Applicability of Soft computing technique for Crime Forecasting: A Preliminary Investigation

Anand Kumar Shrivastav
Research Scholar, Department of Computer Science
Mewar University, Chittorgarh, India
shrivastav.anand@gmail.com

Dr. Ekata
Associate Professor, Department of Applied Science
Krishna Institute of Engineering & Technology, Ghaziabad, India
ekata4@rediffmail.com

Abstract— **Soft Computing is a multidisciplinary field originally proposed by Dr. Lotfi Zadeh. It is the fusion of methodologies designed to model solution to real world problems that are otherwise difficult to model mathematically. Fuzzy Logic, Neural Network and Evolutionary Computation are main constituents of Soft Computing. Soft computing techniques have been successfully implemented in past to solve a number of problems including forecasting. Forecasting of Crime is an interesting subject, though meagerly used in law enforcement. Fuzzy time series methods for forecasting of student enrollment at University of Alabama have been used by many researchers in past. In this paper, we investigate the applicability of fuzzy time series technique for crime prediction.**

Keywords- Soft computing, Time series, Crime, Fuzzy logic, Forecasting

I. Introduction

The increasing volume of crimes has been one of the serious issues for police. In order to strengthen the public security, the police maintains crime database. Police use this database to examine criminal phenomena and related factors of past incidents, so that crime prevention strategies can be framed and implemented. The analysis of crime data may help to understand the behavior of the trend over time and future values may be predicted from past observations. Time series forecast is one of the most important tools for research in the field of social sciences and forecasting has become one of the most integral parts of everyday life in all societies. Crime prediction is an emerging approach in criminal justice studies and criminological research. By reviewing historical data over time, researchers, criminologists and justice practitioners can better understand the pattern of past behavior of crime variables and better predict the future crime behavior [1]. The time series forecasting methods have been used in very wide areas including finance, business, science and engineering, and many interdisciplinary areas. Normally, forecasting is made from historical data. The research activity in the area of combined application of intelligent computing technologies was initiated by Zadeh (1994) who first used the term soft computing, which he defined as a “collection of methodologies that aim to exploit the tolerance for imprecision and uncertainty to achieve tractability, robustness, and low solution cost” [2]. Soft computing is a collection of techniques which uses the human mind as a model and aims at formalizing our cognitive processes, which are subject to operate in an uncertain and imprecise environment. The soft computing approach helps in study of, complex phenomena and builds a low cost and analytic model with complete solutions. The main constituents of soft computing are Fuzzy logic, Neural network and Evolutionary computing. The forecasting models can be made using any of these constituents of soft computing or their combination. The term “fuzzy” was introduced by Zadeh (1965) in his paper on fuzzy sets, where a new mathematical discipline, fuzzy logic, based on the theory of fuzzy sets, was presented [3]. Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation [4]. The Fuzzy Time Series concept was first proposed by Song and Chissom [5]. The high computational overheads are the main drawback of the fuzzy time series model developed by Song and Chissom. The forecasting methods based on fuzzy time series, have been used by many researchers in last two decades [5], [6], [7], [8], [10], [11], [12], [13] and [14]. While most researchers have concentrated on the University of Alabama student enrollment problem, others have applied fuzzy time series technique to forecast the TAIFEX and road accident etc.

The present paper is focused at probing the applicability of fuzzy time series technique in crime prediction. We have utilized historic data of crime incidents (cases of murder in Delhi City) available at www.ncrb.gov.in , to build and test our model. While most researchers have partitioned the universe of discourse into seven intervals, we have partitioned the universe of discourse into five, ten and twenty intervals in order to feel the effect of partitioning. This study is aimed to get some reliable prediction for crime incidents for a lead year. The prediction may help the police and other law enforcement agencies in planning the crime prevention strategies. The rest of this paper is organized as follows. In Section 2, we review the concepts of time series and fuzzy time series. In Section 3, we present method to forecast crime enrollments based on fuzzy time series and show the experimental results. The conclusions are discussed in Section 4.

II. Basic Concepts and Definitions

Time Series

A time series is a time-ordered sequence of observation values of a physical or financial variable made at equally spaced time intervals Δt, represented as a set of discrete values x_1, x_2 , x_3 , ..., etc. Traditionally, time series analysis is defined as a branch of statistics that generally deals with the structural dependencies between the observation data of random phenomena and the related parameters. The observed phenomena are indexed by time as the only parameter; therefore, the name time series is used. The analysis of a given time series is primarily aimed at studying its internal structure (autocorrelation, trend, seasonality, etc.), to gain a better understanding of the dynamic process by which the time series data are generated [15].

Fuzzy time series

The Fuzzy time series forecasting was originally conceived by Song & Chissom and they proposed it in a series of papers to forecast student enrollments at the University of Alabama [5], [6], [9]. The Fuzzy time series models developed by Song and Chissom were associated with high computational overheads due to complex matrix operations. To reduce the computational overhead, Chen [7] simplified the process and proposed a simplified model that includes only simple arithmetic operations.

Definition 1: Fuzzy Time Series

Let $Y(t)$ (t= …, 0,1,2,….), a subset of real numbers, be the universe of discourse on which fuzzy sets $f_i(t)$ (i = 1,2,…) are defined. If $F(t)$ is a collection of $f(t)$(i=1,2,…), then $F(t)$ is called a fuzzy time series on $Y(t)$ (t = …, 0,1,2,….).

Definition 2: Fuzzy Relation

If there exists a fuzzy relationship $R(t\text{-}1), t)$ such that $F(t)$=F(t-1) x $R(t\text{-}1), t)$, where x represents an operator, then $F(t)$ is said to be caused by $F(t\text{-}1)$. The relationship between $F(t)$ and $F(t\text{-}1)$ is denoted by: $F(t\text{-}1)\rightarrow F(t)$. Examples of operators from literature are the max-min composition, the min-max composition and arithmetic operator. If $F(t\text{-}1) = A_i$ and $F(t) = A_j$, the logical relationship between $F(t)$ and $F(t\text{-}1)$ is denoted by $A_i \rightarrow A_j$, where A_i is called the left hand side and A_j, the right hand side of the fuzzy relation. The variable t denotes the time. For example, if t = 1995, the fuzzy relationship between $F(t)$ and $F(t\text{-}1)$ is given by $F(1994)\rightarrow F(1995)$. Note that the right hand side of the fuzzy relation represents the future fuzzy set (forecast). Its crisp counterpart is denoted as $Y(t)$.

Definition 3: *N*-Order Fuzzy Relations

Let $F(t)$ be a fuzzy time series. If $F(t)$ is caused by $F(t\text{-}1)$, $F(t\text{-}2)$, $F(t\text{-}3)$,……, $F(t\text{-}n)$, then this fuzzy relationship is represented by: $F(t\text{-}n)$,…..,$F(t\text{-}2)$, $F(t\text{-}1)\rightarrow F(t)$. It is called an n-order fuzzy time series. The n-order concept was first introduced by Chen. N-order based fuzzy time series models are referred to as high order models.

Definition 4: Time-Invariant Fuzzy Time Series

Suppose $F(t)$ is caused by $F(t\text{-}1)$ only and is denoted by $F(t\text{-}1)\rightarrow F(t)$, then there is a fuzzy relationship between $F(t)$ and $F(t\text{-}1)$ which is expressed as the equation: $F(t) = F(t\text{-}1) \times R(t\text{-}1, t)$.
The relation R is referred to as a first order model of $F(t)$. If $R(t\text{-}1, t)$ is independent of time t, that is for different times t_1 *and* t_2, $R(t_1, t_1\text{-}1)=R(t_2, t_2\text{-}1)$, *t*hen $F(t)$ is called a time-invariant fuzzy time series. Otherwise it is called a time-variant fuzzy time series.

Definition 5: Fuzzy Relationship Group

Relationship with the same fuzzy set on the left hand side can be further grouped into a relationship group. Relationship groups are also referred to as fuzzy logical relationship groups.
Suppose there are relationships such that

$A_i \rightarrow A_{j1}$.
$A_i \rightarrow A_{j2}$.
. . . .
. . . .
$A_i \rightarrow A_{jn}$.

then they can be grouped into a relationship group as follows: $A_i \rightarrow A_{j1}, A_{j2}$.. .. .. A_{jn}.
The same fuzzy set cannot appear more than once on the right hand side or the relationship group. The step-by-step procedure proposed by Chen[7] is listed as:

1. Partition the universe of discourse into equally lengthy intervals.
2. Define fuzzy sets on the universe of discourse.
3. Fuzzify historical data.
4. Identify fuzzy relationships (FLR's).
5. Establish fuzzy relationship groups (FLRG's).
6. Defuzzify the forecasted output.

III. METHOD USED

We have followed the method of [7] and [14] for crime prediction. While [7] and [14] have partitioned the universe of discourse into seven and ten intervals respectively, we have used three different partitioning schemes (i.e. five intervals, ten intervals and twenty intervals) to probe the effect of partitioning on accuracy of forecast results, if any. For analysis purpose, seventeen years (1995-2011) actual crime data of Delhi city under the crime head "murder" have been used. There are various factors that may influence the happening of a particular type of crime in a city. These factors may affect the forecast result, however, in this study; these factors have not been considered and these have been left for future studies.

Step 1: Define the universe of discourse U and partition it into several even and equal length intervals u_1, u_2, ..., and u_n. The universe of discourse U is defined as [D_{min} –D_1, D_{max} + D_2] where D_{min} and D_{max} are the minimum and maximum of cases registered, respectively and D_1, D_2 are two proper positive numbers. From Table-1, $D_{min(delhi)}$= 369, $D_{max(delhi)}$= 523. We have, $D_{1(delhi)}$=69 and $D_{2(delhi)}$=77. Accordingly, the universe of discourse is, $U_{(Delhi)}$ =[300, 600]. The universe of discourse $U_{(Delhi)}$ is partitioned into 5, 10 and 20 equal length intervals and are shown in Table-2. The actual cases registered under the crime head murder and their respective positions in partitioned intervals are shown in Table-1.

Table-1
Actual cases and their position in partitioned intervals

Year	**Cases registered**	**Scheme-I** Five intervals	**Scheme-II** Ten intervals	**Scheme-III** Twenty intervals
1995	413	360-420	390-420	405-420
1996	426	420-480	420-450	420-435
1997	475	420-480	450-480	465-480
1998	523	480-540	510-540	510-525
1999	514	480-540	510-540	510-525
2000	477	420-480	450-480	465-480
2001	450	420-480	450-480	450-465
2002	438	420-480	420-450	435-450
2003	401	360-420	390-420	390-405
2004	401	360-420	390-420	390-405
2005	369	360-420	360-390	360-375
2006	396	360-420	390-420	390-405
2007	411	360-420	390-420	405-420
2008	451	420-480	450-480	450-465
2009	444	420-480	420-450	435-450
2010	453	420-480	450-480	450-465
2011	438	420-480	420-450	435-450

Table-2
Partitioning of universe of discourse

Scheme-I (Five intervals)	u_1 = [300, 360], u_2 = [360, 420], u_3=[420, 480], u_4= [480, 540], u_5= [540, 600].
Scheme-II (Ten intervals)	u_1 = [300, 330], u_2 = [330, 360], u_3=[360, 390], u_4= [390, 420], u_5= [420, 450], u_6=[450, 480], u_7=[480, 510], u_8=[510, 540], u_9=[540, 570], u_{10}=[570, 600].
Scheme-III (Twenty intervals)	u_1 = [300, 315], u_2 = [315, 330], u_3=[330, 345], u_4= [345, 360], u_5= [360, 375], u_6=[375, 390], u_7=[390, 405], u_8=[405, 420], u_9=[420, 435], u_{10}=[435, 450], u_{11}=[450, 465], u_{12}=[465, 480], u_{13}=[480, 495], u_{14}=[495, 510], u_{15}=[510, 525], u_{16}=[525, 540], u_{17}=[540, 555], u_{18}=[555, 570], u_{19}=[570, 585], u_{20}=[585, 600].

Step 2: Get a statistics of the distribution of the actual registered cases (historical enrollments) in each interval. Sort the intervals based on the number of historical enrollment data in each interval from the highest to the lowest. Find the interval having the largest number of historical enrollment data and divide it into four sub-intervals of equal length. Find the interval having the second largest number of historical enrollment data and divide it into three sub-intervals of equal length. Find the interval having the third largest number of historical enrollment data and divide it into two sub-intervals of equal length. Find the interval with the fourth largest number of historical enrollment data and let the length of this interval remain unchanged. If there are no data distributed in an interval, then discard this interval. The distribution of actual cases registered in partitioned intervals is given in Table-3. The re-divided and renamed universe of discourse is shown in Table-4.

Table-3
Distribution of registered cases in partitioned intervals

Frequency of data					
Scheme-I (Five Intervals)	Cases	Scheme-II (Ten Intervals)	Cases	Scheme-III (Twenty Intervals)	Cases
300-360	0	300-330	0	300-315	0
360-420	6	330-360	0	315-330	0
420-480	9	360-390	1	330-345	0
480-540	2	390-420	5	345-360	0
540-600	0	420-450	4	360-375	1
---	---	450-480	5	375-390	0
---	---	480-510	0	390-405	3
---	---	510-540	2	405-420	2
---	---	540-570	0	420-435	1
---	---	570-600	0	435-450	3
---	---	---	---	450-465	3
---	---	---	---	465-480	2
---	---	---	---	480-495	0
---	---	---	---	495-510	0
---	---	---	---	510-525	2
---	---	---	---	525-540	0
---	---	---	---	540-555	0
---	---	---	---	555-570	0
---	---	---	---	570-585	0
---	---	---	---	585-600	0

Table-4
Re-partitioned universe of discourse

Scheme-I (Five intervals)	u_1 = [420-435], u_2 = [435-450], u_3=[450-465], u_4= [465-480], u_5 = [360-380], u_6 = [380-400], u_7=[400-420], u_8 = [480-510], u_9 = [510-540].
Scheme-II (Ten intervals)	u_1= [360-390], u_2 = [390-397.5], u_3 = [397.5-405], u_4=[405-412.5], u_5= [412.5-420], u_6 = [450-457.5], u_7 = [457.5-465], u_8=[465-472.5], u_9= [472.5-480], u_{10} = [420-430], u_{11} = [430-440], u_{12}=[440-450], u_{13} = [510-525], u_{14} = [525-540].
Scheme-III (Twenty intervals)	u_1=[360- 367.5], u_2=[367.5-375], u_3=[390- 393.75], u_4=[393.75-397.5], u_5=[397.5-401.25], u_6=[401.25-405], u_7=[405- 410], u_8=[410-415], u_9=[415-420], u_{10}=[420- 427.5], u_{11}=[427.5-435], u_{12}=[435- 438.75], u_{13}=[438.75 - 442.5], u_{14}=[442.5-446.25], u_{15}=[446.25-450], u_{16}=[450- 453.75], u_{17}=[453.75-457.5], u_{18}=[457.5-461.25], u_{19}=[461.25-465], u_{20}=[465- 470], u_{21}=[470-475], u_{22}=[475-480], u_{23}=[510- 515], u_{24}=[515-520], u_{25}=[520-525].

Step 3: Define each fuzzy set A_i based on the re-divided intervals and fuzzify the actual registered cases shown in Table 1, where fuzzy set A_i denotes a linguistic value of the cases registered represented by a fuzzy set, and $1 \leq i \leq 9$, $1 \leq i \leq 14$ and $1 \leq i \leq 25$. Then, fuzzify the actual registered cases in Table 1. For simplicity, the membership values of fuzzy set A_i either are 0, 0.5 or 1. We have not displayed the membership value 0. Fuzzy sets are shown in Table-5.

Table-5
Defining Fuzzy set

	Scheme-I Five intervals ($1 \le i \le 9$)	Scheme-II Ten intervals ($1 \le i \le 14$)	Scheme-III Twenty intervals ($1 \le i \le 25$)
$A_1 =$	$1/u_1 + 0.5/u_2$	$1/u_1 + 0.5/u_2$	$1/u_1 + 0.5/u_2$
$A_2 =$	$0.5/u_1 + 1/u_2 + 0.5/u_3$	$0.5/u_1 + 1/u_2 + 0.5/u_3$	$0.5/u_1 + 1/u_2 + 0.5/u_3$
$A_3 =$	$0.5/u_2 + 1/u_3 + 0.5/u_4$	$0.5/u_2 + 1/u_3 + 0.5/u_4$	$0.5/u_2 + 1/u_3 + 0.5/u_4$
$A_4 =$	$0.5/u_3 + 1/u_4 + 0.5/u_5$	$0.5/u_3 + 1/u_4 + 0.5/u_5$	$0.5/u_3 + 1/u_4 + 0.5/u_5$
$A_5 =$	$0.5/u_4 + 1/u_5 + 0.5/u_6$	$0.5/u_4 + 1/u_5 + 0.5/u_6$	$0.5/u_4 + 1/u_5 + 0.5/u_6$
$A_6 =$	$0.5/u_5 + 1/u_6 + 0.5/u_7$	$0.5/u_5 + 1/u_6 + 0.5/u_7$	$0.5/u_5 + 1/u_6 + 0.5/u_7$
$A_7 =$	$0.5/u_6 + 1/u_7 + 0.5/u_8$	$0.5/u_6 + 1/u_7 + 0.5/u_8$	$0.5/u_6 + 1/u_7 + 0.5/u_8$
$A_8 =$	$0.5/u_7 + 1/u_8 + 0.5/u_9$	$0.5/u_7 + 1/u_8 + 0.5/u_9$	$0.5/u_7 + 1/u_8 + 0.5/u_9$
$A_9 =$	$0.5/u_8 + 1/u_9$	$0.5/u_8 + 1/u_9$	$0.5/u_8 + 1/u_9 + 0.5/u_{10}$
$A_{10} =$		$0.5/u_9 + 1/u_{10} + 0.5/u_{11}$	$0.5/u_9 + 1/u_{10} + 0.5/u_{11}$
$A_{11} =$		$0.5/u_{10} + 1/u_{11} + 0.5/u_{12}$	$0.5/u_{10} + 1/u_{11} + 0.5/u_{12}$
$A_{12} =$		$0.5/u_{11} + 1/u_{12} + 0.5/u_{13}$	$0.5/u_{11} + 1/u_{12} + 0.5/u_{13}$
$A_{13} =$		$0.5/u_{12} + 1/u_{13}$	$0.5/u_{12} + 1/u_{13} + 0.5/u_{14}$
$A_{14} =$		$0.5/u_{13}$	$0.5/u_{13} + 1/u_{14} + 0.5/u_{15}$
$A_{15} =$			$0.5/u_{14} + 1/u_{15} + 0.5/u_{16}$
$A_{16} =$			$0.5/u_{15} + 1/u_{16} + 0.5/u_{17}$
$A_{17} =$			$0.5/u_{16} + 1/u_{17} + 0.5/u_{18}$
$A_{18} =$			$0.5/u_{17} + 1/u_{18} + 0.5/u_{19}$
$A_{19} =$			$0.5/u_{18} + 1/u_{19} + 0.5/u_{20}$
$A_{20} =$			$0.5/u_{19} + 1/u_{20} + 0.5/u_{21}$
$A_{21} =$			$0.5/u_{20} + 1/u_{21} + 0.5/u_{22}$
$A_{22} =$			$0.5/u_{21} + 1/u_{22} + 0.5/u_{23}$
$A_{23} =$			$0.5/u_{22} + 1/u_{23} + 0.5/u_{24}$
$A_{24} =$			$0.5/u_{23} + 1/u_{24} + 0.5/u_{25}$
$A_{25} =$			$00.5/u_{24} + 1/u_{25}$

Step 4: Establish fuzzy logical relationships based on the fuzzified enrollments:

$A_j \rightarrow A_q$

$A_j \rightarrow A_r$,

.. .. ..

where the fuzzy logical relationship " $A_j \rightarrow A_q$ " denotes " if the fuzzified enrollments of year n-1 is A_j , then the fuzzified enrollments of year n is A_q ". The fuzzy logical relationships are shown in Table 6.

Table-6
Fuzzy Logic Relationship

Scheme-I (Five Intervals)	**Scheme-II (Ten Intervals)**	**Scheme-III (Twenty Intervals)**
$A_3 \rightarrow A_4$, $A_4 \rightarrow A_7$, $A_7 \rightarrow A_9$, $A_9 \rightarrow A_9$, $A_9 \rightarrow A_7$, $A_7 \rightarrow A_6$, $A_6 \rightarrow A_5$, $A_5 \rightarrow A_3$, $A_3 \rightarrow A_3$, $A_3 \rightarrow A_1$, $A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$, $A_3 \rightarrow A_6$, $A_6 \rightarrow A_5$, $A_5 \rightarrow A_6$, $A_6 \rightarrow A_5$	$A_1 \rightarrow A_2$, $A_5 \rightarrow A_6$, $A_9 \rightarrow A_7$, $A_9 \rightarrow A_8$, $A_2 \rightarrow A_4$, $A_6 \rightarrow A_{12}$, $A_{12} \rightarrow A_9$, $A_{12} \rightarrow A_{13}$, $A_3 \rightarrow A_1$, $A_3 \rightarrow A_3$, $A_7 \rightarrow A_3$, $A_{13} \rightarrow A_{12}$, $A_{13} \rightarrow A_{13}$, $A_4 \rightarrow A_9$, $A_8 \rightarrow A_9$	$A_8 \rightarrow A_{10}$, $A_{10} \rightarrow A_{22}$, $A_{22} \rightarrow A_{25}$, $A_{25} \rightarrow A_{23}$, $A_{23} \rightarrow A_{22}$, $A_{22} \rightarrow A_{16}$, $A_{16} \rightarrow A_{12}$, $A_{12} \rightarrow A_5$, $A_5 \rightarrow A_5$, $A_5 \rightarrow A_2$, $A_2 \rightarrow A_4$, $A_4 \rightarrow A_7$, $A_7 \rightarrow A_{16}$, $A_{16} \rightarrow A_{14}$, $A_{14} \rightarrow A_{16}$, $A_{16} \rightarrow A_{12}$

Step 5: Establish fuzzy logical relationships groups. Fuzzy logical relation groups are shown in Table-7.

Table-7
Fuzzy relation groups

Scheme-I (Five Intervals)		**Scheme-II (Ten Intervals)**		**Scheme-III (Twenty Intervals)**	
Group-1	$A_1 \rightarrow A_2$	Group-1	$A_1 \rightarrow A_2$	Group1	$A_2 \rightarrow A_4$
Group-2	$A_2 \rightarrow A_3$	Group-2	$A_2 \rightarrow A_4$	Group2	$A_4 \rightarrow A_7$
Group-3	$A_3 \rightarrow A_1$, $A_3 \rightarrow A_3$, $A_3 \rightarrow A_4$, $A_3 \rightarrow A_6$	Group-3	$A_3 \rightarrow A_1$	Group3	$A_5 \rightarrow A_2, A_5$
Group-4	$A_4 \rightarrow A_7$	Group-4	$A_4 \rightarrow A_9$	Group4	$A_7 \rightarrow A_{16}$
Group-5	$A_5 \rightarrow A_3$, $A_5 \rightarrow A_6$	Group-5	$A_5 \rightarrow A_6$	Group5	$A_8 \rightarrow A_{10}$
Group-6	$A_6 \rightarrow A_5$	Group-6	$A_6 \rightarrow A_{12}$	Group6	$A_{10} \rightarrow A_{22}$
Group-7	$A_7 \rightarrow A_6$, $A_7 \rightarrow A_9$	Group-7	$A_7 \rightarrow A_3$	Group7	$A_{12} \rightarrow A_5$
Group-8	$A_9 \rightarrow A_7$, $A_9 \rightarrow A_9$	Group-8	$A_8 \rightarrow A_9$	Group8	$A_{14} \rightarrow A_{16}$
		Group-9	$A_9 \rightarrow A_7, A_8$	Group9	$A_{16} \rightarrow A_{12}, A_{14}$
		Group-10	$A_{12} \rightarrow A_9, A_{13}$	Group10	$A_{22} \rightarrow A_{16}, A_{25}$
		Group-11	$A_{13} \rightarrow A_{12}, A_{13}$	Group11	$A_{23} \rightarrow A_{22}$
				Group12	$A_{25} \rightarrow A_{23}$

Step 6: We have followed the principles stated by Chen [7] to defuzzify the forecasted output. According to this principle

1. If there exists a one-to-one relationship in the relationship group of A_j, say $A_j \rightarrow A_k$, and the highest degree of belongingness of A_k occurs at interval u_k, then the forecasted output of F (t) equals the midpoint of u_k.
2. If A_j is empty, i.e. $A_j \rightarrow \emptyset$, and the interval where A_j has the highest degree of belongingness is u_j, then the forecasted output equals the midpoint of u_j.
3. If there exists a one-to-many relationship in the relationship group of A_j, say $A_j \rightarrow A_1$, A_2 …, A_n, and the highest degrees of belongingness occurs at set u_1 , u_2 ,…, u_n , then the forecasted output is computed as the average of the midpoints m_1 , m_2,…,m_n of u_1 , u_2,…,u_n . This equation can be expressed as:

$$\frac{m_1 + m_2 + .. m_n}{n}$$

The forecast so obtained along with actual registered cases is shown in Table-8.

Table-8
Actual registered cases versus forecast

Year	Actual registered Cases	Forecast		
		Scheme-I (Five Intervals)	Scheme-II (Ten Intervals)	Scheme-III (Twenty Intervals)
1995	413			
1996	426	416.25	425	423.75
1997	475	485	476.25	477.50
1998	523	498.75	485.625	487.19
1999	514	498.75	496.875	512.50
2000	477	485	496.875	477.50
2001	450	450	476.25	487.19
2002	438	436.66	440	440.62
2003	401	416.25	401.25	399.37
2004	401	416.25	388.125	385.31
2005	369	380	388.125	385.31
2006	396	400	393.75	395.62
2007	411	416.25	408.75	407.50
2008	451	450	453.75	451.88
2009	444	436.66	440	440.62
2010	453	450	453.75	451.87
2011	438	436.66	440	440.62
	MSE	-0.3425	0.164063	0.165625
	MAD	8.25125	9.445313	7.993125

To verify the accuracy of the forecast results, Mean Forecast Error (MFE) and Mean Absolute Deviation (MAD) have been calculated as:

$$MFE = \frac{\sum_{i=1}^{n}(e_i)}{n}; \text{ and } MAD = \frac{\sum_{i=1}^{n}|e_i|}{n}$$

The MFE and MAD are shown in Table-8. It is obvious from the values of MFE and MAD that the five interval model tends to slightly over forecast with an average absolute error of 8.25 units, the ten interval model is working correctly with an average absolute error of 9.44 units, and the twenty interval model is working correctly with an average absolute error of 7.99 units. The forecasted values are very close to the actual values.

IV. CONCLUSION

In this paper, we have investigated applicability of Fuzzy Time series technique for crime forecasting. The scope of present work is limited to probe the suitability of fuzzy time series method to predict the crime and to provide practical computational techniques with simple algorithm and higher degree of output accuracy. In this study, seventeen years actual historic crime data of Delhi city have been used. We have implemented the Fuzzy Time Series method on three different sets of universe of discourses obtained by partitioning it into five intervals (Scheme-I), ten intervals (Scheme-II) and twenty intervals (Scheme-III). The computed forecast shows good accuracy. The forecast results obtained from Scheme-II and Scheme-III are satisfactory while scheme-I results tends to slightly over forecast. There are various factors that may influence the happening of a particular type of crime. It is assumed that if those factors are also taken care of, then the fuzzy time series method could produce better results and can be used as a tool for effective crime prevention strategies. This dimension has been left for future studies.

REFERENCES

[1] G. Solomon Osho, Ayodeji Daramola; “An Empirical Infusion of Time Series Crime Forecasting Modeling into Crime Prevention Approach for Use by Law Enforcement Agencies: An Econometrics Perspective”; National Forum of Multicultural issues Journal, Volume 7, Number 1, 2010.

[2] Zadeh LA (1994) Soft computing and fuzzy logic. IEEE Software, Nov.: 48-58.

[3] Zadeh LA (1965) Fuzzy Sets, Information and Control 8: 338-353

[4] L.A. Zedah, “Fuzzy Logic, Neural Networks, and Software Computing”, Communications of the ACM, 37 (3), 77, March 1994.

[5] Q. Song and B. S. Chissom, “Forecasting Enrollment with Fuzzy Time Series-Part I,” Fuzzy Sets and Systems, Vol. 54, No. 1, 1993, pp. 1-9.

[6] Q. Song, B. S. Chissom, “Forecasting enrollments with fuzzy time series: Part II”, Fuzzy Sets and systems, vol. 62: pp. 1-8, 1994.

[7] S.M. Chen, “Forecasting enrollments based on fuzzy time series”, Fuzzy Sets and Systems 81 (1996), pp. 311-319.

[8] Z. Ismail, R. Efendi, “Enrollment forecasting based on modified weight fuzzy time series”, Journal of Artificial Intelligence, vol. 4(1), pp. 110-118, 2011],

[9] Song, B. S. Chissom, “Fuzzy time series and its models”, Fuzzy Sets and systems, vol. 54, pp. 269-277, 1993.

[10] C. H. Cheng, T. L. Chen and C. H. Chiang, “Trend Weighted Fuzzy Time Series Model for TAIEX Fore-casting,” Proceedings of the 13th International Conference on Neural Information Processing , Part-III, Lecture Notes in Computer Science, Hong Kong, Vol. 4234, 3-6 October 2006, pp. 469-477.

[11] J. R. Hwang, S. M. Chen, C. H. Lee, “Handling forecasting problems using fuzzy time series”, Fuzzy Sets and Systems, vol. 100, pp. 217-228, 1998.

[12] S. Melike, K.Y. Degtiarev, “Forecasting enrollments model based on first- order fuzzy time series”, Proceedings of World Academy of Science, Engineering and Technology, vol. 1, pp. 132-135, 2005.

[13] Stevenson, Porter, “Fuzzy time series forecasting using percentage change as the universe of discourse”, Proceedings of World Academy of Science, Engineering and Technology, vol. 55, pp. 154-157, 2009] .

[14] S. M. Chen and C. C. Hsu, “A New Method to Forecast Enrollment Using Fuzzy Time Series”, International Journal of Applied Sc ience and Engineering, Vol. 3, No. 2, 2004, pp. 234-244.

[15] Ajoy K. Palit, Dobrivoje Popovic, “Computational Intelligence in Time Series Forecasting Theory and Engineering Applications”, Springer-Verlag London Limited 2005.

[16] Lazim Abdullah, Chai Yoke Ling, “Intervals in Fuzzy Time Series Model Preliminary Investigation for Composite Index Forecasting”, ARPN Journal of Systems and Software Vol.2, No.1, January 2012.

[17] Hemant Kumar Pathak, Prachi Singh, “A New Bandwidth Interval Based Forecasting Method for Enrollments Using Fuzzy Time Series”, Applied Mathematics, 2011, 2, 504-507.

[18] Preetika Saxena, Kalyani Sharma, Santhosh Easo; “Forecasting Enrollments based on Fuzzy Time Series with Higher Forecast Accuracy Rate”; Int. J. Computer Technology & Applications,Vol 3 (3), 957-961 MAY-JUNE 2012

[19] T. A. Jilani, S. M. A. Burney, “M-factor high order fuzzy time series forecasting for road accident data”, In IEEE-IFSA 2007, World Congress, Cancun, Mexico, June 18-21, Forthcoming in Book series Advances in Soft Computing, Springer-Verlag, 2007.

[20] T. A. Jilani, S. M. A. Burney, C. Ardil, “Multivariate high order fuzzy time series forecasting for car road accidents”, International Journal of Computational Intelligence, vol. 4, no. 1, pp. 15-20, 2007.