Time Series Prediction using Ensembles of Neuro-Fuzzy Models with interval type-2 and type-1 fuzzy integrators

Jesus Soto, Patricia Melin, Oscar Castillo

Abstract—This paper describes an architecture for Ensembles of Neuro-Fuzzy models with interval type-2 and type-1 fuzzy integrators, with emphasis on its application to the prediction of time series, where the objective is obtained the goal is to minimize the prediction error. The time series that was considered is the Mackey-Glass. The methods used for the integration of the ensembles of neuro-fuzzy (we used the ANFIS models "adaptive network based fuzzy inference system") are: integration by average, the integration by weighted average, interval type-2 and type-1 fuzzy inference systems (FIS) integrators. The performance obtained with this architecture overcomes several standard statistical approaches and neural network models reported in the literature by various researchers.

I. INTRODUCTION

In the state of the art of artificial intelligence it is mentioned that hybrid systems that combine fuzzy logic, neural networks, genetic algorithms and expert systems provide the most efficient methods for solving a variety of problems.

Each of these techniques has particular computational properties that make them ideal for solving certain problems. One of these combinations corresponds to hybrid neuro-fuzzy systems that combine artificial neural network techniques and fuzzy inference techniques.

System modeling based on conventional mathematical tools (e.g., differential equations) is not well suited for dealing with ill-defined and uncertain systems. By contrast, a fuzzy inference system employing fuzzy if-then rules can model the qualitative aspects of human knowledge and reasoning processes without employing precise quantitative analyses [1].

This fuzzy modeling or fuzzy identification, first explored systematically by Takagi and Sugeno [2], has found numerous practical applications in control, prediction and inference. However, there are some basic aspects of this approach, which are in need of better understanding. More specifically:

1) No standard methods exist for transforming human knowledge or experience into the fuzzy rule base of a fuzzy inference system.

2) There is a need of effective methods for tuning the membership functions (MF’s) so as to minimize the output error measure or maximize a performance index.

In this perspective, the aim of this paper is to suggest a new architecture for Ensembles of ANFIS models with integration by type-1 and interval type-2 FIS, which can serve as a basis for constructing a set of type-1 and type-2 fuzzy if-then rules with appropriate membership functions to generate the stipulated input-output pairs. The architecture of adaptive fuzzy inference systems proposed in [3] uses the gradient descent backpropagation [4] and least squares to achieve the learning ability of ANFIS.

This paper reports the results of the simulations, in which the Ensembles of neuro-fuzzy models with interval type-2 and type-1 fuzzy integrators and were used to predict chaotic time series of Mackey-Glass [5,6], therefore the results obtained to each integration models were evaluated for minimize the prediction error, by the metric for the calculation used the root mean square error (RMSE).

The selection of the time series for the simulations was based on the fact that these time series are widely quoted in the literature by different researchers [5,7-13], which allows to compare results with other approaches such as neural networks and linear regression.

In the next section we describe a brief overview of ANFIS, the proposed architecture of ensemble of ANFIS and Ensemble learning. Section III presents the interval type-2 fuzzy logic systems, Section IV presents the chaotic time series that we used for the experiments. Section V presents the simulations and the results obtained with different methods of integration that are used in this work. Section VI presents the conclusions.

II. ANFIS MODELS AND PROPOSED ARCHITECTURE OF ENSEMBLES OF ANFIS

This section presents a brief review of ANFIS, to give us an idea of how the operation of ANFIS really is. In the second step we consider the Ensemble learning, for a better understanding of how they are applied in the proposed method. In the last part we describe the architecture or structure that we have proposed for the Ensembles of ANFIS.

A Brief Overview of ANFIS

The proposed architecture is based on "ANFIS" as referred to by R. Jang [3,4,10], which stands for adaptive neuro fuzzy inference system, and describes how to decompose the parameter set to facilitate the hybrid learning rule for ANFIS architectures representing both Sugeno and Tsukamoto fuzzy models.

For simplicity, consider a fuzzy inference system with
two inputs $x$, $y$ and one output $z$. However consider two fuzzy if-then rules of Takagi-Sugeno type [14]:

Rule 1: If $x$ is $A_1$ and $y$ is $B_1$, then $f_i = p_1 + p_2 y + r_1$

Rule 2: If $x$ is $A_2$ and $y$ is $B_2$, then $f_i = p_3 + p_4 y + r_2$

Fig. 1 (a) and (b) illustrate the reasoning mechanism and the corresponding ANFIS architecture, respectively. The operations of the nodes in the same layer of ANFIS are of the same family of functions, as described below. (In subsequent, $O_i^j$ denotes the $i$th output node of the layer $j$.

Layer 1: Each node in this layer corresponds to a linguistic label and the output node is equal to the value of membership in this linguistic label. The parameters of a node can change the shape of the membership function used to characterize the linguistic label. For example, the function of the $i$th node is given by

$$O_i^1 = \mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x-c}{a_i}\right)\right]^m}$$

where $x$ is the input node; $A_i$ is a linguistic label like, Small, Large, etc. Associated with this node, and $\{a_i, b_i, c_i\}$ is the set of parameters. Parameters in this layer are called premise parameters.

Layer 2: each node in this layer calculates the firing power of each rule:

$$O_i^2 = \omega_i = \mu_{A_i}(x) \times \mu_{B_i}(y), i = 1, 2. \quad (2)$$

Layer 3: the $i$th node of this layer calculates the ratio of the firing strength of the $i$th rule to the sum of all the firing strengths:

$$O_i^3 = \tilde{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, i = 1, 2 \quad (3)$$

Layer 4: node $i$ in this layer has the following node function

$$O_i^4 = \tilde{\omega}_i f_i = \tilde{\omega}_i [(p_i x + q_i y) r_i] \quad (4)$$

Layer 5: the single node in this layer computes the overall output as the sum of all input signals:

$$O_i^5 = \Sigma \tilde{\omega}_i f_i = \Sigma \tilde{\omega}_i [(p_i x + q_i y) r_i] \quad (5)$$

In this way we have built an adaptive network (Fig. 1 (b), which is functionally equivalent to an inference system (Fig. 1 (a)), so it is called ANFIS, referring to fuzzy inference systems based on adaptive networks. The basic learning rule of ANFIS is the gradient descent backpropagation [12], which calculates the error rates (defined as the derivative of the squared error for each output node) recursively from the output to the input nodes.

This is the same learning rule used in neural networks [13]. In the basic architecture of Fig. 1 it may be noted that given the values of premise parameters, the total output $f$ can be expressed as linear combinations of the parameters accordingly:

$$f = \tilde{\omega}_1 f_1 + \tilde{\omega}_2 f_2 = (\tilde{\omega}_1 p_1 + (\tilde{\omega}_2 q_1 + (\tilde{\omega}_2) r_1$$

As a result we have a hybrid learning algorithm [10], which combines the gradient descent and least-squares estimation. More specifically in, the forward step of the hybrid learning algorithm, functional signals (output nodes) are processed towards layer 4 and the parameters of consequence are identified by least squares. In the backward step the premise parameters are updated by gradient descent.

Proposed the Architecture of the Ensemble of ANFIS

In this paper, the proposed the architecture of Ensemble of ANFIS is illustrated in Fig. 2.

![Fig. 2. The architecture of Ensemble of ANFIS](image).

This architecture is divided into 5 sections, where the first section represents the data base to simulate in the Ensemble of ANFIS, which in our case is the data base of the Mackey-Glass [5,10] time series. In the second section, training and validation is performed sequentially in each ANFIS, where the number of ANFIS to be trained can be from 1 to n depending on what the user wants to test, but in our case we are dealing with a set of 3 ANFIS in the Ensemble.

In the third section we have to generate the results of each ANFIS trained in the previous section and in the fourth section we integrate the overall results of each ANFIS, such integration will be done by direct average and weighted average and finally the outcome or the final prediction of the Ensemble ANFIS learning is obtained.

The main idea of Ensemble learning in ANFIS, is that each ANFIS has different ways to train and thus be simulated and this makes something like an expert system, i.e., give different viewpoints and then predict the time series, then take decisions based on the results of each ANFIS and reach the conclusion, then integrating the results and obtaining the best prediction of the time series is being simulated, which can avoid future unexpected events.
A. Ensemble Learning

The Ensemble consists of a learning paradigm where multiple component learners are trained [15] for a same task, and the predictions of the component learners are combined for dealing with future instances [2,17]. Since an Ensemble is often more accurate than its component learners, such a paradigm has become a hot topic in recent years and has already been successfully applied to optical character recognition, face recognition, scientific image analysis, medical diagnosis [16].

In this paper, the learning of the total output of each Ensemble was obtained with the integrator by the average, the weighted average, type-1 and interval type-2 fuzzy integrators, just as the output of each ANFIS is calculated with the root mean squared error RMSE (9).

III. INTERVAL TYPE-2 FUZZY LOGIC SYSTEM

The interval type-2 fuzzy system was proposed for Type-2 fuzzy logic and can handle uncertainty because it can model and reduce it to the minimum of their effects. Also, if all the uncertainties disappear, type-2 fuzzy logic reduces to type-1 fuzzy logic, in the same way that, if the randomness disappears, the probability is reduced to the determinism [14,18]. Fuzzy sets and fuzzy logic are the foundation of fuzzy systems, and have been developed looking to model the form in which the brain manipulates inexact information. Type-2 fuzzy sets are used to model uncertainty and imprecision; originally they were proposed by Zadeh [19,20] and they are essentially “fuzzy–fuzzy” sets in which the membership degrees are type-1 fuzzy sets [4,21-25] shown in Fig. 3.

The basic structure of a type-2 fuzzy system implements a nonlinear mapping of input to output space. This mapping is achieved through a set of type-2 if-then fuzzy rules, each of which describes the local behavior of the mapping.

The uncertainty is represented by a region called footprint of uncertainty (FOU). When $\mu_3(x,u) = 1$, $\forall u \in I_x \subseteq [0,1]$; we have an interval type-2 membership function, as shown in Fig. 4.

The uniform shading for the FOU represents the entire interval type-2 fuzzy set and it can be described in terms of an upper membership function $\mu_u(x)$ and a lower membership function $\mu_l(x)$.

A FLS described using at least one type-2 fuzzy set is called a type-2 FLS. Type-1 FLSs are unable to directly handle rule uncertainties, because they use type-1 fuzzy sets that are certain [26,27]. On the other hand, type-2 FLSs, are very useful in circumstances where it is difficult to determine an exact certainty value, and there are measurement uncertainties [25].

IV. CHAOTIC TIME SERIES

The problem of predicting future values of a time series has been a point of reference for many researchers. The aim is to use the values of the time series known at a point $x = t$ to predict the value of the series at some future point $x = t + P$. The standard method for this type of prediction is to create a mapping from $D$ points of a $\Delta$ spaced time series, is $(x(t - (D - 1) \Delta), ..., x(t - \Delta), x(t))$, to a predicted future value $x(t + P)$. To allow a comparison with previous results in this work [12,13,17,22,28] the values $D = 4$ and $\Delta = P = 6$ were used.

One of the chaotic time series data used is defined by the Mackey-Glass [5,6] time series (5), whose differential equation is given by:

$$\frac{dx(t)}{dt} = \frac{0.2x(t - \tau)}{1 + x^{10}(t - \tau)} - 0.1x(t) \quad (7)$$

For obtaining the values of the time series at each point, we applied the Runge-Kutta method for the solution of equation (7). The integration step was set at 0.1, with initial condition $x(0) = 1.2, \tau = 17, x(t)$ is then obtained for $0 \leq t \leq 1200$, which is illustrated in Fig. 5 (We assume $x(t) = 0$ for $t < 0$ in the integration.)
From the Mackey-Glass time series we extracted 800 pairs of data points [10,11,29,30], similar to [28]. We predict $x(t)$ from the three past values of the time series, that is, $x(t-18)$, $x(t-12)$, and $x(t-6)$. Therefore the format of the training data is

$$\begin{array}{c}
\{x(t-18), x(t-12), x(t-6), x(t)\} \\
\end{array}$$

(8)

Where $t = 19$ to 818 and $x(t)$ is the desired prediction of the time series. The first 400 pairs of data are used to train the ANFIS, while the other 400 pairs of data are used to validate the model identification.

V. SIMULATIONS RESULTS

This section presents the results obtained through experiments on the architecture of ensembles of ANFIS with the type-1 FLS integrator and an interval type-2 FLS integrator, which show the performance that was obtained from each experiment to simulate the Mackey-Glass time series.

In these best Ensembles we have a set of 3 ANFIS for learning, it is noteworthy that the type of membership functions was assigned differently to each ANFIS and the desired goal error was assigned to each Ensemble of 0.01 to 0.000001 and the prediction error of each ANFIS is calculated by the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (a_i - p_i)^2}{n}}$$

(9)

where $a_i$ correspond the real data of time series, $p_i$ correspond of the output of each fuzzy integrators, $t$ is the sequence time series, and $n$ is the numbers of data points of time series.

A. Simulation results of Integration by Average

This integration method is the simplest and most straightforward and consists in the sum of the results generated by each ANFIS (Outputs) divided by the number of ANFIS, and the disadvantage is that there are cases in which the prognosis is not good. The result of this time series in shown in Fig. 6, shows the comparison between the desired prediction and output data generated by the integration by average. Therefore the error of forecast obtained for this method is 0.0174.

B. Simulation results of Integration by Weighted Average

This method of integration by weighted average, in which a weight was given depending on the results obtained from each ANFIS, these weights were assigned manually, where the Lowest error between the ANFIS output receives a weight of 0.50, the ANFIS that had on intermediate error 0.30 and the biggest error was assigned a weight of 0.20, thus obtaining 100% of the weights assigned among the models of the ANFIS Ensemble. The result of this time series are shown in Fig. 7, which shows the comparison between the desired prediction and output data generated by the integration by weighted average, which shows the prediction error between the desired data and data predicted by integrative of weighted average. The error obtained for this method is 0.0173.

C. Simulation results of Integration by Type-1 FIS

For the design of this integration method, we rely on the Mamdani type fuzzy model, so this model will receive as input each of the outputs generated by integrating the Ensemble of ANFIS. To design the fuzzy inference system we went about assigning different membership functions for the inputs and outputs of the system to observe the performance of the system. The best FIS generated for the integration of the ensemble is as follows: The FIS was assigned Generalize Bell membership functions in the inputs (prediction 1, 2 and 3) and outputs (forecast) system, so that prediction error obtained is 0.1112.

The result of this time series is shown in Fig. 8, and shows
the comparison between the desired prediction and actual output data generated by the integration by type-1 FIS.

![Fig. 8. Desired Prediction and Output prediction of integrator by type-1 FIS Mamdani.](image)

### D. Simulation results of Integration by Interval Type-2 FIS

The design of the integrator by interval type-2 FIS is also represented as Mamdani fuzzy model, which was assigned different membership functions to observe the performance of the system and to prove which of the two types of FIS is best for the forecast time series. Therefore this model will receive as input each of the outputs generated by integrating the Ensemble ANFIS. The best FIS generated for the integration of the ensemble is as follows: This FIS was assigned Gaussian membership functions in inputs (prediction 1, 2 and 3) and outputs (forecast) system, so that prediction error obtained 0.0242. The result of this time series shown in Fig. 9, which shows the comparison between the desired prediction and output data generated by the integration by type-2 FIS.

![Fig. 9. Desired Prediction and Output prediction of integrator by interval type-2 FIS Mamdani.](image)

Table I shows the best 5 results of 20 experiments that we tested for the architectures of ensembles of ANFIS, which can be seen in the ensemble 2 is the best of all of them, because for the integration by average obtained an error of 0.0174, for integration by weighted average the error obtained is of 0.0173, for the integration of type-1 FIS the error obtained was 0.12043 and integration interval type-2 FIS the obtained prediction error was 0.04933. Therefore we considered that interval type-2 FIS is better than type-1 FIS because the handling of uncertainty in the membership function is better and therefore there range of dataset of the time series.

<table>
<thead>
<tr>
<th>Ensemble</th>
<th>Integration by Average</th>
<th>Integration by Weighted Average</th>
<th>Integration by Type-1 FIS</th>
<th>Integration by Type-2 FIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0183</td>
<td>0.12073</td>
<td>0.0494</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0174</td>
<td>0.0173</td>
<td>0.04933</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0176</td>
<td>0.12070</td>
<td>0.04957</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0183</td>
<td>0.0182</td>
<td>0.04973</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0178</td>
<td>0.0175</td>
<td>0.12067</td>
<td>0.0496</td>
</tr>
</tbody>
</table>

### VI. CONCLUSION

In conclusion we can say that the results obtained with the proposed architecture of Ensembles of ANFIS have been good and positive in predicting time series (like the Mackey-Glass), as it has managed to minimize the prediction error of the time series.

We conclude that the results obtained with the architectures of ensembles of ANFIS are good, since we achieved 98.5% of accuracy in the forecast with the Mackey-Glass time series and on the opposite side it was obtained prediction error of 1.5%. Therefore, we can conclude that type-2 FIS integrator is better than the type-1 FIS integrator, because in most of the experiments that were performed with the proposed architecture of ensembles of ANFIS, the interval type-2 FIS gave better prediction errors than the type-1 FIS. Therefore the proposal offers efficient results in the prediction of such time series, which can help us make decisions and avoid unexpected events in the future.

The future work will consist in optimizing the architecture of the ensembles of ANFIS, since the objective of this research is to find the best architecture and to know the optimum size that each ensemble of ANFIS will have. The optimization methods will use genetic algorithms (GA) and particle swarm optimization (PSO). The other future work will consist in optimize the membership functions and fuzzy rules of the ANFIS model to minimize the prediction error, also to use GA and PSO, as using these methods we can statistically compare which gives better performance in predicting the time series Mackey-Glass.

### REFERENCES

