International Journal of Information Technology and Knowledge Management
July-December 2010, Volume 2, No. 2, pp. 205-210

SOFT COMPUTING TECHNIQUES: AN APPLICATION TO SHORT TERM FORECAST OF POTATO PRODUCTION

R. K Srivastava

The paper presents a method to give short term forecast of agricultural production based on fuzzy time series. The historical data of potato production of Bareilly district (INDIA) have been taken to implement the model, as such time series data obtained through statistical estimators, involve vagueness. The study uses the fuzzy sets theory Zadeh [1965] and fuzzy time series models introduced by Song and Chissom [1993] and Chen [1996]. The proposed method is more efficient and provides better forecast in comparison to Chen's method, as it uses the domain specific knowledge of the problem in terms of dependency relations. The forecast to potato production have also been obtained by developing an Artificial Neural Network model using Back propagation algorithm. The study is aimed to find the potato forecast for a lead year by using a fuzzy time series model and back propagation algorithm for the forecast. The forecasted potato production, obtained through these two soft computing techniques, have been compared and their performance has been examined.

Keywords: Fuzzy Time Series; fuzzy Set; Production; Forecasting; Linguistic Value; Back Propagation Algorithm

1. Introduction

Future values of a time series data of many processes are neither exactly governed by mathematical function nor by probability distribution. In such processes, the soft computing techniques: Fuzzy time series and artificial neural network, be preferred as these use the relation of dependency, a generalization of function A time series is a sequence of observations taken sequentially in time with an intrinsic feature that the typically adjacent observations are dependent. The time series analysis is concerned with techniques for analysis of such dependency. Both the models: stochastic and dynamic in the time series analysis use the probability structure of sequence of observations to construct the forecast function. The unique features of fuzzy time series models are that it uses the relation, a generalization of functions. Based upon the fuzzy set theory introduced by Zadeh [1965], Song and Chissom [1993] presented the definition of fuzzy time series and outlined its modeling by means of fuzzy relation equations and approximate reasoning. Song and Chissom [1993] applied the fuzzy time series model to forecast the enrollments of the university of Alabama. Chen [1996] proposed an alternative simplified method of defuzzification using arithmetic operations. Hurang [2001] proposed a heuristic model by integrating problem specific heuristic knowledge with Chen's model to improve the forecasting. All these models have been implemented to forecast the enrollments of the University of Alabama. In the present work the proposed model have been implemented on the historical potato crop yield forecast, a highly non linear process, where data in general contain imprecision. The study is aimed to get some reliable forecast for potato production for a lead year. This production forecast may help the farmers as well as the local agro based industries in their business planning.

Dept. of Computer Science, Bareilly College, Bareilly-243001, INDIA
Email: rk_srivastava17@rediffmail.com

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:

$$F(t) = F(t-1) \circ R(t, t-1) \quad (2)$$

or

$$F(t) = (F(t-1) \cup F(t-2) \cup \ldots \cup F(t-m)) \circ R_0 (t, t-m) \quad (3)$$

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

$$F(t) = (F(t-1) \times F(t-2) \times \ldots \times F(t-m)) \circ R_a(t, t-m) \quad (4)$$

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 R_a (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 $R_a(t_1, t_1 - m) = R_a(t_2, t_2 - m)$ 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,

$R(t, t-1) = R$,

$R_a(t, t-m) = R_a(m)$,

$R_0(t, t-m) = R_0(m)$

In general at different times t_1 and t_2, $R(t_1, t_1 - 1) \neq R(t_2, t_2 - 1)$, $R_a(t_1, t_1 - m) \neq R_a(t_2, t_2 - m)$ and $R_\circ(t_1, t_1 - m) \neq R_\circ(t_2, t_2 - m)$. 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.

Potato Production 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 production of potato for the Bareilly district(India) in various years starting from 1987-88 to 2001-2002 with assumption that it includes some vagueness incurred due to statistical sampling.

The algorithm for application of fuzzy time series in potato production forecasting comprises of the following steps:

Table 1

Year	*Actual production (MT)*	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	*Fuzzified production*
1987-88	100216	0	0	0	0.5	1	0.5	0	0	0	A_5
1988-89	101139	0	0.5	0	0.5	1	0.5	0	0	0	A_5
1989-90	100926	0	0.5	0	0.5	1	0.5	0	0	0	A_5
1990-91	82178	0	0.5	1	0.5	0	0	0	0	0	A_3
1991-92	87816	0	0.5	1	0.5	0	0	0	0	0	A_3
1992-93	60235	1	0.5	0	0	0	0	0	0	0	A_1
1993-94	94423	0	0	0.5	1	0.5	0	0	0	0	A_4
1994-95	82691	0	0.5	1	0.5	0	0	0	0	0	A_3
1995-96	105457	0	0	0	0.5	1	0.5	0	0	0	A_5
1996-97	149863	0	0	0	0	0	0	0	0.5	1	A_9
1997-98	68587	1	0.5	0	0	0	0	0	0	0	A_1
1998-99	135441	0	0	0	0	0	0	0.5	1	0.5	A_8
1999-2000	122567	0	0	0	0	0	0.5	1	0.5	0	A_7
2000-2001	100140	0	0	0	0.5	1	0.5	0	0	0	A_5
2001-2002	105378										

Step-1: Let D_{min} and D_{max} be minimum and maximum production. 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 = [60000,150000]. Further the universe of discourse U is partitioned into nine intervals of equal length as follows:

u_1 = [60000, 70000], u_2 = [70000, 80000],

u_3 = [80000, 90000], u_4 = [90000,100000],

u_5 = [100000, 110000], u_6 = [110000, 120000]

u_7 = [120000, 130000], u_8 = [130000,140000],

u_9 = [140000, 150000]

Step 2: Fuzzy sets $A_1, A_2, A_3, \ldots A_9$ 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

A_7 = very very good, A_8 = excellent, A_9 = outstanding

are to be defined. $u_1, u_2, \ldots, u_9$ are chosen as elements of these fuzzy sets. The membership grades of $u_1, u_2, \ldots, u_9$ to each A_i (i = 1,2,..., 9) will decide that how well each u_k (k = 1,2,...9) belong to u_i. We have determined the membership of each element in all the fuzzy sets Ai (i = 1,2,...9) and are expressed as:

$A_1 = \{u_1/1, u_2/.5, u_3/0, u_4/0, u_5/0, u_6/0, u_7/0, u_8/0, u_9/0\}$

$A_2 = \{u_1/.5, u_2/1, u_3/.5, u_4/0, u_5/0, u_6/0, u_7/0, u_8/0, u_9/0\}$

$A_3 = \{u_1/0, u_2/.5, u_3/1, u_4/.5, u_5/0, u_6/0, u_7/0, u_8/0, u_9/0\}$

$A_4 = \{u_1/0, u_2/0, u_3/.5, u_4/1, u_5/.5, u6/0, u_7/0, u_8/0, u_9/0\}$

$A_5 = \{u_1/0, u_2/0, u_3/0, u_4/.5, u_5/.1, u_6/.5, u_7/0, u_8/0, u_9/0\}$

$A_6 = \{u_1/0, u_2/0, u_3/0, u_4/0, u_5/.5, u_6/1, u_7/.5, u_8/0, u_9/0\}$

$A_7 = \{u_1/0, u_2/0, u_3/0, u_4/0, u_5/0, u_6/.5, u_7/1, u_8/.5, u_9/0\}$

$A_8 = \{u_1/0, u_2/0, u_3/0, u_4/0, u_5/0, u_6/0, u_7/.5, u_8/1, u_9/.5\}$

$A_9 = \{u_1/0, u_2/0, u_3/0, u_4/0, u_5/0 , u_6/0, u_7/0, u_8/.5, u_9/1\}$

Where u_i (i = 1, 2...9) is the element and the number below '/ ' is the membership of u_i to A_j (j = 1,2,..9)

Step 3: Fuzzify the historical production data to find out the equivalent fuzzy set to each year's production using the step- 2. The equivalent fuzzy set to each year's production are shown in table-1.

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

Proposed Model		Chen's Model	
$A_1 \rightarrow A_4$	$A_1 \rightarrow A_8$	$A_1 \rightarrow A_4$	$A_1 \rightarrow A_8$
$A_3 \rightarrow A_1$	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_1$	$A_3 \rightarrow A_3$
$A_3 \rightarrow A_5$	$A_4 \rightarrow A_3$	$A_3 \rightarrow A_5$	$A_4 \rightarrow A_3$
$A_5 \rightarrow A_3$	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_3$	$A_5 \rightarrow A_5$
$A_5 \rightarrow A_5$	$A_5 \rightarrow A_9$	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_5$
$A_7 \rightarrow A_5$	$A_8 \rightarrow A_7$	$A_5 \rightarrow A_9$	$A_7 \rightarrow A_5$
$A_9 \rightarrow A_1$		$A_8 \rightarrow A_7$	$A_9 \rightarrow A_1$

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

Proposed Model		Chen's Model	
Group-1:	$A_1 \rightarrow A_4, A_8$	Group-1:	$A_1 \rightarrow A_4, A_8$
Group-2:	$A_3 \rightarrow A_1, A_3, A_5$	Group-2:	$A_3 \rightarrow A_1, A_3, A_5$
Group-3:	$A_4 \rightarrow A_3$	Group-3:	$A_4 \rightarrow A_3$
Group-4:	$A_5 \rightarrow A_3, A_5, A_5, A_9$	Group-4:	$A_5 \rightarrow A_3, A_5, A_9$
Group-5:	$A_7 \rightarrow A_5$	Group-5:	$A_7 \rightarrow A_5$
Group-6:	$A_8 \rightarrow A_7$	Group-6:	$A_8 \rightarrow A_7$
Gropu-7:	$A_9 \rightarrow A_1$	Group-7:	$A_9 \rightarrow A_1$

Step 6: Potao Production forecast from the computed fuzzified potato production is carried by following rules:

(1) If the fuzzified production of the year i is A_j, and there is only one fuzzy logical relationship in the fuzzy logical relationship groups as $A_j \rightarrow A_k$, 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 production of year i + 1 is m_k.

(2) If the fuzzified production 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 production for year i+1 is $(\Sigma^p_{x=1} m_x)/ p$

(3) If the fuzzified production of a year i is A_j, and no logical relationship is found in logical reationship groups, whose current state of production 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 production of year i + 1 is m_j.

Based on the above rules, the forecasted potato production for various years have been calculatedfor both: Chen's and proposed model are given against the actual production in the Table-2.

Table 2

Year	*Actual production (MT)*	*Forecasted potato production by implementing Chen's model*	*Forecasted potato production by proposed model*
1987-88	100216		
1988-89	101139	111666	110000
1989-90	100926	111666	110000
1990-91	82178	111666	110000
1991-92	87816	85000	85000
1992-93	60235	85000	85000
1993-94	94423	115000	115000
1994-95	82691	85000	85000
1995-96	105457	85000	85000
1996-97	149863	111666	110000
1997-98	68587	65000	65000
1998-99	135441	115000	115000
1999-2000	122567	125000	125000
2000-2001	100140	105000	105000
2001-2002	105378	111666	110000

With the above comparison of actual production of potato of Bareilly district with the forecasted production, one can conclude that the forecasted results are very close to that of actual result. The Computed Mean Square Error(MSE) of the forecast by Chen's model is obtained to be 16285600 where as the MSE of the forecast by proposed method comes out to be 12791522 and is much lower than Chen's model as it do not uses the production data of lead year in fuzzy table for making lead year forecast.

4. PRODUCTION FORECASTING: 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 () come close to some target outputs($\bar{Y}$) 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.

5.1. Back Propagation Algorithm (BPA)

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 all the 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

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

The net input to output layer is of the form

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

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.

Method I: In this method, the actual production obtained from historical data have been used for the training in the development process of the ANN model and forecasted production obtained through fuzzy time series have 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: Production of year 1987-88 to 1996-97 as input set and fuzzy output of 1997-98 as desired output.

Step2: Production of year 1987-88 to 1997-98 as input set and fuzzy output of 1998-99 as desired output.

Step3: Production of year 1987-88 to 1998-99 as input set and fuzzy output of 1999-2000 as desired output.

Step4: Production of year 1987-88 to 1999-2000 as input set and fuzzy output of 2000-2001 as desired output.

Step5: production of year 1987-88 to 2000-01 as input set and fuzzy output of 2001-02 as desired output.

Method II: In this method, forecasted potato production obtained through the fuzzy time series have been used for the training of ANN model development and actual production have been used for the 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 are as:

Step 1: Fuzzy output of year 1988-89 to 1996-97 as input set and actual production of 1997-98 as desired output.

Step 2: Fuzzy output of year 1988-89 to 1997-98 as input set and actual production of 1998-99 as desired output.

Step 3: Fuzzy output of year 1988-89 to 1998-1999 as input set and actual production of 1999-2000 as desired output.

Step 4: Fuzzy output of year 1988-89 to 1999-2000 as input set and actual production of 2000-2001 as desired output.

Step 5: Fuzzy output of year 1988-89 to 2000-01 as input set and actual production of 2001-02 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: 100, 200, 500 & 1000. 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. 3 and Table. 4.

Table 3

Year	Input (actual production)	Desired output (Fuzzy)	Neural output (by method I)
1987-88	100216		
1988-89	101139		
1989-90	100926		
1990-91	82178		
1991-92	87816		
1992-93	60235		
1993-94	94423		
1994-95	82691		
1995-96	105457		
1996-97	149863		
1997-98	68587	65000	65004.98
1998-99	135441	115000	114356.23
1999-2000	122567	110000	124216.79
2000-2001	100140	105000	104630.39
2001-2002	105378	110000	110177.81

Table 4

Year	Input (fuzzy values)	Desired output (actual production)	Neural output (by method II)
1987-88			
1988-89	110000		
1989-90	110000		
1990-91	110000		
1991-92	85000		
1992-93	85000		
1993-94	115000		
1994-95	85000		
1995-96	85000		
1996-97	110000		
1997-98	65000	68587	68597.86
1998-99	115000	135441	133782.06
1999-2000	125000	122567	121091.89
2000-2001	105000	100140	100026.42
2001-2002	110000	105378	105168.35

6. RESULT & CONCLUSION

The proposed fuzzy time series method has been implemented to have short term forecast for the potato production of Bareilly district (INDIA). Unlike the enrollment of university, the crop production is affected by various parameters. Further as per historical experiences the production of potato is not governed by any deterministic process as its production is highly unpredictive. It is a highly sensitive crop towards the disease and bad weather. A lead year forecast is of much use to the local industries and farmers. We have considered the indirect relation of various parameters for time series data and presume that their relations are time invariant in view of crop production. The motivation of the study is that the historical crop production data are collected through various sampling techniques involving the vagueness. Further the production forecast has also been obtained through ANN using Back propagation algorithm. A comparison of forecasted productions obtained through these models have been compared in table-5.

Table-5

Year	*Actual production*	*Forecasted by Chen's fuzzy time series model*	*Forecasted by proposed fuzzy time series model*	*Forecasted by artificial neural network BPA (method I)*	*Forecasted by artificial neural network BPA (method II)*
1997-98	68587	65000	65000	65004.98	68597.86
1998-99	135441	115000	115000	114356.23	133782.06
1999-00	122567	125000	125000	124216.79	121091.89
2000-01	100140	105000	105000	104630.39	100026.42
2001-02 (lead year)	105378	111666	110000	110177.81	105168.35

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

% error of chen's fuzzy time series model	% error of proposed time series model	% error of neural forecast (method I)	% error of neural forecast (method II)
14.52	14.11	6.13	0.49

The Proposed method for potato crop production forecast uses simplified arithmetic operations similar to chen[1] rather than complicated max-min composition presented in [7,8]. 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 do not require the knowledge of production for the final target year of forecast. In the study the target year for the forecast was considered to be as year 2001-2002 and hence the actual production of the target year 2001-2002 have neither been used to construct the fuzzy logical relationship nor to construct the fuzzy logical relationship groups. The crop production, forecast computed through the proposed method are quite impressive in terms of error estimate, where the historical production 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 historical production data are not accurate. Further, the computations shows that neuro fuzzy approach method II provides much better forecast for potato production forecast.

REFERENCES

[1] Chen Shyi-Ming [1996], "Forecasting Enrollments based on Fuzzy Time Series", Fuzzy Sets and Systems, 81, 311-319.

[2] 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.

[3] Huarng Kunhuang [2001], "Heuristic Models of Fuzzy Time Series for Forecasting", Fuzzy Sets and Systems, 123, 369-386.

[4] Mamdani, E.E.[1977], "Application of Fuzzy Logic to Approximate Reasoning using Linguistic Synthesis", IEEE Trans Comput, 26, 1182-1191.

[5] Shyi-Ming Chen[1996], "Forecasting Enrollments based on Fuzzy Time Series", Fuzzy Sets and Systems, 81, 311-319.

[6] Song Q. and Chissom B.[1993], "Fuzzy Time Series and its Models", Fuzzy Sets and Systems, 54, 269-277.

[7] Song Q. and Chissom B.[1993], Forecasting Enrollments with Fuzzy Time Series -Part I, Fuzzy Sets and Systems, 54, 1-9.

[8] Song Q. and Chissom B.[1994], 'Forecasting Enrollments with Fuzzy Time Series-Part II", Fuzzy Sets and Systems, 64, 1-8.

[9] Wu W [1986], "Fuzzy Reasoning and Fuzzy Relational Equations", Fuzzy Sets and Systems 20, 67-78.

[10] Werbos Paul J. [1990], "Back Propagation through Time: what it does and how to do it", Proceeding of the IEEE, 78, 1550-1559.

[11] Yang CC[1998], "The use of Back Propagation Neural Network for the Sim. & Analysis of Time Series Data", Transcations of the ASAE, 41, 1181-1187.

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