Computers and Electrical Engineering 000 (2017) 1–21

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Crop yield forecasting using fuzzy logic and regression model☆

Bindu Garg, Shubham Aggarwal*, Jatin Sokhal

Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India

ARTICLE INFO

Article history:
Received 10 March 2017
Revised 11 November 2017
Accepted 11 November 2017
Available online xxx

Keywords:
Mean Square Error
Fuzzy time series
Average forecast error rate
Prediction
Wheat yield
Time series data

ABSTRACT

Time-Series data has been of great importance to the research field of prediction models. Over the past two decades, multitudinous fuzzy time series blueprints have been put forth for agricultural yield production. However, most of these predictions were based on 7th interval partitioning. A surprising insight was that nobody gave a sound reason to justify the choice of that particular interval. So, this paper focuses on predicting data values on a large spectrum of fuzzy logic computations based on second and third-degree relationships. This paper showcases work on 4 different types of the fuzzy interval, where each interval is tested with 4 degrees of regression equations. Each of these 16 cases is performed for the fuzzy logic relationship (FLR) 2 and 3 separately. Apart from this, the robustness of algorithm is a testament to an incredible solution for the time series model. In addition to this, A Regression analysis model has been enforced to accomplish the efficient defuzzification operation. To elucidate the process of forecasting, the historical data of wheat yield of University of Agriculture and Technology has been used.

1. Introduction

The forecasting methodology is very befitting in the cases where incertitude related to the future is concrete. The prediction of outcomes in the future is attained through this process. Pertinent data and graphs are deliberated and queried in order to make optimal choices concerning the future. The employment of time series forecasting has come into picture predominantly for two reasons. First, time series data forms a preponderant part of the data existing in business, economic and financial areas. Next, it is indeed facile to appraise time series as many technologies are procurable for evaluation of time series forecast.

The inspiration for this research work came from the authors' previously published research work [24] and study of previous research work conducted in the area of predictive modeling using fuzzy logic. Various authors in the past have used a particular interval of partitioning in their work but have not justified non-inclusion of other intervals. Most researchers used 7th interval partitioning in their research work but have not justified non-inclusion of other intervals. Use of this particular partition by all made the authors of this paper curious to find out how the results vary by changing the interval of partitioning and even within one interval and how the prediction values change with the change in polynomial degree equation

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. R. C. Poonia.

* Corresponding author.
E-mail address: shubhamaggrwl@gmail.com (S. Aggarwal).

https://doi.org/10.1016/j.compeleceng.2017.11.015
0045-7906/

in regression analysis. A substantial segment of the work on time series has been performed to address and unravel the solutions to problems like a number of outpatient visits, healthcare, prediction in information systems forecasting, economic and sales forecasting, analysis of the budget, stock market prediction and fluctuations and business analysis etc. Thus, there exists a persistent demand for forecasting techniques that offer optimal and veracious results. Thus, the need of the hour is the demand of soft computing based forecasting methods which are efficacious and consummate.

This paper explains the novel contribution to the readers in an easy to navigate sections. Section 1 gives a brief introduction on the topic of fuzzy logic and the inspiration for this paper. Section 2 explains the significance of wheat production is explained in brief. In Section 3, literature survey has been done in order to let users understand the progress in soft computing field in the past. Section 4 presents the proposed method is easy to understand steps along with tables to understand the data distribution. The authors have further subdivided the proposed method into four sections in order to show the results of the application algorithm on polynomial degrees 1 to 4. In Section 5, the results obtained are explained through graphical illustrations in order to give a quick glance of comparison among different intervals. Section 6 shows the robustness of the algorithm to substantiate the algorithm's strength. In Section 7, the paper concludes with discussion on future scope of this paper.

2. Significance of wheat production

India is one of the largest producers of wheat in the world (close to one – fifth of the total wheat yield worldwide). Wheat is the staple diet of the people residing in the eastern and southern parts of the country. Also, it is one of the chief grains of the nation and accounts for the largest portion of the agricultural area under its cultivation. Wheat, being a tropical plant, burgeons comfortable in hot climate and India, a tropical country, endorses this argument. Farming of the wheat crop is considered as a very lucrative, remunerative and sustainable venture throughout the world. The Asian continent on its own produces and consumes more than three-fourths of net wheat production. If sources are to be believed, the recent sharp boom in the Asian wheat production will act as a means to lessen poverty. Since improved production and growth in yield of wheat clearly infers that wheat would be easily available to poor people at a lower price.

Henceforth, the relatively reduced cost of wheat will result in incitation of farmers to invest in higher valued crops and thus, bring prosperity and added income to their families, which will also secure enhanced nutrition for the customers. The evaluation and prediction of wheat production is indeed an Augean field. Past work done on wheat bespeaks that there exists a variation in wheat production, which in addition to being really unpredictable, also consumes a lot of precious time. Also, administrators in agriculture and decision-makers should be easily able to comprehend the forecasted results. A very confined and limited research has been performed in the field of soft computing based techniques for crop planning and crop resource harnessing. The scope of the current prediction systems is circumscribed to case studies and laboratories only. Clearly, these systems would prove impracticable in the field of agriculture since they cannot demonstrate the diversity and intricacy of crop production.

3. Related work

Fuzzy time series prediction is a prudent avenue in the areas where information is inexplicit, unclear and approximate. Also, fuzzy time series can tackle circumstances that neither provides the analysis of trends nor the visualization of patterns in time series. Profound research work has been accomplished on forecasting problems using this concept. Vikas [1] proposed different techniques for prediction of crop yields and used the artificial neural network to predict wheat yield. Adesh [2] did a comparative study of different techniques involving neural networks and fuzzy models. Askar [3] also tried to predict crop yield using time series models. Sachin [4,5] worked specifically on rice yield prediction using fuzzy time series model. Narendra [6] tried to predict wheat yield. Pankaj [7] used adaptive neuro-fuzzy systems for crop yield forecasting Wheat Yield Prediction. Fuzzy time series concepts and definitions were invented and presented by Song and Chissom. They also portrayed the concepts and notions of variant and invariant time series [8,9]. Initially, time series data of the University of Alabama was taken and enrollment forecasting was executed, and after some years they also [10] formulated an average autocorrelation function as a measure of dependency. Later, Chen [11,12] depicted simplified arithmetic operations instead of using max-min composition operations that were previously accustomed by Song & Chissom and then, arranged forecasted model using high order fuzzy time series. Singh [13] proposed fuzzy forecasting method, with a slight variation. Lee administered a fuzzy candlestick pattern to enhance forecasting outcomes [14]. Later, a multivariate heuristic model was designed and implemented to obtain highly intricate and complex matrix computations [15]. Research work was performed to ascertain the length of Intervals of fuzzy time series [16]. Event discretization function based forecasting models were put forth [17] and practiced to predict the average duration of stay of a patient [18]. Garg [19,20] developed a forecasting approach by administering the notion of OWA weights. This model proved to be an accomplishment as it downsized forecasting error to a certain extent. Afterwards, Garg [21,22] also put forward an optimized model based on genetic-fuzzy-OWA forecasting. Garg [23] proposed a new prediction model on time series data. Subsequently, the number of outpatient visits in the hospital was demonstrated by Garg [24].

As a matter of fact, the majority of these models were administered for prediction of all other problem domains except wheat production. Keeping this fact in mind, one can put forth a model to predict wheat production for India on the premise of historical time series wheat data. Real-time data of Pantnagar farm, G.B. Pant University of Agriculture & Technology, India

Please cite this article as: B. Garg et al., Crop yield forecasting using fuzzy logic and regression model, Computers and Electrical Engineering (2017), https://doi.org/10.1016/j.compeleceng.2017.11.015