International Journal of Approximate Reasoning 57 (2015) 1–18

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

# Using interval information granules to improve forecasting in fuzzy time series

Wei Lu [a,*], Xueyan Chen [b], Witold Pedrycz [c,d,e], Xiaodong Liu [a], Jianhua Yang [a]

[a] School of Control Science and Engineering, Dalian University of Technology, Dalian City, PR China
[b] School of Mechanical Engineering, Inner Mongolia University for the Nationality, Tongliao City, PR China
[c] Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB, Canada
[d] Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[e] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

*Article history:*
Received 1 April 2014
Received in revised form 10 August 2014
Accepted 3 November 2014
Available online 8 November 2014

*Keywords:*
Fuzzy time series
Interval granules
Unequal-sized intervals
Forecasting

ABSTRACT

In the process of modeling and forecasting of fuzzy time series, an issue on how to partition the universe of discourse impacts the quality of the forecasting performance of the constructed fuzzy time series model. In this paper, a novel method of partitioning the universe of discourse of time series based on interval information granules is proposed for improving forecasting accuracy of model. In the method, the universe of discourse of time series is first pre-divided into some intervals according to the predefined number of intervals to be partitioned, and then information granules are constructed in the amplitude-change space on the basis of data of time series belonging to each of intervals and their corresponding change (trends). In the sequel, optimal intervals are formed by continually adjusting width of these intervals to make information granules which associate with the corresponding intervals become most "informative". Three benchmark time series are used to perform experiments to validate the feasibility and effectiveness of proposed method. The experimental results clearly show that the proposed method produces more reasonable intervals exhibiting sound semantics. When using the proposed partitioning method to determine intervals for modeling of fuzzy time series, forecasting accuracy of the constructed model are prominently enhanced.



## 1. Related works

As fuzzy time series cope with the forecasting problem under uncertain circumstances in which historical data are incomplete or vague, at present it has been widely applied to many fields such as forecasting enrollments, stock index, temperature, etc., showing better forecasting performance. The concept of fuzzy time series was originally proposed by Song and Chissom [30] engaging the formalisms of fuzzy set theory [38]. Compared with traditional time series, the historical data of fuzzy time series are linguistic terms rather than numeric values. The authors presented a detailed approach to forecast fuzzy time series. It comprises five key steps: (1) partitioning the universe of discourse of time series into a collection of intervals, (2) defining fuzzy sets according to partitioned intervals, (3) transforming numeric time series into fuzzy time series, viz., fuzzifying historical data of time series, (4) mining fuzzy relationships from fuzzy time series, and (5) forecasting and

* Corresponding author.
*E-mail address:* luwei@dlut.edu.cn (W. Lu).

http://dx.doi.org/10.1016/j.ijar.2014.11.002
0888-613X/

defuzzifying the forecasting outputs. Based on these five steps, Song and Chissom developed two time series models — the time-invariant model [28] and the time-variant model [29]. Following the work of Song and Chissom, focusing on step (4) and (5), most of researchers focused on various variants of fuzzy time series models with an intent to improve forecasting accuracy. Chen [4] proposed a simplified model to reduce the computation overhead of Song's model [28] by using simple arithmetic operations to replace max–min composition operations in the process of mining fuzzy logical relations and performing prediction. Huarng [13] presented a heuristic model by integrating problem-specific heuristic knowledge with Chen's model. The forecasting accuracy of the model is better than the accuracy of Chen's model [4]. Subsequently, in order to further enhance forecasting accuracy of model, Chen [5] extended his previous work [4] to a high-order time-invariant fuzzy time series model. Huarng and Yu [14] also developed a Type 2 fuzzy time series model to build up forecasting accuracy, in which extra observations are used to enrich or to refine the fuzzy relationships. Yu showed models of refinement relation [35] and weighting scheme [36] for improving forecasting accuracy. Cheng [9] presented a fuzzy time series model which incorporates the adaptive expectation model into forecasting processes to reduce forecasting errors. Wong et al. [33] developed an adaptive time variant fuzzy time series forecasting model to improve forecasting accuracy. In addition, Huarng and Yu [15] exploited neural networks to construct fuzzy time series model. The model was used to forecast stock index and obtained better forecasting results. Egrioglua et al. [11] proposed a hybrid model by synergy of fuzzy c-means clustering algorithm and artificial neural networks.

Although all the above-mentioned models have reached better forecasting accuracy, forecasting performance of these models can also be significantly affected by partitioning the universe of discourse of time series, i.e., as to the same fuzzy time series model, the difference of partitioned intervals generated by adopting different the universe of discourse partitioning method in step (1) can result in different forecasting accuracy. Thus, how to efficiently partition the universe of discourse is critical to forecasting of fuzzy time series. The techniques of partitioning the universe of discourse can be roughly fallen into two categories according to whether the size of partitioned intervals is equal or not — the equal-sized intervals partitioning techniques and the unequal-sized intervals partitioning techniques.

With regard to the first techniques of partitioning the universe of discourse — the equal-sized intervals partitioning technique, one partitions the universe of discourse of time series into a series of intervals with the same size (length). There are some methods supporting the equal-sized partitioning. A simplest method is to directly partition the universe of discourse into equal-sized intervals without any processing, which is frequently applied into the development process of many fuzzy time series models [28,29,4,13,5,14,35,36,9,15,11]. For example, generally, in the process of developing Chen's model [4] to forecast enrollment of university of Alabama, the universe of discourse of enrollment time series [13 000, 20 000] is evenly partition into 7 intervals, and length of each interval is 1000 — [13 000, 14 000], [14 000, 15 000], [15 000, 16 000], [16 000, 17 000], [17 000, 18 000], [18 000, 19 000], and [19 000, 20 000]. Although the method is very simple and easy to use, the formed intervals is unreasonable as it is with strongly subjectivity and arbitrariness. Huarng [12] observed this drawback of the method and proposed two heuristic methods, i.e., distribution- and average-based to determine effective length of partitioned intervals. Li and Chen [20] proposed the natural partitioning based method, which can recursively partition the universe of discourse level by level in an intuitive fashion. Chen and Hsu [6] presented a two-phase partitioning method based on the statistical distributions of the historical data. Egrioglu et al. [10] obtained the optimum equal-sized intervals by converting the problem of determining length of intervals to the problem of a single variable constrained optimization. The above-mentioned the universe of discourse partitioning methods based on equal-sized intervals is applied to model for forecasting enrollments time series according to Chen's modeling method [4], results of which show higher forecasting accuracy than the one reported for the classical Chen's model [4].

With regard to the second techniques of partitioning the universe of discourse — the unequal-sized intervals partitioning techniques, it can divide the universe of discourse of time series into a series of intervals with the unequal size. Generally, this technique can obtain unequal-sized intervals in following two ways. One is to directly generate unequal-sized intervals from data with the use of some advanced data mining technology such as fuzzy c-means clustering (FCM) or optimization technology such as genetic algorithm (GA), particle swarm optimization (PSO) algorithm or alike. Huarng and Yu [16] proposed the "ratio-based" unequal-sized intervals partition method to generate intervals for improving forecasting results. Chen and Tanuwijaya [8] adopted an automatic clustering algorithm to produce unequal-sized intervals in the universe of discourse for building multivariate fuzzy time series model. Chen and Kao [7] proposed a method of partitioning the universe of discourse in which PSO algorithm is exploited to find optimal unequal-sized intervals according to the distribution of historical data of time series. The other is to obtain unequal-sized intervals in supervised learning way. Kuo et al. [19,18] utilized respectively GA and PSO algorithm to determine the unequal-sized intervals so that error of constructed model is minimum. Singh and Borah [27] also utilized PSO algorithm to construct unequal-sized intervals for developing Type-2 fuzzy model of stock time series on basis of the scheme of supervised learning. These the universe of discourse partitioning methods based on unequal-sized intervals are used to forecast enrollments, stock index, temperature, etc., whose results clearly show that using unequal-sized partitioning can produce better forecasting accuracy than the equal-sized intervals [19,7,18].

The aforementioned techniques of partitioning the universe of discourse, including the equal-sized intervals partitioning technique and the unequal-sized intervals partitioning technique, can improve forecasting accuracy when they are used in the process of modeling of fuzzy time series. However, there are also still some deficiencies: