Handling Different Forms of Uncertainty in Regression Analysis: a Fuzzy Belief Structure Approach

Simon Petit-Renaud and Thierry Denœux

Université de Technologie de Compiègne,
HeuDiaSyc-UMR C.N.R.S. 6599, BP 20529, 60205 Compiègne, France
spetit, tdenoeux@hds.utc.fr

Abstract. We propose a new approach to functional regression based on fuzzy evidence theory. This method uses a training set for computing a fuzzy belief structure which quantifies different types of uncertainties, such as nonspecificity, conflict, or low density of input data. The method can cope with a very large class of training data, such as numbers, intervals, fuzzy numbers, and, more generally, fuzzy belief structures. In order to limit calculations and improve output readability, we propose a belief structure simplification method, based on similarity between fuzzy sets and significance of these sets. The proposed model can provide predictions in several different forms, such as numerical, probabilistic, fuzzy or as a fuzzy belief structure. To validate the model, we propose two simulations and compare the results with classical or fuzzy regression methods.

1 Introduction

The problem of modeling a process from observed data has been is an important topic in several disciplines ranging from nonlinear regression to machine learning and system identification. In classical regression analysis, this can often be seen as a problem of global or local functional approximation. Examples of global functional estimators are generalised linear models, smoothing splines [9] and multi-layer perceptrons. In local models, including Generalised Radial Basis Functions, Kernel regressors, regression trees, or memory-based methods such as the "lazy" method [2], model description is divided in several parts of the input space.

Although statistical regression analysis is one of the most widely used techniques to model relationships between variables, it is not well adapted to all kinds of uncertainties which can be encountered in real applications. Evidence theory [3, 14], fuzzy set theory, and fuzzy evidence theory [15, 21] each highlight different forms of uncertainties [12], and, as such, provide valuable alternatives to classical probability theory for data analysis. Several approaches such as fuzzy or interval linear regression [7], fuzzy [20, 17] and neuro-fuzzy inference systems

[11] aim to model vague or fuzzy phenomena in functional estimation. A neural network approach based on evidence theory has also been proposed as a functional approximation method [6].

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 taking into account probabilistic uncertainty of target values is also treated. The formalism of belief structures is used to assess and quantify uncertainty in the output data. A procedure is proposed for computing a belief structure from a training set, adapting a general principle formerly introduced in a discrimination context [5]. The concept of pignistic probabilities allows to transform degrees of belief into probabilistic confidence values that may be attached to predictions. We show how our model can integrate different kinds of uncertainties such as imprecision, discord and ignorance.

To prevent an exponential increase of focal elements during the combination process, a simplification method is also proposed, based on two criteria: similarity between fuzzy sets and weight of these sets. The similar fuzzy sets are aggregated and the "least representative" fuzzy sets are eliminated.

The paper is organized as follows. In Section 2, we briefly recall the basics of evidence theory and its fuzzy extension. In Section 3, we describe the proposed model and a variant using prototypes. The simplification algorithm is presented in Section 4. Finally, we give some numerical results in Section 5 and conclude in Section 6.

2 Fuzzy Belief Structures

In this section, we briefly introduce some of the main ideas behind the Dempster-Shafer theory of evidence [3,14,19,16]. Let Ω be a finite set, and let $S(\Omega)$ be the set of all 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* defined, respectively, as: $\mathrm{bel}(A) = \sum_{\emptyset\neq B\subseteq A} m(B)$ and $\mathrm{pl}(A) = \sum_{B\cap A=\emptyset} m(B) = \mathrm{bel}(\Omega) - \mathrm{bel}(\bar{A})$. The quantity $\mathrm{bel}(A)$ is interpreted as the *total* belief committed to A and $\mathrm{pl}(A)$ as the belief that *might* be committed to A, if further information became available. They can be regarded, respectively, as lower and upper bounds of a set of compatible probabilities. One of these, the *pignistic* probability, is particularly useful for decision making [16]; it is defined as:

$$\mathrm{BetP}(\omega) = \sum_{A\subset\Omega} \frac{m(A)}{|A|}\delta_A(\omega)$$

where δ is the Kronecker symbol and $|A|$ denotes the cardinality of A. Evidence theory can easily be generalized to continuous spaces provided the number of focal elements $|F(m)|$ remains finite [8] (further generalizations are possible, but they are outside the scope of this paper).

One of the most important operations in 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 $m = m_1 \nabla m_2$, may be defined as: $m(A) = \sum_{B \nabla C = A} m_1(B) m_2(C)$, for all $A \in \Omega$. This operation may produce a subnormal belief structure, i.e. one may have $m(\emptyset) > 0$, particularly in the case of the conjunctive sum defined originally by Dempster, where $\nabla = \cap$. The Dempster normalization procedure converts a subnormal belief structure m into a normal one m^* defined as follows: $m^*(A) = \frac{m(A)}{1-m(\emptyset)}$ for $A \neq \emptyset$ and $m^*(\emptyset) = 0$.

Fuzzy extensions of evidence theory have been proposed by different authors [15, 21]. The basic idea is to allow the focal elements of a belief structure to be fuzzy sets. 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 $\mathcal{F}(\Omega)$ denotes 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 subset $F(m)$ of $\mathcal{F}(\Omega)$. 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 [20] proposed a "smooth normalization procedure" (SNP) for converting a subnormal fuzzy belief structure into a normal one. This method generalizes both fuzzy set normalization and Dempster's normalization of crisp belief structures. Assume m to be 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 generalized. 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 concept of pignistic probability can also easily be extended as: $\mathrm{BetP}(\omega) = \sum_{A \in F(m)} \frac{m(A)}{|A|} A(\omega)$.

3 Application to Regression Analysis

3.1 The Proposed Model

In this paper, we propose to apply the tools of fuzzy evidence theory to regression analysis. This approach extends the method formerly proposed by Denœux [4, 5] in the context of pattern 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 $(\boldsymbol{x}_i, \tilde{y}_i)_{i=1,\ldots,n}$ be a training set of input-output data vectors belonging to $(\mathcal{X}_1 \times \ldots \times \mathcal{X}_r) \times \mathcal{F}(\mathcal{Y})$, where each $\mathcal{X}_j$ and $\mathcal{Y}$ are supposed to be continuous real spaces. The output data can thus be either real numbers, intervals, fuzzy numbers or any fuzzy quantities. Let us note $\mathcal{X}$ the Cartesian space $\mathcal{X}_1 \times \ldots \times \mathcal{X}_r$.

Let $\boldsymbol{x}$ denote an arbitrary input vector, and let y be the corresponding unknown output value. 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 expect the corresponding output y to be similar to $\tilde{y}_i$. This may be modeled by assigning to $\tilde{y}_i$ a certain fraction of a unit amount of belief, depending on the distance between $\boldsymbol{x}$ and $\boldsymbol{x}_i$. Since no other hypothesis is made, the remaining mass of belief should be allocated to the whole set $\mathcal{Y}$. We thus obtain a FBS m_i on $\mathcal{Y}$, with focal elements $F(m_i) = \{\tilde{y}_i, \mathcal{Y}\}$, defined as:

$$\begin{cases} m_i(\tilde{y}_i) = \phi[d(\boldsymbol{x}, \boldsymbol{x}_i)] \\ m_i(\mathcal{Y}) = 1 - \phi[d(\boldsymbol{x}, \boldsymbol{x}_i)] \\ m_i(A) = 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \backslash F(m_i), \end{cases} \tag{3}$$

where ϕ is a decreasing function such as $\phi(0) = 1$ 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, and d is a Euclidean distance induced by some symmetric positive definite 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, several kinds of uncertainties are clearly identified: fuzziness, imprecision or nonspecificity (related to the cardinality of $\tilde{y}_i$) and 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 generalization of the conjunctive rule of combination to FBS. Then, the final FBS is: $m = \cap_{i=1}^{n} m_i$. To decrease the computational load, the combination can alternatively be restricted to the k nearest neighbors of $\boldsymbol{x}$ in the training set: $\{\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(k)}\}$. The final FBS then becomes: $m = \cap_{i=1}^{k} m_{(i)}$.

The pignistic probability function induced by m may then be defined as

$$\mathrm{BetP}(y) = \sum_{A \in F(m)} \frac{m(A)}{|A|} A(y),$$

where $|A| = \int_{\mathcal{Y}} A(y)dy$, since $\mathcal{Y}$ is a continuous space. A point estimate of y can be obtained by taking the expected value of the output:

$$\mathbb{E}(Y) = \int_{\mathcal{Y}} u \mathrm{BetP}(u) du = \sum_{A \in F(m)} \frac{m(A)}{|A|} \int_{\mathcal{Y}} u A(u) du.$$

Let us note the center of area of the set A, $y_A^* \triangleq \frac{\int_{\mathcal{Y}} y A(y) dy}{\int_{\mathcal{Y}} A(y) dy}$. Then, $\mathbb{E}(Y) = \sum_{A \in F(m)} m(A) y_A^*$.

In the following, the model defined by Eq. (3) will be referred to as FBS1.

As remarked by Smets (personal communication), if the $\tilde{y}_i$ are crisp numbers, then m_i and m become standard belief structures with only $n+1$ crisp focal elements and the final belief structure can be easily expressed as (FBS2 model):

$$\begin{cases} m(\{y_i\}) = \frac{1}{K}\phi_i(\boldsymbol{x})\prod_{j\neq i}[1-\phi_j(\boldsymbol{x})] & \forall i \in \{1,\ldots,n\} \\ m(\mathcal{Y}) \quad = \frac{1}{K}\prod_{i=1}^{m}[1-\phi_i(\boldsymbol{x})] \\ m(A) \quad = 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus F(m), \end{cases} \tag{4}$$

where $K = \prod_{i=1}^{n}[1-\phi_i(\boldsymbol{x})] + \sum_{i=1}^{n}\phi_i(\boldsymbol{x})\prod_{j\neq i}[1-\phi_j(\boldsymbol{x})]$ is the normalization factor, and $F(m) = \{\{y_1\},\ldots,\{y_n\},\mathcal{Y}\}$.

3.2 Generalization to Conflicting Outputs

We propose to generalize Eq. (3) 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 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).

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 m_i on $\mathcal{Y}$, whose focal elements are $F(m_i) = F(m_{y_i}) \cup \mathcal{Y}$. The belief structure m_i is defined as:

$$\begin{cases} m_i(\tilde{y}_{ij}) = p_{ij}\phi_i(\boldsymbol{x}) & \forall j \in \{1,\ldots,J(i)\} \\ m_i(\mathcal{Y}) \quad = 1-\phi_i(\boldsymbol{x}) \\ m_i(A) \quad = 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus F(m_i). \end{cases} \tag{5}$$

Here again, the number of elements in $F(m)$, which can reach in the worst case $\prod_{i=1}^{n} J(i)$, can be reduced by considering the k nearest neighbors of $\boldsymbol{x}$ in the training set. This model will be referred to as FBS3.

3.3 Approach based on Prototypes

For the sake of simplicity, we will assume in this section the y_i to be real numbers. If the training set is large, it can be summarized by a smaller number of reference vectors or prototypes. Let $C = \{\boldsymbol{c}_i \in \mathcal{X}, i \in \{1,\ldots,I\}\}$, and $E = \{e_j \in \mathcal{Y}, j \in \{1,\ldots,J\}\}$ be two sets of I input and J output prototypes, respectively. These prototypes may have been created by means of any clustering method such as the c-means or 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 $(C_i)_{i=1}^{I}$, characterized by cluster centers $\boldsymbol{c}_i$ and scatter matrices $\Sigma_i^x = \frac{1}{N}\sum_{k=1}^{N} u_{ik}(\boldsymbol{x}_k - \boldsymbol{c}_i)^T(\boldsymbol{x}_k - \boldsymbol{c}_i)$, where u_{ik} denotes the membership degree of $\boldsymbol{x}_k$ to class C_i. In the same way, $\mathcal{Y}$ 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 independent random vectors $(\boldsymbol{X}_k, Y_k)$ on a probability space $((\mathcal{X},\mathcal{Y}),\mathcal{A},\mathcal{P}_{\mathcal{X}\mathcal{Y}})$, and let us denote by p_{ij} the probability $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}(Y \in E_j | X \in C_i)$. 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 assumed to be Gaussian fuzzy numbers, denoted (e_j, σ_j^y) and defined by: $\tilde{e}_j(u) = \exp\left\{-\frac{1}{2}(\frac{u-e_j}{\sigma_j^y})^2\right\}, \forall u \in \mathcal{Y} \subset \mathbb{R}$.

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 Eqs. (3-5):

$$\begin{cases} m_i(\tilde{e}_j) = p_{ij}\phi[d(\boldsymbol{x}, \boldsymbol{c}_i)] \quad \forall j \in \{1, \ldots, J\} \\ m_i(\mathcal{Y}) = 1 - \phi[d(\boldsymbol{x}, \boldsymbol{c}_i)] \\ m_i(A) = 0 \quad \forall A \in \mathcal{F}(\mathcal{Y}) \setminus \{\tilde{e}_1, \ldots \tilde{e}_J, \mathcal{Y}\}. \end{cases} \tag{6}$$

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

Remark that, if we "fuzzify" each component c_{ij} of $\boldsymbol{c}_i$ by Gaussian fuzzy numbers $\tilde{c}_{ij} = (c_{ij}, \sigma_{ij}^x) \in \mathcal{X}_j$, 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) belonging 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:

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

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

Another problem that remains to be solved in this approach is the determination of the optimal number of prototypes. The problem is essentially the same as that of fuzzy system identification, for which numerous methods for finding the optimal number of clusters have been suggested in the literature [10]. Various criteria have been proposed, based on empirical fuzzy covariance matrices

$$J_i = \sum_{k=1}^{n} \frac{u_{ik}(\boldsymbol{x}_k - \boldsymbol{c}_i)(\boldsymbol{x}_k - \boldsymbol{c}_i)^t}{\sum u_{ik}}$$

or within-classes scatter measures:

$$SSW = \sum_{i=1}^{m} \sum_{k=1}^{n} u_{ik} d^2(\boldsymbol{x}_k, \boldsymbol{c}_i).$$

In particular, a simple but efficient approach is to consider the classical "explained inertia ratio" SSW/SST, where $SST = \sum_{k=1}^{n} d^2(\boldsymbol{x}_k, \bar{\boldsymbol{x}})$ and $\bar{\boldsymbol{x}}$ is the global mean vector. This ratio is a decreasing function of I, a suitable value $^I*$ may be chosen by imposing an acceptable minimal threshold.

Once an input prototype $\boldsymbol{c}_i$ has been determined, we propose to induce the corresponding output prototype as the Nadaraya-Watson Gaussian kernel regressor evaluated at $\boldsymbol{c}_i$:

$$e_i = \frac{\sum_{k=1}^{n} y_k \exp[-(\boldsymbol{c}_i - \boldsymbol{x}_k)^T \Lambda (\boldsymbol{c}_i - \boldsymbol{x}_k)]}{\sum_{k=1}^{n} \exp[-(\boldsymbol{c}_i - \boldsymbol{x}_k)^T \Lambda (\boldsymbol{c}_i - \boldsymbol{x}_k)]},$$

where Λ is a symmetric definite positive matrix to be estimated. For the sake of simplicity, Λ can be assumed to be diagonal: $\Lambda = (\lambda_j)_{j=1}^r$. Optimal values of the λ_j can be obtained by minimizing the classical cross validation criterion for a given test set.

4 Simplification Procedure

A well-known drawback of evidence theory based methods is that they are generally time consuming due the exponential increase of focal elements encountered when combining large numbers of belief structures. We recall that the number of focal elements can reach $|\mathcal{F}(m)| = 2^J$ with FBS4. This implies different kinds of problems. Practically, the computation of the outputs becomes too time-consuming. Moreover, even if the results can be computed, some of individual focal elements of the final FBS may not have a clear interpretation. Some of them could represent approximately the same output, and the resulting information would be so complex that it would become difficult to exploit.

So, our problem is to control the complexity of the output structure. For that purpose, we propose to define a simplification procedure for any (crisp or fuzzy) belief structure. Some authors have already proposed various methods in case of classical belief structures [1]. Here we propose two ways of simplifying a FBS m: elimination of least representative focal elements and aggregation of similar focal elements. During combination, for a chosen threshold w, if $m(F_i) < w$, F_i is eliminated and the FBS m is re-normalized according by. In the aggregation phase, two steps can be defined: choosing which focal elements have to be replaced and replacing two focal elements by a new one. The first question requires the definition of a measure of *similarity* between fuzzy sets. The second one is an information fusion problem. In the following, we will focus on the aggregation phase, for which a hierarchical clustering approach is proposed.

Let m be the initial FBS with focal elements $\{F_1, \dots, F_n\}$. A new FBS m' is obtained by replacing the 2 nearest focal elements F_i and F_j of m by a new one F' and keeping the other focal elements. Then $F(m') = (F(m) \cup F') \backslash (F_i \cup F_j)$.

Among similarity measures of fuzzy sets that have been defined in literature [13, 22], we use the following set-based measure $S(A, B) = \frac{|A \cap B|}{|A \cup B|}$. Then F_i and F_j are chosen such as $S(F_i, F_j) = \max_{k,l} S(F_k, F_l)$. Different means of aggregating the similar focal elements can be used, depending on the binary operator ∇: $F' = F_i \nabla F_j$ with $m'(F') = m(F_i) + m(F_j)$. We can separate these operators in 3 classes: disjunctive, neutral and conjunctive. To keep as much information as possible, it seems preferable to take the union of the two sets $F' = F_i \cup F_j$ as the new focal element. Moreover, this choice leads to a cautious approximation, since m' is less informative than m in terms of the nonspecificity N, defined as $N(m) = \sum_{F \in F(m)} m(F) \log_2 |F|$. It is easy to verify that $N(m') \leq N(m)$, which proves that m' is less specific than m. However, another possibility is to take a more neutral operator by considering the weighted average of the sets, so that:

$$F'(\omega) = \frac{m(F_i)F_i(\omega) + m(F_j)F_j(\omega)}{m(F_i) + m(F_j)}$$

We then have an interesting analogy with Ward's method in hierarchical clustering [18].

This iterative aggregation procedure stops as soon as $\max_{k,l} S(F_k, F_l) < \delta$, where δ is a given threshold.

5 Results

In order to demonstrate the validity and predictive capabilities of our method, we present the results for several simulations. In the case of crisp data, prediction accuracy was measured by the mean squared error (MSE) $E = \frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2$. To describe the information content of the outputs, two classical evidential uncertainty measures [12] were used: nonspecificity, which represents set imprecision, and strife:

$$S(m) = -\sum_{A \in F(m)} m(A) \log_2 \sum_{B \in F(m)} m(B) \frac{|A \cap B|}{|B|}.$$

Example 1: In this regression problem [6], $x \in \mathbb{R}$ 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 neighbors 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 interquartile interval from the calculated pignistic distribution BetP is shown in Figure 1. These results are compared to a model based on fuzzy systems proposed by Yager in [20]. The regions of low density are well reflected by a large value of nonspecificity and ignorance $m(\mathcal{Y})$ and a very large interquartile interval. On the contrary, Yager's model does not reflect well this uncertainty, since the nonspecificity does not increase in those regions. As expected, and contrary to the results in [6], the regions of large variance are characterized by a larger confidence interval and a greater nonspecificity than regions of small variance.

In order to be sure that our method was efficient in terms of accuracy, we compared it with several well-known regression methods (cf. Figure 2). Table 1 shows the mean squared errors for two classical non parametric methods, the Smoothing Cubic Splines (CS) and the Nadaraya-Watson kernel regressor (NW), a memory-based method, called the "lazy" method [2], Yager's model and two versions or our method: FBS4 and the simplified one (FBS2). The MSE have been calculated only in the region of high density for two cases: the first one corresponds to the region of well specified data, with small variance ($x \in [-2.5, -0.7]$). The other one also includes the region with high variance ($x \in [-2.5, -0.7] \cup [0.8, 2.6]$). As shown in Table 1, the performances of our method (FBS2 or FBS4) are roughly equivalent to those of standard methods when applied to this classical problem. However, one can notice that the spline methods seem to be better adapted to the case of homogeneous data, in this example. However, this global method cannot cope with heterogeneous data, as our method seem to do.

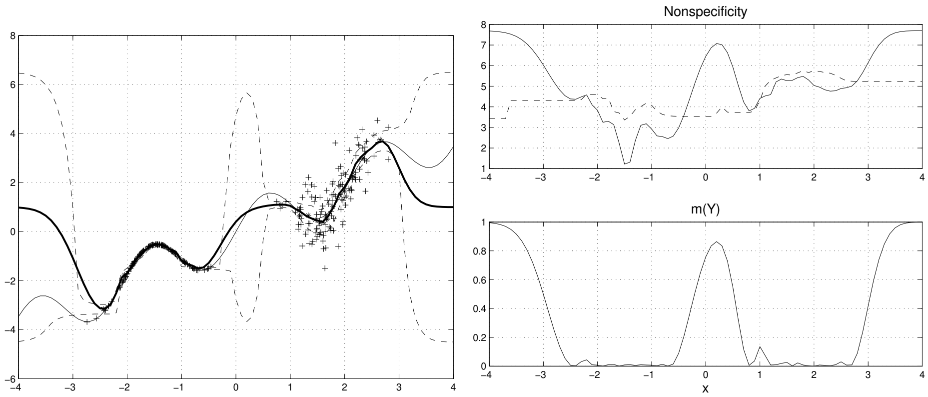

Fig. 1. Example 1. Left: real data (-); estimates (-); interquartile interval of BetP(y) (- -); training set (+) ; **Right**: nonspecificity with FBS4 (-) and with Yager's method (- -), and mass $m(\mathcal{Y})$ assigned to the whole space by FBS4.

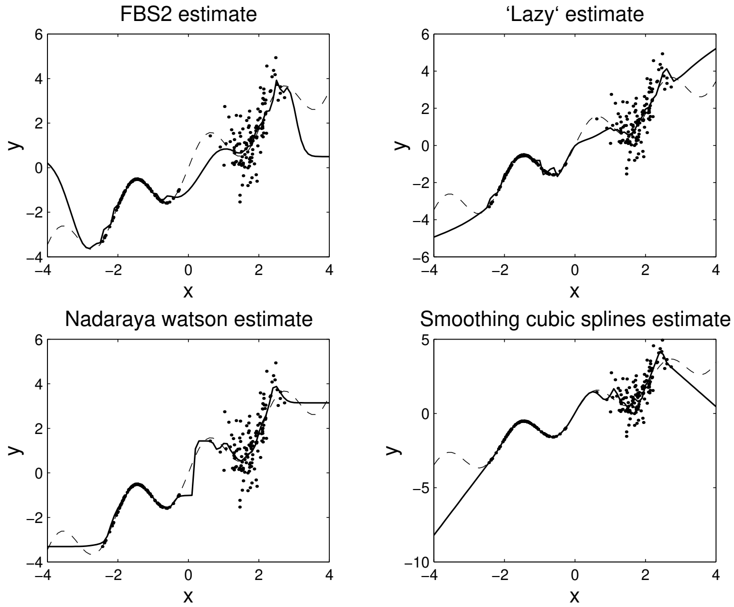

Fig. 2. Example 1.: training set (.); real data (–); punctual estimation for various models (-)

Regression method	CS	NW	"Lazy"	Yager	FBS2	FBS4
MSE 2	0.035	0.016	0.019	0.010	0.013	0.011
MSE 1	10^{-5}	0.003	0.013	0.008	0.011	0.010

Table 1. Mean squared error for various models: MSE 1 for $x \in [-2.5, -0.7]$, MSE 2 for $x \in [-2.5, -0.7] \cup [0.8, 2.6]$.

Example 2 : We consider a set of input-output data generated from the Monod equation $y = \frac{ax}{b+x} + \epsilon(x)$, $x \in [0, 2]$, where $\epsilon(x) \sim \mathcal{N}(0, 0.001)$ if $x \leq 1$ and $\epsilon(x) \sim \mathcal{N}(0, 0.01)$ if $x > 1$. If $x \in [0.5, 1.5[$, $(a, b) = (.35, .5)$, $(a, b) = (.35, .5)$ or $(.45, .45)$ with the same probability $p_{i1} = p_{i2} = .5$ and $(a, b) = (.35, .5)$ if $x \in [0, 2] \backslash [0.5, 1.5[$ (see Figure 3). We used the FBS4 version, with $I = 20$, $J = 30$ and $k = 3$ neighbors, and the prototypes were generated as in Example 1, using the two target functions, depending on the parameter values a and b. We used the "weighted average" version of our simplification method with different aggregation thresholds δ during combination. When there is no simplification, $\delta = 1$. We note that even in this extreme case, the number of focal elements $|F(m)|$ is much less than 2^{30} since most estimated probabilities $\widehat{p_{ij}}$ are equal to 0. In the other extreme case, $\delta = 0$ and all the focal elements are aggregated (i.e. $|F(m)| = 1$) reducing the FBS as a simple fuzzy set. Figure 4 shows that $|F(m)|$ increases in the regions of output discord and quickly decreases with δ. In Figures (3-4), we also show some results for a special case where $\delta = 0.5$. The 4 regions can be clearly characterized by the uncertainty measures. The regions (1-2) ($x \in [0, 1]$) of larger variance are characterized by a larger nonspecificity and the conflict in the output value is well represented by a larger *strife* in regions (2-3) ($x \in [0.5, 1.5]$). The pignistic probability calculated for 4 typical values ($x = 0.2; 0.6; 1.3; 1.6$) summaries well the four encountered situations.

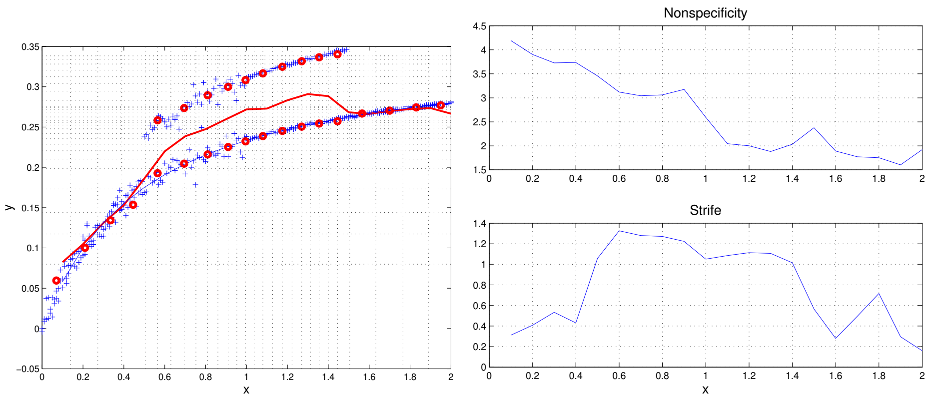

Fig. 3. Example 2. Left: training set (+); point estimation (-), prototypes (o) and classes (:) **Right**: nonspecificity and strife with FBS4 (-).

6 Conclusion

In this paper, a new approach to regression based on fuzzy evidence theory has been proposed. This approach allows to take into account different kinds of output data, such as intervals, fuzzy numbers, or more generally fuzzy belief

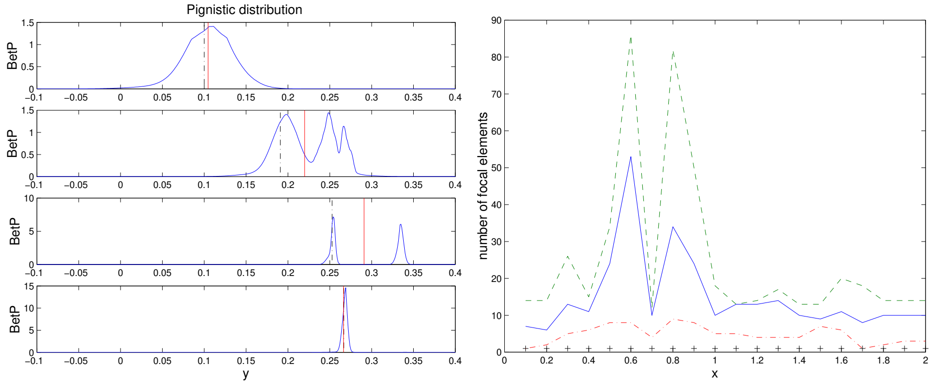

Fig. 4. Example 2. Left: pignistic distribution for 4 different inputs: x = 0.2 - 0.6 -1.3 -1.6 (-); real value (–); point estimation (-) **Right**: Number of focal elements for different values of δ: 1 (–), 0.5(-), 0.2 (-.-) and 0 (+).

structures. In terms of accuracy or point estimation, our method can also cope with classical regression problems as well as statistical methods. Moreover, since it can be viewed as a local functional approximation method, estimations can even be more accurate than those provided by global models. Thus, this method can be seen as a generalisation of regression analysis for non standard data. Two kinds of output data information are offered: a representation of uncertainty that can take several forms (imprecision, indetermination, ignorance) and point estimation accuracy. A well-known drawback of evidence theory based methods is that they are generally time consuming due the exponential increase of focal elements encountered when combining large numbers of belief structures. This problem can be minimized by clustering training data, or by considering only neighboring information in the learning set. In our method, we propose to approximate the belief structures by aggregating similar focal elements. This simplification method can be used for two different objectives. *During* combination, computing time is decreased and *after* combination, the interpretability of the results is enhanced.

References

1. M. Bauer. Approximations for decision making in the Dempster-Shafer theory. In *Uncertainty in Artificial Intelligence*, pages 339–344, Saarbrücken, 1996.
2. H. Bersini, G. Bontempi, and M. Birattari. Is readibility compatible with accuracy? from neuro-fuzzy to lazy learning. In *Fuzzy-Neuro Systems '98*, pages 10–25, Munich, 1998.
3. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, AMS-38:325–339, 1967.
4. 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.

5. T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. *Pattern Recognition*, 30(7):1095–1107, 1997.
6. 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.
7. 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.
8. J.W. Guan and D.A. Bell. A generalization of the Dempster-Shafer theory. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pages 10–25, Chambéry, France, 1993.
9. W. Härdle. *Applied nonparametric regression*. Cambridge University Press, 1990.
10. H. Hellendoorn and D. Driankov. *Fuzzy model identification: selected approaches*. Springer, 1997.
11. J.S.R Jang, C.T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice-Hall, 1997.
12. G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: Theory and applications*. Prentice Hall, Upper Sadddle River, N.J., 1995.
13. M. Setnes, R. Babuška, U. Kaymak, and H. R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man and Cybernetics - part B: Cybernetics*, 28:376–386, 1998.
14. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
15. P. Smets. The degree of belief in a fuzzy event. *Information sciences*, 28:45–62, 1982.
16. P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.
17. T. Takagi and M. Sugeno. Fuzzy identification of systems an its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15:116–132, 1985.
18. J. H. Ward. Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, 58:236–244, 1963.
19. R. R. Yager, M. Fedrizzi, and J. Kacprzyk. *Advances in the Dempster-Shafer theory of evidence*. John Wiley and sons, 1994.
20. 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.
21. J. Yen. Generalizing the Dempster-Shafer theory to fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 20:559–569, 1990.
22. R. Zwick, E. Carlstein, and D. V. Budescu. Measures of similarity among fuzzy concepts: a comparative analysis. *International Journal of Approximate Reasoning*, 1:221–242, 1987.