Adaptive Intuitionistic Fuzzy Inference Systems of Takagi-Sugeno Type for Regression Problems

Petr Hájek, Vladimír Olej

Institute of System Engineering and Informatics
Faculty of Economics and Administration, University of Pardubice, Studentská 84
532 10 Pardubice, Czech Republic
{petr.hajek, vladimir.olej}@upce.cz

Abstract. Recently, we have proposed a novel intuitionistic fuzzy inference system (IFIS) of Takagi-Sugeno type which is based on Atanassov's intuitionistic fuzzy sets (IF-sets). The IFIS represent a generalization of fuzzy inference systems (FISs). In this paper, we examine the possibilities of the adaptation of this class of systems. Gradient descent method and other special optimization methods are employed to adapt the parameters of the IFIS in regression problems. The empirical comparison of the systems is provided on several well-known benchmark and real-world datasets. The results show that by adding non-membership functions, the average errors may be significantly decreased compared to FISs.

Keywords: Intuitionistic fuzzy sets, intuitionistic fuzzy inference systems, adaptation, regression.

1 Introduction

Several generalizations of Zadeh's fuzzy set theory [1] have been developed to handle imprecision and uncertainty in a better way, e.g. IF-sets [2,3], *L*-fuzzy sets [4], interval-valued fuzzy sets (grey sets) [5], or interval-valued IF-sets [6], see [7] for a review. It was proven that interval-valued fuzzy set theory is equivalent to IF-set theory, which is equivalent to the vague set theory. In a similar manner, FISs have also been extended as interval-valued FISs [8], interval-valued type-2 FISs (IT2FISs) [9, 10] or IFIS [11, 12, 13].

The concept of IF-sets can be viewed as an alternative approach to define a fuzzy set in cases where available information is not sufficient for the definition of an imprecise concept by means of a conventional fuzzy set. This may hold true for many real world applications. For example, the uncertainty presented in air quality evaluation was modelled using hierarchical IFISs of Mamdani type in [13].

Recently, we presented a novel IFIS of Takagi-Sugeno type for time series prediction [12, 14]. It was shown that MIN *t*-norm (Gödel *t*-norm) provided the lowest error in terms of root mean squared error (RMSE) on testing data [14]. Furthermore, we developed IFIS with various levels of uncertainty represented by intuitionistic fuzzy index (IF-index). This class of systems presents a strong

possibility to express uncertainty (in the presence of imperfect facts and imprecise knowledge) and provides a good description of object attributes by means of membership functions $\mu_A(x)$ and non-membership functions $\nu_A(x)$.

In this paper, we compare several optimization algorithms used to adapt FISs and IFISs on selected artificial and real world datasets. We will also analyze the behaviour of the IFIS when the weights of the subsystems outputs (with membership functions $\mu_A(x)$ and non-membership functions $\nu_A(x)$) change. The rest of this paper is organized as follows. First, we briefly introduce FISs and IFIS of Takagi-Sugeno type. In this section, the difference between IFIS and related IT2FISs is mentioned. Next, we present the methods used for the adaptation of IFISs. In section 4, regression datasets are described. Finally, the results of experiments are provided and analyzed.

2 IFIS of Takagi-Sugeno Type

The FIS of Takagi-Sugeno type [15] is composed of the following steps. In the fuzzification process, the input variables are compared with the membership functions $\mu_A(x)$. Next, operators (AND, OR, NOT) are applied within the if-then rules. Thus, firing weight of each if-then rule is obtained. Further, the outputs of each if-then rule are generated. The output of each rule is a linear combination of input variables plus a constant term. The final output is the weighted average of each if-then rule's output.

Let $x \in X$, $X=\{x_1,x_2, \dots ,x_i, \dots ,x_n\}$ be input variables defined on the universes $X_1,X_2, \dots ,X_i, \dots ,X_n$ and let y be an output variable defined on the universe Y. Then FIS has n input variables and one output variable. Then the k-th if-then rule R^k, $k=1,2, \dots ,N$, in the FIS of Takagi-Sugeno type can be defined in following form

$$R_k\text{: if } x_1 \text{ is } A_{1,k} \text{ AND } x_2 \text{ is } A_{2,k} \text{ AND } \dots \text{ AND } x_i \text{ is } A_{i,k} \text{ AND } \dots \text{ AND } x_n \text{ is } A_{n,k} \text{ then } y_k=f(x_1,x_2, \dots ,x_n),\ i=1,2, \dots ,n, \quad (1)$$

where $A_{1,k},A_{2,k}, \dots ,A_{i,k}, \dots ,A_{n,k}$ represent fuzzy sets and $f(x_1,x_2, \dots ,x_n)$ can be a linear or polynomial function. In further considerations, we assume a linear function defined as $y_k=a_{1,k}x_1+a_{2,k}x_2+ \dots +a_{i,k}x_i+ \dots +a_{n,k}x_n+b$.

The FIS of Takagi-Sugeno type was designed in order to achieve higher computational effectiveness. This is possible as the defuzzification of outputs is not necessary. Its advantage lies also in involving the functional dependencies of output variable on input variables. The output level y_k of each the k-th if-then rule R_k is weighted by $w_k=\mu(x_1)$ AND $\mu(x_2)$ AND … AND $\mu(x_m)$. The final output y of the FIS Takagi-Sugeno type is the weighted average of all N if-then rule R_k outputs y_k, $k=1,2, \dots ,N$, computed as follows

$$y = \frac{\sum_{k=1}^{N} y_k w_k}{\sum_{k=1}^{N} w_k}. \quad (2)$$

Let a set X be a non-empty fixed set. An IF-set A in X is an object having the form $A=\{\langle x,\mu_A(x),\nu_A(x)\rangle|x\in X\}$, where the function $\mu_A:X\rightarrow[0,1]$ defines the degree of membership function $\mu_A(x)$ and the function $\nu_A:X\rightarrow[0,1]$ defines the degree of non-membership function $\nu_A(x)$, respectively, of the element $x\in X$ to the set A, which is a subset of X, and $A\subset X$, respectively; moreover for every $x\in X$, $0\leq\mu_A(x)+\nu_A(x)\leq 1$, $\forall x\in X$ must hold [2,3]. The amount $\pi_A(x)=1-(\mu_A(x)+\nu_A(x))$ is called the hesitation part, which may cater to either membership value or non-membership value, or both. For each IF-set in X, we will call $\pi_A(x)$ as the IF-index of the element x in set A. It is obvious that $0\leq\pi_A(x)\leq 1$ for each $x\in X$. The value denotes a measure of non-determinancy. The IF-indices $\pi_A(x)$ are such that the larger $\pi_A(x)$ the higher the hesitation margin of the decision maker. The k-th if-then rule R^{μ}_k in FIS$^{\mu}$ and R^{ν}_k in FIS$^{\nu}$ are defined as follows

$$\begin{aligned} R^{\mu}_k&\text{: if } x_1 \text{ is } A^{\mu}_{1,k} \text{ AND } x_2 \text{ is } A^{\mu}_{2,k} \text{ AND ... AND } x_i \text{ is } A^{\mu}_{i,k} \text{ AND ... AND } x_n \text{ is}\\ &A^{\mu}_{n,k} \text{ then } y^{\mu}_k = a^{\mu}_{1,k}x_1+a^{\mu}_{2,k}x_2+ \ldots +a^{\mu}_{i,k}x_i+ \ldots +a^{\mu}_{n,k}x_n+b^{\mu},\\ R^{\nu}_k&\text{: if } x_1 \text{ is } A^{\nu}_{1,k} \text{ AND } x_2 \text{ is } A^{\nu}_{2,k} \text{ AND ... AND } x_i \text{ is } A^{\nu}_{i,k} \text{ AND ... AND } x_n \text{ is}\\ &A^{\nu}_{n,k} \text{ then } y^{\nu}_k = a^{\nu}_{1,k}x_1+a^{\nu}_{2,k}x_2+ \ldots +a^{\nu}_{i,k}x_i+ \ldots +a^{\nu}_{n,k}x_n+b^{\nu}. \end{aligned} \quad (3)$$

The output y^{μ} of FIS$^{\mu}$ (the output y^{ν} of FIS$^{\nu}$) is defined in the same way as presented in equation (2) for firing weights w^{μ}_k and w^{ν}_k, respectively. The output y of the IFIS represents a combination of y^{μ} and y^{ν}

$$y=(1-\beta)y^{\mu}+\beta y^{\nu}, \quad (4)$$

where y is the output of the IFIS, y^{μ} is the output of the FIS$^{\mu}$ using the membership function $\mu_A(x)$, y^{ν} is the output of the FIS$^{\nu}$ using the non-membership function $\nu_A(x)$, and β is the weight of the output y^{ν}. In prior studies [13,14], it was assumed that $\beta=\pi$.

Like IFISs, IT2FISs also represent a generalization of the FISs. The membership degree of an interval type-2 fuzzy set is an interval (also known as the footprint of uncertainty) bounded from the above and below by two type-1 membership functions, $\bar{A}(x)$ (upper membership function) and $\underline{A}(x)$ (lower membership function). When related to IF-sets, $\underline{A}(x)=\mu_A(x)$ and $\bar{A}(x)=\mu_A(x)+\pi_A(x)$. This fact implies the differences between IFISs and IT2FISs. When applying AND operator (MIN t-norm) in if-then rules, the firing intervals $w_k=[\overline{w}_k, \underline{w}_k]$ are obtained. Notice that $\underline{w}_k=w^{\mu}_k=$ MIN$(\mu(x_1),\mu(x_2), \ldots ,\mu(x_m))$ but $\overline{w}_k=$MIN$(\mu(x_1)+\pi(x_1),\mu(x_2)+\pi(x_2), \ldots ,\mu(x_m)+\pi(x_m))$ while $w^{\nu}_k=$MIN$(\nu(x_1),\nu(x_2), \ldots ,\nu(x_m))$. Recently, IT2FISs have been adapted using Karnik-Mendel algorithm and its variants [16,17,18], back-propagation algorithm [19], fuzzy C-means algorithm [20], or genetic algorithms [21,22].

3 IFIS of Takagi-Sugeno Type Adaptation

There has been developed many approaches to adapt FISs in the literature, e.g. neural networks, genetic algorithms, Kalman filter, see e.g. [23,24]. In this paper, we use the

following algorithms to adapt IFIS (and FISs): subtractive clustering algorithm (SCA) [25], Moore-Penrose pseudo-inverse (MPPI), Kalman filter [26], Kaczmarz algorithm [27,28], and gradient algorithm [23].

The identification of IFISs consists of the following steps. First, cluster centres are found to construct the number N of if-then rules and the antecedents of the if-then rules. Second, optimize the consequents of the rules, i.e. parameters $a^{\mu}_{1,k}, a^{\mu}_{2,k}, \ldots, a^{\mu}_{n,k}, b^{\mu}$ and $a^{\nu}_{1,k}, a^{\nu}_{2,k}, \ldots, a^{\nu}_{n,k}, b^{\nu}$. The number of cluster centres c is equal to the number of if-then rules N, $c=N$, and $k=1,2,\ldots,N$ is the index of the cluster centre. It is determined automatically using a SCA. The radius of influence of a cluster r_a is considered the most important parameter in establishing the number of cluster centres c. A large r_a results in fewer clusters, while a small r_a generates a large number of clusters and, thus, can lead to model over-fitting.

In the subtractive clustering algorithm, each data point is considered as a potential cluster centre. A measure of potential of data point $\mathbf{x}_k$ is defined as follows

$$P_k = \sum_{j=1}^{n} e^{-\alpha\left\|\mathbf{x}_k - \mathbf{x}_j\right\|}, \tag{5}$$

where $\alpha=4/r_a^2$. Thus, a data point $\mathbf{x}_k$ with many neighbouring points $\mathbf{x}_j$ has a high potential value P_k. The data point with the highest potential represents the cluster centre of the first cluster. Then, an amount of potential from each data point is subtracted as a function of its distance from the first cluster centre

$$P_k \Leftarrow P_k - P_1^* e^{-\beta\left\|\mathbf{x}_k - \mathbf{x}_1^*\right\|^2}, \tag{6}$$

where $\mathbf{x}_1^*$ is the centre of the first cluster, P_1^* is the potential of $\mathbf{x}_1^*$, and $\beta=4/r_b^2$. The positive constant r_b represents the radius defining the neighbourhood that will have measurable reductions in potential P_k.

Let each input vector $\mathbf{x}_k^*$ is decomposed into two component vectors $\mathbf{y}_k^*$ and y_k^*, where $\mathbf{y}_k^*$ contains first n elements of $\mathbf{x}_k^*$ (input data) and y_k^* contains the output component (for a multi-input and single-output system). Each cluster centre $\mathbf{x}_k^*$ is considered an if-then rule. The output y is represented by a weighted average of the output of each rule (2). Since we use a IFIS of Takagi-Sugeno type of the first order, i.e. $f(x_1,x_2,\ldots,x_m)$ is a linear function, we can compute the y_k^* as follows

$$y_k^* = \mathbf{G}_k \mathbf{y}_k + h_k, \tag{7}$$

where $\mathbf{G}_k$ is a constant vector and h_k is a constant. The estimation of the parameters of the given model can be understood as least squares estimate (LSE) in the form $\mathbf{AX}=\mathbf{B}$, where $\mathbf{B}$ is a matrix of output values, $\mathbf{A}$ is a constant matrix, and $\mathbf{X}$ is a matrix of parameters to be estimated.

Let P is the set of linear parameters and $\mathbf{X}$ is an unknown vector whose elements are parameters in P. Then, we seek the optimal solution of $\mathbf{X}$ using a LSE $\mathbf{X}^*$. In this process, the squared error $\|\mathbf{AX}-\mathbf{B}\|^2$ is minimized. The LSE $\mathbf{X}^*$ can be calculated using the pseudo-inverse of $\mathbf{X}$ as follows

$$\mathbf{X}^*=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B}, \quad (8)$$

where $\mathbf{A}^T$ is a transpose of $\mathbf{A}$, and $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudo-inverse of $\mathbf{A}$ if $\mathbf{A}^T\mathbf{A}$ is non-singular. A generalization of the inverse matrix $\mathbf{A}^{-1}$ is called MPPI $\mathbf{A}^+$.

A sequential method of LSE (also known as recursive LSE) can be used to compute $\mathbf{X}^*$ [23]. This method is more efficient, especially for a low number of linear parameters. Let $\mathbf{a}_i^T$ is the i-th input vector of matrix $\mathbf{A}$ and $\mathbf{b}_i^T$ is the i-th element of $\mathbf{B}$. Then $\mathbf{X}$ can be calculated using the following formulas (online version)

$$\mathbf{X}_{i+1}=\mathbf{X}_i+\mathbf{S}_{i+1}\mathbf{a}_{i+1}(\mathbf{b}^T_{i+1}-\mathbf{a}^T_{i+1}\mathbf{X}_i),\ \mathbf{S}_{i+1}=1/\lambda[\mathbf{S}_i-(\mathbf{S}_i\mathbf{a}_{i+1}\mathbf{a}^T_{i+1}\mathbf{S}_i)/(\lambda+\mathbf{a}^T_{i+1}\mathbf{S}_i\mathbf{a}_{i+1})], \quad (9)$$

where $\mathbf{S}_i$ is the covariance matrix, λ is the forgetting factor, and $\mathbf{X}^*=\mathbf{X}_n$. The initial conditions are $\mathbf{X}_0=0$ and $\mathbf{S}_0=\gamma\mathbf{I}$, where γ is a positive large number and $\mathbf{I}$ is the identity matrix. A small λ shows on the fast effects of old data decay. The LSE of $\mathbf{X}$ can be interpreted as the Kalman filter for the process [26]

$$\mathbf{X}(k+1)=\mathbf{X}(k),\ \mathbf{Y}(k)=\mathbf{A}(k)\mathbf{X}(k)+e, \quad (10)$$

where e is noise, $\mathbf{X}(k)=\mathbf{X}_k$, $\mathbf{Y}(k)=\mathbf{b}_k$ and $\mathbf{A}(k)=\mathbf{a}_k$. Therefore, the sequential method of the LSE presented above is referred to as the Kalman filter algorithm. Kalman filters have been used to optimize the output function parameters FISs of Takagi-Sugeno type [29], to extract if-then rules from a given rule base of FISs [30], and to tune the input membership functions [31]. Another method used to compute the LSE $\mathbf{X}^*$ is the Kaczmarz algorithm. This algorithm is based on the following formula

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \frac{\mathbf{b}_k - \mathbf{a}_k\mathbf{X}_k^T}{\mathbf{a}_k^T\mathbf{a}_k}\mathbf{a}_k . \quad (11)$$

Another possibility to adapt an IFIS is represented by gradient algorithms. Let the i-th error is defined as $E_i=(y_i-o_i)^2$, where y_i is the actual output and o_i is the predicted output. Then, the total error E is given as $E=\Sigma E_i$. The derivative of the overall error measure E with respect to a generic parameter α is

$$\frac{\partial E}{\partial \alpha} = \sum_{i=1}^{n}\frac{\partial E_i}{\alpha}. \quad (12)$$

The update formula for the generic parameter α is defined as

$$\Delta\alpha = -\eta\frac{\partial E}{\partial \alpha}, \text{ where } \eta = \frac{h}{\sqrt{\sum_{\alpha}\left(\frac{\partial E}{\partial \alpha}\right)^2}}, \quad (13)$$

where η is the learning rate, h is the step size and the length of each gradient transition in the parameter space. Usually, we can change the value of h to vary the speed of convergence. Some practical difficulties associated with gradient descent are slow convergence and ineffectiveness at finding a good solution [31].

4 Experimental Results

The following regression datasets were selected for the modelling using IFISs: Friedman benchmark function [32], daily electricity energy [33], stock prices [34], and auto MPG dataset [35], for details see Table 1.

Table 1. Description of regression datasets.

	Dataset			
	Friedman	Energy	Stock	AutoMPG
Origin	Artificial	Real world	Real world	Real world
Input variables	5	6	9	5
Real / integer / nominal	5/0/0	6/0/0	9/0/0	2/3/0
Instances	1200	365	950	392

Friedman dataset is a synthetic benchmark dataset where the instances are generated using the following method. Generate the values of n=5 input variables, x_1,x_2, ... ,x_5 independently each of which uniformly distributed over [0.0, 1.0]. Obtain the value of the target variable y using the equation $y=10(\sin(\pi)x_1x_2)+20(x_3-0.5)^2+10x_4+5x_5+e$, where e is a Gaussian random noise $N(0,1)$.

The Energy problem involves predicting the daily average price of TkWhe electricity energy in Spain. The data set contains real values from 2003 about the daily consumption in Spain of energy from hydroelectric, nuclear electric, carbon, fuel, natural gas and other special sources of energy.

In the stock prices dataset, the data provided are daily stock prices from January 1988 through October 1991, for ten aerospace companies. The task is to approximate the price of the 10th company given the prices of the rest.

The AutoMPG dataset concerns city-cycle fuel consumption in miles per gallon (Mpg), to be predicted in terms of 1 multi-valued discrete and 5 continuous attributes (two multi-valued discrete attributes (Cylinders and Origin) from the original dataset are removed).

Datasets were divided into training and testing data in relation 1:1. This division was realized five times. The quality of regression was measured by RMSE on testing data. We conducted the experiments in Matlab Fuzzy Logic Toolbox using the adaptation algorithms available at [36].

The parameters of FISs and IFISs are set in the following way. The initial setting of FISs and IFISs was conducted using SCA. The number N of if-then rules (and the numbers of membership and non-membership functions at the same time) depends on the choice of parameter r_a in SCA primarily. In order to avoid over-fitting, we tested different values of r_a={0.1,0.2, … ,0.9}. For the IFISs, the IF-index is set on π=0.3 (medium level of hesitancy recommended in [14]) and, therefore, the membership and non-membership functions are defined in the following way

$$\mu(x) = 0.7e^{-\frac{(x-b)^2}{2\sigma^2}}, \quad \nu(x) = (1-\pi(x)) - \mu(x) = 0.7 - 0.7e^{-\frac{(x-b)^2}{2\sigma^2}}. \tag{14}$$

The settings of FISs and IFISs parameters are presented in Table 2. In our experiments, we used the following settings of adaptation methods:
- Gradient algorithm: maximum number of epochs was set to 500, step size h=0.01, step increasing rate to 1.1, and step decreasing rate to 0.9;
- Kaczmarz algorithm: maximum number of sweeps was set to 10;
- Kalman filter: data forgetting factor was set to λ=1.0 and its increasing factor to 1.0;
- MPPI: no parameters;

Table 2. Parameters of FISs and IFISs.

Dataset	Friedman	Energy	Stock	AutoMPG
Type of $\mu_A(x)$ and $\nu_A(x)$	Gaussian	Gaussian	Gaussian	Gaussian
Radius r_a	0.6	0.7	0.5	0.7
N of if-then rules	26	7	12	4
t-norm	MIN	MIN	MIN	MIN

Experiments were realized for various weights β={0.0,0.1, … ,1.0} of the subsystems with non-membership functions $\nu_A(x)$. Concretely, if β=0.0 the output of the IFIS is determined only by the subsystem with membership functions $\mu_A(x)$ while for β=1.0 only the output of the subsystem with non-membership functions $\nu_A(x)$ has impact on the IFIS output. The results of the experiments for β with minimum RMSE on testing data is presented in Table 3. The FISs and IFISs with adaptation achieved significantly lower RMSE (values in bold based on paired sample t-test at p=0.05) than those without adaptation (SCA). For all the datasets, using adapted IFISs resulted in the decrease in RMSEs.

Table 3. RMSE and its standard deviation on benchmark datasets.

	Dataset			
Method	Friedman	Energy	Stock	AutoMPG
FIS-SCA	1.538±0.020	0.494±0.030	2.378±0.584	3.405±0.254
IFIS-SCA	1.538±0.020 (β=0.0)	0.494±0.030 (β=0.0)	2.378±0.584 (β=0.0)	3.405±0.254 (β=0.0)
FIS-gradient	**1.353**±0.026	7.443±1.579	1.423±0.227	3.702±0.211
IFIS-gradient	**1.332**±0.032 (β=0.1)	4.776±2.776 (β=0.5)	1.402±0.219 (β=0.1)	3.684±0.195 (β=0.1)
FIS-Kaczmarz	1.759±0.099	0.619±0.246	1.888±0.267	3.991±0.355
IFIS-Kaczmarz	1.581±0.159 (β=0.3)	0.505±0.034 (β=0.1)	1.708±0.340 (β=0.2)	3.880±0.554 (β=0.1)
FIS-Kalman	1.412±0.031	0.760±0.551	**0.944**±0.046	**2.881**±0.113
IFIS-Kalman	**1.353**±0.039 (β=0.2)	**0.437**±0.031 (β=0.6)	**0.914**±0.044 (β=0.2)	**2.825**±0.100 (β=0.3)
FIS-MPPI	1.411±0.047	0.474±0.038	**0.943**±0.045	**2.866**±0.113
IFIS-MPPI	**1.352**±0.049 (β=0.2)	**0.435**±0.027 (β=0.6)	**0.913**±0.044 (β=0.2)	**2.821**±0.101 (β=0.3)

For comparison, FISs and IFISs with corresponding adaptation methods are shown. In the case of Friedman dataset, the dependence of RMSE (testing data) on β is

depicted in Fig. 1 (SCA) and Fig. 2 (gradient algorithm). When IFIS is identified using SCA only (i.e. it is not adapted subsequently), RMSE increases with rising β for all datasets (minimum RMSE for IFIS-SCA is achieved for β=0.0). In the case of gradient algorithm it is possible to reduce RMSE when combining the outputs of both subsystems with β=0.1. For the Friedman dataset (the most complex dataset in terms of N and the numbers of membership and non-membership functions), gradient algorithm provided the best result but for the Energy dataset this algorithm did not converge and for the Stock and AutoMPG datasets it was outperformed by the Kalman filter and MPPI, respectively. For the Energy dataset, the behaviour of IFIS differs from that one for the other datasets. The lowest RMSE is achieved for the Kalman filter and MPPI with β=0.6, i.e. both subsystems (with membership and non-membership functions) contribute with a similar magnitude.

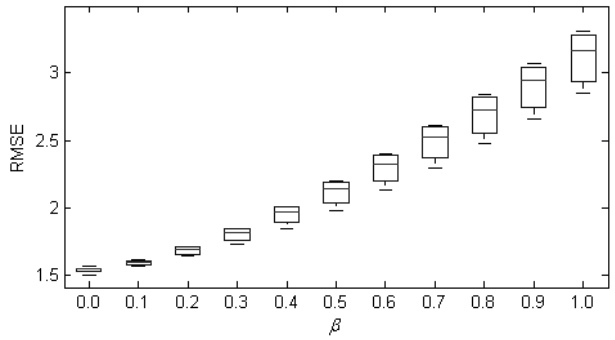

Fig. 1. Relation between β and RMSE for IFIS identified by SCA on Friedman dataset.

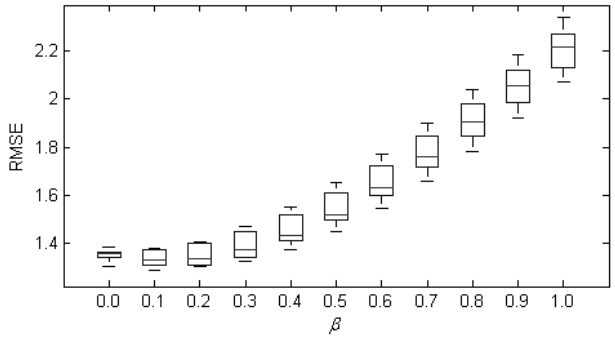

Fig. 2. Relation between β and RMSE for IFIS adapted by gradient algorithm on Friedman dataset.

5 Conclusion

In this paper, we have introduced IFIS of Takagi-Sugeno type. We have proposed several methods to adapt the parameters of this class of systems. We conclude that Kalman filter and MPPI are suitable to adapt the linear functions in the consequents of the IFIS if-then rules. However, gradient algorithm may perform better than these

methods for more complex problems. The antecedents parts of if-then rules, membership functions $\mu_A(x)$ and non-membership functions $\nu_A(x)$ were identified using SCA. For all the tested datasets, it was possible to significantly decrease RMSE on testing data when both an appropriate adaptation method was applied and the subsystems with membership and non-membership functions were combined properly.

In further research we plan to use a wider range of benchmark datasets to test the relation between data complexity and behaviour of IFISs (the value of IF-index especially). We also suggest comparing the two related concepts, IFISs and IT2FISs of Takagi-Sugeno type on regression problems (using also other performance measures such as R^2 or MAE) to show the differences between these two generalizations of FISs. Finally, IFISs should be tested on data with additional noise since generalizations of FISs have been proven to be especially beneficial in the cases where data was corrupted by measurement or estimated noise [37].

References

[1] Zadeh, L.A.: Fuzzy Sets. Inform. and Control 8, 338-353 (1965)
[2] Atanassov, K.T.: Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 20, 87-96 (1986)
[3] Atanassov, K.T.: Intuitionistic Fuzzy Sets. Physica-Verlag, Heidelberg New York (1999)
[4] Goguen, J.: L-fuzzy Sets. J. Math. Anal. Appl. 18, 145-174 (1967)
[5] Sambuc, R.: Fonctions Φ-floues. Application a L'aide au Diagnostic en Pathologie Thyroidienne. Ph.D. Thesis, Univ. Marseille, France (1975)
[6] Atanassov, K.T., Gargov, G.: Interval Valued Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 31, 343-349 (1989)
[7] Deschrijver, G., Kerre, E.E.: On the Relationship Between Some Extensions of Fuzzy Set Theory. Fuzzy Sets and Systems 133, 227-235 (2003)
[8] Gorzalczany, M.B.: A Method of Inference in Approximate Reasoning Based on Interval Valued Fuzzy Sets. Fuzzy Sets and Systems 21, 1-17 (1987)
[9] Karnik, N.N., Mendel, J.M.: Introduction to Type-2 Fuzzy Logic Systems. In: IEEE FUZZ Conf., Anchorage, AK (1998)
[10] Starczewski, J., Rutkowski, L.: Connectionist Structures of Type-2 Fuzzy Inference Systems. In: Wyrzykowski, R. et al. (eds.) PPAM 2001. LNCS, vol. 2328, pp. 634-642. (2001)
[11] Montiel, O., Castillo, O., Melin, P., Sepúlveda, R.: Mediative Fuzzy Logic: A New Approach for Contradictory Knowledge Management. Soft Computing 20, 251-256 (2008)
[12] Olej, V., Hajek, P.: IF-Inference Systems Design for Prediction of Ozone Time Series: The Case of Pardubice Micro-region, In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) 20th Int. Conf. on Artificial Neural Networks, Part I, Thessaloniki, Greece, pp. 1-11. Springer Berlin Heidelberg New York (2010)
[13] Olej, V., Hájek, P.: Possibilities of Air Quality Modelling Based on IF-sets Theory. Computers and Simulation in Modern Science. Selected Papers from WSEAS Conferences, In: Mastorakis, N., Demiralp, N., Mladenov, M. (eds.) pp.90-100. WSEAS Press (2010)
[14] Olej, V., Hajek, P.: Comparison of Fuzzy Operators for IF-inference Systems of Takagi-Sugeno Type. In: Iliadis, L.S., Maglogianis, I., Papadopoulos, H. (eds.) 12th EANN/7th AIAI Joint Conferences, Engineering Applications of Neural Networks/Artificial Intelligence Applications and Innovations, Part II, Corfu, Greece, pp. 92-97. Springer Berlin Heidelberg New York (2011)

[15] Takagi, T., Sugeno M.: Fuzzy Identification of Systems and its Applications to Modeling and Control. IEEE Trans. on Syst. Man and Cybern. 15(1), 116-132 (1985)
[16] Wu, D., Mendel, J.M.: Enhanced Karnik-Mendel Algorithms. IEEE Transactions on Fuzzy Systems 17(4), 923-934 (2009)
[17] Wu, D., Nie, M.: Comparison and Practical Implementation of Type-Reduction Algorithms for Type-2 Fuzzy Sets and Systems. In: IEEE Int. Conf. on Fuzzy Systems, Taipei, Taiwan (2011)
[18] Yeh, C.Y., Jeng, W.H., Lee, S.J.: An Enhanced Type-reduction Algorithm for Type-2 Fuzzy Sets. IEEE Trans. on Fuzzy Systems, (2011) in press.
[19] Lee, Ch.H., Lee, Y.H.: Nonlinear System Identification Using Takagi-Sugeno-Kang Type Interval-Valued Fuzzy Systems via Stable Learning Mechanism. IAENG International Journal of Computer Science 38(3), pp. 1-11 (2011)
[20] Uncu, O., Turksen, I.B.: Discrete Interval Type-2 Fuzzy System Models Using Uncertainty in Learning Parameters. IEEE Trans. on Fuzzy Systems 15(1), 90-106 (2007)
[21] Lee, Ch.H., Hong, J.L., Lin, Y.Ch., Lai, W.Y.: Type-2 Fuzzy Neural Network Systems and Learning. International Journal of Computational Cognition 1(4), 79-90 (2003)
[22] Celikyilmaz, A., Turksen, I.B.: Genetic Type-2 Fuzzy Classifier Functions. In: 10th Annual Meeting of the North American Fuzzy Information Processing Society, pp. 1-6. NAFIPS (2008)
[23] Jang, J.S.R.: ANFIS: Adaptive-Network-based Fuzzy Inference Systems. IEEE Transactions on Systems, Man, and Cybernetics 23(3), 665-685 (1993)
[24] Alcala, R., Casillas, J., Cordon, O., Herrera, F., Zwir, S.J.I.: Learning and Tuning Fuzzy Rule-based Systems for Linguistic Modeling and their Applications. In: Leondes, C.T. (ed.) Knowledge-Based Systems 3, 889-933 (1999)
[25] Chiu, S.: Fuzzy Model Identification Based on Cluster Estimation. J. of Intell. and Fuzzy Syst. 2, 267-278 (1994)
[26] Jang, J.S.R.: Fuzzy Modeling Using Generalized Neural Networks Kalman Filter Algorithm. In: 8th National Conference on Artificial Intelligence (AAAI-91), 762-767 (1991)
[27] Kaczmarz, S.: Approximate Solution of Systems of Linear Equations. Int. J. Control 53, 1269-1271 (1993)
[28] Strohmer, T., Vershynin, R.: A Randomized Kaczmarz Algorithm with Exponential Convergence. Journal of Fourier Analysis and Applications 15(2), 262-278 (2007)
[29] Ramaswamy, P., Riese,M., Edwards, R., Lee, K.: Two Approaches for Automating the Tuning Process of Fuzzy Logic Controllers. In: IEEE Conf. on Decision and Control, San Antonio, TX, 1753-1758 (1993)
[30] Wang, L., Yen, J.: Extracting Fuzzy Rules for System Modeling Using a Hybrid of Genetic Algorithms and Kalman Filter. Fuzzy Sets and Systems 101, 353-362 (1998)
[31] Simon, D.: Training Fuzzy Systems with the Extended Kalman Filter. Fuzzy Sets and Systems 132(2), 189-199 (2002)
[32] Friedman, J.: Multivariate Adaptative Regression Splines. Annals of Statistics 19(1), 1-141 (1991)
[33] Java Software Tool Named KEEL, Knowledge Extraction Based on Evolutionary Learning. http://www.keel.es, 2012.
[34] Altay, G.H., Bilkent, I.U.: University Function Approximation Repository. http://funapp.cs.bilkent.edu.tr, 2000.
[35] Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann (1993)
[36] Sidelnikov, P.K.A.: Sugeno-type FIS Output Tuning. http://www.mathworks.ch/Matlabcentral/fileexchange/28458-sugeno-type-fis-output-tuning, 2010.
[37] Liang, Q., Mendel, J.M.: Interval Type-2 Fuzzy Logic Systems: Theory and Design. IEEE Transactions on Fuzzy Systems 8(5), 535-550 (2000)