B A D A N I A   O P E R A C Y J N E   I   D E C Y Z J E

Nr 4 2003

Przemysław GRZEGORZEWSKI*

APPROXIMATION OF A FUZZY NUMBER PRESERVING ENTROPY-LIKE NONSPECIFITY

The problem of the interval approximation of fuzzy numbers is discussed. A measure of uncertainty, called entropy-like nonspecifity is proposed and interval approximation operator preserving this non-specifity measure is suggested.

1. Introduction

Fuzzy sets are effectively applied in modelling and processing imprecise information. However, sometimes we have to approximate the given fuzzy set by a crisp one. If we then use a defuzzification operator which replaces a fuzzy set by a single number we generally loose too much important information. Therefore, an interval approximation of a fuzzy set is often advisable. Such approximation are used in many areas as fuzzy pattern recognition, fuzzy image processing, fuzzy algebra, etc. (see, e.g., [9], [10], [12]).

In this paper, we restrict our attention to the most important subfamily of all fuzzy sets, i.e., to fuzzy numbers. The problem of interval approximation of fuzzy numbers was discussed by Chanas [1] and Grzegorzewski [5], [6]. In the present paper, we propose a new method for the interval approximation preserving the amount of information delivered by the fuzzy number under study. In particular, we suggest an interval approximation operator which preserves entropy-like non-specifity measure.

* Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, e-mail: pgrzeg@ibspan.waw.pl

2. Interval approximation of a fuzzy set number

A fuzzy subset A of the real line $\mathbf{R}$ with membership function $\mu_A : \mathbf{R} \to [0, 1]$ is called a fuzzy number if it normal, fuzzy convex, its membership function μ_A is upper semicontinuous and its support is a closed interval. A space of all fuzzy numbers will be denoted by $\mathbf{F}(\mathbf{R})$, while $\mathbf{P}(\mathbf{R})$ will be a family of all closed intervals on the real time. Let us also recall that $core\ A = \{x \in \mathbf{R} : \mu_A(x) = 1\}$, and $supp\ A = cl(\{x \in \mathbf{R} : \mu_A(x) > 0\})$, where cl is the closure operator.

Sometimes we have to replace a fuzzy number with a crisp subset of the real line. More precisely, we have to find an operator $C : \mathbf{F}(\mathbf{R}) \to \mathbf{P}(\mathbf{R})$ which transforms fuzzy numbers into closed intervals of the real line. Although we can do this in many ways, such interval approximation of a fuzzy number should fulfil at least two natural requirements comprised in the following definition:

Definition 1. *An operator* $C : \mathbf{F}(\mathbf{R}) \to \mathbf{P}(\mathbf{R})$ *is called an interval approximation operator if for any* $A \in \mathbf{F}(\mathbf{R})$

(C1) $C(A) \subseteq suppA$

(C2) $coreA \subseteq C(A)$.

The definition given above leads to a very broad family of operators. However, it seems desirable to specify some additional requirements. For instance, it is natural to expect that if two fuzzy numbers A and B are close – in some sense – then their interval approximations are also close. This means that the interval approximation of a fuzzy number should fulfil the following continuous-type condition.

(C3) $\forall(A, B \in \mathbf{F})\ \ \forall(\varepsilon > 0)\ \ \exists(\delta > 0)\ \ d(A, B) < \delta \Rightarrow d(C(A), C(B)) < \varepsilon$,

where $d : \mathbf{F}(\mathbf{R}) \to [0, +\infty)$ denotes a metric defined in the family of all fuzzy numbers. Such interval approximation operator which satisfies (C3) is called the continuous interval approximation operator.

Different method for finding interval approximations of fuzzy sets are used. The easiest way is to substitute a fuzzy number either by its support, i.e., $C_0(A) = suppA$ or by its core, i.e., $C_1(A) = coreA$. However, using these methods all information due to fuzziness of the object under discussion is neglected. Hence these methods cannot be recommended to practitioners. Another operator, probably best known and the most popular in practice, is

$$C_{0.5}(A) = \{x \in \mathbf{R} : \mu_A(x) \geq 0.5\}. \tag{1}$$

This operator seems to be a compromise between two extremes C_0 and C_1. Moreover, it has quite a natural interpretation: any $x \in \mathbf{R}$ belongs to the approximation interval $C_{0.5}(A)$ of a fuzzy number A if and only if its degree of belongingness to A is not smaller than its degree of belongingness to the complement of A (i.e., belongs to

A rather to $\neg A$). $C_{0.5}(A)$ is sometimes called in literature the nearest ordinary set of a fuzzy set A. Unfortunately, these three operators are, in general, not continuous.

In [5] and [6], Grzegorzewski suggested a simple and natural continuous approximation operator which is the nearest interval approximation of fuzzy number with respect to L^2-type metric (see [4]). It is worth noting that this nearest interval approximation operator is equivalent to the expected interval of a fuzzy number, considered in a different context independently by Dubois and Prade [3] and Heilpern [7]. The problem of the interval approximations was also considered by Chanas [1], who investigated an operator which was the best wit respect to the Hamming distance

$$d(A, B) = \int_{-\infty}^{\infty} |\mu_A(x) - \mu_B(x)| \, dx, \tag{2}$$

where μ_A and μ_B are membership functions of A and B, respectively.

The papers mentioned above were focused mainly on "geometric" properties of the approximation operators. However, when we approximate one model with another one, this basically means that we want to replace one type of information with an equal amount of information of another type. In other words, we want to convert information of one type to another one while, at the same time, preserving its amount. This expresses the spirit of the so called principle of information invariance (see [11]). Therefore, it seems desirable that an interval approximation $C(A)$ of a fuzzy number A should contain the same (or at least similar) amount of information as the initial fuzzy set A. Here, one can considered different measures of uncertainty, information, specifity or nonspecifity, etc. If we denote such a measure by I, then we can write this natural requirement in the following way

(C4) $I(A) = I(C(A))$.

Below we propose an interval approximation operator of fuzzy numbers which fulfil all desirable conditions (C1)–(C4).

3. Entropy-like nonspecifity measure

A concept of information is intimately connected with the concept of uncertainty. In the framework of probability theory the Shannon entropy [13] is commonly used to measure the amount of information contained in a probability distribution. Although for about three hundred years uncertainty was conceived solely in terms of probability theory, now it is also described by fuzzy set theory, possibility theory, evidence theory and the theory of fuzzy measures. Different types of uncertainty are now recognized in these theories and new tools for handling uncertainty are worked out. Among them one can find the concept of *specifity*, originally introduced by Yager [14], to measure the degree to which fuzzy subset contains one and only one element. This measure can also be used to indicate the degree to which a possibility distribution allows one and only one element as being possible. Conversely, Higashi and Klir [8] described nonspecifity of a fuzzy set, using the so called, *U-uncertainty*. The notion of entropy is also considered in the framework of fuzzy sets: starting from Zadeh's definition [15], through the concept of nonprobabilistic entropy due to De Luca and Termini [2]. For more details connected with the types of uncertainty and uncertainty measures we refer the reader to [11].

Many of the uncertainty measures mentioned above were considered generally for discrete universes of discourse. Below we suggest another nonspecifity measure which seems to be appropriate for our purposes and for dealing with fuzzy numbers (i.e., in the continuous domains). Let us start from the general definition of a noncpecifity measure.

Definition 2. *A nonspecifity is any function* $N : \mathbf{F}(\mathbf{R}) \to [0, +\infty)$ *satisfying the following conditions*

(a) $N(A) = 0 \Leftrightarrow A$ *is a crisp real number*,

(b) $A \subseteq B \Rightarrow N(A) \le N(B)$.

Thus, a nonspecifity measure indicates a grade of nonspecifity of a fuzzy number. The lowest nonspecifity (equal to zero) is received by crisp real numbers and it increases together with enlarging cores and supports. One can easily find many nonspecifity measures, but we will consider here only one such measure, as defined below:

Definition 3. *A function* $H : \mathbf{F}(\mathbf{R}) \to [0, +\infty)$ *given as follows*

$$H(A) = \frac{1}{\ln 2} \int_{-\infty}^{\infty} \mu_A(x) \ln(1 + \mu_A(x))\, dx \tag{3}$$

is called the entropy-like nonspecifity measure.

A function proposed above is called "entropy-like" measure since its form resembles Shanon's entropy. We immediately get.

Proposition 1. *H is a nonspecifity measure.*

Proof: We have to prove that a function H given by (3) satisfies both conditions (a) and (b) from Definition 2. Suppose firstly that A is a crisp real number. Then there exist such $x_0 \in \mathbf{R}$ that $\mu_A(x_0) = 1$ and $\mu_A(x) = 0$ for all $x \in \mathbf{R} \setminus \{x_0\}$. Hence $H(A) = 0$. Conversely, let $H(A) = 0$. This means that $\mu_A(x) = 0$ for $x \in \mathbf{R} \setminus E$ and $\mu_A(x) \neq 0$ for $x \in E$, where E is a set of measure zero. However, since A is a fuzzy number, it should be normal, fuzzy convex, it should have upper semicontinuous membership function and support which is a closed interval. Therefore, set E should have the following form $E = \{x_0\}$, where $\mu_A(x_0) = 1$. Thus, finally, A is a crisp real number and condition (a) holds.

To prove (b) let us assume that we have two fuzzy numbers A and B such that $A \subseteq B$. This means that $\mu_A(x) \le \mu_B(x)$ $\forall x \in \mathbf{R}$, where μ_A and μ_B denote membership functions of A and B, respectively.

Thus, we get

$$H(A) = \frac{1}{\ln 2} \int_{-\infty}^{\infty} \mu_A(x) \ln(1 + \mu_A(x))\, dx \le \frac{1}{\ln 2} \int_{-\infty}^{\infty} \mu_B(x) \ln(1 + \mu_B(x))\, dx = H(B),$$

which completes the proof. ®

Proposition 2. *For any fuzzy number A*

$$|\, core\, A \,| \le H(A) \le |\, supp A|$$

and we get H(*A*) = |*supp A*| = |*core A*| *if and only if A is a crisp interval*.

Proof: It is easily seen that

$$H(A) = \frac{1}{\ln 2} \int_{-\infty}^{\infty} \mu_A(x) \ln(1 + \mu_A(x))\, dx \le \frac{1}{\ln 2} \int_{suppA} \mu_A(x) \ln(1 + \mu_A(x))\, dx$$

$$\le \frac{1}{\ln 2} \int_{suppA} \ln 2 dx = |\, supp A|.$$

On the other hand,

$$H(A) = \frac{1}{\ln 2} \int_{-\infty}^{\infty} \mu_A(x) \ln(1 + \mu_A(x))\, dx \ge \frac{1}{\ln 2} \int_{core\, A} \ln 2 dx = |\, core A|,]$$

which proves the proposition. ®

Among desired properties of the entropy-like nonspecifity measure one can distinguish its continuity. We have

Proposition 3. *Let A and B denote two arbitrary fuzzy numbers. Then* $\forall\, \dot{\varepsilon} > 0)$ $\exists(\delta > 0)$

$$d(A, B) < \delta \Rightarrow |\, H(A) - H(B) \,| < \varepsilon\, .$$

The proof is straightforward, because function (3) is continuous.

4. Interval approximation of a fuzzy number preserving nonspecifity

It is known that for any fuzzy number A there exist four numbers $a_1, a_2, a_3, a_4 \in \mathbf{R}$ and two functions $f_A, g_A : \mathbf{R} \to [0, 1]$, where f_A is a nondecreasing and g_A is a nonincreasing function, such that we can describe a membership function μ_A in the following manner

$$\mu_A(x) = \begin{cases} 0 & \text{if } x < a_1 \\ f_A(x) & \text{if } a_1 \le x < a_2 \\ 1 & \text{if } a_2 \le x \le a_3 \\ g_A(x) & \text{if } a_3 < x \le a_4 \\ 0 & \text{if } a_4 < x. \end{cases} \tag{4}$$

Now, let us consider the following operator $C_H : \mathbf{F}(\mathbf{R}) \to \mathbb{P}(\mathbf{R})$

$$C_H(A) = \left[a_2 - \frac{1}{\ln 2} \int_{a_1}^{a_2} f_A(x) \ln(1 + f_A(x))\, dx,\ a_3 + \frac{1}{\ln 2} \int_{a_3}^{a_4} g_A(x) \ln(1 + g_A(x))\, dx \right]. \tag{5}$$

It can be shown that

Proposition 4. *Let A be a fuzzy number with the membership function μ_A given by* (4). *Then an operator C_H given by* (5) *is an approximation operator*.

Proof: We have to prove that the given set (5) fulfils conditions (C1) and (C2). Let us first adopt the following notation:

$$H_f(A) = \frac{1}{\ln 2} \int_{a_1}^{a_2} f_A(x) \ln(1 + f_A(x))\, dx, \tag{6}$$

$$H_g(A) = \frac{1}{\ln 2} \int_{a_3}^{a_4} g_A(x) \ln(1 + g_A(x))\, dx. \tag{7}$$

Since both H_f and H_g are nonnegative we get

$$C_H(A) = [a_2 - H_f(A), a_3 + H_g(A)] \supseteq [a_2, a_3] = coreA,$$

i.e., (C1) holds.

According to (6) and (7), we get $H_f(A) \le a_2 - a_1$ and $H_g(A) \le a_4 - a_3$. Therefore,

$$C_H(A) = [a_2 - H_f(A), a_3 + H_g(A)] \subseteq [a_1, a_4] = supp\ A,$$

and (C2) holds, which completes the proof. ®

Proposition 5. *The interval approximation C_H given by* (5) *preserves entropy-like nonspecifity, i.e., for any fuzzy number A*

$$H(C_H(A)) = H(A).$$

Proof: Using (6) and (7) notation we have

$$C_H(A) = [a_2 - H_f(A), a_3 + H_g(A)].$$

Since $C_H(A)$ is an interval, thus by Proposition 2 we get

$$\begin{aligned} H(C_H(A)) &= |\,suppC_H(A)\,| = a_3 + H_g(A) - (a_2 - H_f(A)) \\ &= H_f(A) + (a_3 - a_2) + H_g(A) \\ &= \frac{1}{\ln 2}\int_{a_1}^{a_2} f_A(x)\ln(1+f_A(x))\,dx + (a_3 - a_2) + \frac{1}{\ln 2}\int_{a_3}^{a_4} g_A(x)\ln(1+g_A(x))\,dx \\ &= \frac{1}{\ln 2}\int_{a_1}^{a_4} \mu_A(x)\ln(1+\mu_A(x))\,dx = H(A), \end{aligned}$$

which is our claim. ®

Now, let us consider two fuzzy numbers A and B which are close with respect to the Hamming distance (2). Then we can show the following theorem.

Proposition 6. *Let A and B be two fuzzy numbers. Then* $\forall(\varepsilon > 0)\ \exists(\delta > 0)$

$$d(A, B) < \delta \Rightarrow d(C_H(A), C_H(B)) < \varepsilon.$$

The proof results immediately from Proposition 3.

Propositions 4–6 actually prove that all desired conditions (C1)–(C4) are fulfilled. However, it is obvious that one can construct several approximation operators based on the entropy-like nonspecifity measure H satisfying (C1)–(C4). Next, two proposition show that our operator is also in some sense the best one.

Proposition 7. *If A is a trapezoidal fuzzy number then the operator $C_H(A)$ produces the nearest interval approximation preserving the entropy-like nonspecifity measure H of a fuzzy number with respect to the Hamming distance* (2).

Proof: Let us assume that $C_H(A) = [c_1, c_2]$, where $c_1 = a_2 - H_f(A)$ and $c_2 = a_3 + H_g(A)$. Suppose $|C_H(A)| = c_2 - c_1 = h$. It is evident that each interval $I_s(A) = [s, s + h]$, such that $a_1 \le s \le a_2$ and $a_3 \le s + h \le a_4$, is an interval approximation of the fuzzy number A and

$$H(I_s(A)) = H(C_H(A)) = H(A).$$

Moreover, it is seen that there exist $s_0 \in [a_1, a_2]$ such that $C_H(A) = I_s(a)$.

Now, let us compute the Hamming distance between the fuzzy number under study A and given interval approximation $I_s(A)$. By (2) we get

$$d(A, I_s(A)) = \int_{-\infty}^{+\infty} |\mu_A(x) - \chi_{Is}(x)|\, dx\,,$$

where $\chi_{Is}(x)$ is a characteristic function of the interval $I_s(A)$. Assuming that the membership function μ_A of A is given by (4) we get

$$d(A, I_s(A)) = \int_{a_1}^{s} f_A(x)\, dx + (a_2 - s) - \int_{s}^{a_2} f_A(x)\, dx$$
$$+ (s + h - a_3) - \int_{a_3}^{s+h} g_A(x)\, dx + \int_{s+h}^{a_4} g_A(x)\, dx\,.$$

To minimize $d(A, I_s(A))$ we compute its derivative with respect to s

$$d'(A, I_s(A)) = 2 f_A(s) - 2 g_A(s+h)\,.$$

The solution of the equation $d'(A, I_s(A)) = 0$ should fulfil the following condition

$$f_A(s) = g_A(s+h)\,. \tag{8}$$

Since A is a trapezoidal fuzzy number, then its membership function is given by

$$\mu_A(x) = \begin{cases} 0 & \text{if} \quad x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{if} \quad a_1 \le x < a_2 \\ 1 & \text{if} \quad a_2 \le x \le a_3 \\ \dfrac{a_4 - x}{a_4 - a_3} & \text{if} \quad a_2 < a \le a_4 \\ 0 & \text{if} \quad a_4 < x \end{cases} \tag{9}$$

and equation (8) reduces to

$$\frac{s - a_1}{a_2 - a_1} = \frac{a_4 - (s+h)}{a_4 - a_3}\,.$$

After some easy computations we get

$$s = \frac{a_2 a_4 - a_3 a_1 + (a_1 - a_2) h}{a_4 - a_3 + a_2 - a_1}\,. \tag{10}$$

Using (6) and (7) for the trapezoidal fuzzy number (9) we get

$$H_f(A) = \frac{a_2 - a_1}{4 \ln 2}\,,$$

$$H_g(A) = \frac{a_4 - a_3}{4\ln 2}.$$

Then substituting

$$h = c_2 - c_1 = a_2 - H_f(A) - (a_3 + H_g(A)) = \frac{1}{4\ln 2}(a_4 - a_3 + a_2 - a_1) + a_3 - a_2$$

into (10) we get $d'(A, I_s(A)) = 0$ for

$$s = a_2 - \frac{a_2 - a_1}{4\ln 2} = c_1.$$

Since this derivative changes also its sign from a minus into plus crossing the point $s = c_1$, we can assert that the Hamming distance reaches its minimum for the interval $I_{c1}(A) = C_H(A)$ which completes the proof. ®

One can also prove the following theorem:

Proposition 8. *If A is a symmetrical fuzzy number then the operator $C_H(A)$ produces the nearest interval approximation preserving the entropy-like nonspecificity measure H of a fuzzy number with respect to the Hamming distance* (2).

Proof: Let $A \in \mathbf{F}(\mathbf{R})$ with a membership function μ_A be given by (4). Since A is symmetrical, then we have

$$\mu_A\left(\frac{a_2 + a_3}{2} - z\right) = \mu_A\left(\frac{a_2 + a_3}{2} + z\right)$$

for any $z \in \mathbf{R}$. Hence

$$f_A\left(\frac{a_2 + a_3}{2} - z\right) = g_A\left(\frac{a_2 + a_3}{2} + z\right) \tag{11}$$

for any $z \in \mathbf{R}$.

From the proof given above we know that each interval $I_s(A) = [s, s + h]$ such that $h = |C_H(A)| = |c_2 - c_1|$ is an interval approximation of A. Since in our case $H_f(A) = H_g(A)$, hence

$$h = 2H_f(A) + a_3 - a_2.$$

We also know that the nearest interval approximation with respect to the Hamming distance should fulfil (8). Combining (8) with (11) we obtain.

$$s = \frac{a_2 + a_3}{2} - z,$$

$$s + h = \frac{a_2 + a_3}{2} + z$$

and after simple calculations we get

$$s = a_2 - H_f(A) = c_1 .$$

Since this derivative changes also its sign from a minus into plus crossing the point $s = c_1$, hence the Hamming distance reaches its minimum for the interval $I_{c1}(A) = C_H(A)$, which is a desired results. ®

References

[1] CHANAS S., *On the interval approximation of a fuzzy number*, Fuzzy Sets and Systems, 2001, 122, 353–356.

[2] DE LUCA A., TERMINI S., *A definition of a nonprobabilistic entropy in the setting of fuzzy set theory*, Information and Control, 1972, 20, 301–312.

[3] DUBOIS D., PRADE H., *The mean value of a fuzzy number*, Fuzzy Sets and Systems, 1987, 24, 279–300.

[4] GRZEGORZEWSKI P., *Metrics and orders in space of fuzzy numbers*, Fuzzy Sets and Systems, 1998, 97, 83–94.

[5] GRZEGORZEWSKI P., *Interval aggregation in data mining*, [in:] Flexible Query Answering Systems, Larsen H.L., Kacprzyk J., Zadrożny S., Adreasen T., Christianses H. (eds.), Springer, 2000, 465–474.

[6] GRZEGORZEWSKI P., *Nearest interval approximation of a fuzzy number*, Fuzzy Sets and Systems, 2002, 130, 321–330.

[7] HEILPERN S., *The expected value of a fuzzy number*, Fuzzy Sets and Systems, 1992, 47, 81–86.

[8] HIGASHI M., KLIR G.J., *Measures of uncertainty and information based on possibility fistributions*, Intern. J. of General Systems, 1983, 9, 43–58.

[9] KANDEL A., *Fuzzy Techniques in Pattern Recognition*, Wiley, New York, 1982.

[10] KAUFMANN A., *Introduction to the Theory of Fuzzy Subsets*, Academic Press, 1998.

[11] KLIR G.J., WIERMAN M.J., *Uncertainty-Based Information*, Physica-Verlag, 1998.

[12] PAL S.K., DUTTA MAJUMDER D.K., *Fuzzy Mathematical Approach to Pattern Recognition*, Wiley Eastern Ltd, 1987.

[13] SHANNON C.E., *The mathematical theory of communication*, The Bell System Technical J., 1948, 27, 379–423, 623–656.

[14] YAGER R.R., *Entropy and specificity in a mathematical theory of evidence*, Intern. J. of General Systems, 1983, 9, 249–260.

[15] ZADEH L.A., *Probability measures of fuzzy events*, J. Math. Anal. Appl., 1968, 23, 421–427.

Aproksymacja liczb rozmytych zachowująca entropijną miarę niespecyficzności

Zbiory rozmyte okazały się bardzo pomocne w modelowaniu i efektywnym przetwarzaniu nieprecyzyjnych informacji. Czasem zachodzi jednak konieczność przybliżenia danego zbioru rozmytego za pomocą zbioru nierozmytego. W tym celu stosuje się zazwyczaj defuzyfikację (wyostrzanie), ale metoda ta

niestety często prowadzi do utraty zbyt wielu cennych informacji. W tym przypadku wskazane być może posłużenie się aproksymacją przedziałową.

W niniejszej pracy ograniczymy się do najważniejszej podrodziny zbiorów rozmytych, tzn. do liczb rozmytych. Dla wspomnianej rodziny przedstawiono nową metodę aproksymacji przedziałowej, zachowującą ilość informacji, jaką dostarcza przybliżana liczba rozmyta. Dokładniej, wprowadzona zostanie pewna miara informacji, zwana entropijną miarą niespecyficzności, a następnie wskazana zostanie metoda aproksymacji przedziałowej liczb rozmytych, zachowująca tę miarę informacji.