x - x^2$. Our task is to train the function with the sample

$$A = \{(0,0), (0.25, 0.1875), (0.5, 0.25), (0.75, 0.1875), (1, 0)\}.$$

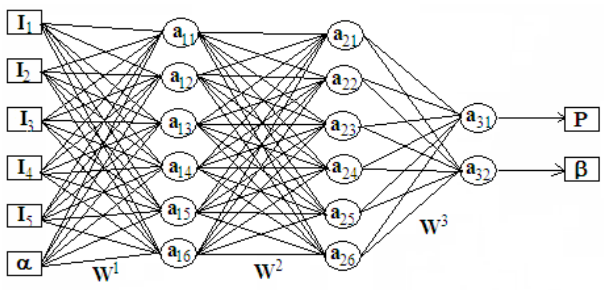

Figure 3: The DNN network for prognosis of the exchange rates

The fuzzified sample is

$$\bar{A} = \{(\langle 0,\alpha\rangle, \langle 0,\alpha\rangle), (\langle 0.25,\alpha\rangle, \langle 0.1875,\alpha\rangle),\\ (\langle 0.5,\alpha\rangle, \langle 0.25,\alpha\rangle), (\langle 0.75,\alpha\rangle, \langle 0.1875,\alpha\rangle),\\ (\langle 1,\alpha\rangle, \langle 0,\alpha\rangle)\},$$

where $\alpha = 0.0001$.

4.2. Prognosis of the exchange rates

We study the variation of the exchange rates of currency RON as a function of 6 currencies (CHF, GBP, HUF, EUR, JPY, USD). The back propagation difusion-neural-network is used for the analysis (see Figure 3). The input of network is the vector $(I_1, I_2, I_3, I_4, I_5, \alpha) =$

$$(CHF/EUR, GBP/EUR, HUF/EUR, 100JPY/EUR, USD/EUR, \alpha)$$

and the otput is $(P, \beta) = (RON/EUR, \beta)$. The activation function of all neurons is

$$tansig(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}.$$

We can use the derivative sample to train a conventional BP-network which is given by Figure 3. The I_1, I_2, I_3, I_4, I_5 and P are the exchange rates of currencies α and β are the accuracy of forecast. The Table 1 contains few elements of the derivative sample.

I_1	I_1	I_1	I_1	I_1	α	P	β
0.6477	1.4109	0.0040	0.7190	0.7384	0.02	0.2543	0.02
0.6464	1.4155	0.0040	0.7209	0.7473	0.02	0.2571	0.02
0.6435	1.4160	0.0040	0.7198	0.7556	0.02	0.2585	0.02

Table 1: The derivative sample

In this model the generator function of quasi-triangular fuzzy nubers is $g(t) = 1 - t^2$. The accuracy ε is ± 0.02. The value of γ is 0.9. The prognosis of the exchange

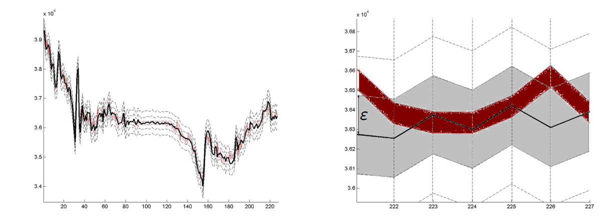

Figure 4: The banded approximation of the exchange rates

rates has precision ± 0.02 on the level 0.9, if the distance between the γ-level set of prognosis and exchange rates P is less than or equal to 0.02.

We can see on the figure 4 that the prediction band $[o_1 - g(\gamma)o_2, o_1 + g(\gamma)o_2]$ and the band $[P - \varepsilon, P + \varepsilon]$ has common point all the time. In conclusion the prognosis has precision ± 0.02 on the level 0.9.

References

[1] Hornik, K., Stinchcombe, M., White, H., Multilayer feedforward networks are universal approximators, *Neural Networks*, Vol. 2(5) (1989), 359–366.

[2] Huang, C. F., Moraga, C., A diffusion-neural-network for learning from small samples, *Internat.J. of Approx. Reason.,* Vol. 35 (2004), 137–161.

[3] Huang, C. F., An approach based on analytic geometry to prove the principle of information diffusion, *Dynamics of Continuous Discrete and Impulsive Systems-Series B-Applications & Algorithms*, Vol. 13 (2006), 146–150.

[4] Kovács, M., A stable embedding of ill-posed linear systems into fuzzy systems, *Fuzzy Sets and Systems,* Vol. 45 (1992), 305–312.

[5] Makó, Z., *Quasi-triangular fuzzy numbers. Theory and applications*, Scientia Publishing House, Cluj-Napoca, (2006).

[6] Menger, K., Statistical metrics, *Proc. Nat. Acad. Sci. USA*, Vol. 28 (1942), 535–537.

[7] Wray, J., Green, G. G. R., Neural networks, approximation theory and finite precision computation, *Neural Networks*, Vol. 8(1) (1995), 31–37.

[8] Zadeh, L. A., Fuzzy Sets, *Information and Control*, Vol. 8 (1965), 338–353.

[9] Zadeh, L. A., The concept of a linguistic variable and its application to approximate reasoning, *Parts 1 and 2, Inform. Sci.,* Vol. 8, 9 (1975), 199–249, 301–357.

Zoltán Makó and Júlia Salamon
Department of Mathematics and Informatics, Sapientia University
P-ta Libertatii nr. 1, RO-530104 Miercurea-Ciuc, Romania