A Novel Method for Rice Production Forecasting Using Fuzzy Time Series

[1]Dr. Sachin Kumar and [2]Narendra Kumar

[1] Prof. CSE Deptt, AKGEC Ghaziabad, India

[2] Research Scholar, Mewar University, India

Abstract

Present study provides some modified techniques for time series based forecasting for forecasting the yield of any crop year. Our study can help in inventory management of rice yield and for management of storage space. We are using the data of previous years and proposing a new method by using the fuzzy time series forecasting technique.

Keywords: Computational Intelligence, Fuzzy time series, Crop yield forecasting, Higher order time series, time series, Crisp value

1. Introduction

India is land of agriculture and there is a great uncertainty about the outcome for any crop year. There are a number of functions which totally depends upon the outcome of crop. We can review the dependencies as follows:
Food Corporation of India (The food bank of nation) procures rice every year for PDS (public distribution system), under which government distributes rice at control rate shops. They ever face a problem of storage. If they have an idea about the yield of any particular year then they can manage storage capacity according to yield. By right storage management government can avoid the misuse of storage. Every year a lot of rice is lost due to no storage if right storage management technique is applied then government can also avoid the loss of rice.

The second and most important thing is inventory of food rice. If we have a rough idea of crop yield for next year then we can purchase the rice from other states in case of shortage and can also sale the excess in case of bumper yield. By this we can minimize the condition of on the spot rush for shortage or excess.

Third thing where forecasting plays key player is the planning of farmer. Farmer is the main player behind the scene. If farmer have an idea about the yield of his crop then according to that outcome he can plan for next crop. He can apply new techniques and precautions timely. Crop forecasting also helps farmers in budget management for the next crop.

Careful study of results can also help in division of land type. Planning department can declare land as fertile or non fertile on the basis of regular studies. By making these type of declarations government can emphasis on most fertile land for agriculture and can use less fertile land for other purposes.
Above is brief idea about the need of crop forecasting. Above discussion proves the importance of impressive forecasting up to some extent. In this paper we are using the data of 20 years for our applications. We are using basic timeseries with modified input for that. We are using fuzzy values for the timeseries to input and after processing these values after de-fuzzyfication outcome is forecasted yield for next year.

2. Literature Review

We have reviewed a number of research articles for our current work. We also have reviewed a number of good books on time series. The main problem in forecasting we have reviewed so far is the accuracy of forecasting. On the basis of accuracy we can rely on that result. There are a good number of papers telling about time series forecasting for rice and wheat as well[2]. Forecasting is also done on the basis of statistical methods like regression, co-relation, median etc, but the problem in center is again accuracy.Fuzzy time series relies on real time fuzzy values as input hence there is a bit more accuracy in pattern and gives more accurate result.

By reviewing a literature we have concluded following terms to use in application:

2.1 Fuzzy Set: There are two types of variables, one is crisp variable and other is fuzzy variable [1]. Crisp variables have fixed value like normal mathematical variable but on contrary fuzzy variables don't have fixed values. Fuzzy variables have membership value in different sets rather than having some fixed value.

We can understand the difference between fuzzy and crisp by following example:

If the speed of car is more than 80 then it is fast then a car running at a speed 81 is fast but a car running at 79.99 is slow in crisp variables. But on other hand fuzzy variables works on range, as car running from 40 to 60 is average speed and 50 to 70 is fast speed. Then we can calculate membership of any input in both sets and on that basis we can determine the result.

2.2 Time series:

Very simple form of time series can be termed as a sequence of data taken at different time stamps. Another definition can also be termed as "series of data with time stamping on is called something as time series." Hence in time series we capture data at specific time to get a pattern and by using that pattern we forecast values after next time stamp of future.

2.3 Fuzzy Time Series:

In normal time series we use statistical methods and forecast or analyze data according to those methods. Fuzzy time series is relatively new idea, in fuzzy time series the data on which time stamping is done is fuzzy variables rather than crisp values[2][3]. First all the variables fuzzified and by defuzzyfication result is calculated.

We have reviewed a number of methods in fuzzy time series. There are various techn iques for de-fuzzyfication on which the result is based. The result varies according to the de-fuzzyfication method. We have compared two methods and also compared our method with those one.

3. Proposed Methodology

The proposed methodology has been discussed up to some extent in upper portion of paper. The main idea behind the concept is to use fuzzy values as input to time series instead of realtime crisp values. Hence total methodology can be divide into further steps:

1. The very first task is to define universal set (universe of discourse) under which complete production history can be covered.
2. The second step is of making partition of the discourse range of by equal length.
3. Define the fuzzy set by making membership functions in respective partitions.
4. Fuzzification of time series data for further calculation comes under next step.
5. After fuzzification we realize the nearest series for each year and calculate a dummy production (DP) for each year.
6. Give the result as this dummy production will be the forecasted result for next year.
7. Take an average for two or three years of DP then give that as result for some more accuracy.

Above is the methodology for calculation of forecasted result. Now we will implement above procedure for our data.

4. Implementation

For the implementation of our proposed methodology we have taken relevant data from a university farm. The data has production details of previous 21 years. This data has records of 1991-92 to 2011-12 rice production per unit land. Now we will use this data as input to our model. Now our procedure includes following steps:

Step 1: Our data has least value 1431 kg. and has maximum value as 3407 kg. , hence we will define a universe of discourse which contains both values inside it. Here we are defining a universe of discourse or universal set as [1400-3500]. This universe can be named as U (universal set).

Step 2: once we have a well defined universal set then we feed this to our fuzzy model for making different memberships. We have defined 7 membership functions (MF's) here for making fuzzy calculations.

As soon as we feed this data to fuzzy model after giving range [1400-3500] with 7 MF's it makes 7 different ranges or partitions.

a1= [1400-1700]
a2= [1700-2000]
a3= [2000-2300]
a4= [2300-2600]
a5= [2600-2900]
a6= [2900-3200]
a7= [3200-3500]

Step 3:

a1= [1400-1700] poor production
a2= [1700-2000] below average production
a3= [2000-2300] average production
a4= [2300-2600] good production
a5= [2600-2900] very good production
a6= [2900-3200] excellent production
a7= [3200-3500] bumper production

The aliases which are given to above intervals are the professional agriculture terms. We are using these terms in our application to give it a more realistic look, and for making it easy to learn.

Step 4: in this step we will break the existing data intervals according to their respective memberships in different intervals.

We have real time data from farm house. That data is as follows:

Year	1991	1992	1993	1994	1995	1996	1997
Yield	2730	2957	2382	2572	2642	2700	2872
Year	1998	1999	2000	2001	2002	2003	2004
Yield	3407	2238	2895	3276	1431	2248	2857
Year	2005	2006	2007	2008	2009	2010	2011
Yield	2318	2617	2254	2910	3434	2795	3000

Now we come to the fuzzy part of this data. We find the fuzzy membership in various sets a1 a2,……a7. Now take a quick look at various techniques for finding fuzzy membership value in any set.

There are a number of ways to assign membership values to fuzzy variables in comparison with the probability density functions to random variables [4]. The methods for assigning membership values are as follows:

1. Intuition
2. Inference
3. Rank Ordering
4. Angular Fuzzy Sets
5. Neural Networks
6. Genetic Algorithm
7. Inductive Reasoning

In general, the triangular membership function can be specified from the formula given below:

$$\mu_{triang}(x) = \begin{cases} 0 & x < L \\ 1 - \dfrac{|C - x|}{(R - L)/2} & L < x < R \\ 0 & x > R \end{cases}$$

Here L and R are the left and right bounds, respectively, and C is the center of the symmetric triangle. Likewise, the trapezoidal membership may be expressed as:

$$\mu_{trapez}(x) = \begin{cases} 0 & x < L \text{ or } x > U \\ \dfrac{x - L}{C - W/2 - L} & L < x < C - W/2 \\ 1 & C + W/2 < x < C + W/2 \\ \dfrac{U - x}{U - (C + W/2)} & C + W/2 < x < U \end{cases}$$

Here L and U are the lower and upper bounds, respectively, C is the center, and W is the width of the top side of the symmetric trapezoid. We can use any of the above formula. We have assigned the fuzzy membership by using above triangular method formula [4][5][6].

The values of different memberships can be taken as follows:

Year	Production	a1	a2	a3	a4	a5	a6	a7
1991	2730	0	0	0.2	0.7	1	0.5	0
1992	2957	0	0	0	0.3	0.8	1	0.5
1993	2382	0	0.3	0.8	1	0.5	0	0
1994	2572	0	0	0.5	1	0.8	0.3	0
1995	2642	0	0	0.3	0.8	1	0.5	0
1996	2700	0	0	0.2	0.7	1	0.5	0
1997	2872	0	0	0	0.5	1	0.8	0.3
1998	3407	0	0	0	0	0.3	0.5	1
1999	2238	0	0.5	1	0.8	0.3	0	0
2000	2895	0	0	0	0.5	1	0.8	1
2001	3276	0	0	0	0	0.3	0.8	1
2002	1431	1	0.5	0.2	0	0	0	0
2003	2248	0	0.5	1	0.8	0.3	0	0
2004	2857	0	0	0	0.5	1	0.8	0.3
2005	2318	0	0.3	0.8	1	0.5	0	0
2006	2617	0	0	0.3	0.8	1	0.5	0
2007	2254	0	0.5	1	0.8	0.3	0	0
2008	2910	0	0	0	0.3	0.5	1	0.8
2009	3434	0	0	0	0	0.3	0.5	1
2010	2795	0	0	0	0.5	1	0.8	0.2
2011	3000	0	0	0	0	0.6	1	0.3

Step 5: in this step the calculation of dummy production is done on the basis of each year's respective production details. Below table is showing the dummy values in DP(dummy production):

Year	Production	a1	a2	a3	a4	a5	a6	a7	DP
1991	2730	0	0	0.2	0.7	1	0.5	0	2723
1992	2957	0	0	0	0.3	0.8	1	0.5	3011
1993	2382	0	0.3	0.8	1	0.5	0	0	2411
1994	2572	0	0	0.5	1	0.8	0.3	0	2489
1995	2642	0	0	0.3	0.8	1	0.5	0	2711
1996	2700	0	0	0.2	0.7	1	0.5	0	2723
1997	2872	0	0	0	0.5	1	0.8	0.3	2789
1998	3407	0	0	0	0	0.3	0.5	1	3250
1999	2238	0	0.5	1	0.8	0.3	0	0	2189
2000	2895	0	0	0	0.5	1	0.8	1	2789
2001	3276	0	0	0	0	0.3	0.8	1	3217
2002	1431	1	0.5	0.2	0	0	0	0	1650
2003	2248	0	0.5	1	0.8	0.3	0	0	2189
2004	2857	0	0	0	0.5	1	0.8	0.3	2789
2005	2318	0	0.3	0.8	1	0.5	0	0	2411
2006	2617	0	0	0.3	0.8	1	0.5	0	2711
2007	2254	0	0.5	1	0.8	0.3	0	0	2146
2008	2910	0	0	0	0.3	0.5	1	0.8	3089
2009	3434	0	0	0	0	0.3	0.5	1	3250
2010	2795	0	0	0	0.5	1	0.8	0.2	2789
2011	3000	0	0	0	0	0.6	1	0.3	2938

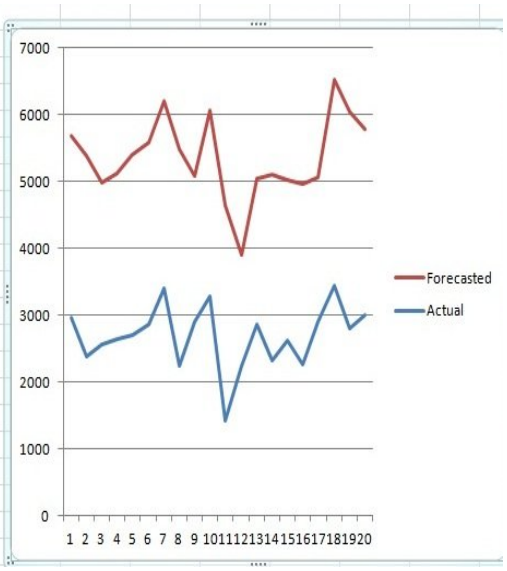
The dummy production DP for year 1991 can be calculated as follows:

Fuzzy vector values for year 1991= {0,0,.2,.7,1,.5,0}. Here for finding the DP we will only consider higher values because they have larger membership in membership function [7][8].

For our current calculation we are using values equal to or more than 0.5.
Now for every partition will use the average for calculation as for a1 [1400-1700], value 1550 will be used for calculation and the same will be followed for other too.
In current scenario for year 1991, a4 has value .7, a5 has 1 and a6 has value as .5.
The formula for DP for any specific year can be termed as weighted mean of respective partitions which have membership value more than .5, hence
DP for 1991 = (.7*2450 + 1*2750 + .5*3050)/(.7+1+.5)
In similar fashion we have find all the DP values for each year.

Step 6: Result can be calculated by taking the dummy production as the forecasted result for next year. Below figure is the obvious result table which has forecasted values. Hence we can find the forecasted value for next year by identifying the pattern of this year's production. Weighted mean method has been used for forecast column.

	Actual	Forecasted
2	2957	2723
3	2382	3011
4	2572	2411
5	2642	2489
6	2700	2711
7	2872	2723
8	3407	2789
9	2238	3250
10	2895	2189
11	3276	2789
12	1431	3217
13	2248	1650
14	2857	2189
15	2318	2789
16	2617	2411
17	2254	2711
18	2910	2146
19	3434	3089
20	2795	3250
21	3000	2789



Comparison between actual production and forecasted production

Year	Production	a1	a2	a3	a4	a5	a6	a7	Dummy	Forecast
1991	2730	0	0	0.2	0.7	1	0.5	0	2723	
1992	2957	0	0	0	0.3	0.8	1	0.5	3011	2723
1993	2382	0	0.3	0.8	1	0.5	0	0	2411	3011
1994	2572	0	0	0.5	1	0.8	0.3	0	2489	2411
1995	2642	0	0	0.3	0.8	1	0.5	0	2711	2489
1996	2700	0	0	0.2	0.7	1	0.5	0	2723	2711
1997	2872	0	0	0	0.5	1	0.8	0.3	2789	2723
1998	3407	0	0	0	0	0.3	0.5	1	3250	2789
1999	2238	0	0.5	1	0.8	0.3	0	0	2189	3250
2000	2895	0	0	0	0.5	1	0.8	1	2789	2189
2001	3276	0	0	0	0	0.3	0.8	1	3217	2789
2002	1431	1	0.5	0.2	0	0	0	0	1650	3217
2003	2248	0	0.5	1	0.8	0.3	0	0	2189	1650
2004	2857	0	0	0	0.5	1	0.8	0.3	2789	2189
2005	2318	0	0.3	0.8	1	0.5	0	0	2411	2789
2006	2617	0	0	0.3	0.8	1	0.5	0	2711	2411
2007	2254	0	0.5	1	0.8	0.3	0	0	2146	2711
2008	2910	0	0	0	0.3	0.5	1	0.8	3089	2146
2009	3434	0	0	0	0	0.3	0.5	1	3250	3089
2010	2795	0	0	0	0.5	1	0.8	0.2	2789	3250
2011	3000	0	0	0	0	0.6	1	0.3	2938	2789
2012										2938

Step 7: step 7 is not necessary but it's something which can be said enhancement to result. We can sum up the dummy production for previous 3-4 years and by averaging them the forecasted value will be generated.

By taking average we can minimize the effect of sudden changes. This average method should only be implemented if there is sudden fall or sudden hike in production for some previous year's production. For example year 2001 has sudden hike and 2002 has sudden fall so this average method can give some near forecasted values.

5. Limitations and future scope

As we have seen in our real scenario in the graph, our forecasted values varying very much similar to actual data. Statistical methods show promising results when data is following some pattern. But due to some sudden hikes and falls in data the result is varying beyond threshold.

The solution for this problem may be by using parameterized method for forecasting. If we use intelligent system and give real time parameter of agriculture to that system, it might be remarkably cool result then.

6. References

[1] Pentlan, J. Mendel. Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3):345{377, Mar 1995.

[2] Adaptive Time-Variant Models for Fuzzy-Time-Series Forecasting. Wai-Keung Wong, Enjian Bai, and Alice Wai-Ching Chu,IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 40, NO. 6, DECEMBER 2010.

[3] Fuzzy Time Series Forecasting With a Probabilistic Smoothing Hidden Markov Model, Yi-Chung Cheng and Sheng-Tun Li, Member, IEEE, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 20, NO. 2, APRIL 2012.

[4] Principles of Soft Computing [page 373], Dr. S.N. Sivanandam, Ms. S.N. Deepa.

[5] Time-Variant Slide Fuzzy Time-series Method for Short-Term Load Forecasting, Xiaojuan Liu, Enjian Bai , Jian'an Fang and Lunhan Luo, 978-1-4244-6585-9/10/$26.00 ©2010 IEEE.

[6] Fractal Based Digital Image Watermarking using Fuzzy c-Mean Clustering. Soheila Kiani, Mohsen Ebrahimi Moghaddam. [2009].

[7] Search for Additive Nonlinear Time Series Casual Models. Tianjiao Chu, Clark Glymour. [2008].

[8] Adaptive Time- Variant models for Fuzzy Time Series forecasting. Wai-Keung Wong, Enjian Bai and Alice Wai-Ching Chu. [2010].

First Author Dr. Sachin Kumar is working as professor in the department of CSE, AKGEC, Ghaziabad. Prior to Join AKGEC, he has worked in KIET & RKGIT Ghaziabad and CCS University, Meerut. He did Ph.D. in Computer Science (CCS University campus) and M.Tech. in Computer Science. He has published/ presented several papers in Journals/ Conferences of repute. He is author of 3 text books. His research interest includes Computer Networks and Wireless Communication.

Second Author Narendra Kumar is doing Ph.D. from Mewar university and working as assistant professor in the department of IT, AKGEC. He did M.Tech. in Computer Science(UPTU). His research interest includes fuzzy logic and digital image processing