Regression Analysis based on Fuzzy Evidence Theory

Simon Petit-Renaud and Thierry Denœux
Université de Technologie de Compiègne
HeuDiaSyc-UMR C.N.R.S. 6599
BP 20529, 60205 Compiègne, France
(E-mail : spetit,tdenoeux@hds.utc.fr)

Abstract

We propose a new approach to functional regression based on the fuzzy evidence theory. This method uses a training set for computing a fuzzy belief structure that quantifies different sort of uncertainties, such as non-specificity, discord in the output data, or low density around the input data. The method can use a very large class of output data, such as real, interval or fuzzy numbers, or, more generally, what we called *fuzzy belief structure* numbers. We show that our approach can be regarded as a kind of a fuzzy system and we present the analogies with the fuzzy model proposed by Yager in [13], which can take output discord into account. The proposed model can provide predictions, in a variety of forms depending on the accuracy of the available information, such as: a crisp output, a fuzzy output, a probability distribution and some information criteria (non-specificity, strife, ignorance degree).

1. Introduction

Although statistical regression analysis is one of the most widely used techniques to model relationships between variables, it does not represent well all kinds of uncertainties which can be encountered in real applications. Evidence theory [1, 8], fuzzy set theory, and fuzzy evidence theory [9, 14], are particularly convenient to this purpose, compared to classical probability theory, and can highlight different forms of uncertainties [6]. Several approaches such as fuzzy linear regression [10, 5], interval regression [7], fuzzy systems aim to model vague or fuzzy phenomena. A neural network approach based on evidence theory [4] has also been proposed as a functional approximation method.

Here, we propose a new regression method, based on fuzzy evidence theory, which can be applied to a large class of output data, such as intervals or fuzzy numbers. A more general case including the probabilistic uncertainty attached to a given output is also treated. The formalism of belief structures allows us to assess and quantify uncertainty in the output data. A training set defines a set of belief structures in a similar manner that Denœux proposed in discrimination [2, 3]. The concept of pignistic probabilities allows to transform degrees of belief into a new type of confidence values that may be attached to predictions. We also show that our model can be interpreted as a fuzzy system. From this point of view, we present the analogy with the model proposed by Yager et al. [12, 13], where the consequent part of the system is a belief structure. We extend this model by introducing the output reference set as a new focal element and modifying the inference rules. Thus, our model can integrate different kinds of uncertainties concerning the data, such as imprecision, discord and degree of confidence.

The paper is organised as follows. in Section 2, we briefly recall the basics of the theory of evidence and its fuzzy extension. In Section 3, we describe the proposed model and a variant using prototypes. The fuzzy system point of view is discussed in Section 4, where a comparison with Yager's model is undertaken. Finally, we give some numerical results and in Section 5 and conclude in Section 6.

2. Fuzzy Belief Structures

In this section, we briefly introduce some ideas of the Dempster-Shafer theory of evidence [1, 8, 11]. Let Ω be a finite set, and $S(\Omega)$ be the set of the subsets of Ω. The fundamental concept for representing uncertainty is that of belief structure, also called basic probability assignment, defined as a function m, from $S(\Omega)$, to $[0,1]$, verifying: $\sum_{A\subseteq\Omega} m(A) = 1$. A belief structure m such that $m(\emptyset) = 0$ is said to be *normal*. Any subset of Ω such as $m(A) > 0$ is called a focal element of m. We will denote by $F(m)$ the set of focal elements of m. The information provided by a belief structure can be represented by a *credibility function* or a *plausibility function*, respectively defined as: $Bel(A) = \sum_{B\subseteq A} m(B)$ and $Pl(A) = \sum_{B\cap A=\emptyset} m(B) = Bel(\Omega) - Bel(\bar{A})$. Bel(A) is interpreted as the *total* belief committed to A and Pl(A) as the belief that *might* be committed to A. They can respectively be regarded as lower and upper bounds of a set of compatible probabilities. Amongst these probabilities, Smets has defined a "neutral" probability, the *pignistic* probability $BetP$ such as $BetP(\omega) = \sum_{A \subset \Omega} \frac{m(A)}{|A|}\delta_A(\omega)$ where δ is the Kronecker symbol and $|A|$ denotes the cardinality of A. Evidence theory can be easily generalised to the case where Ω is infinite if the number of focal elements $|F(m)|$ is finite.

One of the most important operations of the theory is the procedure for aggregating multiple belief structures on the same variable. For *any* binary set operation ∇, the fusion of two belief structures m_1 and m_2, noted $m = m_1 \nabla m_2$, is defined as: $m(A) = \sum_{B \nabla C = A} m_1(B)m_2(C)$, for all A belonging to $\Omega \setminus \emptyset$. This operation may produce a subnormal belief structure, i.e. one can have $m(\emptyset) > 0$, particularly in the case of the conjunctive sum defined originally by Dempster, where $\nabla = \cap$. The Dempster normalisation procedure converts a subnormal belief structure m into a normal one m^* defined as follows: $m^*(A) = \frac{m(A)}{1-m(\emptyset)}$ if $A \neq \emptyset$ and $m(\emptyset) = 0$.

A generalisation of evidence theory has been proposed by different authors [9, 14]. It first consists in allowing the focal elements to be fuzzy subsets. If A is a fuzzy subset of Ω, we will denote by $A(.)$ the membership function of A and by $h(A)$ its height. If one denotes $\mathcal{F}(\Omega)$ the set of fuzzy sets of Ω, a fuzzy belief structure (FBS) is a function m, from $\mathcal{F}(\Omega)$, to $[0, 1]$ such that:

$$\begin{cases} m(A) = 0, \quad \text{if} \quad A \notin F(m) \\ \sum_{A \in F(m} m(A) = 1 \end{cases} \tag{1}$$

for some finite set $F(m) \in \mathcal{F}(\Omega)$. Here again, A is a focal element of m if $m(A) \neq 0$. If at least one of the focal elements is not normal, m is called a subnormal fuzzy belief structure. Yager provided a procedure, called "smooth normalisation procedure (SNP)", for converting a subnormal fuzzy belief structure in a fuzzy belief structure. This method generalises both the fuzzy set normalisation and the classical Dempster belief structure normalisation. Assume m is a subnormal fuzzy belief structure with focal elements $(F_i)_{i=1}^n$. The SNP converts m to a normal FBS m^* with focal elements $(E_i)_{i=1}^n$, such that:

$$\begin{cases} E_i(\omega) = F_i(\omega)/h(F_i) \quad \forall \omega \in \Omega \\ m^*(E_i) = \frac{m(F_i)h(F_i)}{\sum_{i=1}^n m(F_i)h(F_i)} \end{cases} \tag{2}$$

The combination of fuzzy belief structure can also be easily generalised. Whatever the used combination operator ∇, the fuzzy version is employed. Thus, if m_1 and m_2 are two fuzzy belief structures with focal elements $(A_i)_{i=1}^n$ and $(B_i)_{i=1}^p$ respectively, and $m = m_1 \cup m_2$, the focal elements on m are defined as: $C_k = A_i \cup B_j$ where $C_k(\omega) = S(A_i(\omega), B_j(\omega))$ and S is a t-conorm. The corresponding weight is $m(C_k) = \sum_{A_i \cup B_j = C_k} m_1(A_i)m_2(B_j)$.

The pignistic probability can also be easily extended by considering the membership degree to each fuzzy focal element and naturally generalises the classical notion:

$$BetP(\omega) = \sum_{A \subset \Omega} \frac{m(A)}{|A|} A(\omega). \tag{3}$$

3. Application to Regression Analysis

3.1. The Proposed Model

In this paper, we propose a regression method based on fuzzy evidence theory, that can model various kinds of uncertainties. The principle of our approach, which can be viewed as a "regression" one, is analoguous to the one proposed by Denœux [3] in classification.

In the following, we will denote by $\tilde{y}$ a fuzzy quantity, belonging to $\mathcal{F}(\mathcal{Y})$, where $\mathcal{Y}$ is a given reference set. Let $(\boldsymbol{x}_i, \tilde{y}_i)_{i=1,\dots,n}$ be a training set of input - output data vectors belonging to $(\mathcal{X}_1 \times \dots \times \mathcal{X}_r \times \mathcal{F}(\mathcal{Y}))$, where each $\mathcal{X}_j$ and $\mathcal{Y}$ are supposed to be continuous real spaces. Thus the output data can be either real numbers, intervals, fuzzy numbers or any fuzzy quantities. Let us note $\mathcal{X}$ the Cartesian space $\mathcal{X}_1 \times \dots \times \mathcal{X}_r$.

Let us now consider an input vector $\boldsymbol{x}$, for which we wish to provide some information towards its output y, based on the training set. Each pair $(\boldsymbol{x}_i, \tilde{y}_i)$ may be regarded as a piece of evidence inducing some beliefs concerning y. If one learns that $\boldsymbol{x}$ is "close" to $\boldsymbol{x}_i$ according to a relevant metric d, we can hope that the corresponding output y will be similar to $\tilde{y}_i$ and this belief is all the more important than the distance between the two vectors is small. Since no other hypothesis is made, the remaining belief is allocated to the whole set $\mathcal{Y}$. This piece of evidence can be modeled by a FBS in $\mathcal{Y}$, m_i, whose focal elements are $F(m_i) = \{\tilde{y}_i, \mathcal{Y}\}$. The FBS m_i is defined as follows:

$$\begin{cases} m_i(\tilde{y}_i) & = & \phi(d(\boldsymbol{x}, \boldsymbol{x}_i)) \quad \forall i \in \{1, \dots, n\} \\ m_i(\mathcal{Y}) & = & 1 - \phi(d(\boldsymbol{x}, \boldsymbol{x}_i)) \\ m_i(A) & = & 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) - F(m_i) \end{cases} \tag{4}$$

where ϕ is a decreasing function such as $\lim_{d \to 0} \phi(d) = 1$ and $\lim_{d \to \infty} \phi(d) = 0$. In the following, we will denote $\phi_i(\boldsymbol{x}) = \phi(d(\boldsymbol{x}, \boldsymbol{x}_i))$. For example, if ϕ is the exponential function, d is a Euclidean distance induced by some definite positive matrix Σ: $\phi_i(\boldsymbol{x}) = \exp\{-(\boldsymbol{x} - \boldsymbol{x}_i)^T \Sigma (\boldsymbol{x} - \boldsymbol{x}_i)\}$

We note that by this means, two kinds of uncertainties can be clearly identified: imprecision or nonspecificity, which is due to the fuzziness of the focal elements, and the ignorance, represented by the weight assigned to the reference output set $\mathcal{Y}$.

In order to combine the information provided by each piece of evidence, we use the generalisation of the conjunctive rule of combination to FBS. Then, the final FBS is: $m = \cap_{i=1}^n m_i$. and $|F(m)| \leq 2^n$. If no approximation is made, the total number of focal elements can reach 2^n. Another way to decrease the number of calculations is to use the information only provided by the k nearest neighbours of $\boldsymbol{x}$ in the training set: $\{\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(k)}\}$ according to the distance d. Then the final FBS becomes:

$$m = \cap_{i=1}^{k} m_{(i)}$$

The pignistic probability of m is by definition $BetP(y) = \sum_{A \in F(m)} \frac{m(A)}{|A|} A(y)$ where $|A| = \int_{\mathcal{Y}} A(y)dy$ since $\mathcal{Y}$ is a continuous space. A point estimate can be provided by taking the expected value of the output:

$$I\!E(Y) = \int_{\mathcal{Y}} uBetP(u)du = \sum_{A \subset F(m)} m(A) \frac{\int_{\mathcal{Y}} uA(u)du}{|A|}$$

Let us note the center of area of the set A,

$$y_A^* \triangleq \frac{\int_{\mathcal{Y}} yA(y)dy}{\int_{\mathcal{Y}} A(y)dy}$$

Then

$$I\!E(Y) = \sum_{A \in F(m)} m(A) y_A^* \qquad (5)$$

In the following, we will denote the model (4) by **FBS1**.

As remark by Smets (personal communication), if the $\tilde{y}_i$ are fuzzy singletons, the m_i and m become traditional belief structures with only $n+1$ crisp focal elements and the final belief structure, denoted as **FBS2**, can be easily expressed:

$$\left\{ \begin{array}{lcl} m(\{y_i\}) & = & \frac{1}{K} \phi_i(\boldsymbol{x}) \prod_{j \neq i}(1 - \phi_j(\boldsymbol{x})) \quad \forall i \\ m(\mathcal{Y}) & = & \frac{1}{K} \prod_{i=1}^{m}(1 - \phi_i(\boldsymbol{x})) \\ m(A) & = & 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus F(m) \end{array} \right. \qquad (6)$$

where $K = \prod_{i=1}^{m}(1 - \phi_i(\boldsymbol{x}) + \sum_{i=1}^{m} \phi_i(\boldsymbol{x}) \prod_{j \neq i}(1 - \phi_j(\boldsymbol{x})$ is the normalisation factor.

Then the pignistic probability is a discontinuous function:

$BetP(y) = \sum_{i=1}^{m} m(\{y_i\}) \delta_{\{y_i\}}(y) + \frac{m(\mathcal{Y})}{|\mathcal{Y}|} \quad \forall y \in \mathcal{Y}$ and

$$I\!E(Y) = \int_{\mathcal{Y}} yBetP(y)dy = m(\{e_i\})e_i + m(\mathcal{Y}) y_{\mathcal{Y}}^*.$$

Remark: if $\mathcal{Y}$ is a compact on the form $[y_{min}, y_{max}]$, then $y_{\mathcal{Y}}^* = (y_{min} + y_{max})/2$.

3.2. Generalisation to Output Discord

We propose to generalise equation (4) to the case where there is some conflict between different possible outputs for a given input $\boldsymbol{x}_i$. We consider that the knowledge of output y_i can be modeled as a FBS m_{y_i}, such

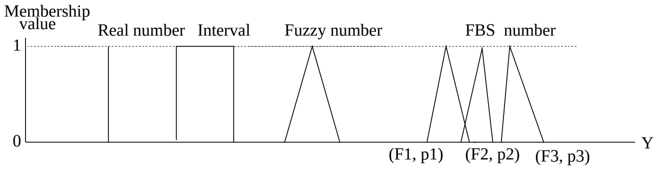

Figure 1: different output representations we have considered

as $F(m_{y_i}) = \{\tilde{y}_{i1}, \ldots, \tilde{y}_{ij}, \ldots, \tilde{y}_{i,J(i)}\}$, where each $\tilde{y}_i$ is a fuzzy quantity, and $m(\tilde{y}_i) = p_{ij}$. We will call such a number m_{y_i} a *fuzzy belief structure* number (FBS number)(see Figure 1).

If we now consider a new vector $\boldsymbol{x}$, $(\boldsymbol{x}_i, m_{y_i})$ defines a piece of evidence concerning the output y of $\boldsymbol{x}$, which can be modeled by a FBS in $\mathcal{Y}$, m_i, whose focal elements are $F(m_i) = F(m_{y_i}) \cup \mathcal{Y}$. The belief structure m_i is defined as follows:

$$\left\{ \begin{array}{lcl} m_i(\tilde{y}_{ij}) & = & p_{ij}\phi_i(\boldsymbol{x}) \quad \forall j \in \{1, \ldots, J(i)\} \\ m_i(\mathcal{Y}) & = & 1 - \phi_i(\boldsymbol{x}) \\ m_i(A) & = & 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus F(m_i) \end{array} \right. \qquad (7)$$

Here again, $|F(m)|$, which can reach $\prod_{i=1}^{n} J(i)$, can be reduced by taking the k nearest neighbours of $\boldsymbol{z}$ in the training set. This model will be referred as **FBS3**.

3.3. Approach Based on Prototypes

For the sake of simplicity, we will assume in this subsection the y_i to be real numbers. If the training set is large, we can summarize the information given by all the vectors by prototypes. Let $C = \{\boldsymbol{c}_i \in \mathcal{X}, i \in \{1, \ldots, I\}\}$, and $E = \{\boldsymbol{e}_j \in \mathcal{Y}, j \in \{1, \ldots, J\}\}$ be two sets of I input and J output prototypes respectively. The prototypes may have been created by means of any clustering method such as the fuzzy c-means algorithms separately or not. If we associate every $\boldsymbol{x} \in \mathcal{X}$ to its nearest prototype in C according to a given distance, the set $\mathcal{X}$ can be partitioned in I classes $(C_i)_{i=1}^{I}$, characterised by the *estimated* cluster centers $\boldsymbol{c}_i$ and the *estimated* cluster dispersion matrix $\Sigma_i^x = \frac{1}{N} \sum_{k=1}^{N} u_{ik} (\boldsymbol{x}_k - \boldsymbol{c}_i)^T (\boldsymbol{x}_k - \boldsymbol{c}_i)$, if we denote by u_{ik} the membership degree of $\boldsymbol{x}_k$ to class C_i. In the same manner, $\mathcal{Y}$ is partitioned in J classes $(E_j)_{j=1}^{J}$. Let us assume that the $(\boldsymbol{x}_k, y_k)$ are a sample coming from N independant random vectors $(\boldsymbol{X}_k, Y_k)$ of a probabilised space $((\mathcal{X}, \mathcal{Y}), \mathcal{A}, \mathcal{P}_{\mathcal{X}\mathcal{Y}})$ and let us denote by p_{ij} the probability $\mathcal{P}_{\mathcal{X}|\mathcal{Y}}(Y \in E_j | X \in C_i)$. The output prototypes e_j can be "fuzzified" as $\tilde{e}_j \in \mathcal{F}(\mathcal{Y})$ by taking the estimated standard deviation σ_j^y of E_j into account. For example, since the variables are supposed to be real and continuous, the $\tilde{e}_j$ can be Gaussian fuzzy numbers, denoted (e_j, σ_j^y) and defined by: $\tilde{e}_j(u) = \exp\left\{-\frac{1}{2}(\frac{u - e_j}{\sigma_j^y})^2\right\}, \forall u \in \mathcal{Y} \subset I\!R$.

For vector $\boldsymbol{x}$, each $\boldsymbol{c}_i$ provides a piece of evidence concerning the output y, modeled by a FBS m_i similar to equations (4) (6) (7):

$$\begin{cases} m_i(\tilde{e}_j) &= p_{ij}\phi(d(\boldsymbol{x},\boldsymbol{c}_i) \quad \forall j \in \{1,\ldots,J\} \\ m_i(\mathcal{Y}) &= 1-\phi(d(\boldsymbol{x},\boldsymbol{c}_i) \\ m_i(A) &= 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus F(m_i) \end{cases} \tag{8}$$

Here, $|F(m)| \leq 2^J$. The number of focal elements is much smaller than in FBS1. This model is refered as **FBS4**.

Remark: if we "fuzzify" each component c_{ij} of $\boldsymbol{c}_i$ by the Gaussian fuzzy numbers $\tilde{c}_{ij} = (c_{ij}, \sigma^x_{ij}) \in \mathcal{X}_j$, and if we denote by $\tilde{\boldsymbol{c}}_i$ the Cartesian product of the $\tilde{c}_{ij}$, we have the following equality for every $x \in \mathcal{X}$: $\tilde{\boldsymbol{c}}_i(\boldsymbol{x}) = \phi(d(\boldsymbol{x},\boldsymbol{c}_i)$ if ϕ is the *exponential* function.

Let us now estimate the probability matrix $P = (p_{ij})$. If $\pi_{ij} = P_{\mathcal{X}\mathcal{Y}}((X,Y) \in C_i \cap E_j)$ and N_{ij} is the random variable counting the pairs of (x_k, y_k) owing to $C_i \times E_j$, by definition $N_{ij} \sim \mathcal{B}(n, \pi_{ij})$. The Strong Law of Large Numbers justifies theoretically the following estimation,

$$\widehat{p}_{ij} = \widehat{P_{Y|X}}(E_j|C_i) = \frac{\widehat{P_{\mathcal{X}\mathcal{Y}}}(E_j \cap C_i)}{\widehat{P_{\mathcal{X}}}(C_i)} = \frac{n_{ij}}{\sum_j n_{ij}}$$

since $\widehat{p}_{ij} \overset{n\to+\infty}{\longrightarrow} \mathcal{P}_{\mathcal{X}|\mathcal{Y}}(Y \in E_j|X \in C_i)$ almost surely.

In a fuzzy probability approach, Yager [12] proposed to take the *normalised* membership degree μ^k_{ij} of each $(\boldsymbol{x}_k, y_k)$ to $\tilde{\boldsymbol{c}}_i \times \tilde{e}_j$, as a natural extension of the preceding frequentist approach to fuzzy sets.

4. Fuzzy System Point of View

4.1. Fuzzy Systems with Probabilistic Uncertainties

Here, we recall briefly the fuzzy model proposed by Yager in [13], that can integrate probabilistic uncertainties in the output. Then we compare it to our model.

A Linguistic Model describes the multiple input, single output system by means of a set of IF-THEN rules with vague predicates. Let $\boldsymbol{x} = x_1,\ldots,x_r$ be the r-dimensional input variable of the base sets $\mathcal{X} = \mathcal{X}_1,\ldots,\mathcal{X}_r$, and y, the output variable taking values in $\mathcal{Y}$. Each input variable x_j is supposed to be described by one of the linguistic values $(B_{ij})_{i=1}^I$ of the base set $\mathcal{X}_j$, while $(D_j)_{j=1}^J$ denote the linguistic values that can be allotted to the output variable. Usually, these values are represented by fuzzy sets of the corresponding base sets. The general model of a fuzzy system is based on a collection of $I * J$ linguistic rules r_{ij} between the variables :

$$r_{ij}: \quad \text{IF} \quad \boldsymbol{x} \text{ is } B_i \quad \text{THEN} \quad y \text{ is } D_j \quad (p_{ij}), \tag{9}$$

where B_i is the cartesian product of the individual fuzzy sets : $B_i = B_{i1} \times \ldots \times B_{iJ}$. Associated with each rule is a rule confidence denoted by $p_{ij} \in [0,1]$. If we impose $\sum_{j=1}^J p_{ij} = 1$, we can introduce the FBS m_{y_i}, with focal elements $(D_j)_{j=1}^J$ such that $m_{y_i}(D_j) = p_{ij}$ and the rule base r_{ij} becomes [13]:

$$R_i: \quad \text{IF} \quad \boldsymbol{x} \text{ is } B_i \quad \text{THEN} \quad y \text{ is } m_{y_i} \tag{10}$$

Let $\wedge$ and T being any t-norm. The classical Mamdani reasoning method can be extended as follows:

1. compute the degree of firing τ_i of each rule R_i: $\tau_i \triangleq \wedge_{j=1}^r B_{ij}(x_j)$
2. compute the output m_i, a FBS on $\mathcal{Y}$ induced by each rule R_i, whose focal elements are fuzzy subsets on Y, such that $F_{ij}(y) \triangleq T(\tau_i, D_j(y))$ and $m_i(F_{ij}) = m_{y_i}(D_j)$
3. combine the rules with the union operator: we obtain a FBS m on $\mathcal{Y}$, such that $m = \cup_{i=1}^I m_i$

A set $E \in F(m)$ is on the form $E = \cup_{i=1}^I F_{ij(i)}$, where $F_{ij(i)} \in F(m_i)$, with the weight $m(E) = \prod_{i=1}^I m_i(F_{ij(i)})$

If we want to obtain a singleton output $\hat{y}$ from the FBS m, one has to define a defuzzification procedure. The taken value is the weighted sum of the defuzzified value y^*_E of each E owing to $F(m)$, say,

$$\hat{y} = \sum_{E \in F(m)} m(E) y^*_E$$

Remark: Another approach is to calculate the pignistic probability $BetP$ on $\mathcal{Y}$ and take the expectation: one can easily show that

$$\mathbb{E}_{BetP}(Y) = \hat{y}.$$

As the m_i and their combination are unnormalised in general, we can use the *Smooth Normalisation Procedure* of m to define a normalised fuzzy output m^* which produces a slightly different defuzzified value.

4.2. Analogy to our Model

If we modify the preceding inference system by taking directly the confidence rule into account for a given $\boldsymbol{x}$, we can model that by means of the "discounting" procedure. In the first step, we transform the FBS m_{y_i} in m'_i that can be denoted $m_{y_i}^{\tau_i(\boldsymbol{x})}$, as follows: $F(m'_i) = \{D_1,\ldots,D_J,\mathcal{Y}\}$ and $m'_i(D_j) = \tau_i(\boldsymbol{x})p_{ij}$, for all i,j and the reference set becomes a new focal element such that: $m'_i(\mathcal{Y}) = 1 - \tau_i$. The aggregation rule is performed with a conjunctive operator, and not a disjunctive one as in Yager's inference method, in order to control the mass allocated to the whole space $\mathcal{Y}$. Then the final FBS is $m' = \cap_{i=1}^{I} m'_i$.

Under certain assumptions (ϕ is the exponentional function, the B_{ij} are Gaussian membership functions and the fuzzy classes are the same in the two cases, and $\wedge$ is the product operator), this inference system corresponds to the model FBS4 (equation 8).

This inference method can be justified as follows. In Yager's inference method, the initial focal elements D_j are transformed by using a multiplication factor and become unnormalised as their height is $\tau_i(\boldsymbol{x})$, but their weights are unchanged. If we re-normalise these focal elements by using the SNP method, we of course reobtain the D_j, but their weight becomes $\tau_i(\boldsymbol{x})p_{ij}$. The remaining weight is allocated to the whole space. This procedure corresponds exactly to our inference method.

The main advantages of our method are the following: first of all, our model can measure global confidence in the output. Secondly, the maximum number of focal elements, in $FBS4$, even if it can be still important, is much less than in Yager's model (2^n). In FBS4, a focal element F of m' is necessary on the form: $\cap_{j \in K \subset \{1,\ldots,J\}} D_j$ or $F = \mathcal{Y}$. The maximum number of focal elements is then less than 2^J.

5. Results

In order to demonstrate the validity and the predictive capabilities of our method, we present the results for several simulations. The following performance indices are used. In the case of crisp data, we use the mean squared error (MSE):

$$E = \frac{1}{2n} \sum_{k=1}^{n} (y_k - \hat{y}_k)^2$$

We use two uncertainty measures of a FBS m defined in fuzzy evidence theory (cf. [6]): a measure of nonspecificity

$$N(m) = \sum_{F \in F(m)} m(F) \log_2 |F|, \qquad (11)$$

which represents set imprecision, and a measure of discord, called *strife*:

$$S(m) = -\sum_{F \in F(m)} m(F) \log_2 \sum_{G \in \mathcal{F}} m(G) \frac{|F \cap G|}{|G|}. \qquad (12)$$

Example 1: We have compared our results with classical regression methods, such as a Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), Gaussian spline (GS) and a Kernel (Nadaraya-Watson) ones, and the Yager model in terms of MSE for the following function :

$$y = 0.25 \cos(\frac{1}{\exp(-x)})$$

	GS	RBF	MLP	NW	Yager	FBS2
MSE	.07	.13	.11	.10	.10	.12

Table 1: Mean squared error for various models

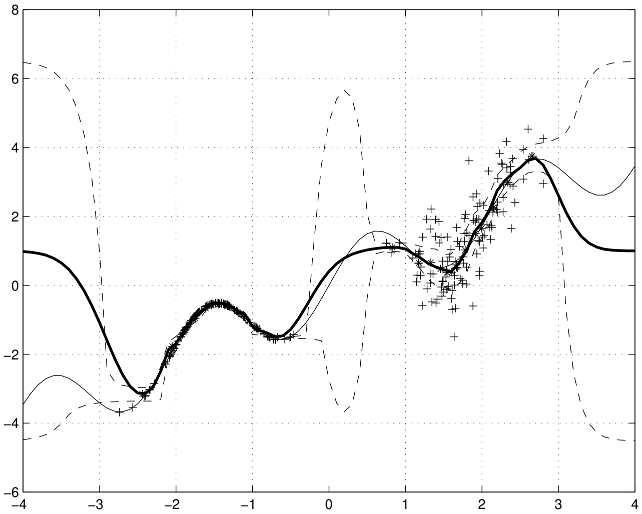

Figure 2: **Example 2.**: real data (-); estimates (**-**); interquartil interval of BetP(y) (- -); training set (*)

$\in [0, 4]$, with a sample size $N = 36$. We used the FBS2 version of our method. The results, given in Table 1, show that the performance index of FBS2 is analoguous to the classical ones.

Example 2: In this regression problem (cf.[4]) $x \in I\!R$ is taken from a mixture of two Gaussian distributions: $f(x) \sim 0.5\mathcal{N}(-1.5, 0.5) + 0.5\mathcal{N}(1.75, 0.5)$ and

$$y = \sin(3x) + x + \epsilon(x),$$

where $\epsilon(x) \sim \mathcal{N}(0, 0.01)$ if $x \leq 0$ and $\epsilon(x) \sim \mathcal{N}(0, 1)$ if $x > 0$. Here, $N = 200$. We used the FBS4 version with $I = J = 25$ prototypes and took the $k = 3$ nearest neighbours for combination. Here, the output prototypes e_i are calculated as output values of the c_i by means of a Nadaraya-Watson regression method using the training set. The interquartil interval from the calculated pignistic distribution $BetP$ is given (cf. Figure 2). The regions of low density are well reflected by a large value of nonspecificity and ignorance $m(\mathcal{Y})$, a very large interquartil interval. On contrary, Yager's model does not reflect well this uncertainty, since the nonspecifity does not increase in those regions. As expected, but contrary to the results in [4], the regions of large variance are characterised by a larger confidence interval and are less specific than regions of small variance (cf. Figure 3).

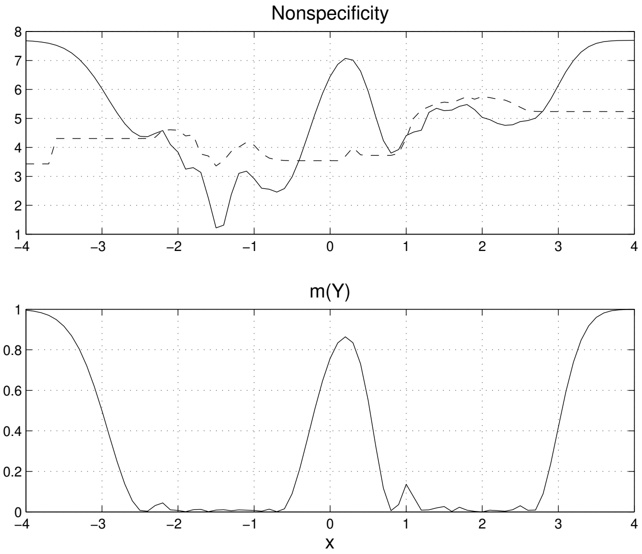

Figure 3: **Example 2.**: nonspecificity with FBS4 (-) and with Yager (- -) and mass of the whole space m(Y) with FBS4

6. Conclusion

In this paper, we focused on a regression method based on the fuzzy evidence theory. We introduced a means to take into account different kinds of output data, such as intervals, fuzzy numbers, or more generally fuzzy belief structures. We showed that our model could be seen as a particular fuzzy system and we presented the analogies with the model proposed by Yager [13], taking output probabilistic uncertainties into account. The drawback of evidence theory based methods is that it is time consuming due the combination of belief structures. We could have minimized the problem by making a clustering of the training set or taking the k "nearest rules or belief structures", but the question needs to be develop. In a next paper, we will focus on that problem of simplification of belief structures, or fuzzy belief structures.

References

[1] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS-38:325–339, 1967.

[2] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

[3] T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7):1095–1107, 1997.

[4] T. Denœux. Function approximation in the framework of evidence theory: A connectionist approach. In *Proceedings of the 1997 International Conference on Neural Networks*, pages 199–203, Houston, 1997.

[5] P. Diamond and H. Tanaka. Fuzzy regression analysis. In R. Slowinski, editor, *Fuzzy sets in decision analysis, operations research and statistics*, pages 349–387. Kluwer academic Publishers, Norwell, 1998.

[6] G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: Theory and applications.* Prentice Hall, Upper Saddle River, N.J., 1995.

[7] Q. Huang L. Huang, B-L Zhang. Robust interval regression analysis using neural networks. *Fuzzy sets and systems*, 97:337–347, 1998.

[8] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[9] P. Smets. The degree of belief in a fuzzy event. *Information sciences*, 28:45–62, 1982.

[10] H. Tanaka, S. Uejima, and K. Asai. Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics*, 12:903–907, 1982.

[11] R. R. Yager, M. Fedrizzi, and J. Kacprzyk. *Advances in the Dempster-Shafer theory of evidence.* John Wiley and sons, 1994.

[12] R. R. Yager and D. P. Filev. *Essential of fuzzy modeling and control.* John Wiley and sons, 1994.

[13] R. R. Yager and D. P. Filev. Including probabilistic uncertainty in fuzzy logic controller modelling using Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25:1221–1230, 1995.

[14] J. Yen. Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20:559–569, 1990.