Neurocomputing 171 (2016) 551–568

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Secondary factor induced stock index time-series prediction using Self-Adaptive Interval Type-2 Fuzzy Sets

Diptendu Bhattacharya [a,*], Amit Konar [b], Pratyusha Das [b]

[a] NationalInstitute of Technology Agartala, Tripura, India
[b] Jadavpur University, Kolkata, India

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
18 June 2015
Accepted 29 June 2015
Communicated by: Dr. Swagatam Das
Available online 11 July 2015

Keywords:
Interval Type-2 fuzzy reasoning
Close price time-series
Secondary factor time-series
Stock index prediction
Self-adaptability

ABSTRACT

The paper introduces an alternative approach to time-series prediction for stock index data using Interval Type-2 Fuzzy Sets. The work differs from the existing research on time-series prediction by the following counts. First, partitions of the time-series, obtained by fragmenting its valuation space over disjoint equal sized intervals, are represented by Interval Type-2 Fuzzy Sets (or Type-1 fuzzy sets in absence of sufficient data points in the partitions). Second, an Interval Type-2 (or type-1) fuzzy reasoning is performed using prediction rules, extracted from the (main factor) time-series. Third, a type-2 (or type-1) centroidal defuzzification is undertaken to determine crisp measure of inferences obtained from the fired rules, and lastly a weighted averaging of the defuzzified outcomes of the fired rules is performed to predict the time-series at the next time point from its current value. Besides the above three main prediction steps, the other issues considered in the paper include: (i) employing a new strategy to induce the main factor time-series prediction by its secondary factors (other reference time-series) and (ii) self-adaptation of membership functions to properly tune them to capture the sudden changes in the main-factor time-series. Performance analysis undertaken reveals that the proposed prediction algorithm outperforms existing algorithms with respect to root mean-square error by a large margin ($\geq 23\%$). A statistical analysis undertaken with paired *t*-test confirms that the proposed method is superior in performance at 95% confidence level to most of the existing techniques with root mean square error as the key metric.

1. Introduction

Prediction of a time-series [1] refers to determining the amplitude of the series at time $t+1$ from its previous m sample values located at time: t, $t-1$, $t-2$, …., $t-(m-1)$ for a finite positive integer m. An m-th order time-series prediction involves all the m previous sample values directly for its forecasting/prediction [2,3]. In this paper, we, for the sake of simplicity, however, use a first order prediction of time-series, where the $(t+1)$-th sample of the time-series directly depends only on the sample value at time $(t+1-d)$, where d denotes the time-delay, although all the previous m sample values are required to design the prediction rules. There exists a vast literature on prediction of time-series for real processes, including rainfall [4,5], population growth [6], atmospheric temperature [7], university enrollment for students [8–11], economic growth [12] and the like. This paper is concerned with stock index, the time-series of which describing close price [13], is characterized by the following four attributes: non-linear [14], non-deterministic, non-stationary [15] and non-Gaussian jointly.

Designing a suitable model for stock index prediction requires handling the above four characteristics jointly. Although there exist several attempts to model time-series using non-linear oscillators [16], non-linear regression [17], adaptive auto-regression [18], Hzorth parameters [19] and the like, none of these could accurately model these time-series [20] for their inherent limitations to capture all the four characteristics jointly.

The logic of fuzzy sets plays a promising role to handle the above problems jointly. First, the nonlinearity of time-series is modeled by the nonlinearity of membership functions and their nonlinear mapping from antecedent to consequent space of fuzzy production rules. Second, the non-deterministic characteristics of the time-series (that might occur due to randomness in a wide space), is here significantly reduced because of its occurrence in

Abbreviations: CSV, Composite Secondary Variation; CSVS, Composite Secondary Variation Series; FOU, Footprint of Uncertainty; IT2, Interval Type-2; IT2FS, Interval Type-2 Fuzzy Set; IT2MF, Interval Type-2 Membership Function; MFCP, Main factor close price; MFTS, Main Factor Time-Series; MFVS, Main Factor Variation Series; RMSE, Root mean square error; SFTS, Secondary factor time-series; SFVS, Secondary Factor Variation Series; SFVTS, Secondary Factor Variation Time-Series; T1, Type-1; T1FS, Type-1 fuzzy Set; VTS, Variation Time-Series

* Corresponding author.
E-mail addresses: diptendu1@gmail.com (D. Bhattacharya), konaramit@yahoo.co.in (A. Konar), pratyushargj@gmail.com (P. Das).

http://dx.doi.org/10.1016/j.neucom.2015.06.073
0925-2312/

List of symbols used

$A_{i,j}$	Type-1 Fuzzy set for partition P_i of MFTS
$\tilde{A}_i$	Interval Type-2 fuzzy set for partition P_i in MFTS
B_i	Classical set for partition Q_i of MFVS
B_i'	Classical set for partition Q_i for SFVS/CSVS
$c(t)$	Close Price on $t-$th day
c_l	Left End Point Centroid of IT2FS
c_r	Right End Point Centroid of IT2FS
c	Centroid of an IT2FS
c'	Measured value of $c(t)$ in centroid calculation
c^x	Type-1 Centroid of $\mu_{A_x}(c(t))$
m_A	Mean values of the distributions of RMSE obtained by algorithms A
P_i	ith Partition for Close Price Time Series (of MFTS)
Q_i	ith Partition for Variation Series (of MFVS/SFVS/CSVS)
s_A	Standard deviation of the respective samples obtained by algorithms A
$V_M^d(t)$	Main Factor Variation Series with delay d
$V_{S^i}^d$	ith Secondary Factor Variation Series with delay d
$V_S^d(t)$	Composite Secondary Variation Series with delay d
W_{S^i}	Weight for ith secondary factor
$\mu_A(x)$	Type-1 Membership function of linguistic variable x in fuzzy set A
$\overline{\mu}_{\tilde{A}}(x)$	Upper Membership Function of IT2FS $\tilde{A}$
$\underline{\mu}_{\tilde{A}}(x)$	Lower Membership Function of IT2FS $\tilde{A}$
Δ_{S^i}	Total Difference Variation for CSVS of ith secondary factor
$\hat{\Delta}_{S^i}$	Normalized value of Δ_{S^i} for CSVS of ith secondary factor

one of a few equal sized partitions of the universe of discourse. Third, the non-stationary characteristics of the time-series that offers a correlation of signal frequencies with time [15] is avoided in fuzzy modeling by time-invariant models of membership functions [8]. Lastly, the non-Gaussian feature may be relaxed as locally Gaussian within small intervals (partitions) of the time-series. Thus, fuzzy sets are capable of capturing the uncertainty/imprecision in time-series prediction that might arise because of the above four hindrances.

The inherent power of fuzzy sets to model uncertainty of time-series has attracted researchers to employ fuzzy logic in time-series prediction. Song et al. [8–11] pioneered the art of fuzzy time-series prediction by representing the time-series value at time $t-1$ and time t as fuzzy membership functions (MFs) and connected them by fuzzy implication relations for all possible time t in the time-series. If there exist n possible discrete values of time t, then we would have $n-1$ possible fuzzy implication relations. Song et al. combined all these implication relations into a single relation R by taking union of all of these relations. The prediction involves first fuzzifying the crisp value of the time series at time t and then using composition rule of inference to determine the MF of the predicted time series at time $t+1$ using R as the composite time-invariant implication relation. Lastly, they defuzzified the result to obtain the crisp value of the time-series at time $t+1$.

The fundamental deviation in the subsequent work by Chen [21] lies in grouping of rules having common antecedents. Thus during the prediction phase, only few rules whose antecedent match with the antecedent of the fuzzified time-series value at time t only, need to be fired to obtain multiple MFs of the inferred consequences, one for each fired rule, an averaging type of defuzzification of which yields the actual inference at time $t+1$. Hwang et al. considered a variation time-series [22] by taking the difference of two consecutive values of the time-series, and used max-product compositional rule of inference to predict the inference of the variation at time $t+1$ from its previous values. A weighted average type of defuzzification was used to obtain the predicted value of the time-series at time $t+1$. Cai et al. [23] introduced genetic algorithm to determine the optimal weight matrix for transitions of partitions of a given time-series from each day to its next day, and used the weight matrix to predict the time-series at time $t+1$ from its value at time t. In [7], Chen et al. extended the work of Hwang et al. by first introducing a concept of secondary factors in the prediction of main factor time-series. There exists a vast literature on time-series prediction using fuzzy logic. A few of these that deserve special mention includes adaptive time-variant modeling [24], adaptive expectation modeling [25], Fibonacci sequence [26], Neural networks [27,28], Particle Swarm Optimization [29] based modeling, fuzzy cognitive maps and fuzzy clustering [30], bi-variate [31,32] and multi-variate [33–37] modeling and high order fuzzy multi-period adaptation model [38] for time-series prediction.

Most of the traditional works on stock index prediction developed with fuzzy logic [39] employ type-1 (T1) fuzzy reasoning to predict future stock indices. Although T1 fuzzy sets have proved their excellence in automated reasoning for problems of diverse domains, including fuzzy washing machines [40,41], fuzzy color TV [42] etc., they have limited power to capture the uncertainty of the real world problems [43]. Naturally, T1 fuzzy logic is incompetent to stock (and general time-series) prediction problems. The importance of Interval Type-2 Fuzzy Set (IT2FS) over its type-1 counterpart in chaotic time-series prediction has already been demonstrated by Karnik and Mendel [44]. There exist a few recent works attempting to model stock prediction problem using type-2 fuzzy sets [45,46]. These models aim at representing a single (interval) type-2 membership function (MF), considering three distinct stock data items, called close, high and low prices [13]. Here too, the authors partitioned each of the above three time-series into intervals of equal size, and represented each partition as T1 fuzzy set. They constructed fuzzy If-Then rules describing transitions of stock index price from one day to the next day for each of the above time series. During prediction, they identified a set of rules containing antecedent fuzzy sets corresponding to current stock prices, obtained union and intersection of the consequents of the rules to derive (interval) type-2 fuzzy inferences and employed center average defuzzifiers to predict the stock price for the next day. Bagestani and Zare [46] extended the above work by adaptation of the structure of the membership functions and weights of the defuzzified outputs to optimize root mean square error. In addition, the latter work employed center of gravity defuzzifier in place of center average defuzzifier used previously. The present paper is an extension of the seminal work of Chen et al. [47] by the following counts.

1. In order to represent the close price $c(t)$ within a partition (interval), we represent each short duration contiguous fluctuation of $c(t)$ in a given partition of the universe of $c(t)$ by a type-1 MF, and take union of all these type-1 MFs within a partition to represent it by an Interval Type-2 Fuzzy Set (IT2FS). Under special circumstances, when a partition includes one or a few contiguous data points only, we represent the partition by a type-1 MF only.
2. The antecedent and consequent of fuzzy prediction rules of the form $A_i \rightarrow A_j$ (extracted from the consecutive occurrence of data