Prediction Intervals for Mechanical Property Forecasting with Improved ANFIS

Siyi Sun
School of Mathematics and Physics
The University of Science and Technology Beijing
Beijing, China
sssummer341@xs.ustb.edu.cn

***Abstract*—Application of Adaptive-Network-based Fuzzy Inference System (ANFIS) for point forecast has burgeoned in recent years. While the performance of ANFIS drops when the uncertainty level increases in dataset, and point forecasts also suffer from unreliable and uninformative problems. Prediction interval (PI) has been proposed as a powerful tool to address these drawbacks. In spite of plentiful application of ANFIS for constructing PIs, the use in mechanical property forecasting is limited. Compared with traditional computational expensive algorithms, in this paper, we combine Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) to minimize the cost function and optimize the parameters of ANFIS. The proposed novel approach is applied to a real mechanical property forecasting case study. The outcomes show modified PSO-based ANFIS not only computational less expensive, but also with good quality of constructing better PIs.**

Keywords—prediction interval, modified PSO-based ANFIS, mechanical property forecasting

I. Introduction

Applied to online dynamic control of products property, optimization of product composition, new products design and other aspects, steel property forecasting plays a crucial role during steel production process. Mechanical property forecasting is capable of predicting steel property before smelting. Based on feedback of forecast results, setting proper technological parameters to reduce cost and enhance the effectiveness is feasible and more precise. Adopting this method can cut down the amount of property research experiments sharply and shorten new products development cycle and delivery cycle. Most existing forecasting methods are point forecasting with drawbacks such as low accuracy and no information about the prediction errors, the uncertainty of the outcomes. This paper proposes a new method to predict steel property.

Performance of most point forecasting methods drops significantly once level of uncertainty increases, such as using dataset with characteristics of multivalued, sparse or noisy, when addressing prediction and regression problems. Also, point forecasting brings out only the point prediction values without any further information about the accuracy of the prediction. Compared with the ameliorable results of point forecasts, Prediction Interval (PI) performs better not only for its good quality to provide more reliable and informative outcomes which use confidence level ($(1-\alpha)\%$) as an index to evaluate the accuracy, but also for its additional characteristic of quantifying the uncertainty. Prediction Interval constructs a range which targets most probably lie within [1-5].

Existing mechanical property forecasting methods can be roughly classified into two categories: mechanism model-based and data-driven. The difficulty of mechanism modeling which is difficult increases with the complexity of organizational structure evolution. Mechanism modeling with unsatisfactory accuracy is only for single steel species and infeasible to be applied to multiple steel species. As a result, data-driven methods are being used more and more widely. Several traditional methods of constructing PI such as delta, Bayesian, Generalized Likelihood Uncertainty Estimation (GLUE), Bootstrap and mean-variance are still not widely applied to real-world problems for its drawbacks in some aspects [6-10].

The paper is organized as follows. Section Two introduces previous PI related work. Section Three lists the PI assessment measures. The methodology of modified PSO-based ANFIS to construct PIs is illustrated in Section Four. Data and experiment are presented in Section Five. Section Six concludes this paper and introduce future work.

II. Related work

First, the implementation difficulties like the calculation of Jacobian and Hessian matrix for each parameter exist when adopting delta and Bayesian methods. The requirements of updating matrix during each iteration is computationally expensive. Second, special assumptions about data distribution also complicate the computation. Delta method assumes that the distribution of noisy obeys normal distribution, Bootstrap method assumes that a less biased estimation of unknown true regression will be offered by an ensemble of NN models as well. These assumptions result in the enormous quantity of computational requirements which are hard to be realized easily.

LUBE Method is proposed to handle problems mentioned above by developing a feed-forward neural network (or universal approximators) with two outputs to directly generate the upper and lower bounds of PIs. It makes no assumption about the data distribution and avoid the calculation of derivatives such as Jacobian matrix and Hessian matrix. Here, we summarized the advantages of some LUBE Methods provided in previous literature:

Khosravi et al.. proposed Adaptive Neurofuzzy Inference Systems (ANFIS) with Simulated Annealing (SA) Algorithm [4]. In spite of sharing the similar structure with neural networks (NNs), ANFIS has its own primary benefit that the consequent parameters used in hidden layers are linear. Different from NN which has high computational mass, ANFIS which is less computational expensive can generate

978-1-7281-7106-7/20/$31.00 ©2020 IEEE
DOI 10.1109/CDS49703.2020.00025

high quality PIs in a short time. SA method is a Derivative-free optimization method which means requiring no information about gradient and higher derivatives of the cost function. SA are not limited by the existence of local minima by using cooling temperature to effectively reduce the probability to accept a new solution which increases the cost function.

H. Quan et al. proposed Neural Networks (NN) with Particle Swarm Optimization (PSO). PSO-based LUBE method is very efficient in constructing PIs with good performance by replacing width index with a new one and using PSO with mutation operator to minimize the cost function and adjust NN parameters. PSO is powerful for optimize the parameters, especially NN connection weights. Mutation operator is capable to obtain variety in Differential Evolution Algorithm (DEA). After integrating mutation operator into PSO, the improvement of searching ability can assist to jump out of local optima to the greatest extend.

H. Zhang et al. proposed Artificial Neural Networks (ANN) with Shuffled Complex Evolution Algorithm (SCE-UA) [3]. When addressing peculiar problems with hydrological processes, we hope to obtain PIs with high degree of symmetry relating to the target variables, which is difficult to realize for the nonlinearity and complexity. ANN combined with SCE-UA can cope with this problem most greatly, because SCE-UA has a good performance in handling high parameter dimensionality without relying on the information of an explicit expression for the objective function or the derivatives (Derivative-free optimization method which we interpreted in 1).

I.M. Galván et al. proposed Neural Networks (NN) with Multiple Objective Particle Swarm Optimization (MOPSO). Several improved versions of PSO has been applied to multi-objective optimization, where MOPSO is one of the most popular versions. Users have to commit on the relative importance (i.e. the weight which results depend on) of each goal before the optimization. A different weight will be adopted during the re-run processes if final PI does not satisfy the requirement of users. Since a Pareto front of non-dominated solution can be provided after a single run of MOPSO, it reduces the need of defining the weight that determines the trade-off between two goals. MOPSO shares another advantage that the prediction module can be adjusted based on different level of PINC with no repeat processes of optimization algorithm.

III. Prediction Intervals Indices

In order to assess the performance of PI, there are some evaluation indices describing the quality of PI through different statistical aspects.

Prediction Interval Coverage Probability (PICP) indicating the probability that the targets lie within the constructed PIs is calculated as:

$$\text{PICP} = \frac{1}{n}\sum_{i=1}^{n} c_i \quad (*\,100\%) \tag{1}$$

where

$$c_i = \begin{cases} 1, y_i \in [L_i, U_i] \\ 0, otherwise \end{cases}$$

L_i and U_i are lower and upper bound of i^{th} PI (use PI_i represents it).

It is easy to get the conclusion that if, $\text{PI}_i = \infty$, PICP can get the maximum value. But in this situation, PI tells no useful information about the prediction of target y_i. In order to solve the contradictory trend, we added Prediction Interval Normalized Average Width (PINAW) as basic index to improve the validity of conveying overall information about targets.

The PINAW (the same as NMPIW) is denoted as:

$$\text{PINAW} = \frac{1}{nR}\sum_{i=1}^{n}(U_i - L_i) \quad (*\,100\%) \tag{2}$$

where R is the range of target variable and calculated as

$$\text{R} = y_{max} - y_{min}.$$

Quan et al.. (2014b) used another index other than PINAW to evaluate the validity of PI which is called Prediction Internal Normalized Root-mean-square Width (PINRW). Different from PINAW, PINRW gave unequal weights to different prediction errors during the calculation with the purpose of getting more effective interval by enlarging the wider error terms.

It is defined as:

$$\text{PINRW} = \frac{1}{R}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(U_i - L_i)^2} \quad (*\,100\%) \tag{3}$$

where all parameters remain unchanged as stated before.

H. Zhang et al.. (2015) first imported target values at all the time steps to construct Prediction Interval Average Relative Width (PIARW). The benefit of this index is increasing the further information included into the evaluation indices. It denoted as follow:

$$\text{PIARW} = \frac{1}{n}\sum_{i=1}^{n}\frac{U_i - L_i}{y_i} \quad (*\,100\%). \tag{4}$$

PI which has high PICP, small PINAW or PINRW or PIARW is the optimal choice, but getting all of them at the same time is impossible. Based on this conflict, there are some cost functions try to judge the PI from various aspects using both coverage probability and width of intervals such as Coverage Width-based Criterion (CWC) defined as:

$$CWC = PINAW(1 + \gamma(PICP)e^{-\eta(PICP-\mu)}) \tag{5}$$

where η and μ are controlling hyper-paremeters determine the penalty of discontent outcomes. μ is the confidence level related with PI and can be set up to $((1-\alpha)\%)$. η magnifies the error between the PICP and μ exponentially. During the test period, $\gamma(PICP)$ is a piecewise function in relation to required PICP:

$$\gamma(PICP) = \begin{cases} 0, & PICP \geq \mu \\ 1, & PICP < \mu \end{cases}.$$

IV. Prediction interval construction based on ANFIS

A. *Adaptive-Network-based Fuzzy Inference System*

ANFIS is famous for its outstanding quality to construct an input-output mapping based on the combination of human knowledge and prescriptive input-output data pairs. In a fuzzy inference system, human knowledge and reasoning processes can be qualitative represented by adopting fuzzy if-then rules obtained via NN. In this article, we consider Takagi-Sugeno inference system with two inputs, x and y, and one output f_{out} for simplicity

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + p_1y + r_1$
Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + p_2y + r_2$

where A_i and B_i are the linguistic labels of inputs x and y, respectively, and p_i, q_i and r_i are optimal parameters of the output function. The fuzzy reasoning is illustrated in Fig. 1(a), and the basic structure is shown in Fig. 1(b).

As we described earlier, ANFIS uses NN to obtain if-then rules. Searching for the parameters can be viewed as a nonlinear regression model as follows

$$z = F(X, \omega^*) + \epsilon \tag{6}$$

where X and z are, respectively, the set of inputs (m independent variables, such as x and y in our model) and the corresponding outputs (dependent variables). $F(\cdot)$ with the parameter set ω^* (such as p_i, q_i and r_i in our model) is the true nonlinear regression function and $\epsilon \sim N(0, \sigma^2 \mathrm{I})$, I is identity matrix. The estimated parameters are denoted as $\widehat{\omega}$, optimizing $\widehat{\omega}$ to minimize the sum of squared error (SSE) cost function adopting backpropagation or hybrid training techniques:

$$SSE = (z - \hat{z})^T (z - \hat{z})$$

where $\hat{z}$ is the vector of predicted values.

Employing the standard methods to construct the asymptotic PIs for nonlinear regression models is a feasible approach to address the parameters estimation problem. $F(X, \omega^*)$ can be predicted by a first-order Taylor series in the neighborhood of the true set of parameters ω^*

$$\hat{z} = F(X, \omega^*) + J(\widehat{\omega} - \omega^*) \tag{7}$$

where the i, jth element in the Jacobian matrix J is $\frac{\partial F(X)}{\partial \omega_j}$, evaluated at the true parameter vector ω^*. Given the estimated parameter $\widehat{\omega}$, the predicted value of F at X_0 is given by

$$\widehat{z_0} = F(X_0, \omega^*) + g_0^T (\widehat{\omega} - \omega^*) \tag{8}$$

where g_0 is a vector whose ith entry is $\frac{\partial F(X_0)}{\partial \omega_j}$, calculated at the true parameter vector ω^*. As mentioned before, $\epsilon \sim N(0, \sigma^2 \mathrm{I})$ are independent, the $(1-\alpha)\%$ PI for $\widehat{z_i}$ is

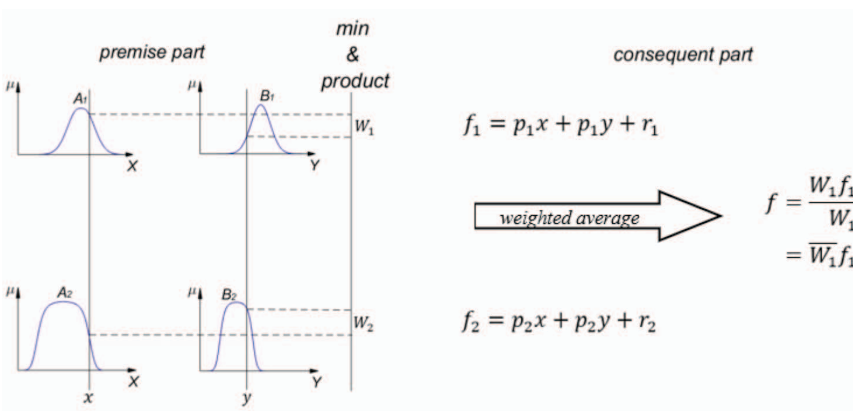

Fig. 1.(a) Takagi-Sugeno fuzzy reasoning

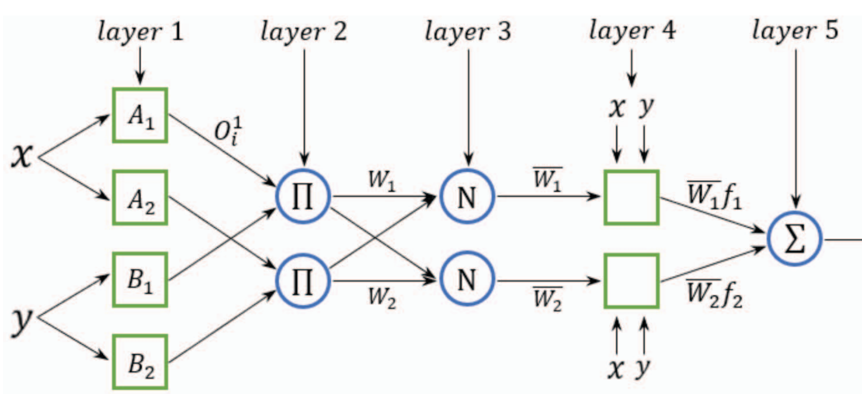

Fig. 1.(b) Equivalent ANFIS

$$\widehat{z_0} \pm t_{n-p}^{1-\frac{\alpha}{2}} s \sqrt{1 + g_0^T (J^T J)^{-1} g_0}. \qquad (9)$$

Here, n is the number of training samples and p is the number of NN parameters. $t_{n-p}^{1-\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of a cumulative t-distribution function with $n-p$ degrees of freedom. J is the Jacobian matrix described earlier. Since $E[SSE] = \sigma^2(n-p)$, an asymptotical unbiased estimation s of σ given as follows:

$$s^2 = \frac{SSE}{n-p}$$

Especially, adding a relatively small value $(J^T J + \lambda I)^{-1}$ to (9) can cope with the singularity problem. Since the following part of each rule in ANFIS model is a linear combination of the input variables and a constant, derivatives can be less calculative expensive compared with NN models.

B. PSO-Based ANFIS

The ANFIS models complicated nonlinear systems by employing fuzzy MFs to segment input dimensions; the input space is split into certain local region and covered by MFs with overlapping. Adopting simple local models, the number of MFs and layers in ANFIS largely determine the ANFIS approximation ability by carrying out space partitioning. Usually $\mu_{A_i}(x)$ is bell-shaped ranging from 0 to 1 such as:

$$\mu_{A_i}(x) = \frac{1}{1+[(\frac{x-c_i}{a_i})^2]^{b_i}} \qquad (10)$$

or the Gaussian function

$$\mu_{A_i}(x) = exp\,[-(\frac{x-c_i}{a_i})^2] \qquad (11)$$

where $\{a_i, b_i, c_i\}$ (or $\{a_i, c_i\}$) is the parameter set.

Traditional approaches to search for parameters more calculative expensive than the particle swarm optimization (PSO). PSO is a population-based search algorithm with good quality of calculating. It uses fitness function to select the optimal position for individual particle and for global particles. In this paper, we combine GA operator and PSO algorithm to optimize the premise and consequent

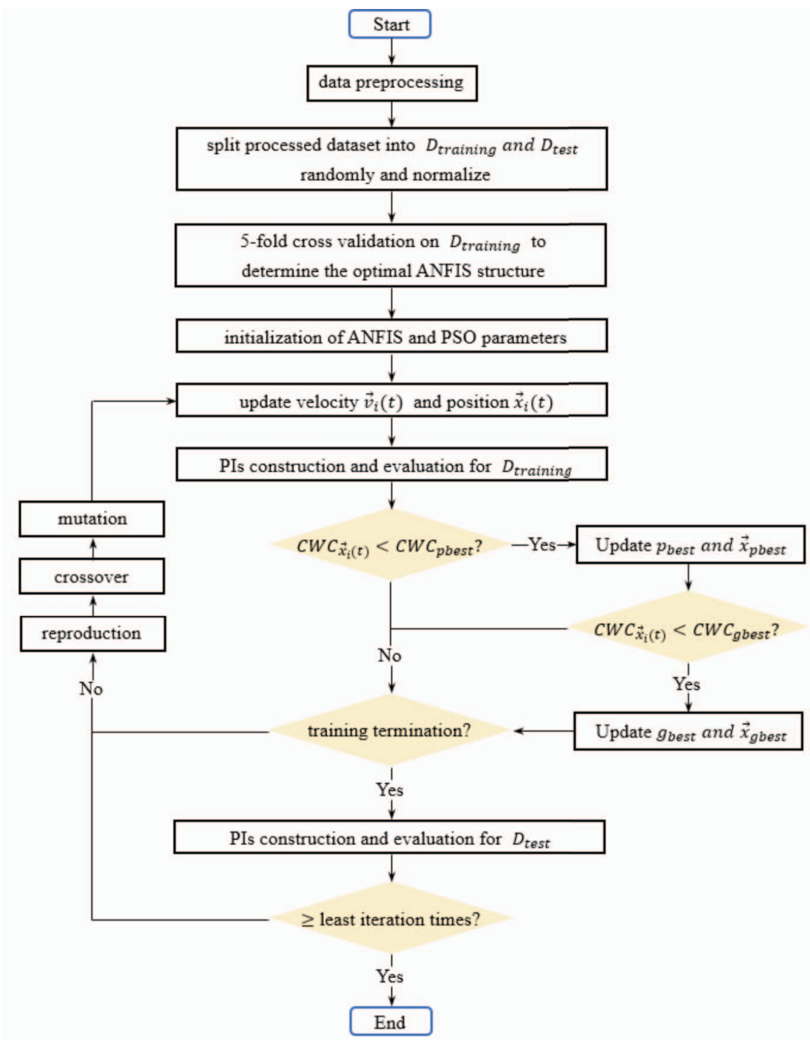

Fig. 2. Modified PSO-based ANFIS for construction and evaluation of PIs

parameters. Fig.2 illustrates the flow chart of our modified PSO-based ANFIS approach.

V. Experiment

A. Dataset

One real-world mechanical property forecasting case study with 26500 samples is conducted and employed to validate the effectiveness of the method demonstrated in Section Four. Sample points of mechanical properties lie at the end of the strip. Filter parameters as follow: select 16 chemic al ingredients including C, Mn, P, S, Cr, Al, O, Nb, V, Ti and Mo, temperature of strip tail at the inlet of each finishing mi ll, rolling force and significant technological parameters such as the thickness of raw material, tapping temperature, terminal rolling thickness, speed and temperature and coiling temperature. Screen out 68 parameters relating to mechanical properties of hot rolled strip steel totally as inputs, then forecast yield strength as output.

The number of hidden neuron for dataset YIELD is searched between [1, 20] for the complexity of problem.

B. Methodology of Case Study

Certain necessary data preprocessing, such as detecting outliers and normalization, is performed prior to modelling for the sake of avoiding the impact on subsequent modeling resulted from data abnormality and dimensional inconsistence.

Apply 5-fold cross validation to the dataset to train and validate each candidate ANFIS. Adopting mean value of CWC for five validation sets as the measurement of the quality of each candidate ANFIS structure. Employing structure with the smallest mean value of CWC in the consequent training. The basic ANFIS structure is shown above. The difference between each candidate ANFIS structure is the number of hidden neurons. After determine the number of hidden neurons, construct PIs and adopt PSO method to optimize the parameter set of ANFIS. When the required CWC and the training termination is met, the training process terminates. The updated $\vec{x}_{gbest}$ is used as parameter of ANFIS to construct the optimal PIs and calculate the PICP, PINAW, PINRW and CWC. These indices are employed to evaluate the quality when apply PIs to test samples. All parameter sets of PSO and the CWC cost function are shown in TABLE I.

TABLE I. Parameters for PSO and CWC

Parameter	Numerical value
PSO	
c_1	1.22
c_2	1.49
ω_{max}	0.9
ω_{min}	0.7
CWC	
μ	0.9
η	200

C. Test Results

Fig. 3(a)-(d) consisting of two diagrams shows the outcomes for four times of D_{test} samples. Left is the point forecast output of the PSO-based ANFIS model and expectedoutput of the true dataset. Right is the constructed PIs for YIELD and the targets of the dataset. The calculated PI indices are shown in TABLE II.

TABLE II. PI Characteristics for test set of the Mechanical Property Forecasting Case Study

Times \ Targets	PICP(%)	PINAW(%)	PINRW(%)	CWC
1	98.00	32.08	33.63	32.08
2	99.00	27.12	29.74	27.12
3	99.00	38.66	40.22	38.66
4	99.00	32.04	33.12	32.04

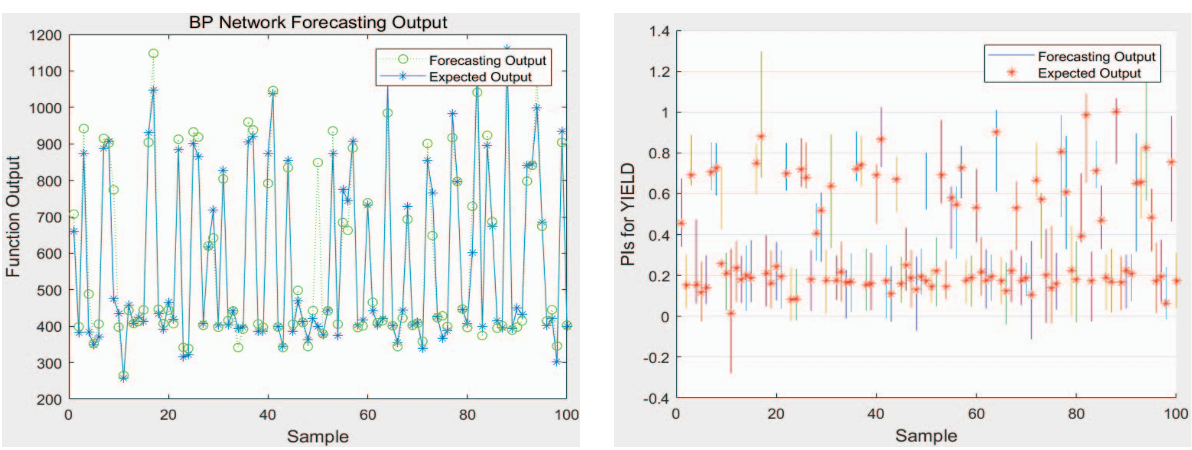

Fig. 3.(a)

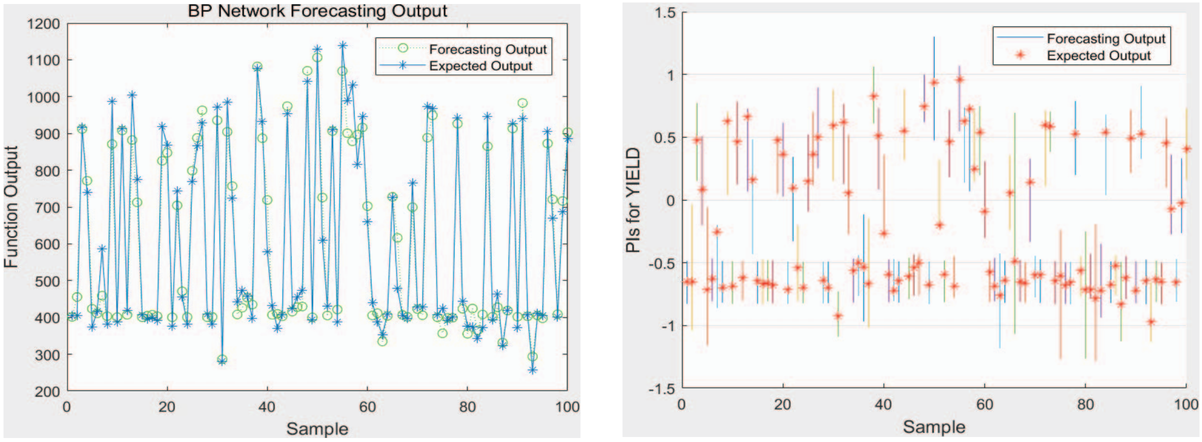

Fig. 3.(b)

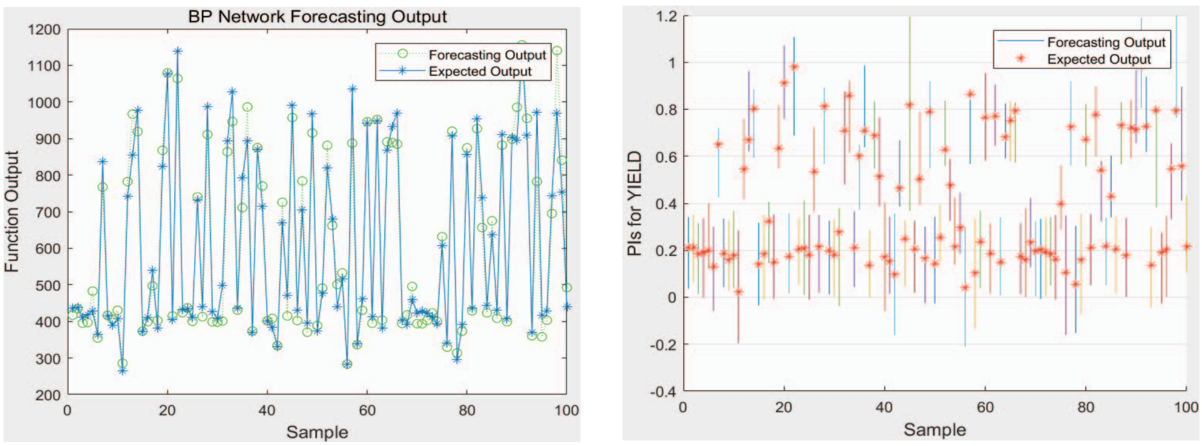

Fig. 3.(c)

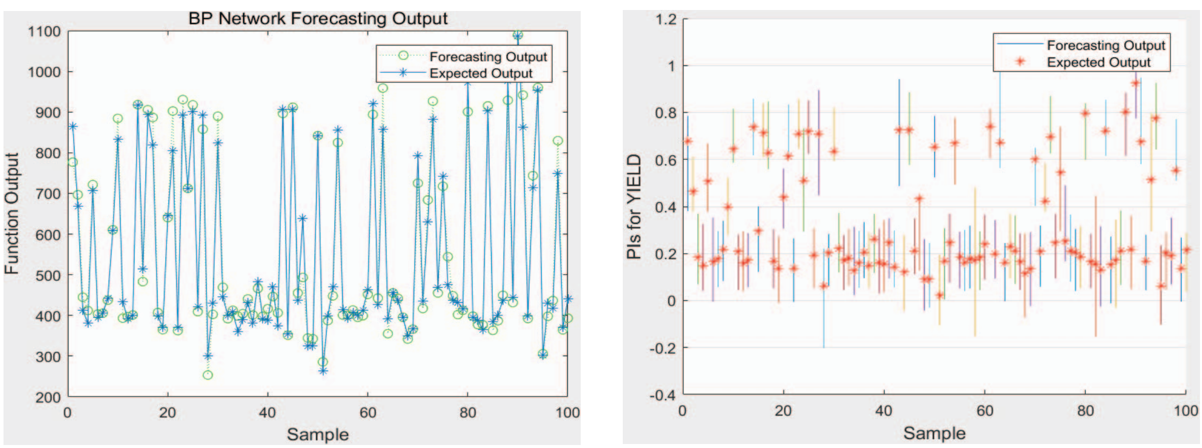

Fig. 3.(d)

For each run of the test data, almost every sample lie within the constructed PIs, which also can be implied from the PICP indices. As we can see in Tab. 2, for all outcomes, we attain PICP $\geqslant 95\%$ which satisfy the assigned confidence level. The satisfactory results indicate that using CWC as cost function and adopting PSO in optimizing the parameters can improve the quality of PIs.

From another aspect, PINAWs of all runs of the test data are less than 40%, which means that all constructed PIs conveys highly effective information about the targets. We can find in right diagrams that most sample are predicted to lie within a narrow interval, but certain samples can have much wider range. These large PIs indicates the inconsistence between training set and test set which leads to the increase of PINAW. Thus, PINAW can effectively express the uncertainty of the dataset. In order to evaluate the stability and effectiveness of the results obtained by proposed approach, we randomly choose 1900 samples as the training set and 100 samples as the testing set from totally 26500 samples up to 20 times. In spite of this, we use median values of PICP, PINAW, PINRW and CWC instead of the greatest results to ensure the outcomes are convincing.

Therefore, modified PSO-based ANFIS possesses the ability to generate narrower width but higher coverage probability PIs in mechanical property forecasting.

VI. Conclusion

Because of the unreliability and limit of information conveyed by point forecasting, PI is proposed in this article to generate a range which targets most likely lie within. PI is a promising tool in mechanical property forecasting for its good quality of providing more reliable and informative outcomes and quantifying the uncertainty. By adopting the modified PSO algorithm in constructing of PIs and optimizing for ANFIS models in this paper, the calculation is less expensive and the outcomes are more satisfactory. The uncertainty of PIs is represented by PICP and some other characteristics. In order to evaluate the quality of constructed PIs comprehensively, CWC consisting of PICP and PINAW is employed.

Four real datasets with sixteen inputs and one output of mechanical properties is used to validate the method in this paper. An ANFIS with one hidden layer is adopted to prove the effectiveness of the method and accuracy of PIs constructed via the novel approach. All datasets obtain the content outcomes with high PICP over 95% and low PINAW less than 40%. Thus, the experiments illustrate that modified PSO-based ANFIS outperforms traditional algorithms in constructing PIs for mechanical property forecasting.

By applying the method proposed in this research to real-world steel produce process, the amount of property research experiments can be cut down sharply and new products development cycle and delivery cycle can be shortened. Furthermore, candidate structures with more hidden layers of ANFIS will be studied in future work to improve the quality of PI.

References

[1] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, “Lower upper bound estimation method for construction of neural network-based prediction intervals,” IEEE Trans. Neural Netw., vol. 22, pp. 337–346, December 2010.

[2] A. Khosravi, S. Nahavandi, and D. Creighton, “Construction of optimal prediction intervals for load forecasting problems,” IEEE Trans. Power Syst., vol. 25, pp. 1496-1503, September 2010.

[3] H. Zhang, J. Zhou, L. Ye, X. Zeng, and Y. Chen, “Lower upper bound estimation method considering symmetry for construction of prediction intervals in flood forecasting,” Water. Resour. Manag., vol. 29, pp. 5505-5519, September 2015.

[4] A. Khosravi, S. Nahavandi, and D. Creighton, “Prediction interval construction and optimization for adaptive neurofuzzy inference systems,” Trans. Fuzzy Syst., vol. 19, pp. 983-988, November 2011.

[5] J.-S. Jang, “ANFIS: Adaptive-network-based fuzzy inference system,” IEEE Trans. Syst., Man, Cybern., vol. 23, pp. 665–685, May/Jun 1993.

[6] J.-S. Jang, “Fuzzy modeling using generalized neural networks and kalman filter algorithm,” AAAI, Anaheim, CA, USA, vol. 91, pp. 762–767, Augest 1991.

[7] M. A. Hosen, A. Khosravi, S. Nahavandi, and L. Sinnott, “Prediction interval-based ANFIS controller for nonlinear processes,” Int. Joint Conf. Neural Networks, pp. 4901–4907, 2016.

[8] T. Xie, G. Peng, and H. Wang, “Interval Construction and Optimization for Mechanical Property Forecasting with Improved Neural Networks,” UKCI., Springer, Cham, pp. 223-224, September 2019.

[9] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, “Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting,” IEEE Trans. Power Syst., vol. 26, pp. 137-144, March 2011.

[10] V. S. Ghomsheh, M. A. Shoorehdeli, and M. Teshnehlab, “Training ANFIS structure with modified PSO algorithm,” IEEE MCCA., Athens, Greece, pp:1-6, June 2007.

[11] H. Quan, D. Srinivasan, and A. Khosravi, “Particle swarm optimization for construction of neural network-based prediction intervals,” Neurocomputing, vol. 127, pp. 172-180, March 2014.

[12] J. T. G. Hwang and A. A. Ding, “Prediction intervals for artificial neural networks,” J. Amer. Stat. Assoc., vol. 92, pp. 748–757, 1997.

[13] R. D. de Veaux, J. Schumi, J. Schweinsberg, and L. H. Ungar, “Prediction intervals for neural networks via nonlinear regression,” Technometr., vol. 40, pp. 273–282, 1998.