WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia FUZZ IEEE

Prediction Interval Construction Using Interval Type-2 Fuzzy Logic Systems

Abbas Khosravi, Saeid Nahavandi, Doug Creighton, Reihaneh Naghavizadeh
Centre for Intelligent Systems Research
Deakin University, Geelong, VIC, 3217, Australia
Email: {abbas.khosravi,saeid.nahavandi,douglas.creighton@deakin.edu.au} and naghavizade_r@yahoo.com

***Abstract*—This study proposes a novel non-parametric method for construction of prediction intervals (PIs) using interval type-2 Takagi-Sugeno-Kang fuzzy logic systems (IT2 TSK FLSs). The key idea in the proposed method is to treat the left and right end points of the type-reduced set as the lower and upper bounds of a PI. This allows us to construct PIs without making any special assumption about the data distribution. A new training algorithm is developed to satisfy conditions imposed by the associated confidence level on PIs. Proper adjustment of premise and consequent parameters of IT2 TSK FLSs is performed through the minimization of a PI-based objective function, rather than traditional error-based cost functions. This new cost function covers both validity and informativeness aspects of PIs. A metaheuristic method is applied for minimization of the non-linear non-differentiable cost function. Quantitative measures are applied for assessing the quality of PIs constructed using IT2 TSK FLSs. The demonstrated results for four benchmark case studies with homogenous and heterogeneous noise clearly show the proposed method is capable of generating high quality PIs useful for decision-making.**

***Index Terms*—type-2 fuzzy logic, prediction intervals, confidence level.**

I. Introduction

Fuzzy logic systems (FLSs) are universal approximators with the capability of identifying and approximating nonlinear relationships and patterns hidden in data to any desired degree of accuracy. They are transparent models and can be understood and interpreted by the human decision makers. Traditional FLSs use precise type-1 fuzzy sets (T1 FS) in their structure. Once the type-1 membership functions (MF) are selected and adjusted, all uncertainties disappear. To remedy this critical problem, Zadeh introduced the type-2 fuzzy set (T2 FS) as an extension of the concept of an ordinary T1 FS [1]. T2 FSs are characterized by a three dimensional fuzzy membership function (MF) including a footprint of uncertainty, and as such can outperform type-1 fuzzy logic systems (T1 FLSs) with the same number of rules [2]. The structure of a type-2 fuzzy logic system (T2 FLS) is very similar to the structure of a T1 FLS. The only difference is the output processing block. In the T1 FLSs, this only includes a defuzzifier that produces a crisp output. The output processing block in T2 FLSs includes a type reduction unit that transforms a T2 FS into a T1 FS. This type-reduced set is then defuzzified into a crisp value [3].

The computational complexity of general T2 FLSs hinders their widespread use in practical applications. Interval type-2 fuzzy logic systems (IT2 FLSs), a restricted class of T2 FLSs, were introduced to reduce the computational requirements needed for type reduction [4]. IT2 FLSs are characterized by secondary MFs that only take the value of one over their domain. In the light of their application and further studies conducted in the recent years [5] [6] [7], it is already known that IT2 FLSs are a promising tool for processing uncertain data. Quite often, both qualitative and quantitative (numerical) uncertainties in data may translate into rule uncertainties, ruining performance of a model. IT2 FLSs with additional degrees of freedom make it possible to manage and minimize the effects of uncertainties, even better than traditional T1 FLSs [8]. The recent literature is rich in application of these methods in different fields such as control [9], path planning [2], decision-making [10], and forecasting [11], [12].

Research works on FLSs and their application for point forecasts and prediction are abundant. The focus of this paper is on developing theories and a framework for construction of Prediction intervals (PIs) using IT2 TSK FLSs. PIs are a promising statistical tool for quantification of uncertainties associated with predictions and forecasts. A $(1-\alpha)\%$ PI for a future observation, y_f, has the form $[L_n, U_n]$ where $P(L_n \leq y_f \leq U_n) \xrightarrow{P} (1-\alpha)\%$ as the sample size $n \to \infty$. L_n and U_n are the lower and upper bounds of the PI respectively. Also $(1-\alpha)\%$ is the prescribed confidence level associated with PIs.

For the case of neural networks (NNs), several methods have been proposed in the literature for quantification of associated uncertainties [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] [23]. A comprehensive review of these methods can be found in [24]. These methods can be broadly divided into parametric and non-parametric methods. In the former group, methods usually assume that the quantitative uncertainty (noise) is independently and identically distributed and follows a known distribution with unknown parameters. Non-parametric methods construct PIs without making any special assumption about the data and noise distributions. So, they have more flexibility and are more suitable for practical applications than parametric methods.

The major contribution of this paper is the introduction of a novel method for construction of PIs using IT2 TSK FLSs. In fact, the work presented in this paper is a first study on how PIs with a prescribed confidence level can be derived

using IT2 TSK FLSs. In contrast to parametric methods, the proposed method requires no special assumption about data and its distribution to construct PIs. It also does not require calculation of complicate matrices and derivatives required by traditional PI construction methods. It is possible to build an empirical distribution for data by constructing PIs for different confidence levels ranging between zero to one. As a distribution-free method, the IT2 TSK FLS-based method for construction of PIs is robust against violations of the normality assumption and offers a promising method for rapid construction of reliable PIs.

The proposed method is applied to construct PIs for four synthetic benchmark case studies. The level of uncertainty in data is controlled by adding homogeneous and heterogeneous noise. It is demonstrated that the proposed method generates high quality PIs, which are theoretically valid and practically informative. Also, the performance of the proposed method is stable in different replicates of experiments, which indicates it effectively and efficiently handles prevailing uncertainties in data.

The rest of this article is organized as follows: Section II introduces IT2 TSK FLSs. PI assessment measures and indices are briefly discussed in Section III. Section IV describes the proposed method for PI construction using IT2 TSK FLSs. Simulation results are demonstrated in Section V. Finally, we draw conclusions in Section VI.

II. Interval Type-2 TSK Fuzzy Logic Systems

The kernel of a FLS is its knowledge rule base. The TSK type fuzzy models utilize fuzzy values in the antecedent part of their rules, linear functions in the consequent part of their rules. IT2 TSK FLSs are TSK fuzzy models, where antecedent MFs are IT2 FSs or consequent parameters are intervals. Of particular interest is the case where the consequents are interval T1 FSs [8]. This is the most general form of IT2 TSK FLSs with the greatest degrees of freedom. The extra flexibility of this type allows for efficient handling of uncertainties and minimizing their effects on the quality of forecasts.

In an IT2 TSK FLS with a rule base of M rules in which each rule has p antecedents, let the lth rule be denoted by R^l,

$$\begin{aligned} R^l:\ &If\ x_1\ is\ \tilde{F}_1^l,\ x_2\ is\ \tilde{F}_2^l, \cdots,\ and\ x_p\ is\ \tilde{F}_p^l,\ then \\ &Y^l \ = \ C_0^l + \sum_{i=1}^{p} C_i^l\ x_i \end{aligned} \tag{1}$$

where $l = 1, ..., M$. $\tilde{F}_i^l$ is the ith IT2 FS ($i = 1, ..., p$) composed of a lower and upper bound MF,

$$\mu_{\tilde{F}_i^l}(x_i) = [\underline{\mu}_{\tilde{F}_i^l}(x_i),\ \bar{\mu}_{\tilde{F}_i^l}(x_i)] \tag{2}$$

C_i^l is also an interval T1 FS, where its center and spread are c_i^l and s_i^l respectively,

$$C_i^l = [c_i^l - s_i^l,\ c_i^l + s_i^l] \tag{3}$$

where $i = 0, ..., p$. Given an input $x = (x_1, x_2, ..., x_p)$, the result of the input and antecedent operations (firing strength) is an interval type-1 set, $F^l = [\underline{f}^l, \bar{f}^l]$, where,

$$\underline{f}^l(x) = \underline{\mu}_{\tilde{F}_1^l}(x_1) * \underline{\mu}_{\tilde{F}_2^l}(x_2) * \cdots * \underline{\mu}_{\tilde{F}_p^l}(x_p) \tag{4}$$

$$\bar{f}^l(x) = \bar{\mu}_{\tilde{F}_1^l}(x_1) * \bar{\mu}_{\tilde{F}_2^l}(x_2) * \cdots * \bar{\mu}_{\tilde{F}_p^l}(x_p) \tag{5}$$

where * represents a t-norm. It is assumed that the singleton fuzzifier is used in obtaining (4) and (5).

Y^l in (1) is the output from the lth If-Then rule, which is a T1 FS, $Y^l = [y_L^l, y_R^l]$. y_L^l and y_R^l are evaluated as,

$$y_L^l = \sum_{i=1}^{p} c_i^l\ x_i + c_0^l\ -\ \sum_{i=1}^{p} s_i^l\ |x_i| - s_0^l \tag{6}$$

$$y_R^l = \sum_{i=1}^{p} c_i^l\ x_i + c_0^l\ +\ \sum_{i=1}^{p} s_i^l\ |x_i| - s_0^l \tag{7}$$

The final output of the IT2 TSK FLS model is obtained through combining the outcomes of M rules,

$$\begin{aligned} Y = [y_L, y_R] = \\ &\int_{y^1 \in [y_L^1, y_R^1]} \cdots \int_{y^M \in [y_L^M, y_R^M]} \\ &\int_{f^1 \in [f_L^1, f_R^1]} \cdots \int_{f^M \in [f_L^M, f_R^M]} \frac{1}{\frac{\sum_{l=1}^{M} f^l y^l}{\sum_{l=1}^{M} f^l}} \end{aligned} \tag{8}$$

y_L and y_R in (8) can be calculated using the iterative Karnik-Mendel (KM) procedure [25]. It transfers a T2 FS into a T1 FS using the concept of center of sets. In the KM algorithm, y_L and y_R are calculated as below,

$$y_L = \frac{\sum\limits_{l=1}^{L} \bar{f}_i^l y_L^l + \sum\limits_{l=L+1}^{M} \underline{f}_i^l y_L^l}{\sum\limits_{l=1}^{L} \bar{f}_i^l + \sum\limits_{l=L+1}^{M} \underline{f}_i^l} \tag{9}$$

$$y_R = \frac{\sum\limits_{l=1}^{R} \underline{f}_i^l y_R^l + \sum\limits_{l=R+1}^{M} \bar{f}_i^l y_R^l}{\sum\limits_{l=1}^{R} \underline{f}_i^l + \sum\limits_{l=R+1}^{M} \bar{f}_i^l} \tag{10}$$

where L and R are switch points which can be computed by the KM algorithm. Finally, the defuzzified crisp output from the IT2 TSK FLS is the mean of y_L and y_R,

$$y = \frac{y_L + y_R}{2} \tag{11}$$

The design of an IT2 TSK FLS model consists of the selection of MFs for inputs and the determination of the parameters of premise (MF parameters) and consequent (linear coefficients) parameters.

III. Performance Index for Prediction Intervals

PIs can be examined from two perspectives: their validity and their informativeness. Here, a few measures are described for evaluation of the quality of PIs from these perspectives.

Once constructed, the most important characteristic of PIs is their actual coverage probability. PI coverage probability (PICP) is measured by counting the number of target values covered by the constructed PIs,

$$PICP = \frac{1}{n}\sum_{t=1}^{n} c_t \tag{12}$$

where, $c_t = 1$ if $y_t \in [L_t, U_t]$, otherwise $c_t = 0$. L_t and U_t are the lower and upper bounds of the t-th PI respectively. PICP is a measure of validity of PIs constructed with an associated confidence level. PIs are always constructed with a prescribed confidence level. Theoretically, $PICP \geq (1-\alpha)\%$. Otherwise, constructed PIs are not reliable and should be discarded.

A satisfactory PICP greater than the nominal confidence level can be easily achieved by generating excessively wide PIs. However, narrow PIs are practically more informative than wide PIs, and therefore more useful for decision-making. So, another measure is required to examine how much informative PIs are. PI normalized averaged width (PINAW) assesses PIs from this aspect and measures how wide they are,

$$PINAW = \frac{1}{nR}\sum_{t=1}^{n} (U_t - L_t) \tag{13}$$

where R is the range of the underlying target. PINAW is the average width of PIs as a percentage of the underlying target range. If the minimum and maximum values of targets are used for the lower and upper bounds of PIs, all targets will be bracketed by PIs. Therefore, we will have $PICP = 100\%$ and $PINAW = 100\%$.

While PICP is related to the correctness (validity) of intervals, PINAW is a measure of their informativeness. A combinational coverage width-based criterion (CWC) is used for simultaneous evaluation of the quality of constructed PIs from both perspectives,

$$CWC = PINAW\left(1 + \gamma\, e^{-\eta(PICP-\mu)}\right) \tag{14}$$

where $\gamma = 0$ if $PICP \geq \mu$; otherwise $\gamma = 1$. η and μ in (14) are two hyperparameters controlling the location and amount of CWC jump. The design of CWC is based on two principles: (i) if PICP is less than the nominal confidence level, $(1-\alpha)\%$, CWC should be large regardless of the width of PIs (measured by PINAW), and (ii) if PICP is greater than or equal to its corresponding confidence level, then PINAW should be the influential factor. The exponential term in (14) is eliminated whenever $PICP \geq \mu$ and CWC becomes equal to PINAW. Further information about these measures and hyperparameters η and μ can be found in [14], [19], [22], [23], [24], [26], [27].

IV. PI-based Training

This section specifically investigates how PIs can be constructed for targets predicted by an IT2 TSK FLS. The method for construction of PIs will be based on the recent developments in the field of IT2-FLSs. The Karnik-Mendel (KM) and Wu-Mendel algorithms for type reduction compute the left and right end points needed to characterize IT2 FSs. These two end points can be considered as upper and lower bounds of an interval, where the corresponding target lies somewhere between them. The issue is that this interval cannot be called a PI, as it ignores the concept of the confidence level. The confidence level is a prescribed probability associated with PIs. As defined before, in an infinite run of the PI construction method, the coverage probability of PIs should asymptotically approach the nominal confidence level. How the current intervals (left and right end points obtained from the type reduced set) can be linked to the concept of confidence level is the question to be answered.

The key idea here is to train IT2 TSK FLS models based on the characteristics of intervals that they are going to generate. Put in other words, training should be done using a PI-based cost function, rather than traditional error-based cost functions. The most important characteristic of PIs is their coverage probability. $PICP \geq (1-\alpha)\%$ is the essential condition of valid PIs. While wide PIs can easily satisfy this requirement, those are less informative and therefore, practically less useful. Ideally, PIs should be valid and as narrow as possible. To construct such PIs, CWC can be applied as the cost function for training of IT2 TSK FLSs. The rational is that CWC covers both important characteristics of PIs. Using it, premise and consequent parameters of IT2 TSK FLSs can be adjusted such that the constructed PIs using the end points of the type-reduced set satisfy the coverage probability requirement, i.e., $PICP \geq (1-\alpha)\%$. During the training, if PIs are invalid with a PICP less than the associated confidence level, CWC takes a large exponential value. Otherwise, CWC becomes equal to PINAW and the training algorithm tries to make PIs narrower. It is important to note that CWC does not allow excessively narrowing PIs which may result in an unsatisfactorily low PICP. In fact, reaching a tradeoff between validity and informativeness of PIs is the goal of training process.

The proposed method for construction of PIs using IT2 TSK FLSs is as follows:

1) Available data samples are split into two training (D_{train}) and test (D_{test}) sets.
2) An IT2 TSK FLS model is built and its parameters are initialized. The initialization of MF parameters is performed such that the range of each input is covered by its IT2 MFs.
3) Premise and consequent parameters of the IT2 TSK FLS model are then adjusted through minimization of CWC as the cost function. In each iteration/epoch of the optimization process, PICP, PINAW, and the cost function value is calculated for the PIs constructed using the new set of parameters. Optimization contin-

ues till no further improvement is achieved for several iterations.

In mathematical terms, the training procedure can be described as follows. Let w be the set of all premise and consequent parameters of the IT2 TSK FLS model. According to this, the training procedure will be,

$$w_{opt} = arg \ \min_{w} CWC \tag{15}$$

subject to constraints imposed on parameters of MFs. These constraints depend on the type of employed MFs in the model and are often in the form of linear inequalities for matrices.

4) Once training is over, the trained model is used for construction of PIs for test samples in D_{test}.

CWC as the cost function is nonlinear in the parameters space and has several sharp peaks and deep valleys. This is due to the fact that CWC is highly sensitive to model parameters. A small change in a parameter may result in invalid PIs with a low PICP, which consequently causes an astronomical increment in the CWC value. Besides, CWC is discontinuous with regard to the parameters of IT2 TSK FLS models. Therefore, it cannot be minimized using traditional gradient descent-based methods that rely on derivatives of the cost function. Besides, those methods are highly likely to be trapped in local minima of the cost functions.

On the other hand, stochastic gradient-free global optimization methods can be effectively applied for minimization of CWC as the cost function. These methods avoid falling in local minima and do not require calculation of computationally expensive derivatives. Examples of these methods are genetic algorithm, simulated annealing, and particle swarm optimization.

V. Numerical Results

Four synthetic benchmark case studies with the homogenous and heterogeneous noise are considered in this study to examine performance of the proposed method. The first case study (CS1) is the two dimensional version of Rastrigin's function [28], [29], which is based on the function of De Jong with the addition of cosine modulation in order to produce frequent local minima. One thousand random data points are generated with a uniform distribution between -5.12 and 5.12, i.e., $-5.12 \leq x_{k,i} \leq 5.12, i = 1, 2, \cdots, 1000, k = 1, 2$. The function has the following form,

$$y_i = 20 + \sum_{k=1}^{2} [x_{k,i}^2 + 10cos(x_{k,i})] + \epsilon_i \tag{16}$$

where $\epsilon_i \sim \mathcal{N}(0, 5)$. Rastrigin's function is a typical example of a non-linear multimodal function and quite often is used as a benchmark case study in the field of optimization.

The datasets for the other three synthetic case studies come from a univariate function with heterogeneous noise. One thousand random data points, $x_1, x_2, \cdots, x_{1000}$, are generated with a uniform distribution on the interval $[-10, 10]$.

TABLE I

Parameters used in the experiments

Parameter	Numerical Value
Number of MFs	3
α	0.1
μ	0.9
η	50
T_0	10
Geometric cooling schedule	$T_{k+1} = 0.9\ T_k$
D_{train}	80% of samples
D_{test}	20% of samples

Target values, $y_1, y_2, \cdots, y_{1000}$, are then calculated using the following model,

$$y_i = g(x_i) + \epsilon_i \tag{17}$$

where $g(x_i) = x_i^2 + sin(x_i) + 2$. ϵ_i also follows a Gaussian distribution with mean zero and a non-constant variance $\frac{g(x_i)}{\tau}$. As the additive noise variance is not constant, it is heterogeneous. Three values are considered for τ: 10, 5, and 1. Hereafter we refer to these three case studies as CS2 ($\tau = 10$), CS3 ($\tau = 5$), and CS4 ($\tau = 1$) respectively. The smaller the τ, the greater the noise variance. So, the level of uncertainty in the data samples of CS4 is higher than the level of uncertainty in the data samples of CS2 and CS3.

The available samples in each case study are first split into training set, D_{train}, and test set, D_{test}. These account for 80% and 20% of samples, respectively. IT2 TSK FLS is first trained using training samples, where the fitness function (CWC) is minimized using the simulated annealing (SA) method. Then, the quality of PIs constructed for the test set is quantitatively examined. To avoid misleading results and subjective judgment, experiments are repeated ten times for each case study and averaged results are reported.

Table I summarizes parameters used for implementing the proposed method and conducting experiments. PIs are constructed with a 90% confidence level ($\alpha = 0.1$). μ is set to $1 - \alpha$, because the prescribed level of confidence of PIs is $(1 - \alpha)\%$. Also, η is selected to be 50 in order to highly penalize PIs with a coverage probability lower than the nominal confidence level. SA uses a geometric cooling schedule with a cooling factor of 0.9. Also, the initial temperature is set to 10 to allow for uphill movements and to avoid getting trapped in local minima. MF parameters are initialized such that to cover the range of inputs. A random initialization is applied for consequent parameters.

Gaussian MF with a fixed mean and uncertain standard deviations are considered for the antecedent parts of the IT2 TSK FLS models:

$$\mu_{\tilde{F}_i^l}(x_{i,k}) = exp\left[-\frac{1}{2}\left(\frac{x_{i,k} - m_i^l}{\sigma_i^l}\right)^2\right] = \mathcal{N}(m_i^l, [\sigma_{i,1}^l, \sigma_{i,2}^l]) \tag{18}$$

where $i = 1, ..., p$, $l = 1, ..., M$, and k indicates the sample index. Fig. 1 shows three MFs for the first input of case

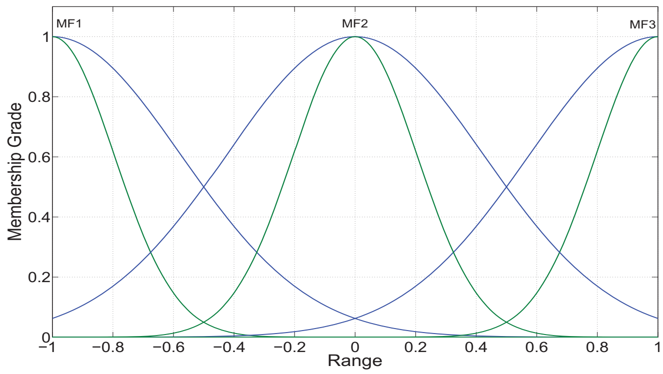

Fig. 1. The IT2 Gaussian MFs with fixed mean and uncertain standard deviation for input 1 in case study 1.

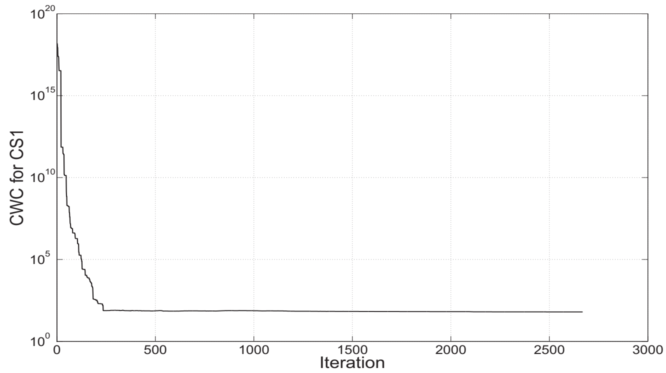

Fig. 2. The minimization of CWC during the training for CS1.

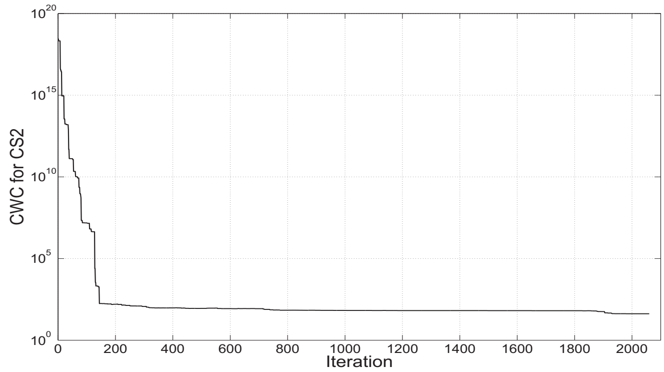

Fig. 3. The minimization of CWC during the training for CS2.

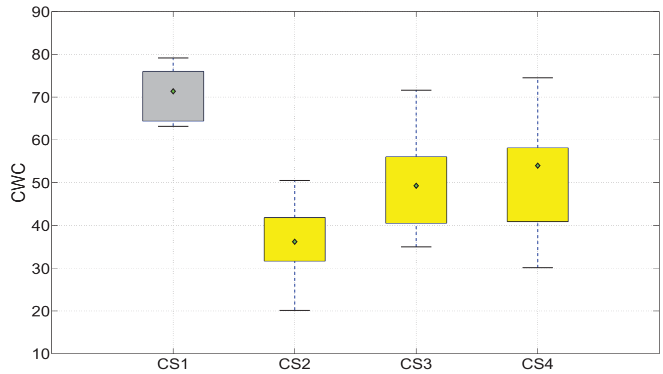

Fig. 4. The box plot of CWC_{test} for four case studies.

study one. The uncertain standard deviation helps to capture the non-stationary behavior of the targets.

Fig. 2 and 3 show the cost function versus the iteration number for the first two case studies. In the early iterations when the cooling temperature is large, the SA method allows for uphill movements. So, there are several solutions resulting in the increment of CWC. As the optimization process continues, these transitions gradually vanish and the SA method becomes greedy. During these iterations, SA continuously decreases and converges to its minimum. The optimization in both cases terminates when no further improvement is achieved for several iterations. It is important to notice that there is a huge difference between the initial and final values of the cost function. This is due to the fact that the initial PIs have a PICP much lower than the nominal confidence level. This results in an astronomical CWC value. During the optimization, the quality of PIs is significantly improved and their coverage probability moves towards the confidence level. So, the cost function continuously falls. Upon the termination of the optimization process, $PICP_{train}$ is greater than 90% and CWC is equal to PINAW.

Table II summarizes simulation results obtained for test samples of four case studies. All three measures introduced in Section III are listed in this table for a precise analysis. With the exception of only one case, we always have $PICP \geq 90\%$. Therefore, all constructed PIs are theoretically valid and reliable. The validity is the key feature of constructed PIs. It indicates that regardless of homogeneity or heterogeneity of the noise, IT2 TSK FLS models can efficiently handle uncertainties and generate high quality PIs. Furthermore, it also indicates the proposed PI-based training method is effective in transforming IT2 TSK FLS models into a reliable tool for rapid construction of PIs, rather than construction of simple intervals without any indication of their accuracy.

For the first case study with the homogeneous noise, $PICP_{test}$ is at least 1.5% greater than the nominal confidence level. As the level of uncertainty increases from CS2 to CS4, the gap between PICP and the nominal confidence level becomes smaller. The median values of $PICP_{CS2}$, $PICP_{CS3}$, and $PICP_{CS4}$ are 93.50%, 91.75%, and 91.50%, respectively. Also, the mean values of $PICP_{CS2}$, $PICP_{CS3}$, and $PICP_{CS4}$ are 93.10%, 91.65%, and 91.45%, respectively. So, there is a reduction in the mean and median values of PICP from CS2 to CS4. This signifies that a higher level of uncertainty in data makes achieving a satisfactory coverage probability more problematic. However, the proposed method here performs well and all PIs constructed for test samples in different replicates of case studies are valid.

We have $CWC_{test} = PINAW_{test}$, because $PICP_{test} \geq$

TABLE II
PICP, PINAW, AND CWC FOR TEST SAMPLES OF FOUR CASE STUDIES

Case Study	PICP	PINAW	CWC
CS1	93.00	70.15	70.15
	92.00	64.43	64.43
	92.00	74.40	74.40
	93.50	75.98	75.98
	92.50	63.84	63.84
	95.00	71.68	71.68
	92.50	78.68	78.68
	92.50	63.18	63.18
	93.50	72.11	72.11
	91.50	79.16	79.16
CS2	94.00	41.81	41.81
	96.00	38.96	38.96
	94.50	34.52	34.52
	90.50	43.09	43.09
	94.00	40.79	40.79
	96.00	31.67	31.67
	93.00	27.50	27.50
	90.00	50.53	50.53
	90.50	32.86	32.86
	92.50	20.13	20.13
CS3	93.50	70.97	70.97
	92.00	42.95	42.95
	92.50	40.11	40.11
	91.00	43.79	43.79
	92.00	42.75	42.75
	91.50	48.95	48.95
	92.50	40.54	40.54
	90.00	71.64	71.64
	91.50	34.96	34.96
	90.00	56.02	56.02
CS4	91.00	36.97	36.97
	89.50	45.73	104.44
	91.50	58.10	58.10
	91.50	74.51	74.51
	93.00	41.81	41.81
	92.00	30.10	30.10
	92.00	42.72	42.72
	93.00	52.30	52.30
	91.00	57.81	57.81
	90.00	40.90	40.90

90%. Therefore, the analysis of results based on these two measures is the same. Fig. 4 displays the box plot of CWC for ten replicates of four case studies. The mean values of CWC are also shown by a green diamond. The inter quartile range of CWC for the first case study is small. This indicates that the performance of the proposed method for construction of PIs is stable in the case of homogeneous noise. According to these box plots for CS2-CS4, the PIs become wider in response to a higher level of uncertainty in data. For all statistics of $PINAW$ shown in Fig. 4, we have $CS_2 \leq CS_3 \leq CS_4$. These include mean, median, the 25th percentile (the lower quartile), the 75th percentiles (the upper quartile), and the inter quartile range. The proposed method identifies the higher level of uncertainty in the data and generates wider PIs to accommodate more uncertainties.

These results indicate that the IT2 TSK FLS models trained using the proposed method can generate valid PIs that can successfully cover all numerical uncertainties present in data.

VI. Conclusion

A novel method for construction of prediction intervals using IT2 TSK FLSs is proposed in this paper. The method interprets the left and right end points of the type reduced set as the lower and upper bounds of a prediction interval. The confidence level concept is linked to this interval through applying a prediction interval-based training algorithm. In this algorithm, IT2 TSK FLS models are trained through minimization of a prediction interval-based cost function, rather than minimization of traditional error-based cost functions. A simulated annealing method is applied for adjusting premise and consequent parameters of IT2 TSK FLS models and minimization of the nonlinear, nondifferentiable cost function. Demonstrated results for four case studies with homogeneous and heterogeneous noise show the proposed method generates high quality prediction intervals with a coverage probability greater than the nominal confidence level. Furthermore, as the level of uncertainty in data increases, the method effectively widens prediction intervals to accommodate extra uncertainties and retain the validity of prediction intervals.

Acknowledgment

This project is supported by the Australian Research Council (Discovery Grant DP120102112) and the Centre for Intelligent Systems Research (CISR) at Deakin University.

References

[1] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning–i," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[2] H. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 4, pp. 524–539, 2004.

[3] N. Karnik, J. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 6, pp. 643–658, 1999.

[4] Q. Liang and J. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 535–550, 2000.

[5] J. Mendel, "Type-2 fuzzy sets and systems: an overview," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20–29, 2007.

[6] H. Hagras, "Type-2 FLCs: A new generation of fuzzy controllers," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 30–43, 2007.

[7] R. John and S. Coupland, "Type-2 fuzzy logic: A historical view," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 57–62, 2007.

[8] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, Englewood Cliffs, NJ, 2001.

[9] D. Wu and W. Wan Tan, "Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 8, pp. 829–841, 2006.

[10] D. Wu and J. Mendel, "Computing with words for hierarchical decision making applied to evaluating a weapon system," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 441–460, 2010.

[11] A. Khosravi, S. Nahavandi, and D. Creighton, "Short term load forecasting using interval type-2 fuzzy logic systems," in *IEEE International Conference on Fuzzy Systems (FUZZ)*, 2011, pp. 502–508.

[12] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Interval type-2 fuzzy logic systems for load forecasting: A comparative study," *IEEE Transactions on Power Systems, DOI: 10.1109/TPWRS.2011.2181981, Accepted on 24-Dec-2011*.

[13] J. T. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 748–757, 1997.

[14] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problem," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1496–1503, 2010.

[15] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University, Press, Oxford, 1995.

[16] D. J. C. MacKay, "The evidence framework applied to classification networks," *Neural Computation*, vol. 4, no. 5, pp. 720–736, 1992.

[17] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[18] T. Heskes, "Practical confidence and prediction intervals," in *Neural Information Processing Systems*, T. P. M. Mozer, M. Jordan, Ed., vol. 9. MIT Press, 1997, pp. 176–182.

[19] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "A lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 1045–9227, 2011.

[20] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 983–988, 2011.

[21] D. Nix and A. Weigend, "Estimating the mean and variance of the target probability distribution," in *IEEE International Conference on Neural Networks*, 1994.

[22] A. Khosravi, E. Mazloumi, S. Nahavandi, D. Creighton, and J. Van Lint, "A genetic algorithm-based method for improving quality of travel time prediction intervals," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1364–1376, 2011.

[23] A. Khosravi, E. Mazloumi, S. Nahavandi, D. Creighton, and J. W. C. van Lint, "Prediction intervals to account for uncertainties in travel time prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 537–547, 2011.

[24] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "A comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

[25] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, no. 1-4, pp. 195–220, Feb. 2001.

[26] A. Khosravi, S. Nahavandi, and D. Creighton, "A prediction interval-based approach to determine optimal structures of neural network metamodels," *Expert Systems with Applications*, vol. 37, pp. 2377–2387, 2010.

[27] ——, "Load forecasting and neural networks: A prediction interval-based perspective," *B.K. Panigrahi et al. (Eds.): Computational Intelligence in Power Engineering, SCI 302*, pp. 131–150, 2010.

[28] H. Muhlenbein, M. Schomisch, and J. Born, "The parallel genetic algorithm as function optimizer," *Parallel Computing*, vol. 17, no. 6-7, pp. 619–632, 1991.

[29] A. Torn and Z. A., *Global Optimization*. Lecture Notes in Computer Science, N 350, Springer-Verlag, Berlin, 1989.