Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-016-0548-5

ORIGINAL ARTICLE

Rainfall and financial forecasting using fuzzy time series and neural networks based model

Pritpal Singh[1]

Received: 6 January 2015 / Accepted: 12 May 2016

Abstract In this study, the author presents a new model to deal with four major issues of fuzzy time series (FTS) forecasting, viz., determination of effective lengths of intervals (i.e., intervals which are used to fuzzify the numerical values), repeated fuzzy sets, trend associated with fuzzy sets, and defuzzification operation. To resolve the problem of determination of length of intervals, this study suggests the application of an artificial neural network (ANN) based algorithm. After generating the intervals, the historical time series data set is fuzzified based on FTS theory. In part of existing FTS models introduced in the literature, each fuzzy set is given equal importance, which is not effective to solve real time problems. Therefore, in this model, it is recommended to assign weights on the fuzzy sets based on their frequency of occurrences. In the FTS modeling approach, fuzzified time series values are further used to establish the fuzzy logical relations (FLRs). To determine the trends associated with the fuzzy sets in the corresponding FLR, this article also introduces three trend-based conditions. To deal repeated fuzzy sets and trend associated with them, this study proposes a new defuzzification technique. The proposed model is verified and validated with real-world time series data sets. Empirical analyzes signify that the proposed model has the robustness to deal one-factor time series data sets very efficiently than existing FTS models. Experimental results show that the proposed model also outperforms over the conventional statistical models.

Keywords Fuzzy time series (FTS) · Artificial neural network (ANN) · Fuzzification · Indian summer monsoon rainfall (ISMR)

1 Introduction

Fuzzy time series (FTS) is a modeling approach, in which historical values of time series are used to forecast the future values by employing fuzzy set concept. Forecasting using FTS is applied to several areas, including forecasting university enrollments [3], sales, road accidents and financial forecasting [7, 8, 12, 53]. In a conventional time series, the recorded values of a special dynamic process are represented by crisp numerical values. However, in a FTS model, the recorded values of a special dynamic process are represented by linguistic values. Based on FTS concept, first forecasting model was introduced by [52]. They presented the FTS model by means of fuzzy relational equations involving max-min composition operation and applied the model to forecast the enrollments in the University of Alabama. In 1996, Chen [9] used simplified arithmetic operations, avoiding the complicated max-min operations and their method produced better results. Later, many studies provided some improvements to the existing methods in terms of effective lengths of intervals [46], fuzzification [13], fuzzy logical relationships [1] and defuzzification techniques [45, 47]. Detail description of the FTS modeling approach can be found in these articles [6, 33, 43, 44, 50, 57].

The literature review reveals that ANN and fuzzy logic theory have great capability in solving many real life problems, especially when it comes to the process of complex decision making [2, 20, 32, 58, 59]. So, it is beneficial to hybridize ANN and fuzzy set techniques

✉ Pritpal Singh
pritpal@tezu.ernet.in; drpritpalsingh82@gmail.com

1 Smt. Chandaben Mohanbhai Patel Institute of Computer Applications, CHARUSAT Campus, Anand 388421, Gujarat, India

together by substituting the demerits of one technique by the merits of another technique. These two techniques can be fused as:

- Application of ANN for designing fuzzy logic based systems, and
- Application of fuzzy logic for designing ANN based systems.

Many researchers proposed various hybridize based models to solve complex decision making problems, such as rough-fuzzy hybridization scheme for case generation [36], genetic fuzzy systems and ANN for stock price forecasting [19], genetic algorithm and rough set theory for stock price forecasting [14], hybridization of ANN and FTS for TAIEX forecasting [22, 63], hybridization of self organizing map (SOM) and wavelet transform for performing image compression [23], hybridization of SOM and fuzzy set for monitoring the millennium development goals [40]. Sarlin [41] developed three models for predictions of the global financial crisis, viz., a conventional statistical model, a back-propagation neural network (NN) and a neuro-genetic (NG) model that uses a genetic algorithm for choosing the optimal NN configuration.

In FTS modeling approach, determination of effective lengths of intervals for the historical time series data set is a major scientific research issue. To resolve this problem, self-organizing feature maps (SOFM), an ANN based technique, is hybridized with the FTS modeling approach. Hence, the proposed model is an ensemble approach of these two techniques. A recent study of Singh and Borah [48] suggested that the predictability of Indian Summer Monsoon Rainfall (ISMR) in seasonal time scale is possible using time series analysis, which also motivated us to design the model using FTS modeling approach. Hence, the proposed model is initially applied to seasonal rainfall time series data set of the north-eastern region of India. To demonstrate the effectiveness of the model, it is further applied to stock index data sets.

This paper is organized as follows. In Sect. 2, various problems corresponding to FTS model and effective measures to resolve these problems have been discussed. In Sect. 3, some basic concepts of fuzzy set theory has been explained with an overview of FTS. Application of ANN for designing the proposed model has been discussed in Sect. 4. A brief description of the data set is provided in Sect. 5. In Sect. 6, the proposed neuro-fuzzy hybridized model has been presented. In Sect. 7, statistical parameters for analyzing the performance of the model has been discussed. The performance of the model has been assessed and presented in the results Sect. 8. The application of the proposed model on the stock index data sets is discussed in Sect. 9. The conclusions and future works are discussed in Sect. 10.

2 Problem definitions

Literature review reveals that accuracy of FTS modeling approach depends on four major factors, viz., (a) Lengths of intervals, (c) Handling of repeated fuzzy sets, (d) Handling of trend associated with fuzzy sets, and (e) Defuzzification operation. In this study, problems associated with these four factors are tried to resolve based on the hybridization of ANN with FTS. The main purpose of this hybridization is explained next.

Factor 1 (Lengths of intervals). For fuzzification of time series data set, determination of lengths of the intervals of the historical time series data set is very essential. Most of the FTS models [9, 52] maintain the fixed lengths of intervals without providing any specific reason. Experimental result of Huarng [21] shows that the lengths of intervals always affect the results of forecasting. Therefore, self-organizing feature maps (SOFM), an ANN based technique, is used in this study to create effective lengths of intervals.

Factor 2 (Handling of repeated fuzzy sets). Most of the researchers [1, 47] have given equal importance to each of the fuzzy sets involved in FTS modeling approach, which is not an effective way to solve real time problems. Because, each fuzzy set represents various uncertainties involved in the domain. Hence, there are two possible ways to determine the weights for the fuzzy sets as: (i) assign their weights based on human interpretation, and (ii) assign weights based on their occurrences in the fuzzified time series data set. Assignment of weights based on human-knowledge is not an acceptable solution for real world problems as human-interpretation varies from one to another. Therefore, second way is considered, where all the fuzzy sets are given importance based on their frequency of occurrences. For example, if the fuzzy set A_i $(i = 0, 1, 2, \ldots, n)$ occurs 2 times in the fuzzified time series data set, then its weight is assigned as 2 in terms of percentage (i.e., 2 %). Sometimes, the nature of the event is very dynamic. Therefore, it is advantageous to assign weight in %, because it will help to capture the variation in datum in the particular time instance.

Factor 3 (Handling of trend associated with fuzzy sets). In the FTS modeling approach, fuzzified time series values are further used to establish the fuzzy logical relations (FLRs). These FLRs are represented in the form of "$A_i \rightarrow A_j$" (i and $j = 1, 2, \ldots, n$), where "A_i" and "A_j" represent *previous state* and *current state* of the FLR, respectively. Here, each i and j represents the indices of the fuzzy sets A_i and A_j, respectively. The *current state* of FLRs can exhibit three different trends with respect to the *previous state*, which can be determined based on the following three conditions:

Condition 1 If the index value (i) of previous state's fuzzy set (A_i) is less than the index value (j) of current state's fuzzy set (A_j), i.e., $i<j$, then the trend of current state's fuzzy set "A_j" will be *upward* ($\uparrow$).

Condition 2 If the index value (i) of previous state's fuzzy set (A_i) is greater than the index value (j) of current state's fuzzy set (A_j), i.e., $i>j$, then the trend of current state's fuzzy set "A_j" will be *downward* ($\downarrow$).

Condition 3 If the index value (i) of previous state's fuzzy set (A_i) is equal to the index value (j) of current state's fuzzy set (A_j), i.e., $i=j$, then the trend of current state's fuzzy set "A_j" will be *unchanged* ($=$).

Perviously, trends represented by the FLRs are not considered by researchers during final decisions making. In this model, these three trends (i.e., "$\uparrow$", "$\downarrow$" and "$=$") represented by the fuzzy sets are used in obtaining the forecasting results.

Factor 4 (Defuzzification operation). Song and Chissom [52] adopted the following method to forecast enrollments of the University of Alabama:

$$Y(t) = Y(t-1) \circ R, \tag{1}$$

where $Y(t-1)$ is the fuzzified enrollment of year $(t-1)$, $Y(t)$ is the forecasted enrollment of year t represented by fuzzy set, "$\circ$" is the max-min composition operator, and "R" is a fuzzy relation indicating fuzzy relationships between FTS. This method takes a large amount of computations to derive the fuzzy relation R, and the max-min composition operations of Eq. 1 will take a large amount of computation time when the fuzzy relation R of Eq. 1 is very big [9]. In 1996, Chen[9] used simplified arithmetic operations for defuzzification operation by avoiding this complicated max-min operations, and their method produced better results than Song and Chissom models [52]. Most of the existing FTS models have used Chen's defuzzification method [9] to acquire the forecasting results. However, forecasting accuracy of these models are not good enough, so it is required to adopt a better method that can be employed for the defuzzification operation. So, to resolve this problem, a new "Frequency-Based Defuzzification Technique" for the defuzzification operation, is introduced in this study.

3 Fuzzy sets and FTS—A brief overview

In 1965, Zadeh [64] introduced the fuzzy sets theory involving continuous set membership for processing data in the presence of uncertainty. He also presented fuzzy arithmetic theory and its application [65]. Next, a basic concepts of FTS are reviewed from Singh [43].

Definition 1 (Fuzzy Set) [64]. A fuzzy set is a class with varying degrees of membership in the set. Let U be the universe of discourse, which is discrete and finite, then fuzzy set A can be defined as follows.

$$A = \{\mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \ldots\} = \Sigma_i \mu_A(x_i)/x_i, \tag{2}$$

where μ_A is the membership function of A, $\mu_A\colon U \to [0,1]$, and $\mu_A(x_i)$ is the degree of membership of the element x_i in the fuzzy set A. Here, the symbol "$+$" indicates the operation of union and the symbol "/" indicates the separator rather than the commonly used summation and division in algebra, respectively.

When U is continuous and infinite, then the fuzzy set A of U can be defined as:

$$A = \int \mu_A(x_i)/x_i, \forall x_i \in U \tag{3}$$

where the integral sign "$\int$" stands for the union of the fuzzy singletons, $\mu_A(x_i)/x_i$.

The fuzzy time series (FTS) concept was proposed in Song and Chissom [52]. The definition of FTS is briefly reviewed as follows:

Definition 2 (Fuzzy Time Series (FTS)) [52]. Let $Y(t)(t=0,1,2,\ldots)$ be a subset of R and the universe of discourse on which fuzzy sets $\mu_i(t)(i=1,2,\ldots)$ are defined and let $F(t)$ be a collection of $\mu_i(t)(i=1,2,\ldots)$. Then, $F(t)$ is called a FTS on $Y(t)(t=0,1,2,\ldots)$.

In Definition 2, $F(t)$ is a function of time t, and $\mu_i(t)$ are linguistic values of $F(t)$, where $\mu_i(t)(i=1,2,\ldots)$ are represented by fuzzy sets, and the values of $F(t)$ can be different at different times, because the universe of discourse can be different at different times. FTS can be divided into two categories which are the time-invariant FTS and the time-variant FTS.

If $F(t)$ is caused by $F(t-1)$, i.e., $F(t-1) \to F(t)$, then this relationship can be represented as follows:

$$F(t) = F(t-1) \circ R(t, t-1) \tag{4}$$

where $R(t,t-1)$ is fuzzy relationship between $F(t)$ and $F(t-1)$. Here, R is the union of fuzzy relations and "$\circ$" is max-min composition operator. It is also called the first-order model of $F(t)$.

Definition 3 (Fuzzy time-variant and time-invariant series) [11]. Let $F(t)$ be a FTS, and $R(t,t-1)$ be a first–order model of $F(t)$. If $R(t,t-1) = R(t-1,t-2)$ for any time t, and $F(t)$ only has finite elements, then $F(t)$ is referred as a time-invariant FTS. Otherwise, it is referred as a time-variant FTS.

Definition 4 (Fuzzy logical relationship) [9, 52]. Assume that $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between $F(t)$ and $F(t-1)$ is referred as a fuzzy logical relationship (FLR), which can be represented as:

$$A_i \rightarrow A_j, \tag{5}$$

where A_i and A_j refer to the *previous state* and *current state* of the FLR, respectively.

Definition 5 (Fuzzy logical relationship group) [9, 52]. Assume the following FLRs:

$$A_i \rightarrow A_{k1},$$
$$A_i \rightarrow A_{k2},$$
$$\ldots$$
$$A_i \rightarrow A_{km}.$$

Chen [9] suggested that the FLRs having the same previous state can be grouped into a same fuzzy logical relationship group (FLRG). So, based on the Chen's model, these FLRs can be grouped into the same FLRG as:

$$A_i \rightarrow A_{k1}, A_{k2}, \ldots, A_{km}.$$

4 ANN and its application for creation of intervals

ANN is a computational model that is inspired by the human brain [51]. ANN is composed of a large number of interconnected *nodes* or *neurons*, which usually operate in parallel, and are configured in regular architectures. Researchers employ ANN in various forecasting problems such as electric load forecasting [55], short-term precipitation forecasting [27], credit ratings forecasting [28], tourism demand forecasting [29] etc., due to its capability to discover complex nonlinear relationships [15] in the observations.

Data clustering is a popular approach for automatically finding classes, concepts, or groups of patterns [30]. Time series data are pervasive across all human endeavors, and their clustering is one of the most fundamental applications of data mining [24]. In literature, many data clustering algorithms [60] have been proposed, but their applications are limited to the extraction of patterns that represent points in multidimensional spaces of fixed dimensionality [61]. In our proposed model, a distance-based clustering algorithm, i.e., the self-organizing feature maps (SOFM) is employed for determining the intervals of the historical time series data set by clustering them into different groups. The SOFM is developed by Kohonen [25], which is a class of neural networks with neurons arranged in a low dimensional (often two-dimensional) structure, and trained by an iterative unsupervised or self-organizing procedure [31, 35, 42]. The SOFM converts the patterns of arbitrary dimensionality into the response of one-dimensional or two-dimensional arrays of neurons, i.e., it converts a wide pattern space into a feature space. The neural network performing such a mapping is called feature map. The training process of the SOFM NN can be found in [49].

To effectively fuzzify the time series data set, it is essential to partition it into different intervals. In this study, the SOFM NN is used to partition the historical data into intervals of different lengths. These intervals are presented in Sect. 6. In this work, MATLAB 7.2 [34] is used for implementing the SOFM NN to determine the intervals of time series data set.

5 Description of data set and its importance of forecasting

The Indian economy is based on agriculture and its agricultural products, and crop yield is heavily dependent on the summer monsoon (JuneSeptember) rainfall. Therefore, any decrease or increase in annual rainfall will always have a severe impact on the agricultural sector in India. About 65 % of the total cultivated land in India are under the influence of rain-fed agriculture system [54]. Therefore, prior knowledge of the monsoon behavior (during which the maximum rainfall occurs in a concentrated period) will help the Indian farmers and the Government to take advantage of the monsoon season. This knowledge can be very useful in reducing the damage of crops during the less rainfall in the monsoon season. Therefore, forecasting the monsoon is a major scientific issue in the field of monsoon meteorology. The ensemble of statistics and mathematics has increased the accuracy of forecasting of the Indian summer monsoon rainfall (ISMR) up to some extent. But the non-linear nature of the ISMR, its forecasting accuracy is still below the satisfactory level.

In 2002, Indian Meteorological Department (IMD) fails to predict the deficit of rainfall during the summer monsoon season, which led to considerable concern in the meteorological community [16]. In 2004, drought has been again observed in the country with a deficit of more than 13 % rainfall [17], which could not be predicted by any statistical or dynamical model.

The forecasting of the ISMR started more than 100 years ago [48]. Mathematical and statistical models require complex computing power [26]. Therefore, many researchers have paid attention to apply ANN in the ISMR forecasting [4, 18]. In literature, several types of neural networks are available, but usually feed-forward and back-propagation neural networks are used in the ISMR forecasting [48].

The north-eastern region of India refers to the easternmost region of India consisting of the neighboring seven states as: Arunachal Pradesh, Assam, Nagaland, Meghalaya, Manipur, Tripura and Mozoram. All these states are shown in Fig. 1.

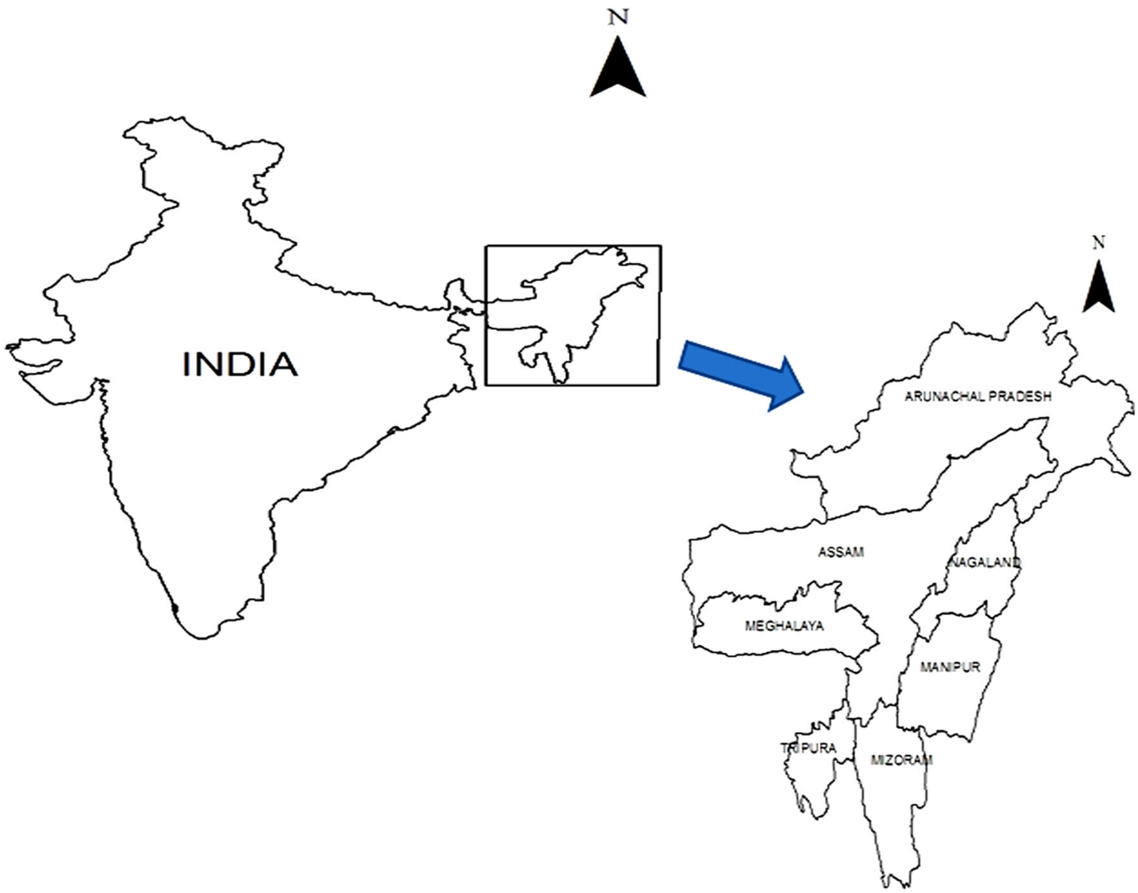

Fig. 1 Map showing the north-eastern region of India

In this region, monsoon rainfall starts in the month of June and ends in the month of September. July and August are mid of the monsoon season. For modeling purpose, time series data set of all these individual months are obtained from Pathasarathy et al. [38] for the period 1901–1990. The seasonal rainfall data set is prepared by taking the mean (June, July, August and September) of all these time series data set. This data set is shown in Table 1. For simulating purpose, this data set is divided into the training set from the period 1901–1960 and the testing set from the period 1961–1990. In this study, the seasonal rainfall forecasting is considered as the main forecasting objective.

6 Proposed Neuro-Fuzzy Hybridized Model

Most of the existing FTS models as discussed earlier use the following six common steps to deal with the forecasting problems:

Step 1 Partition the universe of discourse into intervals.
Step 2 Define linguistic terms for each of the interval.
Step 3 Fuzzify the time series data set.
Step 4 Establish the FLRs based on Definition 4.
Step 5 Construct the FLRGs based on Definition 5.
Step 6 Defuzzify and compute the forecasted values.

In this section, an improved FTS forecasting model is presented, which is based on the hybridization of ANN and FTS concept. The proposed model is referred to as "Neuro-fuzzy hybridized model for time series forecasting". The proposed model is verified by forecasting the seasonal rainfall values in the north-eastern region of India from the period 1901–1990. This model is initially trained with the data set for the period 1901–1960. Each step of the training process is explained next.

Step 1. Define the universe of discourse U of the historical time series data set, and partition it into intervals of different lengths.
[Explanation] Define the universe of discourse U for the time series training data set. Assume that $U = [M_{min}, M_{max}]$, where M_{min} and M_{max} are the minimum and maximum values of the time series training data set. In this data set, the universe of discourse is: $U = [1229.4, 1793.2]$.

Table 1 Mean seasonal rainfall data set (in cm.) for the north-eastern region of India [38]

Year	Rainfall	Year	Rainfall	Year	Rainfall	Year	Rainfall	Year	Rainfall	Year	Rainfall
1901	1361.70	1916	1422.40	1931	1440.40	1946	1412.20	1961	1260.10	1976	1352.70
1902	1525.70	1917	1414.20	1932	1372.30	1947	1626.80	1962	1245.90	1977	1385.00
1903	1406.20	1918	1793.20	1933	1424.70	1948	1462.40	1963	1443.50	1978	1331.00
1904	1329.40	1919	1329.30	1934	1482.00	1949	1478.90	1964	1415.50	1979	1347.10
1905	1551.90	1920	1337.40	1935	1491.90	1950	1499.40	1965	1417.10	1980	1214.30
1906	1505.00	1921	1515.50	1936	1441.60	1951	1489.00	1966	1540.10	1981	1225.10
1907	1338.50	1922	1497.80	1937	1373.40	1952	1474.90	1967	1271.70	1982	1222.90
1908	1354.20	1923	1350.20	1938	1472.70	1953	1536.20	1968	1539.40	1983	1328.50
1909	1436.10	1924	1455.30	1939	1459.00	1954	1406.50	1969	1369.60	1984	1484.10
1910	1530.30	1925	1241.10	1940	1365.50	1955	1438.50	1970	1491.70	1985	1308.80
1911	1415.90	1926	1439.90	1941	1430.10	1956	1605.00	1971	1373.80	1986	1245.00
1912	1308.10	1927	1483.10	1942	1485.10	1957	1262.10	1972	1207.50	1987	1590.90
1913	1451.30	1928	1398.60	1943	1454.50	1958	1259.10	1973	1411.80	1988	1642.80
1914	1286.50	1929	1460.90	1944	1388.40	1959	1229.40	1974	1695.60	1989	1544.40
1915	1457.90	1930	1406.60	1945	1388.20	1960	1530.30	1975	1277.40	1990	1311.80

Now, by applying the SOFM NN, the universe of discourse U is partitioned into intervals of different lengths as: $a_0, a_1, \ldots,$ and a_n. Then, time series values are assigned to their corresponding intervals. Centroid of each interval is recorded by taking the mean of upper-bound and lower-bound of the interval. The resulting intervals and centroids for the data set are shown in Table 2.

Step 2. Define linguistic terms for each of the interval. Assume that the historical time series data set is distributed among n intervals (i.e., $a_0, a_1, \ldots,$ and a_n). Therefore, define n linguistic terms $A_0, A_1, \ldots, A_n$, which can be represented by fuzzy sets, as shown below:

$$\begin{aligned}
A_0 &= 1/a_0 + 0.5/a_1 + 0/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_n, \\
A_1 &= 0.5/a_0 + 1/a_1 + 0.5/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_n, \\
A_2 &= 0/a_0 + 0.5/a_1 + 1/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_n, \\
&\vdots \\
A_n &= 0/a_0 + 0/a_1 + 0/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0.5/a_{n-1} + 1/a_n. \qquad (6)
\end{aligned}$$

Here, the maximum degree of membership of fuzzy set A_i occurs at interval a_i and $0 \leq i \leq n$. **[Explanation]** There are 43 linguistic terms are defined for the rainfall data set as: $A_0, A_1, \ldots, A_{42}$, because total 43 intervals are generated. All these defined linguistic terms are shown as follow:

$$\begin{aligned}
A_0 &= 1/a_0 + 0.5/a_1 + 0/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_{42}, \\
A_1 &= 0.5/a_0 + 1/a_1 + 0.5/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_{42}, \\
A_2 &= 0/a_0 + 0.5/a_1 + 1/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0/a_{n-1} + 0/a_{42}, \\
&\vdots \\
A_{42} &= 0/a_0 + 0/a_1 + 0/a_2 + \ldots + 0/a_{n-2} \\
&\quad + 0.5/a_{n-1} + 1/a_{42}. \qquad (7)
\end{aligned}$$

In Eq. 7, for example, A_0 represents a linguistic term, which denotes a fuzzy set $= \{a_0, a_1, a_2, \ldots, a_{42}\}$. This fuzzy set consists of 43 members with different degree of membership values $= \{1, 0.5, 0, \ldots, 0\}$. Similarly, the linguistic term A_1 denotes the fuzzy set $= \{a_0, a_1, a_2, \ldots, a_{42}\}$, which also consists of 43 members with different degree of membership values $= \{0.5, 1, 0.5, \ldots, 0\}$. The descriptions of remaining linguistic terms, viz., $A_2, A_3, \ldots, A_{42}$, can be provided in a similar manner.

Each interval bears a weight (in %) equal to the number of elements in the interval. For example, if the interval a_6 contains two elements, then its weight will be 2%. The main advantage of assigning the weight in % is that it will help to calculate the variation in rainfall amount in the time series data set.

Table 2 Intervals produced by the SOFM NN for the training data set with their corresponding centroids

Interval	Centroid
$a_0 = [1229.40, 1229.40]$	1229.40
$a_1 = [1241.10, 1241.10]$	1241.10
$a_2 = [1259.10, 1259.10]$	1259.10
$a_3 = [1262.10, 1262.10]$	1262.10
$a_4 = [1286.50, 1286.50]$	1286.50
$a_5 = [1308.10, 1308.10]$	1308.10
$a_6 = [1329.30, 1329.40]$	1329.30
$a_7 = [1337.40, 1338.50]$	1338.00
$a_8 = [1350.20, 1350.20]$	1350.20
$a_9 = [1354.20, 1354.20]$	1354.20
$a_{10} = [1361.70, 1361.70]$	1361.70
$a_{11} = [1365.50, 1365.50]$	1365.50
$a_{12} = [1372.30, 1373.40]$	1372.80
$a_{13} = [1388.20, 1388.40]$	1388.30
$a_{14} = [1398.60, 1398.60]$	1398.60
$a_{15} = [1406.20, 1406.60]$	1406.40
$a_{16} = [1412.20, 1415.90]$	1414.10
$a_{17} = [1422.40, 1422.40]$	1422.40
$a_{18} = [1424.70, 1424.70]$	1424.70
$a_{19} = [1430.10, 1430.10]$	1430.10
$a_{20} = [1436.10, 1436.10]$	1436.10
$a_{21} = [1438.50, 1438.50]$	1438.50
$a_{22} = [1439.90, 1441.60]$	1440.60
$a_{23} = [1451.30, 1451.30]$	1451.30
$a_{24} = [1454.50, 1455.30]$	1454.90
$a_{25} = [1457.90, 1459.00]$	1458.50
$a_{26} = [1460.90, 1462.40]$	1461.70
$a_{27} = [1472.70, 1472.70]$	1472.70
$a_{28} = [1474.90, 1474.90]$	1474.90
$a_{29} = [1478.90, 1478.90]$	1478.90
$a_{30} = [1482.00, 1482.00]$	1482.00
$a_{31} = [1483.10, 1485.10]$	1484.10
$a_{32} = [1489.00, 1491.90]$	1490.50
$a_{33} = [1497.80, 1499.40]$	1498.60
$a_{34} = [1505.00, 1505.00]$	1505.00
$a_{35} = [1515.50, 1515.50]$	1515.50
$a_{36} = [1525.70, 1525.70]$	1525.70
$a_{37} = [1530.30, 1530.30]$	1530.30
$a_{38} = [1536.20, 1536.20]$	1536.20
$a_{39} = [1551.90, 1551.90]$	1551.90
$a_{40} = [1605.00, 1605.00]$	1605.00
$a_{41} = [1626.80, 1626.80]$	1626.80
$a_{42} = [1793.20, 1793.20]$	1793.20

Step 3. Fuzzify the time series data set in accordance with the intervals in Step 1.

[Explanation] To fuzzify the time series data set, the triangular membership rule [39] is applied to each interval in each defined fuzzy set. In this study, it is assumed that the maximum membership value of one year's observation occurs at interval a_i and $0 \leq i \leq n$, and the fuzzified value for that particular year is considered as A_i. For example, the actual rainfall value for year 1958 is 1259.10, where 1259.10 belongs to the interval a_2 with the highest degree of membership value 1, so it is fuzzified to A_2.

In Eq. 7, interval a_0 corresponds to linguistic terms A_0 and A_1 with degree of membership values 1 and 0.5, respectively, and remaining fuzzy sets with degree of membership value 0. For ease of computation, the degree of membership values of fuzzy set $A_j (j = 0, 1, 2, \ldots, 42)$ are considered as either 0, 0.5 or 1, and $0 \leq j \leq 42$. Similarly, interval a_1 corresponds to linguistic terms A_0, A_1 and A_2 with degree of membership values 0.5, 1, and 0.5, respectively, and remaining fuzzy sets with degree of membership value 0. The descriptions of remaining intervals, viz., $a_2, a_3, \ldots, a_{42}$, can be provided in a similar manner. The fuzzified rainfall values for the training data set are shown in Table 3.

Step 4. Establish the FLRs between the fuzzified time series values.

[Explanation] Based on Definition 4, FLRs between two consecutive fuzzified rainfall values are established. For example, in Table 3, fuzzified rainfall values for Years 1901 and 1902 are A_{10} and A_{36}, respectively. So, a FLR between A_{10} and A_{36} can be established as: $A_{10} \rightarrow A_{36}$. In this way, FLRs for the fuzzified rainfall values are obtained, which are presented in Table 4.

Step 5. Establish the FLRGs.

[Explanation] Based on Definition 5, the FLRs can be grouped into a FLRG. For example, in Table 4, there are two FLRs with the same previous state, $A_6 \rightarrow A_{39}$ and $A_6 \rightarrow A_7$. Therefore, these FLRs can be grouped into the FLRG as: $A_6 \rightarrow A_{39}, A_7$. The FLRGs for the remaining FLRs are presented in Table 5.

The FLRGs shown in Table 5 are in the form of $A_i \rightarrow A_{j0}, A_{j1}, \ldots, A_{jp}$; where left hand side and right hand side of the expression represent *previous state* and *current state*, respectively. Now, each fuzzy set associated with the current state of the FLRGs are classified into different trends as per the conditions provided in Factor 3. For example, consider the "FLRG 16: $A_{15} \rightarrow A_6, A_{21}, A_{22}$". Here, fuzzified rainfall values A_6, A_{21} and A_{22} of the current state can be classified into the following trends as:

Table 3 Fuzzified rainfall values for the training data set with their corresponding centroids and weights

Year	Rainfall	Fuzzified Rainfall	Centroid	Weight (in %)	Year	Rainfall	Fuzzified Rainfall	Centroid	Weight (in %)
1901	1361.70	A_{10}	1361.70	1	1931	1440.40	A_{22}	1440.60	3
1902	1525.70	A_{36}	1525.70	1	1932	1372.30	A_{12}	1372.80	2
1903	1406.20	A_{15}	1406.40	3	1933	1424.70	A_{18}	1424.70	1
1904	1329.40	A_6	1329.30	2	1934	1482.00	A_{30}	1482.00	1
1905	1551.90	A_{39}	1551.90	1	1935	1491.90	A_{32}	1490.50	2
1906	1505.00	A_{34}	1505.00	1	1936	1441.60	A_{22}	1440.60	3
1907	1338.50	A_7	1338.00	2	1937	1373.40	A_{12}	1372.80	2
1908	1354.20	A_9	1354.20	1	1938	1472.70	A_{27}	1472.70	1
1909	1436.10	A_{20}	1436.10	1	1939	1459.00	A_{25}	1458.50	2
1910	1530.30	A_{37}	1530.30	2	1940	1365.50	A_{11}	1365.50	1
1911	1415.90	A_{16}	1414.10	3	1941	1430.10	A_{19}	1430.10	1
1912	1308.10	A_5	1308.10	1	1942	1485.10	A_{31}	1484.10	2
1913	1451.30	A_{23}	1451.30	1	1943	1454.50	A_{24}	1454.90	2
1914	1286.50	A_4	1286.50	1	1944	1388.40	A_{13}	1388.30	2
1915	1457.90	A_{25}	1458.50	2	1945	1388.20	A_{13}	1388.30	2
1916	1422.40	A_{17}	1422.40	1	1946	1412.20	A_{16}	1414.10	3
1917	1414.20	A_{16}	1415.10	3	1947	1626.80	A_{41}	1626.80	1
1918	1793.20	A_{42}	1793.20	1	1948	1462.40	A_{26}	1461.70	2
1919	1329.30	A_6	1329.30	2	1949	1478.90	A_{29}	1478.90	1
1920	1337.40	A_7	1338.00	2	1950	1499.40	A_{33}	1498.60	2
1921	1515.50	A_{35}	1515.50	1	1951	1489.00	A_{32}	1490.50	2
1922	1497.80	A_{33}	1498.60	2	1952	1474.90	A_{28}	1474.90	1
1923	1350.20	A_8	1350.20	1	1953	1536.20	A_{38}	1536.20	1
1924	1455.30	A_{24}	1454.90	2	1954	1406.50	A_{15}	1406.40	3
1925	1241.10	A_1	1241.10	1	1955	1438.50	A_{21}	1438.50	1
1926	1439.90	A_{22}	1440.60	3	1956	1605.00	A_{40}	1605.00	1
1927	1483.10	A_{31}	1484.10	2	1957	1262.10	A_3	1262.10	1
1928	1398.60	A_{14}	1398.60	1	1958	1259.10	A_2	1259.10	1
1929	1460.90	A_{26}	1461.70	2	1959	1229.40	A_0	1229.40	1
1930	1406.60	A_{15}	1406.40	3	1960	1530.30	A_{37}	1530.30	2

$A_{15} \rightarrow A_6(\downarrow), A_{21}(\uparrow), A_{22}(\uparrow).$

In this way, all the trends of fuzzified rainfall values for the training data set are recorded for the future consideration, which are also shown in Table 5. If the same fuzzy set appears more than once in the current state of the FLRG, it is included only once in the FLRG.

Step 6. Defuzzify and compute the forecasted values from the fuzzified time series data set. To defuzzify the fuzzified time series data set and to obtain the forecasted values, a "Frequency-Based Defuzzification Technique" is proposed. This defuzzification technique employs the FLRGs obtained in **Step 5** to get the forecasted values. The proposed defuzzification technique is designed in such a way that it can deal with the trend exhibited by different fuzzy sets of the corresponding FLRGs. Based on the application of technique, it is categorized as: **Principle 1** and **Principle 2**. The **Principle 1** is given as follows:

- **Principle 1**: For forecasting year $Y(t)$, the fuzzified value for year $Y(t-1)$ is required, where "t" is the current year which we want to forecast. The **Principle 1** is applicable only if there are more than one fuzzified values available in the current state. The steps under **Principle 1** are explained next.

  Step a. Obtain the fuzzified rainfall value for year $Y(t-1)$ as A_i $(i = 0, 1, 2, \ldots, n)$.

  Step b. Obtain the FLRG whose previous state is A_i, and the current state is $A_{j0}, A_{j1}, \ldots, A_{jp}$, i.e., the FLRG is in the form of "$A_i \rightarrow A_{j0}, A_{j1}, \ldots,$

Table 4 A sample of FLRs for the rainfall data set

FLRs
$A_{10} \rightarrow A_{36}$
$A_{36} \rightarrow A_{15}$
$A_{15} \rightarrow A_{6}$
$A_{6} \rightarrow A_{39}$
$A_{39} \rightarrow A_{34}$
$A_{34} \rightarrow A_{7}$
$A_{7} \rightarrow A_{9}$
$A_{9} \rightarrow A_{20}$
$A_{20} \rightarrow A_{37}$
$A_{37} \rightarrow A_{16}$
$A_{16} \rightarrow A_{5}$
$A_{5} \rightarrow A_{23}$
$A_{23} \rightarrow A_{4}$
$A_{4} \rightarrow A_{25}$
$A_{25} \rightarrow A_{17}$
$A_{17} \rightarrow A_{16}$
$A_{16} \rightarrow A_{42}$
$A_{42} \rightarrow A_{6}$
$A_{6} \rightarrow A_{7}$
...
$A_{2} \rightarrow A_{0}$
$A_{0} \rightarrow A_{37}$

A_{jp}". The fuzzy sets associated with the current state of the FLRG may exhibit miscellaneous trends (i.e., $\uparrow$, $\downarrow$ and $=$), because it contains more than one fuzzified values in the current state.

Step c. Find the interval where the maximum membership value of the fuzzy set A_i (previous state) occurs, and let this interval be a_i $(i = 0, 1, 2, \ldots, n)$.

Step d. Compute the variation in rainfall for the fuzzy set A_i as:

$$V_r = \left[\frac{C_i \times W_i}{100}\right] \quad (8)$$

Here, C_i and W_i represent the corresponding centroid of the interval and weight (in %) for the fuzzy set A_i, respectively, where $i = 0, 1, 2, \ldots, n$.

Step e. Find the intervals where the maximum membership values of the fuzzy sets $A_{j0}, A_{j1}, \ldots, A_{jp}$ (current state) occur. Let these intervals be $a_{j0}, a_{j1}, \ldots, a_{j}p$. All these intervals have the corresponding centroids $C_{j0}, C_{j1}, \ldots, C_{jp}$.

Table 5 A list of generated FLRGs

FLRG	Previous state $(t-1) \rightarrow$ Current state (t)
FLRG 1:	$A_0 \rightarrow A_{37}(\uparrow)$
FLRG 2:	$A_1 \rightarrow A_{22}(\uparrow)$
FLRG 3:	$A_2 \rightarrow A_0(\downarrow)$
FLRG 4:	$A_3 \rightarrow A_2(\downarrow)$
FLRG 5:	$A_4 \rightarrow A_{25}(\uparrow)$
FLRG 6:	$A_5 \rightarrow A_{23}(\uparrow)$
FLRG 7:	$A_6 \rightarrow A_7(\uparrow), A_{39}(\uparrow)$
FLRG 8:	$A_7 \rightarrow A_9(\uparrow), A_{35}(\uparrow)$
FLRG 9:	$A_8 \rightarrow A_{24}(\uparrow)$
FLRG 10:	$A_9 \rightarrow A_{20}(\uparrow)$
FLRG 11:	$A_{10} \rightarrow A_{36}(\uparrow)$
FLRG 12:	$A_{11} \rightarrow A_{19}(\uparrow)$
FLRG 13:	$A_{12} \rightarrow A_{18}(\uparrow), A_{27}(\uparrow)$
FLRG 14:	$A_{13} \rightarrow A_{13}(=), A_{16}(\uparrow)$
FLRG 15:	$A_{14} \rightarrow A_{26}(\uparrow)$
FLRG 16:	$A_{15} \rightarrow A_6(\downarrow), A_{21}(\uparrow), A_{22}(\uparrow)$
FLRG 17:	$A_{16} \rightarrow A_5(\downarrow), A_{41}(\uparrow), A_{42}(\uparrow)$
FLRG 18:	$A_{17} \rightarrow A_{16}(\downarrow)$
FLRG 19:	$A_{18} \rightarrow A_{30}(\uparrow)$
FLRG 20:	$A_{19} \rightarrow A_{31}(\uparrow)$
FLRG 21:	$A_{20} \rightarrow A_{37}(\uparrow)$
FLRG 22:	$A_{21} \rightarrow A_{40}(\uparrow)$
FLRG 23:	$A_{22} \rightarrow A_{12}(\downarrow), A_{31}(\uparrow)$
FLRG 24:	$A_{23} \rightarrow A_4(\downarrow)$
FLRG 25:	$A_{24} \rightarrow A_1(\downarrow), A_{13}(\downarrow)$
FLRG 26:	$A_{25} \rightarrow A_{11}(\downarrow), A_{17}(\downarrow)$
FLRG 27:	$A_{26} \rightarrow A_{15}(\downarrow), A_{29}(\uparrow)$
FLRG 28:	$A_{27} \rightarrow A_{25}(\downarrow)$
FLRG 29:	$A_{28} \rightarrow A_{38}(\uparrow)$
FLRG 30:	$A_{29} \rightarrow A_{33}(\uparrow)$
FLRG 31:	$A_{30} \rightarrow A_{32}(\uparrow)$
FLRG 32:	$A_{31} \rightarrow A_{14}(\downarrow), A_{24}(\downarrow)$
FLRG 33:	$A_{32} \rightarrow A_{22}(\downarrow), A_{28}(\downarrow)$
FLRG 34:	$A_{33} \rightarrow A_8(\downarrow), A_{32}(\downarrow)$
FLRG 35:	$A_{34} \rightarrow A_7(\downarrow)$
FLRG 36:	$A_{35} \rightarrow A_{33}(\downarrow)$
FLRG 37:	$A_{36} \rightarrow A_{15}(\downarrow)$
FLRG 38:	$A_{37} \rightarrow A_{16}(\downarrow)$
FLRG 39:	$A_{38} \rightarrow A_{15}(\downarrow)$
FLRG 40:	$A_{39} \rightarrow A_{34}(\downarrow)$
FLRG 41:	$A_{40} \rightarrow A_3(\downarrow)$
FLRG 42:	$A_{41} \rightarrow A_{26}(\downarrow)$
FLRG 43:	$A_{42} \rightarrow A_6(\downarrow)$

Step f. Calculate the average of the centroids $C_{j0}, C_{j1}, \ldots, C_{jp}$ as:

$$A_{centroid} = \left[\frac{C_{j0} + C_{j1} + \ldots + C_{jp}}{p}\right] \quad (9)$$

Here, $p$represents the total number of fuzzy sets associated with the current state of the FLRG.

Step g. Compute the trend value as:

$$T_{value} = \left[\frac{(C_{j0} * V_r) + (C_{j1} * V_r) + \ldots + (C_{jp} * V_r)}{p}\right] \quad (10)$$

In Eq. 10, if any fuzzified rainfall value in the current state exhibits *downward trend* (↓), then the symbol "∗" represents "addition" operation; otherwise it represents "subtraction" operation for *upward trend* (↑). In the considered FLRG "$A_i \rightarrow A_{j0}, A_{j1}, \ldots, A_{jp}$", if any fuzzified rainfall value in the current state exhibits unchanged trend with respect to the previous state fuzzified rainfall value (i.e., $i = j$), then find the interval where the maximum membership value of the fuzzy set A_j occurs. Let this interval be a_j, which has the centroid C_j. This centroid C_j can directly be added to the numerator of left hand side of Eq. 10, without addition or subtraction of V_r value from C_j.

Step h. Compute the forecasted rainfall value as:

$$F_{forecast} = \left[\frac{A_{centroid} + T_{value}}{2}\right] \quad (11)$$

– **Principle 2**: This principle is applicable only if there is only one fuzzified rainfall value in the current state. The steps under **Principle 2** are given as follows:

Step a. Obtain the fuzzified rainfall value for year $Y(t-1)$ as A_i $(i = 0, 1, \ldots, n)$.

Step b. Find the FLRG whose previous state is A_i, and the current state is A_j, i.e., the FLRG is in the form of "$A_i \rightarrow A_j$". The fuzzy set associated with the current state of the FLRG set may exhibit either upward (↑), downward (↓) or unchanged (=) trend, because it contains only one fuzzified value in the current state.

Step c. Find the interval where the maximum membership value of the fuzzy set A_i occurs. Let these interval be a_i $(i = 0, 1, 2, \ldots, n)$.

Step d. Compute the variation in rainfall for the fuzzy set A_i as:

$$V_r = \left[\frac{C_i \times W_i}{100}\right] \quad (12)$$

Here, C_i and $W_i$represent the corresponding centroid of the interval and weight (in %) for the fuzzy set A_i, respectively, where $i = 0, 1, 2, \ldots, n$.

Step e. Compute the trend value as:

$$T_{value} = \left[A_j * V_r\right] \quad (13)$$

In Eq. 13, if the fuzzified rainfall value in the current state exhibits *downward trend* (↓) with respect to the previous state fuzzified rainfall value, then the symbol "∗" represents "addition" operation; otherwise it represents "subtraction" operation for *upward trend* (↑). In the considered FLRG "$A_i \rightarrow A_j$", if the fuzzified rainfall value in the current state exhibits unchanged trend with respect to the previous state fuzzified rainfall value (i.e., $i = j$), then find the interval where the maximum membership value of the fuzzy set A_j occurs. Let this interval be a_j, which has the centroid C_j. This centroid C_j can directly be added to the left hand side of Eq. 13, without addition or subtraction of V_r value from C_j.

Step f. Find the interval where the maximum membership value of the fuzzy set A_j (current state) occurs. Let this interval be C_j.

Step g. Compute the forecasted rainfall value as:

$$F_{forecast} = \left[\frac{C_j + T_{trend}}{2}\right] \quad (14)$$

Based on the proposed model, here two examples have been presented to compute the forecasted values of the summer monsoon rainfall.

- **Example 1:** For forecasting the rainfall amount for the $Y(1927)$, the fuzzified rainfall value for the previous year $Y(1926)$ is required. The fuzzified rainfall value for the year $Y(1926)$ is obtained from Table 3, which is A_{22}. Obtain the FLRG whose previous state is A_{22}. In Table 5, this FLRG is in the form of $A_{22} \rightarrow A_{12}, A_{31}$ (i.e., FLRG 23). Here, **Principle 1** is applicable, because there are more than one fuzzified rainfall values (A_{12} and A_{31}) available in the current state.
  In this FLRG, the fuzzified rainfall values in the current state exhibit two different trends as: $A_{22} \rightarrow A_{12}(\downarrow), A_{31}(\uparrow)$. Now, obtain the interval where the maximum membership value of the fuzzy set A_{22} occurs from Table 2, which is a_{22}. This interval a_{22} has the corresponding centroid 1440.60 ($= C_{22}$). Here, the fuzzy set A_{22} has the weight 3% ($= W_{22}$). Based on Eq. 8, compute the variation in rainfall for the fuzzy set A_{22} as:
  $$V_r = \left[\frac{1440.60 \times 3}{100}\right] = 43.218$$
  Now, find the intervals where the maximum membership values of the fuzzy sets A_{12} and A_{31} occur from Table 2, which are a_{12} and a_{31}, respectively. All these intervals a_{12} and a_{31} have the corresponding centroids 1372.80 ($= C_{12}$) and 1484.10 ($= C_{31}$), respectively. Based on Eq. 9, compute the average of the centroids as:
  $$A_{centroid} = \left[\frac{1372.80 + 1484.10}{2}\right] = 1428.45$$
  Now, Compute the trend value based on Eq. 10 as:
  $$T_{value} = \left[\frac{1372.80 + 43.218}{2}\right] + \left[\frac{1484.10 - 43.218}{2}\right]$$
  $$= 1428.45$$
  Here, "addition (+)" operation is done for downward trend, whereas "subtraction (–)" operation is done for upward trend. Now, calculate the forecasted rainfall value for $Y(1927)$ based on Eq. 11 as:
  $$F_{forecast} = \left[\frac{1428.45 + 1428.45}{2}\right] = 1428.45$$
- **Example 2:** For forecasting the rainfall amount for the $Y(1915)$, the fuzzified rainfall value for the previous year $Y(1914)$ is required. The fuzzified rainfall value for the year $Y(1914)$ is obtained from Table 3, which is A_4. Obtain the FLRG whose previous state is A_4. In Table 5, this FLRG is in the form of $A_4 \rightarrow A_{25}$, (i.e., FLRG 5). Here, **Principle 2** is applicable, because there is only one fuzzified rainfall value (A_{25}) available in the current state.
  In this FLRG, the fuzzified rainfall value in the current state exhibits only a single trend as: $A_4 \rightarrow A_{25}(\uparrow)$. Now, obtain the interval where the maximum membership value of the fuzzy set A_4 occurs from Table 2, which is a_4. This interval a_4 has the corresponding centroid 1286.50 ($= C_4$). Here, the fuzzy set A_4 has the weight 1% ($= W_4$). Based on Eq. 12, compute the variation in rainfall for the fuzzy set A_4 as:
  $$V_r = \left[\frac{1286.50 \times 1}{100}\right] = 12.865$$
  Now, find the interval where the maximum membership value of the fuzzy set A_{25} occurs from Table 1, which is a_{25}. This interval a_{25} has the corresponding centroid 1458.50 ($= C_{25}$). Now, Compute the trend value based on Eq. 13 as:
  $$T_{value} = [1458.50 - 12.865] = 1445.635$$
  Here, "subtraction (–)" operation is done for upward trend. Now, calculate the forecasted rainfall value for the year $Y(1915)$ based on (14) as:
  $$F_{forecast} = \left[\frac{1458.50 + 1445.635}{2}\right] = 1452.07$$

7 Performance analysis parameters

The performance of the model is evaluated with the help of root mean square error (RMSE), average forecasting error rate (AFER), evaluation parameter (δ_r), correlation coefficient (CC), coefficient of determination (CC^2), and tracking signal (TS). All these parameters are defined as follows:

1. The RMSE can be defined as:
$$RMSE = \sqrt{\frac{\sum_{i=1}^{k}(Forecast_i - Actual_i)^2}{k}} \tag{15}$$
2. The AFER can be defined as:
$$AFER = \frac{\sum_{i=1}^{k}(Forecast_i - Actual_i)/Actual_i}{k} \times 100 \tag{16}$$
3. The δ_r can be defined [5] as:
$$\delta_r = \frac{|Forecast_i - Actual_i|}{\sigma} \tag{17}$$
4. The CC can be defined as:

$$CC = \frac{k\sum(Actual_i)(Forecast_i) - (\sum Actual_i)(\sum Forecast_i)}{\sqrt{k(\sum Actual_i^2) - (\sum Actual_i)^2}\sqrt{k(\sum Forecast_i^2) - (\sum Forecast_i)^2}} \quad (18)$$

5. The steps involve in determining the TS can be explained as [56]:
   (a) Compute the mean absolute deviation (M_{ad}):
   $$M_{ad} = \frac{\sum_{i=1}^{k} |(Forecast_i - Actual_i)|}{k} \quad (19)$$
   (b) Compute the running sum of forecast errors (R_{sfe}):
   $$R_{sfe} = \sum_{i=1}^{k}(Forecast_i - Actual_i) \quad (20)$$
   (c) Compute the TS:
   $$TS = \frac{R_{sfe}}{M_{ad}} \quad (21)$$

Here, each $Forecast_i$ and $Actual_i$ is the forecasted and actual value of year i, respectively, and k is the total number of years to be forecasted. In Eq. 17, σ is the standard deviation of the training as well as testing data sets. A value of δ_r less than 1 indicates good forecasting. In Eq. 18, the value of CC is such that $-1 < CC < +1$. The $+$ and $-$ indicate the positive linear correlations and negative linear correlations between the forecasted and actual value of rainfall, respectively. A correlation coefficient (CC) greater than 0.8 is generally considered as strong, whereas CC less than 0.5 is considered as weak. The CC^2 lies between $0 < CC^2 < 1$, and indicates the strength of the linear association between A_i and F_i. A CC^2 value closer to 1 represents strong linear relationship between A_i and F_i.

In Eq. 19, a $M_{ad} > 0$ indicates that forecasting model tends to under-forecast, whereas $M_{ad} < 0$ indicates that forecasting model tends to over-forecast. In Eq. 20, the R_{sfe} indicates the bias in the forecasting. A value of R_{sfe} equal to "0" indicates that positive errors are equal to the negative errors. Based on Eq. 21, TS value is determined which is used to measure how well time series forecasting model is performed. A TS value between -4 and $+4$ indicates that the model is working correctly.

Table 6 Experimental results for forecasting the ISMR based on training and testing data sets

Evaluation criterion	Experimental result (Training data set)	Experimental result (Testing data set)
RMSE	59.42	67.80
AFER	2.41%	3.49%
δ_r	0.4	0.4
CC	0.85	0.86
CC^2	0.72	0.74
M_{ad}	34.55	48.64
R_{sfe}	12.56	5.42
TS	0.36	0.11

Table 7 Additional parameters and their values during training and testing processes of the SOFM NN

Serial number	Additional parameter	Input value
1	Initial weight	0.3
2	Learning rate	0.5
3	Epochs	1000
4	Learning radius	3
5	Activation function	Sigmoid

8 Empirical analysis

The proposed model is trained and tested using the data set mentioned in Sect. 5. For the optimal result, 43 and 20 numbers of intervals for the training and testing data sets are selected, respectively. The results for the training and testing data sets are presented in Table 7, in terms of *RMSE*, *AFER*, average of δ_r, CC, CC^2 and TS. The computed value of δ_r is much less than 1 as shown in Table 6. The CC between actual and forecasted rainfall is close to 1, which indicates the efficiency of the model. The CC^2 values for the training and testing data sets exhibit the strong linear association between actual and forecasted rainfall. TS values are plotted for for the forecasted rainfall, which are depicted in Figs. 2 and 3. From these figures, it is observed that TS values lie between the ranges ± 4, which indicate that the model is working correctly.

During simulation of the SOFM NN, a number of experiments were carried out to set additional parameters [51], viz. initial weight, learning rate, epochs, learning radius and activation function to obtain the optimal results, and the best values have been chosen that exhibit the best behavior in terms of accuracy. The determined optimal values of all these parameters are given in Table 7.

To assess the forecasting performance of the proposed model, it is compared with the existing FTS models [9, 10, 62]. Furthermore, the performance of the proposed model is also compared with various statistical time series models, such as linear regression, quadratic regression, compound regression, and exponential regression [37]. The forecasted results are listed in Table 8. The comparative analyzes show that the proposed model is more precise than existing FTS and linear statistical models.

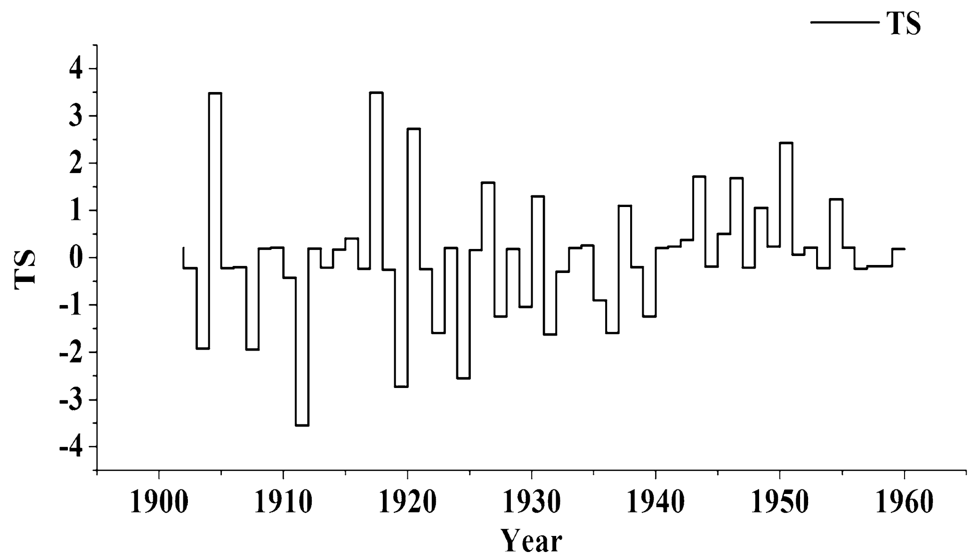

Fig. 2 TS curves for the training data set

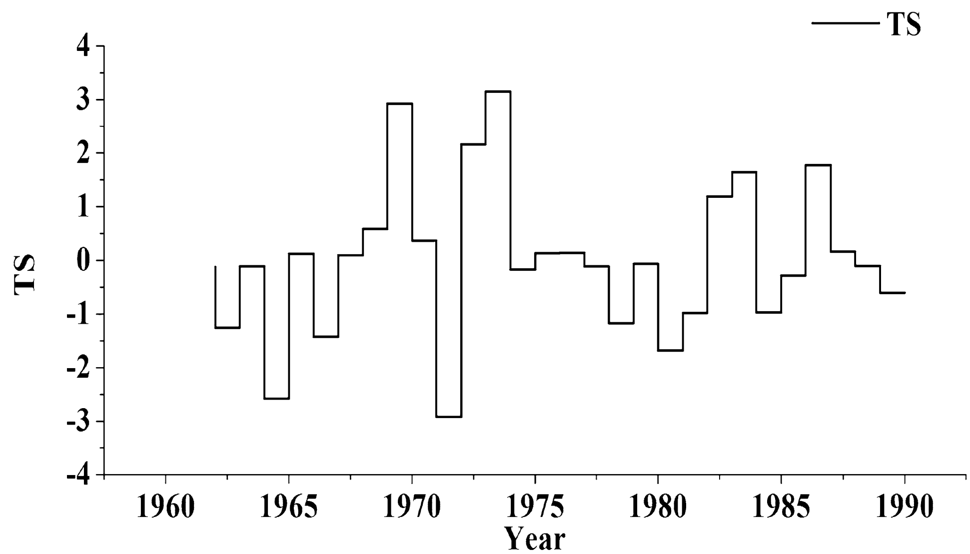

Fig. 3 TS curves for the testing data set

Table 8 Comparison of AFERs for training and testing data sets based the existing FTS models and linear models for the ISMR data set

Model	AFER (Training data set) (%)	AFER (Testing data set) (%)
Chen [9] model	3.80	4.99
Chen [10] model	3.70	4.98
Yu [62] model	3.62	4.96
Linear regression	4.91	7.82
Quadratic regression	4.91	7.82
Compound regression	4.91	7.82
Exponential regression	4.91	7.82
Proposed model	2.41	3.49

9 Stock index forecasting – a case study

Accurate forecasting of stock prices is a challenging task for brokers and investors. Various factors such as social, economic and political influence on the stock market and control the variation of the stock prices. So, investing in the stock market is one of the most exciting and challenging tasks, because losing and profit depends on a good decision of an investor. The gain and lose explain the non-stationary nature of the stock market, which is extremely difficult to predict accurately in advance without an expert system.

In this section, the proposed method is applied for forecasting the daily index of the National Stock Exchange (NSE) (from 1/08/2012 – 28/09/2012) and Taiwan Stock Index Futures (TAIFEX) (from 3/8/1998 – 30/9/1998).

Table 9 A comparison of the AFER values for different methods for forecasting the NSE and TAIFEX data sets

Models	AFER (NSE) (%)	AFER (TAIFEX) (%)
Chen Model [9]	2.89	2.86
Yu Model [62]	2.79	2.67
Chen High-Order Model (3rd-Order) [10]	2.25	2.36
BPNN Model [48]	1.99	1.93
Fuzzy-Neuro Model [47]	1.87	1.76
Proposed Model	1.50	1.65

Table 10 Experimental results for the NSE and TAIFEX data sets

Statistical parameters	Experimental result (NSE)	Experimental result (TAIFEX)
RMSE	249.61	245.22
δ_r	0.25	0.24
CC	0.96	0.97
CC^2	0.92	0.94
M_{ad}	36.24	35.98
R_{sfe}	−4.84	−4.25
TS	−0.13	−0.12

Table 9 shows a comparison of the AFER values for different models for forecasting the stock index data sets of NSE and TAIFEX. The AFER value of the proposed model shown in Table 9 is smaller than the ones of the existing models [9, 10, 47, 48, 62]. In other words, the proposed model outperforms the conventional models presented in articles [9, 10, 47, 48, 62] for forecasting the stock index. This smaller value of AFER indicates that there are very small differences between the actual and forecasted values.

The consistency of the proposed model has been analyzed on the basis of statistical parameters as discussed in Sect. 7. Experimental results are listed in Table 10. All these statistical analyzes signify the robustness of the proposed model, because it takes the decision by considering and resolving major problems of FTS modeling approach, as discussed in Sect. 2.

10 Conclusions and the way ahead

This article presents a novel approach by combining ANN and FTS for building a time series forecasting expert system. The main contributions of this article are presented as follows:

- First, the author shows that the forecasted accuracy of the FTS model can be enhanced by resolving the problems associated with different factors such as determination of lengths of intervals, handling of repeated fuzzy sets, handling of trends associated with fuzzy sets, and defuzzification operation. In this work, the author suggested the use of SOFM NN for determination of effective lengths of intervals of the historical time series data sets. Further, the author also proposed techniques for the assignment of weights for the repeated fuzzy sets based on their frequency of occurrences, determination of trends for the various fuzzy sets, and defuzzification operation.
- Second, the author shows that the seasonal rainfall forecasting problem can be solved effectively by the proposed "Neuro-fuzzy hybridized model".
- Third, the author demonstrates that the proposed model is superior to the existing FTS models [9, 10, 62] for the prediction of rainfall.
- Fourth, the author validates the proposed model using various statistical parameters, and shows that the presented model has the ability to predict the non-linear behavior of the summer monsoon rainfall much more accurately compared to various statistical linear models.
- Fifth, the author shows the applicability of the proposed model for forecasting the stock index data sets. Experimental results show that the proposed model can effectively be utilized as an expert decision making system in stock as well as financial time series forecasting.

A significant drawback of the FTS forecasting model is that increase in the number of intervals of time series data set increases the accuracy rate of forecasting, but decreases the fuzziness of time series data set [49]. Therefore, in this study, results are obtained with minimum number of intervals.

The proposed model has the limitation that it can apply only in one-factor time series data set. Therefore, the author has tried to make the proposed model more generalize, so that it can applicable in different kinds of one-factor time series data set, and can be employed in various domains flexibly. Work is underway to apply the proposed model on other domains in the following way:

1. Apply the proposed model on different regions of seasonal rainfall data set (one-factor), and check its accuracy and performance with different size of intervals, and
2. To test the performance of the model for different types of financial, stocks, insurance and marketing data set (one-factor).

Hence, this study implies that the approaches that have been adopted in the proposed model can be applied to improve the accuracy and performance of FTS forecasting model.

References

1. Aladag CH, Yolcu U, Egrioglu E (2010) A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks. Math Comput Simul 81(4):875–882
2. Ashfaq RAR, Wang XZ, Huang JZX, Abbas H, He YL (2016) Fuzziness based semi-supervised learning approach for Intrusion Detection System (IDS). In Sci. doi:10.1016/j.ins.2016.04.019
3. Askari S, Montazerin N (2015) A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering. Expert Syst Appl 42(4):2121–2135
4. Chakraverty S, Gupta P (2007) Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Comput Appl 17:187–192
5. Chakraverty S, Gupta P (2008) Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Comput Appl 17:187–192
6. Chen MY (2014) A high-order fuzzy time series forecasting model for internet stock trading. Future Gener Comput Syst 37:461–467
7. Chen MY, Chen BT (2014) Online fuzzy time series analysis based on entropy discretization and a fast fourier transform. Appl Soft Comput 14:156–166
8. Chen MY, Chen BT (2015) A hybrid fuzzy time series model based on granular computing for stock price forecasting. Inf Sci 294:227–241
9. Chen SM (1996) Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst 81:311–319
10. Chen SM (2002) Forecasting enrollments based on high-order fuzzy time series. Cybern Sys 33(1):1–16
11. Chen SM, Hwang JR (2000) Temperature prediction using fuzzy time series. IEEE Trans Syst Man Cybern 30:263–275
12. Chen SM, Kao PY (2013) TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines. Inf Sci 247:62–71
13. Cheng C, Chang J, Yeh C (2006) Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost. Technol Forecast Social Change 73:524–542
14. Cheng CH, Chen TL, Wei LY (2010) A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Inf Sci 180(9):1610–1629
15. Czibula G, Czibula IG, Găceanu RD (2013) Intelligent data structures selection using neural networks. Knowl Inf Syst 34:171–192
16. Gadgil S, Srinivasan J, Nanjundiah RS, Kumar KK, Munot AA, Kumar KR (2002) On forecasting the Indian summer monsoon: the intriguing season of 2002. Curr Sci 83(4):394–403
17. Gadgil S, Rajeevan M, Nanjundiah R (2005) Monsoon prediction—Why yet another failure? Curr Sci 88(9):1389–1400
18. Guhathakurta P, Rajeevan M, Thapliyal V (1999) Long range forecasting Indian summer monsoon rainfall by a hybrid principal component neural network model. Meteorol Atmos Phys 71:255–266
19. Hadavandi E, Shavandi H, Ghanbari A (2010) Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. Knowl Based Syst 23(8):800–808
20. He YL, Wang XZ, Huang JZ (2016) Fuzzy nonlinear regression analysis using a random weight network. Inf Sci. doi:10.1016/j.ins.2016.01.037
21. Huarng K (2001) Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets Syst 123:387–394
22. Huarng K, Yu THK (2006) The application of neural networks to forecast fuzzy time series. Phys A Stat Mech Appl 363(2):481–491
23. Kathirvalavakumar T, Ponmalar E (2013) Self organizing map and wavelet based image compression. Int J Mach Learn Cybern 4(4):319–326
24. Keogh E, Lin J (2005) Clustering of time-series subsequences is meaningless: implications for previous and future research. Knowl Inf Syst 8(2):154–177
25. Kohonen T (1990) The self organizing maps. In: Proceedings of the IEEE, vol 78, issue 9. IEEE, pp 1464–1480
26. Krishna KK, Soman MK, Kumar KR (1995) Seasonal forecasting of Indian summer monsoon rainfall: a review. Weather 50:449–467
27. Kuligowski RJ, Barros AP (1998) Experiments in short-term precipitation forecasting using artificial neural networks. Mon Weather Rev 126:470–482
28. Kumar K, Bhattacharya S (2006) Artificial neural network vs. linear discriminant analysis in credit ratings forecast: a comparative study of prediction performances. Rev Acc Finance 5(3):216–227
29. Law R (2000) Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. Tour Manag 21(4):331–340
30. Lee J, Lee YJ (2013) An effective dissimilarity measure for clustering of high-dimensional categorical data. Knowl Inf Syst pp 1–15. doi: 10.1007/s10115-012-0599-1
31. Liao TW (2005) Clustering of time series data—a survey. Pattern Recogn 38(11):1857–1874
32. Lu SX, Wang XZ, Zhang GQ, Zhou X (2015a) Effective algorithms of the moore-penrose inverse matrices for extreme learning machine. Intell Data Anal 19(4):743–760
33. Lu W, Chen X, Pedrycz W, Liu X, Yang J (2015b) Using interval information granules to improve forecasting in fuzzy time series. Int J Approx Reasoning 57:1–18
34. MATLAB (2006) Version 7.2 (R2006). http://www.mathworks.com/. Accessed 1 Nov 2015
35. Meschino GJ, Comas DS, Ballarin VL, Scandurra AG, Passoni LI (2015) Automatic design of interpretable fuzzy predicate systems for clustering using self-organizing maps. Neurocomputing 147:47–59
36. Pal SK, Mitra P (2004) Case generation using rough sets with fuzzy representation. IEEE Trans Knowl Data Eng 16(3):292–300
37. PASW (2012) PASW Statistics 18. http://www.spss.com.hk/statistics/. Accessed 1 Dec 2015
38. Pathasarathy B, Munot AA, Kothawale DR (1994) All India monthly and seasonal rainfall series: 1871–1993. Theor Appl Climatol 49:217–224
39. Ross TJ (2007) Fuzzy logic with engineering applications. John Wiley and Sons, Singapore
40. Sarlin P (2012) Visual tracking of the millennium development goals with a fuzzified self-organizing neural network. Int J Mach Learn Cybernet 3(3):233–245
41. Sarlin P (2014a) On biologically inspired predictions of the global financial crisis. Neural Comput Appl 24(3–4):663–673
42. Sarlin P (2014b) A weighted SOM for classifying data with instance-varying importance. Int J Mach Learn Cybernet 5(1):101–110
43. Singh P (2015a) A brief review of modeling approaches based on fuzzy time series. Int J Mach Learn Cybern, pp 1–24. doi: 10.1007/s13042-015-0332-y
44. Singh P (2015b) Computational intelligence for big data analysis, vol 19, Springer-Verlag/ Heidelberg, chap Big Data Time Series

Forecasting Model: A Fuzzy-Neuro Hybridize Approach, pp 55–71
45. Singh P (2016) High-order fuzzy-neuro-entropy integration-based expert system for time series forecasting. Neural Comput Appl, pp 1–18. doi: 10.1007/s00521-016-2261-4
46. Singh P, Borah B (2013a) An efficient time series forecasting model based on fuzzy time series. Eng Appl Artif Intell 26:2443–2457
47. Singh P, Borah B (2013b) High-order fuzzy-neuro expert system for time series forecasting. Knowl Based Syst 46:12–21
48. Singh P, Borah B (2013c) Indian summer monsoon rainfall prediction using artificial neural network. Stoch Env Res Risk Assess 27(7):1585–1599
49. Singh P, Borah B (2014a) An effective neural network and fuzzy time series-based hybridized model to handle forecasting problems of two factors. Knowl Inf Syst 38(3):669–690
50. Singh P, Borah B (2014b) Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization. Int J Approx Reason 55:812–833
51. Sivanandam SN, Deepa SN (2007) Principles of soft computing. Wiley India (P) Ltd., New Delhi
52. Song Q, Chissom BS (1993) Forecasting enrollments with fuzzy time series—part I. Fuzzy Sets Syst 54(1):1–9
53. Sun B, Guo H, Karimi HR, Ge Y, Xiong S (2015) Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. Neurocomputing 151:1528–1536
54. Swaminathan MS (1998) Padma Bhusan Prof. P. Koteswaram First Memorial Lecture-23rd March 1998. In: Climate and sustainable food security, vol 28, Vayu Mandal, pp 3–10
55. Taylor JW, Buizza R (2002) Neural network load forecasting with weather ensemble predictions. IEEE Trans Power Syst 17:626–632
56. Wang JW, Liu JW (2010) Weighted fuzzy time series forecasting model. Proceedings of the Second international conference on Intelligent information and database systems: part I. Springer-Verlag, Hue, Vietnam, pp 408–415
57. Wang L, Liu X, Pedrycz W (2013) Effective intervals determined by information granules to improve forecasting in fuzzy time series. Expert Syst Appl 40(14):5673–5679
58. Wang XZ (2015) Uncertainty in learning from big data-editorial. J Intell Fuzzy Syst 28(5):2329–2330
59. Wang XZ, Ashfaq RAR, Fu AM (2015) Fuzziness based sample categorization for classifier performance improvement. J Intell Fuzzy Syst 29(3):1185–1196
60. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan G, Ng A, Liu B, Yu P, Zhou ZH, Steinbach M, Hand D, Steinberg D (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37
61. Xiong Y, Yeung DY (2002) Mixtures of ARMA models for model-based time series clustering. IEEE International Conference on Data Mining. Los Alamitos, USA, pp 717–720
62. Yu HK (2005) Weighted fuzzy time series models for TAIEX forecasting. Physica A 349(3–4):609–624
63. Yu THK, Huarng KH (2008) A bivariate fuzzy time series model to forecast the TAIEX. Expert Syst Appl 34(4):2945–2952
64. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
65. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. InformSci 8:199–249