International Journal of Statistics and Applied Mathematics 2021; 6(4): 104-108

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2021; 6(4): 104-108

www.mathsjournal.com
Received: 06-05-2021
Accepted: 16-06-2021

S Selvakumar
Government Arts and Science College, Nagercoil, Tamil Nadu, India

Kasthuri
Erode Arts and Science College, Erode, Tamil Nadu, India

Corresponding Author:
S Selvakumar
Government Arts and Science College, Nagercoil, Tamil Nadu, India

Forecasting enrolments using fuzzy invariant Markov models

S Selvakumar and Kasthuri

Abstract
The forecast predicts the future events of time series. Forecasting plays a key role in various fields such as weather forecasting, economic and business planning, etc. Song and Chissom (1994) proposed fuzzy time series and pioneered this model. In this work, an efficient fuzzy time series forecasting model based on fuzzy clustering to handle forecasting problems and improving forecasting accuracy. Each value (observation) is represented by a fuzzy set. The fuzzy sets are converted into the Transition Probability Matrix (TPM). The invariant Markov models are compared with existing models. The results are displayed numerically and graphically.

Keywords: Fuzzy time series, fuzzy c-means, cluster analysis, markov model, forecasting and forecast error

Introduction
Time series has been mostly used for forecasting. *The time series forecast* can deal with many forecasting problems. It cannot solve forecasting problems in which the historical data are linguistic terms. To address this issue, Song and Chissom (1994) [7] proposed fuzzy time series model for forecasting the enrollments of University of Alabama. And also presented the time-invariant fuzzy time series model and the time-variant fuzzy time series model for forecasting the enrollments of the University of Alabama. Chen (2004) [2] proposed a method to forecast the enrollments of University of Alabama. Chen and Hwang *(*2000) [1] employed a method based on fuzzy time series to forecast the daily temperature. Hwang *et al.* (1998) [4] presented the differences of the enrollments *data, forecast* based fuzzy time series. Milk *et al* (2004) [5] proposed first order time series model for forecasting the enrollments of the University of Alabama. Suresh *et al* (2010) [10] presented distribution based length of interval is used for forecasting the accidents occurred in India. Singh (2007) [6] proposed first order time-invariant model with compared high-order model.
Chen *(*2000) [1] proposed two factors time-variant fuzzy time series model. Sullivan *et al* (1994) [8] proposed Markov model which used linguistic labels with probability distributions. Hongxu Wang *et al* (2014) [3] forecasted using fuzzy time series on students enrollments data. Viswam and Satyanarayana Reddy (2018) [11] predicted short term share market using Autoregressive Integrated Moving Average (ARIMA) model. Wanie *et al* (2020) [12] predicted rainfall data in Tasik Kenyir using neural networks. Moulana *et al* (2020) [13] predicted short term rainfall and long term rainfall using machine learning methods. In this paper, fuzzy c-means and invariant Markov model were used for forecasting enrollments of University of Alabama during 1971 to 1992. The performance of these different models was evaluated using the forecasting accuracy criteria namely, the forecasting error and average forecasting error with Markov model.

Fuzzy Time Series
Song and Chissom (1994) [7] presented the concept of fuzzy time series based on the historical enrollments of the University of Alabama. Fuzzy time series used to handle forecasting problems. They presented the time-invariant fuzzy time series model and the time-variant fuzzy time series model based on the fuzzy set theory for forecasting the enrollments of the University of Alabama.

Markov process

Markov process is one in which the future value is independent of the past values, given the present value.

Markov chain

If for all n,

$$P\{X_n = a_n \mid X_{n-1} = a_{n-1}, \ldots, X_0 = a_0\} = P\{X_n = a_n \mid X_{n-1} = a_{n-1}\},$$

then the process $\{X_n\}, n = 0,1,\ldots,$, is called a Markov chain.

Transition probability matrix

When the Markov chain is homogeneous, the step transition probability is denoted by p_{ij}. The matrix p = $\{p_{ij}\}$ is called a transition probability matrix. The transition probability matrix of a Markov chain is a stochastic matrix. since $p_{ij} \geq 0$ and $\sum_j p_{ij}$ = 1, for all i. that is the sum of all the elements of any row of the transition probability matrix is 1.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots p_{1n} \\ p_{31} & p_{32} & p_{33} & \cdots p_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{n3} & \cdots p_{nn} \end{bmatrix}$$

Fuzzy C-means Clustering

- Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters.
- This method was developed by Dunn in 1973 and improved by Bezdek in 1981.

Fuzzy *C*-means algorithm

Choose the number of clusters, *C* and m, typically *2*

Step 1: Initialise all u_{ij}, membership values randomly - matrix U^0

Step 2: At step *k*: Compute centroids, c_j using

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

Step 3: Compute new membership values, u_{ij} using Step1: Initialize U = $[u_{ij}]$ matrix, $U^{(0)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

Step 4: Update $U^{k+1} \leftarrow U^k$

Step 5: Repeat steps 2-4 until change of membership values is very small,

$U^{k+1} - U^k < \varepsilon$ where ε is some small value, typically 0.01

Step by step forecasting process

Step 1: Define the universe of discourse U is defined as

U= [Dmin-D1, Dmax+D2]

Where Dmin is the minimum value of the first order variation of the historical data, max is the maximum value of the historical data and D1, D2 are two positive integers.

Step 2: Partition the universe of discourse into equal length of intervals.

Step 3: Construct the fuzzy membership function using fuzzy c-means algorithm.

Step 4: Determines the fuzzy logical relationships using fuzzy c-means clustering.

Step 5: Find the fuzzy relation matrix RI by using fuzzy c- means membership coefficients.

Step 6: Defuzzify the group by using Markov matrix and the corresponding intervals.

Step 7: Forecast outputs: Pt+1=pt*Rm The predicted value is calculated the middle value of each interval is Multiplied by the minimum value of corresponding interval and sum.

Step 8: Forecast is obtained by,

Forecasting error = $\frac{|Forecast\ Value - Actual\ Value|}{Actual\ Value * 100\%}$

Average Forecasting error = $\frac{Sum\ of\ the\ forecasting\ error}{Total\ number\ of\ error}$

Computational Results

Step 1: Define the universe of discourse U is defined as
U = [13055-55, 19333+667]

Step 2: Partition the universe of discourse U into the equal intervals:
U_1 = [13000, 14000]
U_2 = [14000, 15000
U_3 = [15000, 16000]
U_4 = [16000, 17000]
U_5 = [17000, 18000]
U_6 = [18000, 19000]
U_7 = [19000, 20000].

Step 3: Construct the fuzzy membership functions and fuzzified variations using fuzzy c-means algorithm.

Table 1: Membership coefficients (in %, rounded)

Year	Actual Enrollments	Cluster 1 (A_1)	Cluster 2 (A_2)	Cluster 3 (A_3)	Cluster 4 (A_4)	Cluster 5 (A_5)	Cluster 6 (A_6)	Cluster 7 (A_7)	Fuzzified variations
1971	13055	62	10	9	7	5	4	3	A_1
1972	13563	89	3	3	2	1	1	1	A_1
1973	13867	67	10	8	6	4	3	2	A_1
1974	14696	15	37	20	13	7	4	3	A_2
1975	15460	1	4	92	2	1	0	0	A_3
1976	15311	3	39	42	9	3	2	1	A_3
1977	15603	3	14	54	21	5	2	2	A_3
1978	15861	1	5	9	80	3	1	1	A_4
1979	16807	1	2	2	3	90	2	1	A_5
1980	16919	1	2	2	4	87	3	1	A_5
1981	16388	4	10	14	32	29	7	4	A_4
1982	15433	1	7	86	4	1	1	0	A_3
1983	15497	1	6	86	5	1	1	0	A_3
1984	15145	1	94	3	1	0	0	0	A_2
1985	15163	1	94	3	1	0	0	0	A_5
1986	15984	1	3	5	85	3	1	1	A_4
1987	16859	0	1	1	1	95	1	1	A_5
1988	18150	0	0	0	0	0	99	0	A_6
1989	18970	2	3	3	4	6	16	66	A_7
1990	19328	1	1	1	1	1	3	92	A_7
1991	19337	1	1	1	1	2	3	91	A_7
1992	18876	3	4	4	5	7	22	56	A_7

Table 1 shows that membership coefficients using fuzzy c-means clustering and fuzzified variation. The highest membership measured as fuzzified varitions.

Step 4: The historical variations of the time series data are fuzzified in order to have the fuzzy logical relations obtained as follows:
Variations in the fuzzy logic relationships
$A_1 \rightarrow A_1$ $A_1 \rightarrow A_2$ $A_2 \rightarrow A_3$ $A_3 \rightarrow A_3$ $A_3 \rightarrow A_5$ $A_5 \rightarrow A_5$

$A_5 \rightarrow A_7$ $A_7 \rightarrow A_4 A_4 \rightarrow A_2$ $A_2 \rightarrow A_1$ $A_1 \rightarrow A_4$ $A_4 \rightarrow A_3$
$A_3 \rightarrow A_4$ $A_4 \rightarrow A_4 A_4 \rightarrow A_7$ $A_7 \rightarrow A_8$ $A_8 \rightarrow A_7$ $A_5 \rightarrow A_5$
$A_5 \rightarrow A_4$ $A_4 \rightarrow A_2$

Step 5: The Transition probability matrix is:

$$\begin{bmatrix} 0.67 & 0.33 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.33 & 0.33 & 0.33 & 0 & 0 & 0 \\ 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0.33 & 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0 & 0.33 & 0.33 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 6: The forecasted values have been obtained by using the computation algorithm.

Table 2: Actual, Forecasted and Forecasting Error Enrollments

Year	Actual Enrollments	Forecasted Enrollments	Forecasting error
1971	13055	---	---
1972	13563	13665	0.75
1973	13867	13665	1.46
1974	14696	15015	2.17
1975	15460	15300	1.03
1976	15311	15300	0.07
1977	15603	15300	1.94
1978	15861	16340	3.02
1979	16807	16995	1.12
1980	16919	16995	0.45
1981	16388	16340	0.29
1982	15433	15300	0.86
1983	15497	15300	1.27
1984	15145	15015	0.86
1985	15163	15015	0.98
1986	15984	16340	2.23
1987	16859	16995	0.81
1988	18150	18500	1.93
1989	18970	19500	2.79
1990	19328	19500	0.89
1991	19337	19500	0.84
1992	18876	19500	3.31

Table 2 Shows that actual and forecasted student enrollments of University of Alabama from 1971 to 1992 and also the difference between actual and estimated value is called forecasting error.

Step 7: Forecast is obtained by:
Forecast error = 29.07
Average Forecast error = 1.38

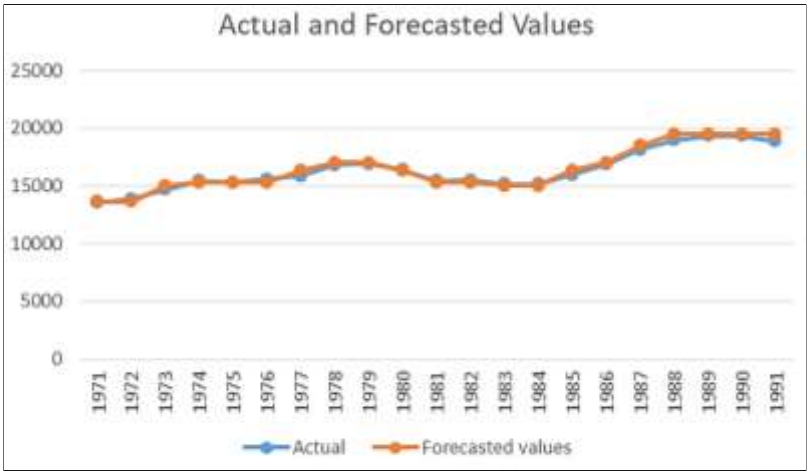

Fig 1: Actual and Forecasted Values of Enrollment Data

Figure 1 shows that the actual and forecasted values of enrollment data. All the actual and forecasted values are closely related to one another.

Step 8: Average forecasting error with Markov Model:

Table 3: Average forecasting error with Markov Model

Song & Chissom time invariant model	Chen's time-invariant model	Malike Sah & Konstain time- invariant model	K.Senthamarai Kannan *et al* time-invariant model	S. Suresh *et al* Markov model	Proposed
3.18%	3.23%	2.42%	1.59%	1.49%	1.38%

Table 3 Shows that average forecasting error is minimum when compared to other model

Conclusion

A modified fuzzy Markov model used for forecasting the enrollments of University of Alabama. Fuzzy c-means used for membership coefficients. The highest membership value is fuzzified variations. It provides higher accuracy when comparing Song & Chissom time invariant model, Chen's time-invariant model, Malike Sah & Konstain time- invariant model, K.Senthamarai Kannan *et al* time-invariant model and S. Suresh *et al* Markov model. It gives minimum average forecasting error when comparing the existing methods. Average forecasting error value is 1.38%.

Reference

1. Chen SM, Hwang JR. Temperature prediction using fuzzy time series, IEEE Transactions on Systems, Man, Cybernetics-Part B: Cybernetics 2000;30:263-275.
2. Chen SM, Hsu C. A new method to forecasting enrollments using fuzzy time series, International Journal of Applied Science and Engineering 2004;3:324-244.
3. Hongxu Wang, Hui Wang, Jianchun Guo, Hao Feng. A Fuzzy Time Series Forecasting Model Based on Yearly Difference of the Student Enrollment Number, International Conference on Soft Computing in Information Communication Technology (SCICT 2014), Atlantis Press 2014,172-175p.
4. Hwang JR, Chen SM, Lee HC. Handling forecasting problems using fuzzy time series, Fuzzy sets and Systems 1998;100:217-228.
5. Melike S, Konstsntin YD. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series, In proc. Int. Conf. on Computational Intelligence, Istanbul, Turkey 2004.
6. Shiva Raj Singh. A Simple Time Variant Method for Fuzzy Time Series Forecasting, Cybernetics and Systems 2007;38:305-321.
7. Song Q, Chissom BS. Forecasting Enrollments with Fuzzy Time Series, Part-1, Fuzzy sets Systems 1994;54:1-9.
8. Sullivan J, Woodall WH. A Comparison of Fuzzy Forecasting and Markov Modeling, Fuzzy Sets Syst 1994;64:279-293.
9. Suresh S, Senthamari Kannan K, Arumugam P. Forecasting model on fuzzy Markov Chain, International journal of Statistics and Systems 2009;4(2):137-147.
10. Suresh S, Senthamari Kannan K. Generalized Rule Based on Fuzzy Time Series, National Conference on Recent Trends in Statistical Research 2010,67-76p.
11. Viswam N, Satyanarayana Reddy G. Stock Market Prediction using Time Series Analysis, International Journal of Statistics and Applied Mathematics 2018;3(1):465-469.
12. Wanie M Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, Ahmed El-Shafie. Rainfall Forecasting Model using Machie Learning Methods, Ain Shams Engineering Journal 2020,1-13p.
13. Moulana Mohammed, Roshitha Kolapalli, Niharika Golla, Siva Sai Maturi. Prediction of Rainfall Using Machine Learning Techniques, International Journal of Scientific & Technology Research 2020;9:3236-3240.