International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 4, Issue 6, November - December 2015 ISSN 2278-6856

A hybrid method of prediction based on Soft Computing Techniques

R.K.Srivastava[1], Atib Khan[2]

[1]Head, Department of Computer ScienceDr Shakuntala Misra Rehabilitation University
Lucknow, India

[2]Research Scholar,Shri Venkateshwar University, Amroha(U.P)

Abstract

In Soft Computing Techniques based prediction various methods have been developed to establish the hybrid model for forecasting the future values, the present paper proposes a hybrid method of soft computing i.e; the combination of artificial neural networks and fuzzy set theory techniques. The study uses the fuzzy set theory Zadeh(1965) and fuzzy time series models introduced by Song and Chissom(1993) and Chess(1993). The prediction has also been obtained by developing an Artificial Neural Network model using back propagation algorithm. The study is aimed to find the forecast for primary education literate population of Lucknow district for a lead year by using a fuzzy time series model and back propagation algorithm. The success of the proposed model has been examined by comparing the result of proposed model with actual primary educated population of Lucknow district.

Keywords: about Fuzzy time series, Fuzzy sets, Linguistic value, Back propagation algorithm

1. INTRODUCTION

Soft computing technique for forecasting emerged as a latest approach for forecasting the future values where exact time series data of many process neither governed by mathematical function nor by probability distribution. Fuzzy set theory is an intellectual adventure in which the philosophy of mathematics, abstraction and idealization are combined. Chen (1996) described a simplified method for time series forecasting using the arithmetic operations. Hureng(2001) developed a heuristic model for time series forecasting using heuristic increasing and decreasing relational to improve the forecast of enrollments. Chen() applied the fuzzy time series model for forecasting the enrollments and suggested to use fuzzy logical relationship group to deal with ambiguity which had been found by him. Yu() presented the refined fuzzy time series model by extending the fuzzy logical relations adding a refinement factor for forecasting and applied it on enrollments forecasting. In the present work the developed model have been implemented on the basis of past 20 years actual data of primary education enrollment of Lucknow district taken from NIC Lucknow website. It is supposed as non linear process where data is generally contains uncertainty and imprecision. The study is aimed to find reliable forecast for enrollment of primary education of Lucknow district for a lead year, which will be beneficial for future planning of basic education department for arrangement of resources in coming years. We have used neural network and fuzzy set theory techniques. ANN have been shown to be more accurate for prediction model, in this study neural network technique of soft computing as feed forward part of back propagation algorithm has been used and analyze their property to simulate complex process. We also present an improved method for fuzzy time series forecasting using a difference parameter as fuzzy relation for forecasting. The developed process has been given in simple computational algorithm and robustness of the model has been examined.

2. Fuzzy time series models

Let Y(t) (t = …, 0,1,2,..), is a subset of R1, be the universe of discourse on which fuzzy sets fi (t) (i = 1,2,…) are defined and F(t) is the collection of fi (i = 1,2,…). Then F(t) is called fuzzy time series on Y(t) (t = …, 0,1,2,…). Further F(t) can be understood as a linguistic variable and fi(t) (i = 1,2,…) as the possible linguistic values of F(t).

Definition 1: Suppose F(t) is caused by a F(t-1) only or by F(t-1) or (F(t-2)or…or F(t-m) (m>0).This relation can be expressed as the following fuzzy relational equation:

$$F(t) = F(t-1) \circ R(t, t-1) \quad (1)$$

or

$$F(t) = (F(t-1) \cup F(t-2) \cup \ldots \cup F(t-m)) \circ R_0 (t, t-m) \quad (2)$$

The equation is called the first order model of F(t).

Definition 2: Suppose F(t) is caused by a F(t-1), F(t-2),…, and F(t-m) (m>0) simultaneously. This relation can be expressed as the following fuzzy relational equation

$$F(t) = (F(t-1) \times F(t-2) \times \ldots \times F(t-m)) \circ R_a(t, t-m) \ldots\ldots..(3)$$

and is called the m^{th} order model of F(t).

Definition 3: If in (2) or (3) or (4), the fuzzy relation R(t, t-1) or Ra (t, t-m) or R_a (t, t-m) of F(t) is dependent of time t , that is to say for different times t_1 and t_2, $R(t_1, t_1 - 1) = R(t_2, t_2 - 1)$, or $Ra(t_1, t_1 - m) = Ra(t_2, t_2 - m)$ or $R_0(t_1, t_1 - m) = R_0(t_2, t_2 - m)$, then F(t) is called a time invariant fuzzy time series .Otherwise it is called a time variant fuzzy time series,

In the case of time invariant fuzzy time series,

$$R(t, t-1) = R,$$
$$Ra(t, t-m) = Ra(m),$$
$$R_0(t, t-m) = R_0(m)$$

In general at different times t_1 and t_2, $R(t_1, t_1 - 1) \neq R(t_2, t_2 - 1)$, $Ra(t_1, t_1 - m) \neq Ra(t_2, t_2 - m)$ and $R°(t_1, t_1 - m) \neq R°(t_2, t_2 - m)$. There are two reasons for this: first, the universes of discourse on which the fuzzy sets are defined may be different at different times: second the value of F(t) at different times may be different.

Depending upon the complexity of the system, fuzzy time series modeling for a forecast process may use type of relations $R(t, t-1)$, $R_a(t, t-m)$, $R_o(t, t-m)$. Several methods Dubois and Parde[1991], Wu[1986] and Mamdani[1977] are available to determine these relations.

4 Primary education enrollment forecasting : proposed model versus Chen's model

Fuzzy time series model deals with situation where the data are linguistic values, in contrast to the conventional time series approaches that typically manipulate numerical data. If data are available in crisp form, it is to be fuzzified before the fuzzy time series methodology can be applied. Fuzzification process starts with defining the universe of discourse U, which contains the historical data and upon which the fuzzy sets are defined. The study deals with the enrollments of number of students of primary classes for the Lucknow district (India) in various years starting from 1993-94 to 2013-2014 with assumption that it includes some vagueness incurred due to statistical sampling.

The algorithm for application of fuzzy time series in enrolment of number of students of primary classes forecasting comprises of the following steps:

Step-1: Let D_{min} and D_{max} be minimum and maximum primary education enrollments. Based upon D_{min} and D_{max}, we define the universe of the discourse U as $[D_{min}-D_1, D_{max}+D_2]$,where D_1 and D_2 are two proper positive numbers and accordingly, the universe of discourse

U= [280000, 520000].

Further the universe of discourse U is partitioned into six intervals of equal length as follows:

u_1 = [280000, 320000] u_2 = [320000,360000]
u_3 = [360000,400000] u_4 = [400000, 440000]
u_5 = [440000, 480000] u_6 = [480000, 520000]

Step 2: fuzzy sets $A_1, A_2, A_3,\ldots A_6$ on universe of discourse, having linguistic values as

A_1= not good, A_2 =not too good

A_3= satisfactory good A_4= good

A_5= fairly good A_6= very good

are to be defined . $u_1,u_2,\ldots, u_6$ are chosen as elements of these fuzzy sets.The membership grades of $u_1, u_2,\ldots, u_6$ to each A_i (i=1,2,…, 6) will decide that how well each u_k (k=1,2,…6) belong to u_i. We have determined the membership of each element in all the fuzzy sets Ai (i= 1,2,…6) and are expressed as:

A_1={ u_1/1, u_2/.5, u_3/0, u_4/0, u_5/0, u_6/0}

A2={ u_1/.5, u_2/1, u_3/.5, u_4/0, u_5/0, u6/0}

A3={ u_1/0, u_2/.5, u_3/1, u_4/.5, u_5/0, u6/0}

A4={ u_1/0, u_2/0, u_3/.5, u_4/1, u_5/.5, u6/0}

A5={ u_1/0, u_2/0, u_3/0, u_4/.5, u_5/.1,u6/.5}

A6= { u_1/0, u_2/0, u_3/0, u_4/0, u_5/.5, u6/1}

u_i (i= 1,2…6) is the element and the number below '/ ' is the membership of u_i to A_j (j= 1,2,..6)

Step 3: Fuzzify the enrollment of number of students of primary classes data to find out the equivalent fuzzy set to each year's enrollments using the step- 2 .

Step 4: Fuzzy logical relationship of the primary education enrollments have been obtained from Table-1, where the fuzzy logical relationship $A_{j1} \rightarrow A_k$ means: if the primary education enrollments of year j is A_j then that of year j+1 is A_k , where A_j is called the current state of primary education enrollments and A_k is called the next state of the primary education enrollments. The fuzzy logical relationship for primary education enrollments are derived as follows

Proposed model		Chen's model	
$A_1 \rightarrow A_1$	$A_1 \rightarrow A_1$	$A_1 \rightarrow A_1$	$A_1 \rightarrow A_3$
$A_1 \rightarrow A_1$	$A_1 \rightarrow A_1$	$A_2 \rightarrow A_1$	$A_2 \rightarrow A_2$
$A_1 \rightarrow A_1$	$A_1 \rightarrow A_3$	$A_3 \rightarrow A_4$	$A_4 \rightarrow A_4$
$A_2 \rightarrow A_1$	$A_2 \rightarrow A_2$	$A_4 \rightarrow A_5$	$A_5 \rightarrow A_5$
$A_2 \rightarrow A_2$	$A_3 \rightarrow A_4$	$A_5 \rightarrow A_6$	$A_6 \rightarrow A_6$
$A_4 \rightarrow A_4$	$A_4 \rightarrow A_4$		
$A_4 \rightarrow A_4$	$A_4 \rightarrow A_4$		
$A_4 \rightarrow A_5$	$A_5 \rightarrow A_5$		
$A_5 \rightarrow A_6$	$A_6 \rightarrow A_6$		
	$A_6 \rightarrow A_6$		

Step 5: Based on the Fuzzy logical relationship, we derive the fuzzy logical relationship groups for primary education enrollments, which comes to be

The equivalent fuzzy set to each year's enrollments are shown in table-

Table- 1

Year	Actual Enrollment Primary Education	A_1	A_2	A_3	A_4	A_5	A_6	Fuzzified Enrollment Primary Education
1993-94	320589	0.5	1	0.5	0	0	0	A_2
1994-95	335765	0.5	1	0.5	0	0	0	A_2
1995-96	345088	0.5	1	0.5	0	0	0	A_2
1996-97	289836	1	0.5	0	0	0	0	A_1
1997-98	289004	1	0.5	0	0	0	0	A_1
1998-99	287881	1	0.5	0	0	0	0	A_1
1999-2000	287881	1	0.5	0	0	0	0	A_1
2000-2001	301008	1	0.5	0	0	0	0	A_1
2001-2002	319959	1	0.5	0	0	0	0	A_1
2002-2003	387560	0	0.5	1	0.5	0	0	A_3
2003-2004	418595	0	0	0.5	1	0.5	0	A_4
2004-2005	426389	0	0	0.5	1	0.5	0	A_4
2005-2006	434856	0	0	0.5	1	0.5	0	A_4
2006-2007	430699	0	0	0.5	1	0.5	0	A_4
2007-2008	439583	0	0	0.5	1	0.5	0	A_4
2008-2009	488425	0	0	0	0.5	1	0.5	A_5
2009-2010	474351	0	0	0	0.5	1	0.5	A_5
2010-2011	505301	0	0	0	0	0.5	1	A_6
2011-2012	510899	0	0	0	0	0.5	1	A_6
2012-2013	512182	0	0	0	0	0.5	1	A_6

Proposed model	Chen's model
Group-1: $A_1 \rightarrow A_1$, A_3, A_1, A_1, A_1, A_1	Group-1: $A_1 \rightarrow A_1$, A_3
Group-2: $A_2 \rightarrow A_1$, A_2, A_2	Group-2: $A_2 \rightarrow A_1$, A_2
Group-3: $A_3 \rightarrow A_4$	Group-3: $A_3 \rightarrow A_4$
Group-4: $A_4 \rightarrow A_4$, A_4, A_4, A4, A5	Group-4: $A_4 \rightarrow A_4$, A_5
Group-5: $A_5 \rightarrow A_5$, A_6	Group-5: $A_5 \rightarrow A_5$, A_6
Group-6: $A_6 \rightarrow A_6$, A_6	Group-6: $A_6 \rightarrow A_6$

Step 6: Primary education enrollment forecast from the computed fuzzified primary education enrollment is carried by following rules:

(1) If the fuzzified enrollment of the year i is A_j , and there is only one fuzzy logical relationship in the fuzzy logical relationship groups as A_j □ A_k , where A_j and A_k are fuzzy sets and the maximum membership value of A_k occurs at interval u_k, and the midpoint of u_k is m_k , then the forecasted enrollment of year i+1 is m_k.

(2) If the fuzzified enrollment of the year i is A_j , and there are several fuzzy logical relationships are defined in logical relationship groups such as:

$$A_j \rightarrow A_{k1},\ A_j \rightarrow A_{k2},\ A_j \rightarrow A_{kp}$$

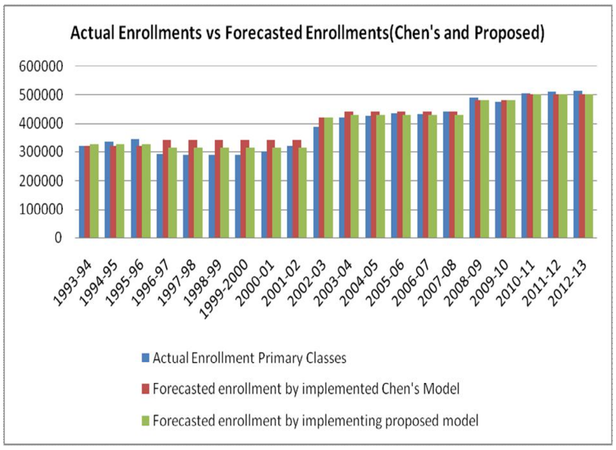
Where A_j, A_{k1}, A_{k2}, …A_{kp} are fuzzy sets, and whose elements posses the maximum membership values at intervals u_1, u_2, …u_p, respectively. If the midpoints of u_1, u_2, …u_p are m_1, m_2, …m_p respectively, then the forecasted enrollment for year i+1 is $(\sum^{p}_{x=1} m_x)/ p$

(3) If the fuzzified enrollment of a year i is A_j, and no logical relationship is found in logical reationship groups , whose current state of primary education enrollments is A_j , where the maximum membership value of A_j occurs at interval u_j ,and the midpoint of u_j is m_j , then the forecasted enrollment of year i+1 is m_j.

Based on the above rules, the forecasted primary education enrollment for various years has been calculated for both: Chen's and proposed model are given against the actual enrollment in the

Table-2

Year	Actual Enrollment Primary classes	Forecasted enrollment by implementing Chen's model	Forecasted enrollment by implementing proposed model
1993-94	320589	320000	326666
1994-95	335765	320000	326666
1995-96	345088	320000	326666
1996-97	289836	340000	313333
1997-98	289004	340000	313333
1998-99	287881	340000	313333
1999-2000	287881	340000	313333
2000-2001	301008	340000	313333
2001-2002	319959	340000	313333
2002-2003	387560	420000	420000
2003-2004	418595	440000	428000
2004-2005	426389	440000	428000
2005-2006	434856	440000	428000
2006-2007	430699	440000	428000
2007-2008	439583	440000	428000
2008-2009	488425	480000	480000
2009-2010	474351	480000	480000
2010-2011	505301	500000	500000
2011-2012	510899	500000	500000
2012-2013	512182	500000	500000



In time series forecasting, the forecasting accuracy of a model is commonly measured in terms of mean square error (MSE) or in terms of percentage error. Lower the mean square error or average percentage error, better the forecasting method.

$$\text{Mean square error} = \frac{\sum^{n}_{i=1}(\text{actual value}_i - \text{forecasted value}_i)^2}{n}$$

and forecasting error as

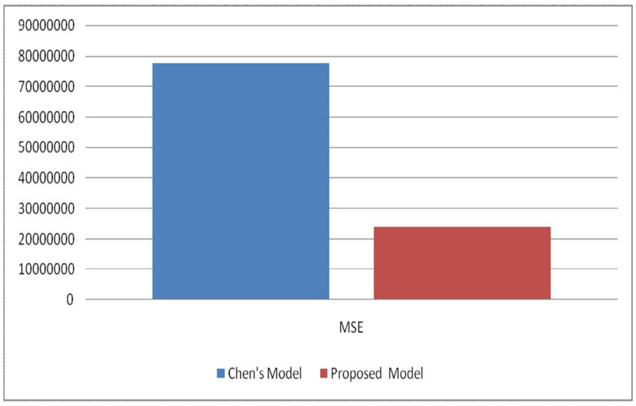
$$\text{Forecasting error(in\%)} = \frac{(\text{forecasted} - \text{actual value})}{\text{actual value}} \times 100$$

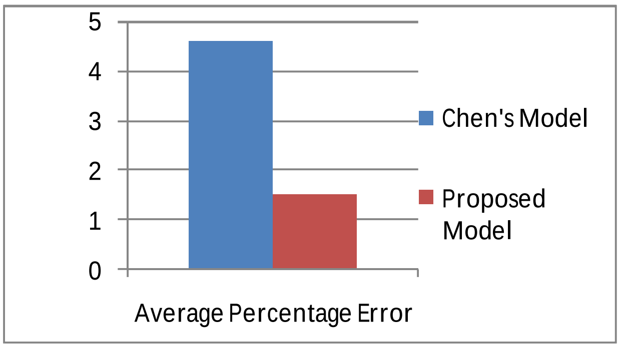
$$\text{Average forecasting error (in\%)} = \frac{\text{sum of forecasting error}}{\text{numbers of errors}}$$

In comparison of accuracy in forecasted values of our proposed model with other model, the mean square error and average forecasted error have been computed and compared with other method in Table-3.

Table-3 Comparison of MSE and Average Error:

Model	Chen's Model	Proposed Model
MSE	777860940	241227955
Average Percentage Error	4.6124	1.507





comparative study of MSE and percentage forecasted error shown in table-3 proves that the prediction by proposed method is having more accuracy than the Chen's model. The MSE and percentage average forecasted error of different developed model are measure of accuracy of predicting methods. The forecast average error is 1.507% in proposed model and 4.624% in Chen's model whereas MSE of proposed model is 241227955 and Chen's model is 777860940, so the MSE obtained in Chen's model is much higher than our proposed model which shows superiority and robustness of the proposed model over Chen's model since the values of the error in proposed model is reducing.

4. Prediction through Artificial Neural Network

Back Propagation through time is a powerful tool of artificial neural network with application to many areas as pattern recognition, dynamic modeling and nonlinear systems. Back propagation algorithm(BPA) provides an efficient way to calculate the gradient of the error function using chain rule of differentiation. The error after initial computation in the forward pass is propagated backward from the output units, layer by layer. BPA, a generalized Delta rule is commonly used algorithm for supervised training of multi layer feed forward artificial neural network. In supervised learning, we try to adapt an artificial network so that the actual outputs ($\overline{Y}$) come close to some target outputs(Y) for a training set, which contains T patterns. The goal is to adapt the parameters of network so that it performs well for pattern from outside the training set.

4.1 Back propagation Algorithm (B P A)

Let the training set be $\{x(k),d(k)\}_{k=1}^{N}$,Where x(k) is the input pattern vector to the network and d(k) is the desired output vector for the input pattern x(k).The output of the j^{th} output unit is denoted by y_j , connections weights from the i^{th} unit in one layer to the j^{th} unit in the layer above are denoted by wij. If m be the no. of output units and dj(k) is the desired output from the j^{th} output unit whose actual output in response to the k^{th} input exempler x(k) is y_j ,for j=1,2,3,……..,m. The sum of squares of the error over the entire output unit for this k^{th} exampler by

$$E(k)=(1/2)\sum_{j=1}^{m}[y_j(k)-dj(k)]^2$$

Error E(k) is affected by the output from unit j at the output layer and is determined by

$$\frac{\partial E(k)}{\partial y_j}=y_j-d_j$$

a) The net input to output layer is of the form

$$S_j=\sum_i y_i^{(1)}w_{ij}-\theta_j$$

Where $y_i^{(1)}$ *is* the output from the i^{th} unit in the first layer below the output layer, w_{ij} is connection weight multiplying $y_i^{(1)}$ and θ_j is the threshold of unit j. The negative of threshold is defined to be the bias. BPA has been applied to develop ANN model of a lead year by two methods.

4.2 ANN based method:

In this method, the number of enrollment of primary education obtained from forecasted data by the method of fuzzy time series, have been used for the training in the development process of the ANN model and forecasted enrollment obtained through fuzzy time series have also been used as desired output for validation of developed model outside the training set. The model have been developed for forecast based on different training sets and validation the steps adopted as :

Step1: Fuzzy output primary education enrollment of year 1993-94 to 2007-2008 as input set and fuzzy output of 2008-2009 as desired output.

Step2: Fuzzy output primary education enrollment of year 1993-94 to 2008-2009 as input set and fuzzy output of 2009-2010 as desired output.

Step3: Fuzzy output primary education enrollment of year 1993-94 to 2009-2010 as input set and fuzzy output of 2010-2011 as desired output.

Step4: Fuzzy output primary education enrollment of year 1993-94 to 2010-2011 as input set and fuzzy output of 2011-2012 as desired output.

Step5: Fuzzy output primary education enrollment of year 1993-94 to 2011-2012 as input set and fuzzy output of 2012-2013 as desired output.

The algorithm has been implemented through C programming language, considering two hidden layers and computations have been made by various iterations levels like: 300 & 500. Out of these, the best suitable forecasted values have been obtained by model with 500 iterations. The results so obtained has been illustrated in Table

Table-4

Year	Forecasted enrollment by implementing proposed model of Fuzzy Time Series	Forecasted enrollment by implementing ANN model
1993-1994	326666	
1994-1995	326666	
1995-1996	326666	
1996-1997	313333	
1997-1998	313333	
1998-1999	313333	
1999-2000	313333	
2000-2001	313333	
2001-2002	313333	
2002-2003	420000	
2003-2004	428000	
2004-2005	428000	
2005-2006	428000	
2006-2007	428000	
2007-2008	428000	
2008-2009	480000	498425.84
2009-2010	480000	487423.39
2010-2011	500000	498965.36
2011-2012	500000	509327.94
2012-2013	500000	514865.05

5. Result & Conclusion:

The proposed neuro fuzzy method has been implemented to have prediction for the primary education enrollment of Lucknow district (INDIA). Unlike the enrollment of university, the primary education enrollment is affected by various parameters. Further as per historical experiences the primary education enrollment of Lucknow district is not governed by any deterministic process as its primary education enrollment is highly unpredictive. Future academic sessions prediction is of much use to the district primary education department. We have considered the indirect relation of various parameters for time series data and presume that their relations are time invariant in view of primary education enrollment. The motivation of the study is that the primary education enrollment data are collected through various sampling techniques involving the vagueness. Further the primary education enrollment forecast has also been obtained through ANN using Back propagation algorithm. A comparison of forecasted primary education enrollments obtained through these models has been compared in table-5.

Table-5

Year	Actual primary education enrollment	Forecasted by Chen's fuzzy time series model	Forecasted by proposed fuzzy time series model	Forecasted by artificial neural network BPA
2008-2009	488425	480000	480000	498425.84
2009-2010	474351	480000	480000	487423.39

2010-2011	505301	500000	500000	498965.36
2011-2012	510899	500000	500000	509327.94
2012-2013	512182	500000	500000	514865.05

The accuracy of forecasting model is measure in terms of less average percentage error The average percentage forecast error of the various model are as follows:

Table-6

% Average error of chen's fuzzy time series model	% Average error of proposed time series model	% Average error of neural forecast
4.61	1.50	0.752

The Proposed method for primary education enrollment forecast uses simplified arithmetic operations similar to chen(1996) rather than complicated max-min composition presented in Song(1993,1994). In view of considering the weights of various fuzzy logical relationships, it uses the repeated fuzzy logical relations to construct the fuzzy logical relationship groups. It has further supremacy over the Chen's method, as it does not require the knowledge of primary education enrollment for the final target year of forecast. In the study the target year for the forecast was considered to be as year 2012-2013 and hence the actual primary education enrollment of the target year 2012-2013 have neither been used to construct the fuzzy logical relationship nor to construct the fuzzy logical relationship groups. The primary education enrollment forecast computed through the proposed method of input data taken for ANN training purpose from forecasted value of fuzzy time series method, mare quite impressive in terms of error estimate, where the primary education enrollment data are not accurate. Further, the computations show that computed data through the proposed method are quite impressive in terms of error estimate, where the primary education enrollment data are not accurate. Further, the computations shows that neuro fuzzy approach method provides much better forecast for primary education enrollment forecast.

References

[1] Chen Shyi-Ming [1996],Forecasting enrollments based on fuzzy time series ,Fuzzy sets and Systems 81,311-319.

[2] Chou-An Chen,(2007), Neuro –Fuzzy Technology as a Predictor of ParathyroidHarmonse Level in Hemodialysis Patients,Tohku J,Exp.Med., 211, pp. 81-87.

[3] Duboi, S. D and H Parde [1991], Fuzzy sets in approximate reasoning Part I : interface with possibility distributions, Fuzzy Sets and Systems 40, 143-202.

[4] Huarng Kunhuang [2001], Heuristic models of fuzzy time series for forecasting, Fuzzy Sets and Systems 123,369-386.

[5] K.K.Shukla & Raghunath, (1999),An Efficient Global Algorithm for Supervised Training of Neural Networks,International Journal of Computers and EE,PERGAMON Press, 25, pp.195-218.

[6] Mamdani, E.E.[1977], Application of Fuzzy logic to approximate reasoning using linguistic synthesis, IEEE Trans Comput 26, 1182-1191.

[7] Shyi-Ming Chen[1996],Forecasting enrollments based on fuzzy time series ,Fuzzy Sets and Systems 81,311-319.

[8] Song Q. and Chissom B.[1993], Fuzzy time series and its models, Fuzzy Sets and Systems ,54,269-277.

[9] Song Q. and Chissom B.[1993],Forecasting enrollments with fuzzy time series -Part I, Fuzzy sets and Systems 54,1-9.

[10] Song Q. and Chissom B.[1994],Forecasting enrollments with fuzzy time series -Part II, Fuzzy Sets and Systems 64, 1-8.

[11] S.R.Singh,(2006), A Robust Method of Forecasting Based on Fuzzy Time Series, Elsevier Applied Mathematics and Computation, 188, pp.472-484.

[12] Tanu Verma, DR. R.K.Srivastava, Artificial Neural Network based Heart Disease Predictive Approach, Volume 4, Issue 3, March 2015, IJAITEM, 2319-4847.

[13] Wu W [1986], Fuzzy reasoning and fuzzy relational equations, Fuzzy Sets and Systems 20, 67-78.

[14] Werbos Paul J. [1990], Back propagation through time: what it does and how to do it,Proceeding of the IEEE,78,1550-1559.

[15] Yang CC[1998],The use of back propagation neural network for the sim. & analysis of time seriesdata Transcations of the ASAE,41,1181-1187 .

[16] Zadeh L. A., [1965], Fuzzy sets, Information & Control ,8,38-353 .