Fuzzy Time Series Analysis and Prediction using Swarm Optimized Hybrid Model

A report submitted in partial fulfilment of the requirements of Bournemouth University for the degree of Master of Philosophy

Submitted by

A.K.M. Kamrul Islam

Bournemouth University

Faculty of Science & Technology

Bournemouth University, United Kingdom

October 2018

Copyright Statement

Dedication

To my parents for opening my eyes to the world

Abstract

Time series forecasting has an extensive trajectory record in the fields of business, economics, energy, population dynamics, tourism, etc. where factor models, neural network models, Bayesian models are exceedingly applied for effective prediction. It has been exemplified in numerous forecasting surveys that finding an individual forecasting model to achieve the best performances for all potential situations is inadequate. Moreover, modern research endeavour has focused on a deeper understanding of the grounds. Rather than aim for designing a single superior model, it focused on the forecasting methods that are effective under certain situations. For instance, due to the qualitative nature of forecasting, a business can come up with diverse scenarios depending on the interpretation of data. Therefore, the organizations never rely on any individual forecasting model solely, rather focused on sets of individual models to attain the best possible knowledge of the future.

The time series forecasting model has a great impact in terms of prediction. Many forecasting models related to fuzzy time series were proposed in the past decades. These models were widely applied to various problem domains, especially in dealing with forecasting problems where historical data are linguistic values. A hybrid forecasting method can be effective to improve forecast accuracy by merging sets of the individual forecasting models. Numerous hybrid forecasting models have been proposed last couple of years that combined fuzzy time series with the evolutionary algorithms, but the performance of the models is not quite satisfactory. In this research, a novel hybrid fuzzy time series forecasting model is proposed that used the historical data as the universe of discourse and the automatic clustering algorithm to cluster the universe of discourse by adjusting the clusters into intervals. Furthermore, the particle swarm optimization algorithm is also examined to improve forecasted accuracy. The proposed method is considered to forecast student enrolment of the University of Alabama. The model achieves a significant improvement in forecast accuracy as compared to state-of-the-art hybrid fuzzy time series forecasting models.

It is obvious from the literature that no forecasting technique is appropriate for all situations. There is substantial evidence to demonstrate that combining individual forecasts produces gains in forecasting accuracy. The addition of quantitative forecasts to qualitative forecasts may reduce forecast accuracy. Individual forecasts are combined based on either the simple arithmetic average method or an artificial neural network. Research has not yet revealed the conditions for the optimal forecast combinations. This thesis provides a few contributions to enhance the existing combination model. A set of Individual forecasting models is used to form a novel combination forecasting model based on the characteristics of resulting forecasts.

All methods derived in this thesis are thoroughly tested on several standard datasets. The related characteristics of the resulting forecasts are observed to have different error decompositions both for hybrid and combination forecasting model. Advanced combination structures are investigated to take advantage of the knowledge of the forecast generation processes.

Publication Resulting from the Report

1. Islam, A.K.M. Kamrul, Bouchachia, Hamid, Cang, Shuang, Yu, Hongnian: ‘Fuzzy time series forecasting based on fuzzy logical relationships, PSO technique and automatic clustering algorithm’, *ICAISE 2016: 18th International Conference on Artificial Intelligence and Software Engineering* (Accepted), 2016.

Table of Contents

List of Figures

List of Tables

Acknowledgement

Writing acknowledgments to all who have contributed to this work is practically impossible. But I would like to explicitly thank a few of the people who helped me along the way.

First and foremost, my appreciation and heartfelt gratitude to my thesis supervisor and mentor, Dr. Hamid Bouchachia, Bournemouth University, for his constant endeavour, step by step guidance and the time he devoted to this research work.

I would also like to express my deepest gratitude to the European Union for their funding of this project in Bournemouth University, UK, as without this it is unlikely that a successful progress of this project would have been achieved. I would like to thank Dr. Suang Cang and Prof. Hongnian Yu for their continued guidance and help throughout.

I am very thankful to the Dean, Faculty of Science & Technology, for the facilities and the work environment provided to me. These facilities enable me to enrich my knowledge and carry out my research.

Many thanks go to all the staff at Bournemouth University, my friends and colleagues who have helped me with this research project.

In particular, I am indebted to my dearest parents for their endless love. I am so lucky that, from childhood, I have been influenced by their strong philosophical background. Without these emotional and mental supports, the achievement of the thesis could not be possible.

A.K.M. Kamrul Islam

October 2018

Chapter 1
Introduction

1 Introduction

Forecasting describes a broad research area concerned with estimation of future events and is important for the organization to plan or adopt the necessary policies. Forecasting can assist to make a better development and decision-making in the country. It can be found in a wide variety of sectors like stock returns, finance analysis, weather news, currency exchange, GDP estimation, tourism demand, etc. Forecasting studies have had a half a century history. Several review articles on forecasting have been published over the last few decade. Forecasting models can be broadly divided into following categories: (i) time-series models (ii) AI models, such as neural networks, fuzzy time-series theory, grey theory, and expert systems (Kayacan et al., 2010). Time series forecasting has been a very active research area since 1950's, and a variety of forecasting approaches have been introduced in the scientific literature and were used in many practical applications. Nonlinear statistical time series models have been proposed with the aim to improve the forecasting performance of nonlinear systems. These include bilinear model, threshold autoregressive model (TAR), smoothing transitions autoregressive model (STAR), autoregressive conditional heteroscedastic model (ARCH) and generalized autoregressive conditional heteroscedastic model (GARCH). These models are known as the second generation of time series models (Chen, 1996).

Various forecasting techniques are available in the literature. All the methods fall into one of two overarching approaches*:* qualitative and quantitative. Quantitative models assume adequate knowledge of an underlying process and are often experts' judgements. Hyndman and Koehler (Hyndman and Koehler, 2006) stated that, every forecasting method is disparate in terms of accuracy, scope, time horizon and cost. The preference of the forecasting model is the key issue that influences on the forecasting accuracy. Moreover, the individual forecasting model is not quite enough under certain situations for the accuracy of prediction. Despite the consensus on the need to develop more accurate forecasts and the recognition of their corresponding benefits, there is no one model that stands out best in terms of forecasting accuracy (Song and Chissom, 1993b). A hybrid forecasting approaches have been proposed by many researchers to improve the forecasting accuracy to outperform individual forecasting approaches (Huang and Jane, 2009). Evolutionary and optimization algorithm in hybrid model could be a good practice to maintain the higher rate of accuracy in prediction. The time series forecasting model, and other models have been combined to improve the forecasting performance of the time series model.

Due to the uncertainty of relying on a single forecasting model, the combination forecasting models have been immensely used to improve the forecasting performance in various sectors. Forecasting literature suggests that, combinations of model are generally found to outperform the specific models being combined, independently of the time horizon considered (Coshall and Charlesworth, 2011). Moreover, the combination of multiple single forecast model provides the best performance in terms of forecasting (Hua et al., 2007, Wong et al., 2007). To avoid the difficulty and risk for model selection, combining forecasting would be another major motivation. In 2003, Zhang stated that the final selected model is not necessarily the best for future uses due to many potential influencing factors such as sampling variation, model uncertainty and structure change. Model selection can be at ease by combining several forecasting models (Stock and Watson, 2004). Bates and Granger (Bates and Granger, 1969) suggested that the combination of models that contain independent information is most likely to improve forecast accuracy. Shen and Huang (2008) performed the forecast by combining the methods like simple average, variance-covariance and mean squared forecast error methods with multiple-step-ahead forecasting horizons and seven single forecasting techniques (autoregressive distributed lag model, error correction model, maximum likelihood error correction model, vector autoregressive model, time-varying parameter model, seasonal naïve model and seasonal autoregressive integrated moving average model). Most of the combination forecasting methods described above are based on the linear combination, but there are some situations when the forecasting problem can be nonlinear. For this nonlinear relationship between inputs and outputs, linear models may provide only an approximate forecast, while nonlinear combination methods may provide more accurate forecasting.

1.1 Aims

The research work aims to develop a more accurate and effective forecasting model by comparison to the current forecasting models with the purpose of achieving better performance. Therefore, the research approach followed by involving five main steps 1) To understand the existing individual, hybrid, and combination forecasting models along with their application in different domains, 2) To identify the key issues that can improve the performance and accuracy of the forecasting model, 3) To design and implement a new hybrid forecasting model that can be evaluated through performance measurement methods, 4) To investigate a set of individual forecasting models and ponder how the models can be extended and combined with the essential features to attain better accuracy. The performance of the forecasting model can be compared in terms of the local machine to a parallel processing environment.

1.2 Objectives

The objectives of this study are-

1. To analyse the forecasting models in terms of time series and artificial intelligence. To comprehend the existing individual forecasting models (ARIMA, SARIMA, SVR, RBF, and ANN, etc.), hybrid forecasting models (ARIMA and ANN, GA and SARIMA, etc.) and the combination forecasting models (simple average, trimmed mean, winsorized mean, median, etc.) in terms of prediction (Song and Chissom, 1993b).
2. To develop an effective hybrid forecasting model using fuzzy time series and particle swarm intelligence that can accurately predict the future observations with higher accuracy rate compared to state-of-the art models.

   A hybrid forecasting model has proposed and the following step by step processes are maintained:

   - Construct a hybrid forecasting model with fuzzy time series, particle swarm optimization and clustering technique.
   - Implement the hybrid forecasting model using a MATLAB simulator with practical datasets.
   - Train and validate the proposed model.
   - Test the hybrid model with appropriate datasets.
   - Forecast error analysis based on error measurement techniques and compare the result with state-of-the art individual or hybrid forecasting models.
3. To propose a combination forecasting model by combining the weights of different individual forecasting models with a data mining algorithm that can select specific model weights from the entire individual model.

   A combination forecasting model is proposed and the following step by step processes are maintained:

   - Establish the combination forecasting model with ARIMA, ANN, RBF, ANFIS, and Naïve Bias models.
   - Implement the combination forecasting model using a MATLAB simulator with practical datasets.
   - Train, validate and test each individual model that applied in the proposed combination forecasting model.
   - Linear and nonlinear forecast combination methods applied to get suitable combination results.

- Forecast error analysis based on error measurement techniques and compare the result with state-of-the art combination forecasting models.

4. To evaluate the performance of the proposed model with real-life datasets.
5. To evaluate the performance of the proposed forecasting model in a local machine.

1.3 Outline of the whole report

The background knowledge and structure of the thesis are relevant in three different areas: time series forecasting, hybrid forecasting method, and forecast combination method. In Chapter 2, the preliminary information will be extended, providing a literature review and discuss the most significant impacts and algorithms for each of the areas. Chapter 3 mainly focused on time series forecasting methods that currently used in different domains. This chapter also investigates the question of how effectively fuzzy time series forecasting methods perform in empirical studies along with the goal to assess the benefit of applying complex forecasting algorithm that usually needs to be identified and fitted by experts. In the same context, a hybrid forecasting model using automatic clustering technique and particle swarm optimization is described and compared the results to state-of-the-art forecasting models. The performance of the proposed method has been extensively verified using publicly available datasets to make the results comparable to state-of-the-art methods available in current literature. It also provided a thorough investigation of prospects to enhance forecast accuracy. Chapter 4 concludes by summarising the results and findings of the project and an outlook on future work ends the thesis.

Chapter 2

Literature Review

2 Literature Review

2.1 Forecasting

Forecasting is the process of constructing a prediction of future events and occurrences based on present and past data sets, trends, scenarios, etc. and is the key element in various sectors of the world. Short-range and long-range planning are essential in terms of future prediction. A forecast is opposed to a prediction as it is based on previous data, whereas prediction is based on instinct or guess. For instance, the late afternoon news broadcasts the weather forecast, not the weather prediction. Forecasting also refers to formal statistical methods like time series and is based on noticeable, observable data and trends. There are different one-off spread factors and seasonal factors are crucial for getting an accurate forecasts (Bowerman and O'Connell, 1993).

2.1.1 Characteristics of a Good Forecast

There are a few attributes that are beneficial to determine a decent forecast:

- Accuracy —accuracy should be maintained in real-time that the analogy can be yielded to alternative forecasts.
- Reliability —good forecast can be obtained from the forecast method if the degree of confidence of the user can be established.
- Timeliness —a certain amount of period is required to respond to the forecast so the forecasting horizon must permit for the period necessary to make alterations.
- Easy to use and understanding —forecasting process should be easy for the user to utilize it efficiently and should be confident and comfortable working with it.
- Cost-effectivity —the cost of getting the forecast should not outweigh the advantages attained from the forecast.

2.1.2 Number of Assumptions in Forecasting

Forecasting is centred on several assumptions characterized below:

- The history will replicate itself. In other words, what has occurred in the earlier period will take place again in the future.
- As the forecast limit reduces, forecast accuracy increases. For instance, a forecast for tomorrow will be more accurate than a forecast for next month; a forecast for next month will be more accurate than a forecast for next year, and a forecast for next year will be more accurate than a forecast for ten years in the future.
- Aggregate forecast is more accurate than forecasting individual items. This means that a company will be able to forecast total demand over its entire spectrum of products more

accurately than it will be able to forecast individual stock-keeping units. For example, General Motors can more accurately forecast the total number of cars needed for next year than the total number of white Chevrolet Impalas with a certain option package (Stock and Watson, 2002).

- Forecasts are infrequently accurate and almost never totally accurate, although some are very close. Therefore, it is sensible to offer a forecast “range.” If one were to forecast demand of 100,000 units for the next month, it is extremely unlikely that demand would equal exactly 100,000. However, a forecast of 90,000 to 110,000 would provide a much larger target for planning (Stock and Watson, 2006).

2.1.3 Forecasting Applications

Forecasting elements have the following applications:

- Forecasting utilization rates for credit cards: build a model based on historical data and use the model to score a current credit card portfolio to determine utilization rates.
- Model loss rates of a group of home equity lines of credit as a function of time.
- An independent system operator, organized for monitoring the electrical grid, has a need to predict electrical usage – the volatility of the daily usage can be thought of as a blend of day-ahead-market volatility and monthly volatility, where the month can be one or more months forward.(Skamarock and Klemp, 2008)

2.1.4 Forecasting Approaches

Two types of forecasting methodologies exist, these are qualitative and quantitative method. (Stock and Watson, 2002)

2.1.4.1 Qualitative Models

Qualitative forecasting methods have the following properties:

- Used when the situation is vague & little data exist
- New products and new technology
- Involves intuition, experience
- ex., Forecasting sales to a new market
- Qualitative methods include Delphi technique, Nominal Group technique, executive opinions, market research

2.1.4.2 Quantitative Models

Quantitative forecasting methods have the following properties:

- Used when the situation is ‘stable’ & historical data exist
- Existing products and current technology

- Heavy use of mathematical techniques
- ex., Forecasting sales of a mature product

2.1.5 Autoregressive Integrated Moving Average (ARIMA) Model

The autoregressive–moving-average (ARMA) models can only be used for stationary time series data. However, in practice many time series such as those related to socio- economic and business show non-stationary behaviour. Time series, which contain trend and seasonal patterns, are also non-stationary in nature. Thus, from application viewpoint ARMA models are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason, the ARIMA model is proposed, which is a generalization of an ARMA model to include the case of non-stationary as well.

In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points. The mathematical formulation of the ARIMA(p,d,q) model using lag polynomials is given below (Zhang, 2003):

$$\varphi(L)(1\text{-}L)^{d}\, y_t \;=\theta(L)\varepsilon_t\,,\ i.e.$$

$$\left(\sum_{i=1}^{p} \varphi_i L^i\, y_t\right)(1\text{- }L)^d y_t = \left(1+\ \sum_{j=1}^{p} \varphi_j L^j\right)\varepsilon_t \qquad (2.1)$$

- Here, p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated and moving average parts of the model, respectively.
- The integer d controls the level of differencing. Generally, d = 1 is enough in most cases. When d = 0, then it reduces to an ARMA(p,q) model.
- L is the lag operator, the φ_i are the parameters of the autoregressive part of the model, φ_j are the parameters of the moving average part and are the error terms. The error terms are generally assumed to be independent that the variables distributed identically are sampled based on a normal distribution using zero mean.
- An ARIMA(p,0,0) is nothing but the AR(p) model and ARIMA (0,0, q) is the MA(q) model.
- ARIMA (0,1,0), i.e. $y_t = y_{t-1} + \varepsilon_t$ is a special one and known as the *Random Walk (RW)* model. It is widely used for non-stationary data, like economic and stock price series.

A useful generalization of ARIMA models is the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which allows non-integer values of the differencing parameter d. ARFIMA has useful application in modelling time series with long memory.
In this model the expansion of the term $(1-L)^{d}$ is to be done by using the general binomial theorem. Various contributions have been made by researchers towards the estimation of the general ARFIMA parameters.

2.1.6 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

The ARIMA model is for non-seasonal non-stationary data. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal difference of applicable order is used to remove non-stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the earlier year and is calculated as $z_t = y_t - y_{t-s}$. For monthly time series $s = 12$ and for quarterly time series $s = 4$. This model is generally termed as the SARIMA $(p,d,q)\times(P,D,Q)^s$ model.

The mathematical formulation of a SARIMA$(p,d,q)\times(P,D,Q)^s$ model in terms of lag polynomials is given below (Tseng and Tzeng, 2002):

$$\Phi_p(L^s)\,\varphi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_Q(L^s)\,\theta_q(L)\varepsilon_t \text{, i.e.}$$

$$\Phi_p(L^s)\,\varphi_p(L)z_t = \Theta_Q(L^s)\,\theta_q(L)\varepsilon_t \qquad (2.2)$$

Here z_t is the seasonally differenced series, P is the seasonal autoregressive order, Q is the seasonal moving average, D is the seasonal difference order, and s is the number of time steps for a single seasonal period.

2.1.7 Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) approach has been hinted as an unconventional technique to time series forecasting and it achieved enormous recognition in the last few years. The basic idea of ANNs was to assemble a model for imitating the intelligence of the human brain into a machine. Like the functionality of a human brain, ANNs try to identify consistencies and patterns in the input data discover from experience and then deliver comprehensive results based on their known preceding knowledge.

Although the development of ANNs was mainly biologically motivated, but afterwards they have been applied in many different areas, especially for forecasting and classification purposes. Below the salient features of ANNs has mentioned, which make them quite a favourite for time series analysis and forecasting (Al-Alawi and Al-Hinai, 1998).

First, ANNs are data-driven and self-adaptive in nature. There is no need to specify a certain model form or to create any *a priori* hypothesis about the statistical dissemination of the data; the desired model is adaptively established based on the features produced from the data. This approach is extremely beneficial for many practical circumstances, where no theoretical assistance is available for an appropriate data initiation process.

Second, ANNs are intrinsically non-linear, which makes them more practical and precise in modelling convoluted data patterns, as contradicted to various traditional linear methodologies, such as ARIMA methods. There are many occasions, which suggest that ANNs made relatively better analysis and forecasting than various linear models.

Finally, as suggested by Hornik and Stinchcombe (1989), ANNs are comprehensive functional approximators. They have shown that a network can approximate any continuous function to any anticipated accuracy. ANNs use parallel processing of the information from the data to approximate a huge class of functions with a high degree of accuracy. Further, they can deal with a situation, where the input data are erroneous, incomplete, or fuzzy.

The most widely used ANNs in forecasting problems are multi-layer perceptron's (MLPs), which use a single hidden layer feed-forward network (FNN). The model is characterized by a network of three layers, viz. input, hidden and output layer, connected by acyclic links. There may be more than one hidden layer. The nodes in numerous layers are also recognized as processing elements. The three-layer feed-forward architecture of ANN models can be diagrammatically illustrated as below:

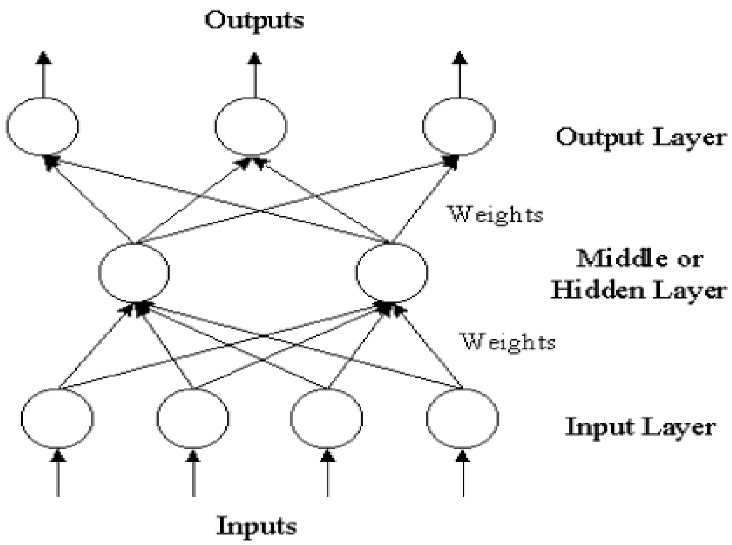

Figure 2.1: The three-layer feed forward ANN architecture(Al-Alawi and Al-Hinai, 1998)

The output of the model is computed using the following mathematical expression:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_j y_{t-i}\right) + \varepsilon_t, \qquad (2.3)$$

Here y_t (i=1,2, ..., p) are the p inputs and y_t is the output. The integers p, q is the number of input and hidden nodes, respectively. α_j (j= 0,1,2,...,q) and β_{ij} (i =0,1,2,..., p; j 0,1,2,...,q) are the connection weights and ε_t is the random shock; α_0 and β_{0j} are the bias terms.

2.1.8 Support Vector Regression

Various stochastic and neural network methods for time series modelling and forecasting has been applied last few years. Despite their own strengths and weaknesses, these methods are quite successful in forecasting applications. Recently, a new statistical learning theory, viz. the *Support Vector Regression (SVR)* has been receiving increasing attention for classification and forecasting. SVR was developed by Vapnik and his co-workers at the AT&T Bell laboratories in 1995. Initially SVR's were designed to represent pattern classification problems, such as optimal character recognition, face identification and text classification, etc. But soon they found wide applications in other domains, such as function approximation, regression estimation and time series prediction problems (Chapelle and Vapnik, 1999).

Vapnik (2002) is based on the *Structural Risk Minimization (SRM)* principle. The objective of SVR is to find a decision rule with the good generalization ability through selecting some certain subset of the training data, called support vectors. In this method, an optimal separating hyper plane is constructed, after nonlinearly mapping the input space into a higher dimensional feature space. Thus, the quality and complexity of SVM solution does not depend directly on the input space.

Another important characteristic of SVR is that the training process is equivalent to solving a linearly constrained quadratic programming problem. Therefore, contrary to other networks' training, the SVR solution is always unique and globally optimal. However, a major disadvantage of SVR is that when the training size is large, it requires an enormous amount of computation which increases the time complexity of the solution. Now we are going to present a brief mathematical discussion about the SVR concept.

2.1.9 Hybrid Forecasting Methods

Hybrid forecasting methods bring together regression, data smoothing, and other methods to generate forecasts that can compensate for the weaknesses of individual methods. For instance, several forecasting methods are excellent at short-term forecasting, but cannot obtain seasonality. Hybrid forecasting methods involve Vanguard dampened trend, a robust hybrid model that instantaneously reveals all trends, cycles, and seasonality in historical data and responds with the most accurate exponential smoothing method. Vanguard Dampened Trend is available across all Vanguard business forecasting applications (Luxhøj et al., 1996).

2.1.10 Forecast combinations

It seems apparent that no forecasting technique is suitable for all situations. There is significant indication to demonstrate that combining individual forecasts produces gains in forecasting accuracy. There is also evidence that adding quantitative forecasts to qualitative forecasts reduces accuracy. Research has not yet revealed the conditions or methods for the best possible combinations of forecasts. Judgmental forecasting usually entails combining forecasts from more than one source. Informed forecasting starts with a set of key assumptions and then employs a combination of historical data and expert opinions. Involved forecasting seeks the opinions of all those directly affected by the forecast (e.g., the sales force would be included in the forecasting process). These techniques generally produce higher quality forecasts than can be attained from a single source.

Combining forecasts offer us a way to compensate for inadequacies in a forecasting technique. By selecting complementary methods, the shortcomings of one technique can be offset by the advantages of another. Since the publication of the seminal paper on forecast combination by Bates and Granger in 1969, research in this area has been active. In general, four main reasons for the potential benefits of forecast combinations have been identified:

- It is implausible to be able to accurately model a actual data generation method in only one model. Single models are most likely to be interpretations of a much more intricate reality, so various models might be complementary to each other and be able to approximate the true process better.
- Even if a single best model is available, a lot of specialist knowledge is required in most cases to discover the appropriate functions and parameters. Forecast combinations help to achieve good results without in-depth knowledge about the application and without time-consuming, computationally complex fine-tuning of a single model.
- It is not always feasible to consider all the evidence an individual forecast is based on into account and establish a superior model, because information may be private, unobserved, or provided by a closed source.
- Individual models may have different speeds to adapt to changes in the data generation process. Those changes are difficult to detect in real-time, which is why a combination of forecasts with different abilities to adapt might perform well.

Forecasting combination techniques recommend an alternative approach to single models' forecasts. Bates and Granger (1969) were the first to propose such techniques to improve the forecasting accuracy of individual models (Salerno et al., 2007). Over the last three decades, these techniques have become highly predominant in the forecasting literature. Numerous authors have sketched the reasons behind the prevalence of these techniques. For instance, Timmermann (2006) points out that combined forecasts allow to well aggregate all relevant information gained in different single model forecasts and they are more robust against a misspecification of the data generating process. Brown and Murphy (1996) note that combination forecasts are more likely to improve forecasting performance when each single model forecast being combined is independent of the other (or uncorrelated). Timmermann (2006) also stresses that combination forecasts are particularly useful when structural breaks are present in the data series. Again, each individual model will process differently the structural breaks.

The conclusions that have been reached regarding these combination techniques vary from one paper to another. For example, Winkler and Makridakis (1983) generate forecasts for 1001 economic time series with different types of data and conclude that more complex combination methods slightly outperform the simple average method for long term forecasting horizons. In the air transportation forecasting literature, Chu (1998) provides monthly forecasts of tourist arrivals to Singapore for the year 1988 using a SARIMA and a sine wave regression model. He applies a version of the variance covariance method adapted for seasonal data. Forecasting performance is evaluated using the MAPE.

He finds that the combined forecast is more accurate than the ones issued from ARIMA and sine wave. Shen et al. (2011) use tourist flows from the United Kingdom to seven major touristic destinations to point out that unequal weighing schemes outperform the simple average method. In contrast, Coshall (2009) reviews tourist departures from the United Kingdom to twelve destinations. He concludes that the performance of different combination methods depends on the forecasting horizon. In this case, the variance-covariance method outperforms simple averaging for one and two years ahead forecasts while the reverse is true for three years ahead forecasts. Finally, Wong et al. (2007) study tourist arrivals to Hong Kong. They find that forecasting performance depends on the number of single model forecasts being combined. Thus, they mention that the best performance is likely to be achieved by combining two or three single model forecasts at most.

2.1.11 Linear Forecast Combination

The linear combination of forecasts computes a combined forecast $\hat{y}^c$ as the weighted sum of m individual forecasts $\hat{y}_1$,$\hat{y}_2$,.., $\hat{y}_m$ as shown below:

$$\hat{y}^c = \sum_{i=1}^{m} \omega_i \hat{y}_i \qquad (2.4)$$

Weights can be estimated in various ways. One easy and often remarkably robust example is the simple average combination with equal weights. A variance-based approach first mentioned by Bates and Granger in (1969) and further extended by Newbold and Granger in 1974 uses the average of the sum of the past squared forecast errors (MSE) over a certain period. Granger and Ramanathan (1984) propose the regression method and treat individual forecasts as regressors in an ordinary least squares regression including a constant. In a rank-based approach, according to Bunn (1975), each combination weight is expressed as the likelihood that the corresponding forecast is going to outperform the others, based on the number of times where it performed best in the past. Gupta and Wilton (1987) additionally consider the relative performance of other models using a matrix with pairwise odd ratios. Each element of the matrix represents the probability that the model of the corresponding line will outperform the model on the corresponding column.

2.1.12 Nonlinear Forecast Combination

Potentially nonlinear relationships among forecasts are not considered in linear forecast combination, providing the main argument for usage of nonlinear combination methods. The most examined nonlinear methods for forecast combination are backpropagation feedforward neural networks, where individual forecasts are input data and the combined forecast is obtained as the output. This method was first mentioned by (Shi et al., 1999). Fuzzy systems for forecast combination can be found following two different paradigms. First, fuzzy systems can be observed as a kind of regime model where two or more different forecasting models can be active at one time. Second, the resulting fuzzy system almost always outperforms or draws level with the individual forecasts and linear forecast combination methods. A self-organizing algorithm based on the Group Method of Data Handling (GMDH) technique presented by Xu (2002) was first proposed by Ivakhnenko (1970).

Individual forecasts are taken as an input variable for the combination algorithm, different transfer functions, usually polynomials, then create intermediate model candidates for the first layer. Iteratively, the best models are selected with an external criterion and used as input variables for the next layer, producing more complex model candidates until the best model is found. Several authors favour the approach of pooling forecasts before combining them. By grouping similar forecasts and subsequently combining the pooled forecasts, several issues like increased weight estimation errors because of a high number of forecasts to combine can be addressed. Research in this area recently started with clustering forecasts based on their recent past’s error variance in and continued with investigations by Riedel and Gabrys (2005) on how to extend and modify the clustering criteria in the context of a big pool of individual forecasts that have been diversified by different methods. The treelike structures of these multi-level and multi-step forecast combinations can be evolved with genetic programming, using the quality of the combined predictions on the validation data as the fitness function to optimize.

2.1.13 Forecast Combination Methods

Simple Average Combination Method

The SA combination method calculates composite forecasts by taking the arithmetic average of individual forecasts. Clemen (1989) conclude that the virtues of this method include impartiality, robustness, and a good track record in economic and business forecasting. It is thus a common choice in forecast combination studies and serves as a useful benchmark. The method can be expressed as,

$$f_{ct} = \sum_{i=1}^{n} \frac{f_{it}}{n} \quad (2.5)$$

where f_{ct} denotes the combined forecast, f_{it} is the i^{th} forecast in time t, and n is the number of forecasts to be combined.

Trimmed Mean Method

In the trimmed average, individual forecasts are combined with a simple arithmetic mean, excluding the worst performing k% of the models. Usually, the value of k is selected from the range of 10 to 30. This method is sensible only when $n \geq 3$ (Luh and Guo, 1999).

Winsorized Average Method

In the Winsorized average, the i^{th} smallest and largest forecasts are selected and set to the $(i + 1)^{th}$ smallest and largest forecasts, respectively (Shete et al., 2004).

Variance–Covariance Method

Bates and Granger (1969) introduce the VACO method. In the two-model forecast combination case, the combined forecasts are given as

$$f_{ct} = wf_1t + (1 - w)\ w\ f_2t \quad (2.6)$$

where f_{ct} is the combination forecast based on the individual forecasts of f_1t and f_2t, and w and (1-w) are the weights assigned to f_1t and f_2t respectively. The weight that minimises the combined forecast variance is

$$w^* = (\sigma_{22}^2 - \sigma_{12}) + \sigma_{22}^2 + \ \sigma_{11}^2\ \sigma_{22}^2 - 2\sigma_{12} \quad (2.7)$$

σ_{11}^2 and σ_{22}^2 are unconditional individual forecast errors, and T is the sample size. According to Fritz et al. (1984), the foregoing formula can be easily extended to include more than two individual forecasts, and the weights can be calculated by

$$W_i = \frac{[\sum_{t=1}^{T} e_{jt}{}^2]}{\sum_{j=1}^{m} [\sum_{t=1}^{T} e_{it}{}^2]} \quad (2.8)$$

Granger and Ramanthan Regression Method.

The regression method developed by Granger and Ramanthan (1984) proceeds by regressing actual values on competing for individual forecasts and a constant term, and then employing least squares parameter estimates to produce a combination forecast:

$$f_{ct} = \hat{\beta}_0 + \hat{\beta}_1 f_{1t} + \hat{\beta}_2 f_{2t} + \cdots + \hat{\beta}_n f_{nt} \quad (2.9)$$

where f_{ct} represents a combined forecast based on a linear combination of k individual forecasts, f_{it} (i = 1,2,. . .,n), and $\hat{\beta}_i$ (i = 0,1,. . .,n) denotes the least squares estimator based on observations up to time t-1, y_{t-1}, that is, the actual values at period t-1. The series of

y_{t-1}, is regressed against the individual forecasts, $f_{i,\ t-1}$. (i = 1,2, . . ., n), and a constant term to determine $\hat{\beta}_i$ (i = 0,1, . . ., n).

Discounted Mean Square Forecast Error Method.

The discounted MSFE method was first proposed by Bates and Granger (1969) for a two-individual- forecast case and subsequently generalised by Newbold and Granger (1974) for an n-individual-forecast combination. The method makes use of the full sample, but weights recent observations more heavily (Diebold and Lopez, 1996). The combination of an n-individual forecasts for period t is given as

$$f_{ct} = \sum_{i=1}^{n} \omega_i f_{it} \tag{2.10}$$

where f_{it} is the forecast for period t from forecasting method i, w_i is the weight assigned to individual forecast f_{it}, and n is the number of individual forecasts.

Shrinkage Method.

Clemen and Winkler (1986), Diebold and Pauly (1990) employed Bayesian shrinkage techniques to allow the incorporation of varying degrees of prior information into the estimation of combination weights. In this shrinkage method, the least squares weights and arithmetic mean emerge as the two extreme cases for the posterior mean. The actual posterior mean combination weights are a matrix-weighted average of those for the two extreme cases. The exact location depends on prior precision, which can be estimated from the data using an empirical Bayesian procedure. Such procedures, which employ shrinkage towards a measure of central tendency (e.g., the arithmetic mean), are increasingly playing a role in forecast combinations.

Although the combination weights are coaxed towards the arithmetic mean, the data are still allowed to speak when they have something to say. The shrinkage method computes the weights as an average of the recursive ordinary least squares estimator of the weights based on the GR method and equal weighting, that is,

$$W_{it} = \lambda\hat{\beta}_{it} + (1-\lambda)(1/n) \tag{2.11}$$

where $^\wedge\beta_{it}$ is the i^{th} estimated coefficient from a recursive ordinary least squares regression, and $\lambda = \max\{0; 1\text{-}k/n(T\text{-}1\text{-}n)\}$, where k is a constant that controls the amount of shrinkage towards equal weighting and k takes a value between 0 and 1. A larger k corresponds to more shrinkage towards equal weighting.

Time-Varying-Parameter Combination Method with the Kalman Filter.

This method utilises the Kalman filter algorithm to estimate the coefficients in the combined regression, which are assumed to follow a random walk process. It has been used by Sessions and Chatterjee (1989), LeSage and Magura (1992), and Stock and Watson (2004).

The TVP combination method begins with the GR regression model with time varying parameters,

$$f_{ct} = \beta_{0t} + \beta_{1t}f_{1t} + \beta_{2t}f_{2t} + \cdots + \beta_{kt}f_{kt} + e_t$$
$$\text{and } \beta_{it} = \beta_{it-1} + \eta_{it} \qquad (2.12)$$

where η_{it} is independent and identically distributed and is uncorrelated with e_t. The Kalman filter approach also facilitates real-time parameter 'updating' and can readily handle both stationary (e.g., autoregressive moving average) and non-stationary (e.g., integrated autoregressive moving average) parameter drifts (Diebold and Lopez, 1996).

Outperformance Method

This method was proposed by Bunn (1975). The weights are the probabilities assessed and revised in a Bayesian manner. Each individual weight is interpreted as the probability that its respective forecast will perform the best (in the smallest absolute error sense) on the next occasion. Each probability is estimated as the fraction of occurrences in which its respective forecasting model has performed the best in the past.

Optimal Method

This pivotal method for combining forecasts was proposed by Bates and Granger (1969). The weights are determined to minimize the combined forecast error variance. Diebold and Lopez (1996) refer to this method as the "variance-covariance" method since the weights are achieved using the covariance matrix of forecast errors. Granger and Ramanathan (1984) demonstrated that the method is equivalent to a least squares regression in which the constant is suppressed, and the weights are constrained to sum to one. This approach involves the covariance matrix of forecast errors to be accurately estimated. In practice, this matrix is often not stationary, in which case it is estimated based on a short history of forecasts and thus the method becomes an *adaptive* approach to combining forecasts.

Optimal (*adaptive*) with independence assumption

The covariance matrix of forecast errors is restricted to be diagonal, comprising just the individual forecast error variances (Bunn 1985).

Optimal (*adaptive*) with restricted weights

As well as the diagonal restriction, individual weights are restricted not to be outside the interval [0,1] (Newbold and Granger 1974).

Regression

The combined forecast is obtained via ordinary least squares (OLS) regression with the inclusion of a constant (Granger and Ramanathan 1984).

Regression with restricted weights

A least-squares regression with the inclusion of a constant is performed, but the weights are constrained to sum to one (Holden *et al.* 1990).

2.2 Time Series

Time series are any univariate or multivariate quantitative data collected over time either by private or government agencies. Common uses of time series data include: 1) modelling the relationships between various time series; 2) forecasting the underlying behavior of the data; and 3) forecasting what effect changes in one variable may have on the future behavior of another variable (Chatfield, 2000).

Time-series are a structured way to represent data. Visually, it's a curve that evolves over time. A time-series is a list of dates, each date being an associated with a value (a number). For example, the daily sales of a product can be represented as a time-series. Forecasting time-series mean that we extend the historical values into the future where measurements are not available yet. Forecasting is typically performed to optimize areas such as inventory levels, production capacity or staffing levels (Palit and Popovic, 2006).

There are two main structural variables that define a time-series forecast:

- The period which represents the aggregation level. The most common periods are month, week, and day in the supply chain (for inventory optimization). Call centres typically rely on the quarter-hour period (for staffing optimization).
- The horizon which represents the number of periods ahead that need to be forecasted. In supply chain, the horizon is typically equal or greater to the lead time (Palit and Popovic, 2006, Timmermann, 2006).

There are several key notions that we should be cognizant of when explaining time series data. These attributes will enlighten how we pre-process the data and select the appropriate modelling technique and parameters. Ultimately, the purpose is to simplify the patterns in the historical data by removing known sources of variation and making the patterns more consistent across the entire datasets. Simpler patterns will generally lead to more accurate forecasts.(Skamarock and Klemp, 2008, Song and Chissom, 1993b)

- Trend: A trend exists when there is a long-term increase or decrease in the data.
- Seasonality: A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week.
- Autocorrelation: Refers to the phenomena whereby values of Y at time t are impacted by previous values of Y at t-i. To find the proper lag structure and the nature of auto correlated values in your data, use the autocorrelation function plot.

- Stationary: A time series is said to be stationary if there is no systematic trend, no systematic change in variance, and if strictly periodic variations or seasonality do not exist

2.3 Fuzzy Time Series

Univariate or multivariate quantitative data collected over time in the past decades represented a time series. Researchers have made significant progress in dealing with time series analysis. Traditional time series methods appeared ineffective in some situations and fuzzy time series performed enormously well. Zadeh first proposed the fuzzy set theory (Bonissone, 1980, Zadeh, 1976, Dubois, 1980) to deal with uncertainty using linguistic terms. Song and Chissom (1993) successfully introduced the fuzzy set concept in time series analysis to propose the fuzzy time series. Chen (1996) improved fuzzy time series forecasting method by introducing simple arithmetic operations.

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$ A fuzzy set in the universe of discourse U can be represented as follows:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \cdots + f_A(u_n)/u_n \qquad (2.13)$$

where f_A denotes the membership function of the fuzzy set A, $f_A: U \rightarrow [0, 1]$ and $f_A(u_i)$, $(1 \leqslant i \leqslant n)$, denotes the degree of membership of u_i in the fuzzy set A and $f_A(u_i) \in [0, 1]$.

2.4 Forecast Performance Measures

The accuracy of the forecast is the degree of familiarity of the statement of the quantity of that quantity's genuine value. The actual value generally cannot be measured at the time the forecast is produced because the statement concerns the future. For most businesses, more accurate forecasts increase their effectiveness to serve the demand while lowering overall operational costs.

To apply a certain model in a real or simulated time series, first the raw data are split up into two parts, viz. the Training Set and Test Set. The observations in the training set are used for constructing the desired model. Often a small subpart of the training set is kept for validation purpose and is known as the Validation Set. Sometimes a pre-processing is done by normalizing the data or taking logarithmic or other transforms (Granger and Pesaran, 2004, Baldwin and Kain, 2006, Cassar, 2014). Once a model is constructed, it is used for generating forecasts. The test set observations are kept for verifying how accurate the fitted model performed in forecasting these values. If necessary, an inverse transformation is applied on the forecasted values to convert them in original scale.

To judge the forecasting accuracy of a model or for evaluating and comparing different models, their relative performance on the test dataset is considered. Due to the fundamental importance of time series forecasting in many practical situations, proper care should be taken while selecting a model. For this reason, various performance measures are proposed in literature (Cassar, 2014) to estimate forecast accuracy and to compare different models. These are also known as performance metrics (Baldwin and Kain, 2006). Each of these measures is a function of the actual and forecasted values of the time series.

Various Forecast Performance Measures

The commonly used performance measures and their important properties are listed below:

2.4.1 Mean Absolute Error (MAE)

The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| \qquad (2.14)$$

As the name suggests, the mean absolute error is an average of the absolute errors, where f_i is the prediction and y_i the true value. Note that alternative formulations may include relative frequencies as weight factors.(Hyndman and Koehler, 2006)

The properties of MAE are:

- It measures the average absolute deviation of forecasted values from original ones.
- It is also termed as the *Mean Absolute Deviation (MAD)*.
- It shows the magnitude of overall error, occurred due to forecasting.
- In MAE, the effects of positive and negative errors do not cancel out.
- MAE does not provide any idea about the direction of errors.
- For a good forecast, the obtained MAE should be as small as possible.
- MAE also depends on the scale of measurement and data transformations.
- Extreme forecast errors are not panelised by MAE.

2.4.2 Mean Absolute Percentage Error (MAPE)

The **mean absolute percentage error** (MAPE), also known as a mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

This measure is given by (Hyndman and Koehler, 2006)

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right| \times 100 \qquad (2.15)$$

The important features are:

- This measure represents the percentage of average absolute error occurred.
- It is independent of the scale of measurement but affected by data transformation.
- It does not show the direction of error.
- MAPE does not penalize extreme deviations.
- In this measure, opposite signed errors do not offset each other.

2.4.3 Mean Squared Error (MSE)

The mean squared error (MSE) of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.

The mathematical definition of this measure is (Hyndman and Koehler, 2006)

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \qquad (2.16)$$

The properties are:

- It is a measure of the average squared deviation of forecasted values.
- As here the opposite signed errors do not offset one another, MSE gives an overall idea of the error occurred during forecasting.
- It penalizes extreme errors occurred while forecasting.
- MSE emphasizes the fact that the total forecast error is in fact much affected by large individual errors, i.e. large errors are much expensive than small errors.
- MSE does not provide any idea about the direction of overall error.
- MSE is sensitive to the change of scale and data transformations.
- Although MSE is a good measure of overall forecast error, but it is not as intuitive and easily interpretable as the other measures discussed before.

2.4.4 Sum of Squared Error (SSE)

The SSE calculates the sum of the squared errors of the prediction function. It is mathematically defined as (Hyndman and Koehler, 2006)

$$\text{SSE} = \frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x_\iota})^2 \qquad (2.17)$$

Where x_t is the actual observation time series and $\hat{x_\iota}$ is the forecasted time series.

The properties of MPE are:

- It measures the total squared deviation of forecasted observations, from the actual values.

- The properties of SSE are same as those of MSE.

2.4.5 Root Mean Squared Error (RMSE)

The square root of the mean of the square of all the error. The use of RMSE is very common and it makes an excellent general-purpose error metric for numerical predictions. Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors. It is mathematically defined as (Hyndman and Koehler, 2006)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \qquad (2.18)$$

The properties of RMSE are:

- RMSE is nothing but the square root of calculated MSE.
- All the properties of MSE hold for RMSE as well.

2.5 Research gap

From the literature review the following research gaps are identified-

- **Individual Forecasting Method vs. Hybrid Forecasting Method-**
  Forecasts are seldom accurate. A lot of individual forecasting model has been employed for forecasting purpose in the last couple of decades. A forecasting method that is suitable for one domain might not be appropriate for another domain. Therefore, the procedure of picking a finest forecasting method in advance is not possible in most cases. Rather than focusing on making forecasting by a single method, combining distinct models can be brought into account to establish forecasts. Both theoretical and empirical outcomes imply that hybrid methods can be an effective and efficient way to improve forecasts. Moreover, hybrid model leads to improve the forecasting accuracy and performance. In forecasting research, several combining schemes have been recommended. In 2010, Wedding and Cios defined a combining methodology based on radial basis function and the Box–Jenkins method. Luxhoj (1996) presented a hybrid econometric and an ANN approach for sales forecasting. In 1998, Zhang and Hann proposed a model to combine several feed-forward neural networks to improve time series forecasting accuracy.
- **Accuracy Effected by Data Interval Length and Forecasting Rules Content-** There are two main factors affecting the forecast accuracy, these are the length of an interval in datasets and the content of forecast rules. The length of the interval is required for partitioning the universe of discourse has a significant impact on the forecasting results. Huarng (2002) proposed mean-based and distribution-based methods to determine the length of interval. Egrioglu et al. (2006) calculated length

of intervals in first order and high order models by using single variable constrained optimization. Huarng and Yu (2006) proposed a method where the length of the interval is not fixed and is exponentially increased with a ratio. Moreover, the universe of discourse has been partitioned based on dynamic length of interval instead of fixed length interval. Kuo et al. (2009) applied differential evolution algorithm to define dynamic interval lengths. Uslu et al. (2009) proposed an approach based on weights formed chronologically. Moreover, there are some findings in the literature that have applied fuzzy clustering techniques in fuzzification stage. In this research, an automatic clustering algorithm is exploited for partitioning datasets rather than considering fixed length interval. Furthermore, a new hybrid forecast model has been proposed based on the fuzzy time series, particle swarm optimization and automatic clustering technique by pondering the two aforementioned factors. The role of the model is to obtain the appropriate content of the two mentioned factors to enhance forecasts accuracy.

- **Model Selection in Combination Forecasting Model-**

  Forecast combination is currently observed as a handy tool in rational forecasting. Combine forecasts are expected to be effective when there is uncertainty and for that the best forecasting method needs to identify. This may be because of encountering a new situation, have a heterogeneous set of time series or expect the future to be especially turbulent. Despite a large literature, it was not obvious a priori, which method would be more accurate. Meade and Islam (Meade and Islam, 1998) compared a selection rule (picking the best-fitting model) against a combined forecast. Using seven forecasting methods on 47 data sets, they found that the combined forecast was more accurate than the best fitting model for 77% of the forecasts. Irrelevant or inadequate models may turn up to have little weight or no weight in the combination, and their inadequacy becomes apparent. The question is now whether looking at those weights, that are typically obtained through auxiliary least-squares based regressions, is informative enough or not. For instance, it can be happened that some weights are negative when forecasts are all on the same side of the true data points. The hold-out sample may also not be large enough to find significant weights. Even a small weight in the forecast combination can be enough to establish better forecast performance. Another model selection strategy is related to the notion of encompassing. Therefore, a model is selected in the final combination if the combination with that model yields more forecast accuracy than a combination without that model.

  In this research, decision tree algorithm will be used for appropriate model selection to combine the weights of the individual forecasting model. A decision tree of forecast method was developed to illustrate the selection of forecast method; it resumes selecting the best forecasting method, according to the time series data pattern. Forecast methods decision tree helps to pre-select alternative methods to forecast future demand. It is necessary determine the error measure for each one, and to choose the one that best fits the data, that is, the one that have the lowest forecast error.

- **The issue of weights in combining forecasts-**

  A widespread concern with combining forecasts is the question of how-to best weight the components, and many scholars have proposed methods for doing so. In fact, the simple average (i.e., assigning equal weights to components) was found to often provide more accurate forecasts than complex approaches to estimating "optimal" combining procedures (Clemen, 1989). Empirical research has repeatedly confirmed these findings. The sophisticated methods included combinations based on principal components, trimmed mean, optimal least squared estimates, and Bayesian shrinkage. The performance of these methods varied over time, across target variables, and across time horizons.

  Simple averages of all available forecasts provided more accurate predictions than sophisticated combination methods, which relied heavily on historical performance for weighing the component forecasts. One reason for the strong performance of equal weights is that the accuracy of the component forecasts varies over time and strongly depends on external effects. Smith and Wallis (2009) studied this question by conducting a Monte Carlo simulation of combinations of two forecasts, and reappraising a published study on different combinations of multiple forecasts of US output growth and concluded that the simple average will be more accurate than estimated "optimal" weights if two conditions are met: (1) the combination is based on a large number of individual forecasts and (2) the optimal weights are close to equality. The reason is that, in such a situation, each forecast has a small weight, and the simple average provides an efficient trade-off against the error that arises from the estimation of weights.

Chapter 3

Hybrid Forecasting Model

3 Hybrid Forecasting Model

3.1 Fuzzy Time Series

In the past decades, researchers have made some progress in dealing with time series analysis. There were some situations where traditional time series methods appeared ineffective and fuzzy mathematics worked tremendously better. Zadeh first proposed the fuzzy set theory (Bonissone, 1980, Zadeh, 1976, Dubois, 1980) to deal with uncertainty using linguistic terms. Song and Chissom (1993) successfully introduced the fuzzy set concept in time series analysis to propose the fuzzy time series. Chen (1996) improved fuzzy time series forecasting method by introducing simple arithmetic operations.

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots., u_n\}$. A fuzzy set in the universe of discourse U can be represented as follows:

$$A = f_{A}(u_1)/u_1 + f_{A}(u_2)/u_2 + \cdots + f_{A}(u_n)/u_n \tag{3.1}$$

where f_A denotes the membership function of the fuzzy set A, f_A: $U \to [0, 1]$ and $f_A(u_i)$, $(1 \leqslant i \leqslant n)$, denotes the degree of membership of u_i in the fuzzy set A and $f_A(u_i) \in [0, 1]$.

From the literature of Song and Chissom (1993), Chen (2002) and Chen and Chung, (2006) the definitions of the fuzzy time series has defined as follows.

Definition 1

Let $Y(t)$ (t = 0, 1, 2, 3, …), be the universe of discourse and a subset of real number by which fuzzy sets $f_i(t)$ are defined. Assume $F(t)$ is a collection of $f_1(t)$, $f_2(t)$, …, then $F(t)$ is called the fuzzy time series definition of $Y(t)$.

Definition 2

Assume $F(t)$ and $F(t-1)$ are fuzzy sets denoted as $F(t-1) \to F(t)$, then the fuzzy logical relationships can be expressed as $F(t) = F(t-1) \circ R(t, t-1)$, where "∘" represents an max min composition operator and $R(t-1, t)$ is the fuzzy relationship between $F(t-1)$ and $F(t)$. Moreover, $F(t)$ said to be occurring by $F(t-1)$ and $F(t-1)$, $F(t)$ refer to the current state and the next state of fuzzy time series, respectively.

Definition 3

Let $R(t-1, t)$ be the first order model of $F(t)$. Assume for time t, $R(t, t-1) = R(t-1, t-2)$, then $F(t)$ is mentioned as time-invariant fuzzy time series otherwise mentioned as time-variant fuzzy time series.

Definition 4

Fuzzy logical relationships can be grouped together according to the same current state of the fuzzy logical relationships. Two examples are illustrated as follows:

- Three first-order Fuzzy logical relationships with the same current state A_i and different next state are $A_i \rightarrow A_j$, $A_i \rightarrow A_k$ and $A_i \rightarrow A_l$, respectively. The first order fuzzy logical relationships can be grouped (G_r) together and represented as follows:

$$G_r : A_i \rightarrow A_j, A_k, A_m \tag{3.2}$$

where *r* is a group label of the fuzzy relationships.

- Two third-order fuzzy logical relationships with the same current state "A_i, A_j, A_k" are $A_i, A_j, A_k \rightarrow A_m$ and $A_i, A_j, A_k \rightarrow A_n$, respectively. The third order fuzzy logical relationships can be grouped (G_r) together and represented as follows:

$$G_r : A_i, A_j, A_k \rightarrow A_m, A_n, \tag{3.3}$$

where *r* is a group label of the fuzzy relationships.

3.2 An automatic clustering algorithm

In this section, an automatic clustering algorithm, (Chen et al., 2008) has been mentioned to cluster the historical enrolment data of University of Alabama into different length of intervals. The steps of the algorithm are described as follows:

Step 1: Numerical Data has sorted in an ascending sequence of n different numerical data. Assume that the set without duplicate data in an ascending data sequence can be shown as follows:

$$d_1,\ d_2,\ d_3,\ \ldots,\ d_i,\ \ldots,\ d_n.$$

From the ascending data sequence, we can calculate the average difference value of the data*:*

$$ave\_dif = \frac{\sum_{i=1}^{n-1} d_{i+1} - d_i}{n-1}, \tag{3.4}$$

where "*ave_dif*" represented the average of the differences between each pair of data in the ascending data sequence.

Step 2: The smallest datum in the ascending data sequence has set into the current cluster. Determine the appropriate place for the numerical datum in the ascending data sequence by following the current cluster using the value of "*ave_dif*". The numerical datum following the datum in the current cluster can be put into the current cluster or needs to be put into a new cluster are measured based on the following principles:

Principle 1: Let, there is only one datum in the current cluster, and it is the first cluster and assumes that d_2 is the adjacent datum of d_1, shown as follows:

$$\{d_1\},\ d_2,\ d_3,\ \dots,\ d_n.$$

If the difference between d_2 and d_1 is less than the average difference (d_2-$d_1 \leqslant ave\_dif$), then set d_2 into the current cluster. Otherwise, generate a new cluster for d_2 and determine it as the current cluster.

Principle 2: Let, there is only one datum d_j in the current cluster and it is not the first cluster. Assume that d_k is the adjacent datum of d_j and d_i is the largest datum in the cluster, which is the antecedent cluster of the current cluster, shown as follows:

$$\{d_1,\ \dots\},\ \dots,\ \{\dots,\ d_i\},\quad ,d_k,\dots,d_n.$$

If the difference between d_k and d_j is less than the average difference (d_k-$d_j \leqslant ave\_dif$) and also less than the difference between d_j and d_i (d_k-d_j<d_j-d_i), then set d_k into the current cluster which d_j belongs to. Otherwise, a new cluster needs to be generated for d_k and assume the new generated cluster to be the current cluster which d_k belongs to.

Principle 3: Let, there is more than one datum in the current cluster and the current cluster is not the first cluster. Let d_i is the largest datum in the current cluster and assume that d_j is the adjacent datum next to d_i, shown as follows:

$$\{d_1,\ \dots\},\ \dots,\ \{\dots\},\ \{\dots,\ d_i\},\ d_j,\ \dots,\ d_n.$$

If the difference between d_j and d_k is less than the average difference (d_j-$d_i \leqslant ave\_dif$) and the difference between d_j and d_i is less than the cluster difference (d_j-$d_i \leqslant clu\_dif$), then set d_j into the current cluster which d_i belongs to. Otherwise, a new cluster needs to be generated for d_j and let the new generated cluster that d_j belongs to be the current cluster, where "*clu_dif*" denotes the average difference of the distances between every pair of adjacent data in the cluster and the value of *clu_dif* is calculated as follows:

$$clu\_dif = \frac{\sum_{i=1}^{n-1}(C_{i+1} - C_i)}{n-1}, \qquad (3.5)$$

Step 3: According to the results of **Step 2,** the contents of these clusters can be adjusted by using the following principles:

Principle 1: If a cluster consists of more than two data, then we retain the smallest datum and largest datum and remove the others.

Principle 2: If a cluster consists of exactly two data, then leave it unchanged.

Principle 3: If a cluster has only one datum d_q, then the difference between d_q and ave_dif values *(i.e.* d_q*-ave_dif)* and the summation between d_q and ave_dif values (i.e. d_q*+ave_dif)* set into the cluster and remove d_q from this cluster. In terms of the following situation, the cluster needs to be adjusted again:

Situation 1: In the first cluster if the situation occurs, then the value of "d_q-*ave_diff*" needs to remove instead of d_q from this cluster.

Situation 2: In the last cluster if the situation occurs, then the value of "d_q+*ave_dif*" needs to be removed instead of d_q from this cluster.

Situation 3: If the value of "d_q-*ave_dif*" is smaller than the smallest value in its antecedent cluster, then undo all the action in **Principle 3**.

Step 4: The clustering results obtained in **Step 3** are assumed as follows:

$\{d_1,d_2\},\{d_3,d_4\},\{d_5,d_6\},\ldots,$(McAfee et al.)$,\{d_s,d_t\},\ldots,\{d_{n-1},d_n\}$.

By the following sub-steps, transform these clusters into contiguous intervals:

Step 4.1: The first cluster $\{d_1,\ d_2\}$ transformed into the interval $[d_1,\ d_2)$.

Step 4.2: If the current interval is $[d_i,d_j)$ and the current cluster is $\{d_k,d_l\}$, then

(1) If d_j is greater than equal to d_k, (i.e. $d_j \geqslant d_k$) then transform the current cluster $\{d_k,d_l\}$ into the interval $[d_k,d_l)$. Let $[d_k,d_l)$ be the current interval and let the next cluster$\{d_m,d_n\}$ be the current cluster.

(2) If d_j is less than d_k, (i.e. $d_j<d_k$) then transform $\{d_k,d_l\}$ into the interval $[d_k,d_l)$ and create a new interval $[d_j,d_k)$ between $[d_i,d_j)$ and $[d_k,d_l)$. Let $[d_k,d_l)$ be the current interval and let the next cluster $\{d_m,d_n\}$ be the current cluster. If the current interval is $[d_i,d_j)$ and the current cluster is $\{d_k\}$, then transform the current interval $[d_i,d_j)$ into $[d_i,d_k)$. Let $[d_i,d_k)$ be the current interval and let the next cluster be the current cluster.

Step 4.3: The current interval and the current cluster need to check repeatedly until all the clusters have been transformed into intervals.

Step 5: Add each three of the intervals obtained in **Step 4** in a new interval and continue this until the last interval come.

3.3 Particle swarm optimization

Particle swarm optimization (PSO) is an optimization approach introduced by (Eberhart and Shi, 2001, Shi and Eberhart, 2001, Kennedy et al., 2001) that can effectively search optimal or near optimal solution of any kind of optimization problems (Poli et al., 2007, Feng et al., 2006, Chen et al., 2007). The PSO contains a swarm of particles like the behaviour of animal such as bird flocking, fish schooling that explore the space of possible solutions to an optimization problem. Particles are initialized randomly and then allowed to fly in the virtual searing space for an optimization problem. Each particle calculates its own fitness and neighbouring particle fitness at the time of optimization. Any Particles can remember its own best position as well as the candidate's position it has been passed so far when moves to another position. At each optimization step, a moving particle (*id)* adjusts its candidate position according to following equations:

$$V_{id} = \omega \times V_{id} + c_1 \times \text{Rand}() \times (P_{id} - X_{id}) + c_2 \times \text{Rand}() \times (P_{gd} - X_{id}) \tag{3.6}$$

$$X_{id} = X_{id} + V_{id} \tag{3.7}$$

where V_{id} *denotes the velocity of the particle id*, ω denotes the inertia weight factor; c_1 and c_2 are acceleration values which represent the self-confidence coefficient and the social confidence coefficient, respectively. The value of ω linearly decreased during the moving process and the c_1 and c_2 are constants in a standard PSO. The symbol X_{id} is the current position of the particle, P_{id} is the personal best position of the particle with best fitness value; P_{gd} is the best one of all personal best positions of all particles that experience a global best fitness value; *d* denotes the dimension of the problem space; *Rand*( ) denotes a function that can generate random real numbers in the range of (0, 1). V_{id} is limited to [-V_{max}, V_{min}] where V_{max} is a constant which determines the resolution of searching regions between the present position and the target position (Chen and Chung, 2006). The standard PSO is described in the following algorithm.

Algorithm 1. The Standard PSO Algorithm

1. randomly initialize positions and velocity of all particles
2. **while** stop condition (the optimal solution is found, or maximum iterations are attained) is not reached **do**
3. **for** each particle *id* **do**
4. evaluate the fitness
5. update local best position and global best position
6. adapt velocity using Eq. (3.6)
7. update the position using Eq. (3.7)
8. **end for**
9. **end while**

Based on the literature review and information collected from the existing forecasting model, a hybrid forecasting model has been proposed and the approaches described below in figure 3.4 with a relevant flowchart.

3.4 Flow Chart of proposed hybrid forecasting model

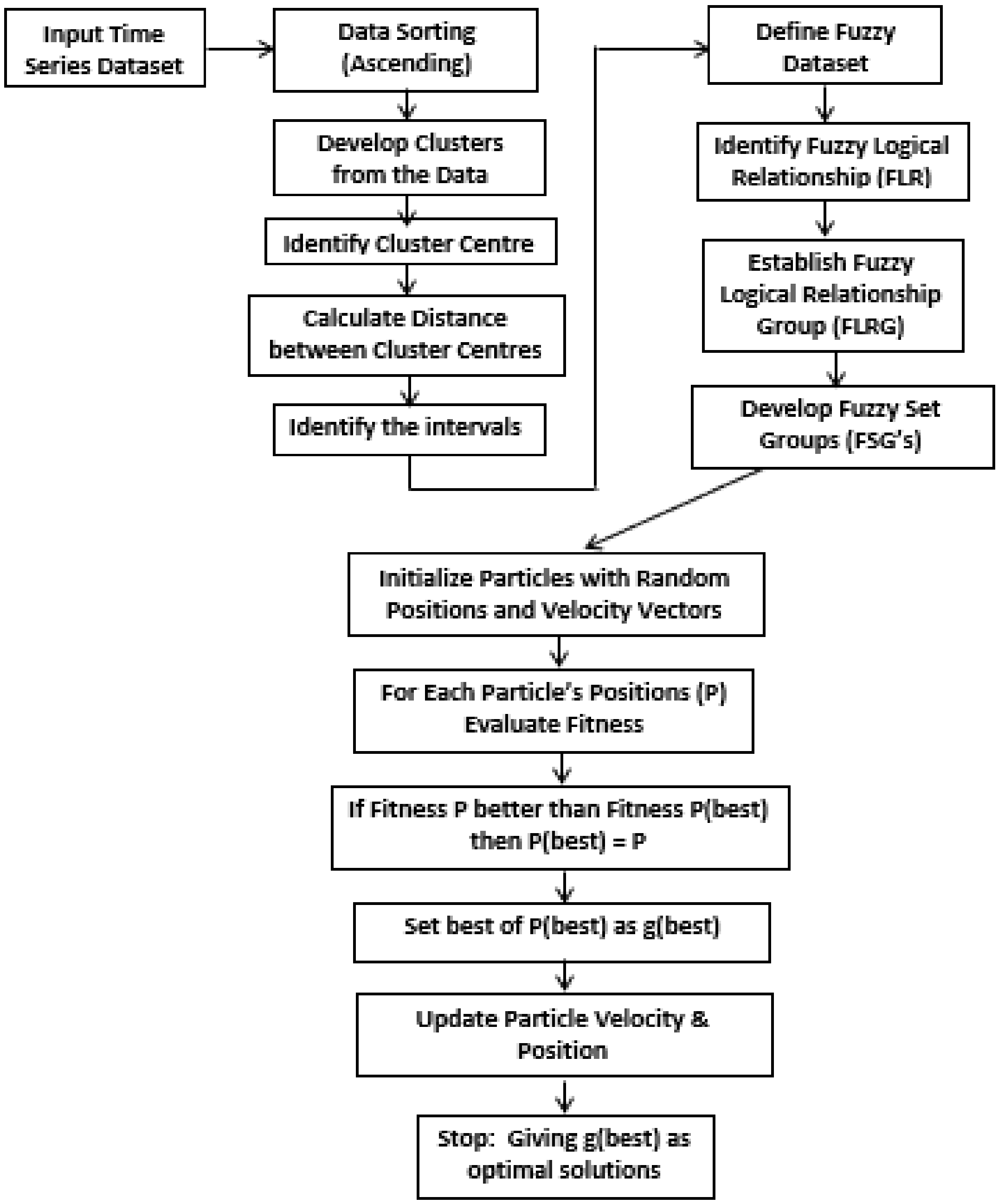

Figure 3.1 Flow chart of the proposed hybrid forecasting model

3.5 Fuzzy time series in forecasting model

In this section, a brief overview of fuzzy time series was introduced to forecast the enrollments of the University of Alabama. Historical enrollments of the University of

Alabama are listed in Table 3.1 (Chen and Chung, 2006). The step by step procedure of the proposed model using fuzzy time series is explained as follows:

Step 1: Define the universe of discourse

Let $Y(t)$ be the historical data on enrollments of the University of Alabama at year t ($1971 \leqslant t \leqslant 1993$). The universe of discourse is defined as $U = [U_{min} - D_{min}, U_{max} + D_{max}]$, where U_{min} and U_{max} the minimum and the maximum enrollment of $Y(t)$, respectively. D_{min} and D_{max} are two positive integer values used to tune the lower bound and upper bound of the U. According to historical data shown in Table 3.1, attained $U_{min} = 13{,}055$ and $U_{max} = 19{,}337$ at year 1971 and 1992, respectively. For getting appropriate intervals, set $D_{min} = 55$ and $D_{max} = 663$ and get the universe of discourse on $U = [13{,}000, 20{,}000]$.

Table 3.1 Historical enrollment of University of Alabama.

Year	**Actual Enrolments**
1971	13 055
1972	13 563
1973	13 867
1974	14 696
1975	15 460
1976	15 311
1977	15 603
1978	15 861
1979	16 807
1980	16 919
1981	16 388
1982	15 433
1983	15 497
1984	15 145
1985	15 163
1986	15 984
1987	16 859
1988	18 150
1989	18 970
1990	19 328
1991	19 337
1992	18 876

Step 2: Partition of U into appropriate intervals by using an automatic clustering technique

Phase 1 From the historical data shown in Table 1 in an ascending sequence and the following sorted results can be obtained:

13055, 13563, 13867, 14696, 15145, 15163, 15311, 15433, 15460, 15497, 15603, 15861, 15984, 16388, 16807, 16859, 16919, 18150, 18876, 18970, 19328, 19337

If there are repeated numerical data, just consider the data only once in the sorted sequence. Based on equation (4), the value of *ave_dif* can be calculated as,

ave_dif=[(13563-13055)+(13867-13563)+(14696-13867)+(15145-14696)+(15163-15145)+(15331-15163)+(15433-15331)+(15460-15433)+(15497-15433)+(15603-15497)+(15861-15331)+(15984-15861)+(16388-15984)+(16807-16388)+(16859-16807)+(16919-16859)+(18150-16919)+(18876-18150)+(18970-18876)+(19328-18970)+(19337-19328)]/21=6282/21=299.

Phase 2 By following the *ave_dif* value and the three principles of **Phase 2**, the clustering results from the ascending data sequence are as follows:

{13055}, {13563}, {13867}, {14696}, {15145, 15163}, {15311, 15433, 15460, 15497}, {15603}, {15861, 15984}, {16388}, {16807, 16859}, {16919}, {18150}, {18876, 18970}, {19328, 19337}.

Phase 3 By performing the three principles of **Phase 3**, the clustering results from Step 2 can be found in the following form:

{13055,13354},{13264,13862},{13568,14166},{14397,14995}, {15145,15163}, {15331,15497}, {15603},{15861,15984}, {16089,16687}, {16807,16859},{16919}, {17851,18449}, {18876,18970}, {19328,19337}.

Phase 4 By following and performing the sub-steps of **Phase 4**, the following intervals can be observed:

u1=[13055,13354), u2=[13354,13862), u3=[13862,14166), u4=[14166,14397), u5=[14397,14995), u6=[14995,15145), u7=[15145,15163), u8=[15163,15331), u9=[15331,15603), u10=[15603,15861), u11=[15861,15984), u12=[15984,16089), u13=[16089,16687), u14=[16687,16807), u15=[16807,16919), u16=[16919,17851), u17=[17851,18449), u18=[18449,18876), u19=[18876,18970), u20=[18970,19328), u21=[19328,19337].

Phase 5 Finally, the intervals by considering the range are as follows:

u1=[13055,14166), u2=[14166,15145), u3=[15145,15603), u4=[15603,16089), u5=[16089,16919), u6=[16919,18876), u7=[18876,19337)

The intervals generation process from the clusters of the historical enrollments of the University of Alabama is shown in Table 3.2

Table 3.2 Clustering Historical Enrollment Datasets of University of Alabama.

Clusters	**Data**	**Lower bound(bs_t)**	**Upper bound(be_t)**	**Middle value(m_t)**
$Cluster_1$	{13055, 13563, 13867}	13055	14166	13610.5
$Cluster_2$	{14696, 15145}	14166	15145	14655.5
$Cluster_3$	{15163, 15311, 15433, 15460, 15497, 15603}	15145	15603	15374
$Cluster_4$	{15861, 15984}	15603	16089	15843
$Cluster_5$	{16388, 16807, 16859, 16919}	16089	16919	16504
$Cluster_6$	{18150, 18876}	16919	18876	17897.5
$Cluster_7$	{18970, 19328, 19337}	18876	19337	18672.5

Step 3: Define all Fuzzy set on historical enrollment data

According to the interval in Step 2, seven linguistic variable A_i ($1 \leqslant i \leqslant 7$) of enrollment can be considered for the seven intervals. The linguistic variable values are A_1 = "not many", A_2 = "not too many", A_3 = "many", A_4 = "many many", A_5 = "very many", A_6 = "too many" and A_7 = "too many many" adopted from Song and Chissom (1993b). Fuzzy set can be represented as $A_i = \delta_1/u_1 + \delta_2/u_2 + \delta_3/u_3 + \delta_4/u_4 + \delta_5/u_5 + \delta_6/u_6 + \delta_7/u_7,$ where the symbol '+' denotes union operator, '/' denotes the membership of u_j which belongs to A_i , u_j ($1 \leqslant j \leqslant 7$) is the element of fuzzy sets, δ_j ($1 \leqslant j \leqslant 7$) is the real number ($1 \leqslant \delta_j \leqslant 7$). In other word, fuzzy set is represented in the form $A_i = \{\delta_j/u_j\}$ where $u_j = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$ with a different membership degree $\delta_j = \{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7\}$. Thus, the definitions of all the fuzzy sets are listed as follows:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7$$
$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7$$
$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + 0/u_7$$
$$A_4 = 0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0/u_6 + 0/u_7$$
$$A_5 = 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 + 0/u_7$$

$$A_6 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6 + 0.5/u_7$$
$$A_7 = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0.5/u_6 + 1/u_7$$

To fuzzify all historical data, the usual method is to assign a corresponding linguistic value of each year's enrollment into an equivalent interval. For example, the historical enrollment of year 1972 is 13,563 which falls within (13,055, 14,166], so it belongs to interval u_1. By considering Eq(3.1), the membership degree of the fuzzy set A_6 and A_7 with values $\delta_6 = 1$ and $\delta_7 = 1$ can be found, these are greater than all other fuzzy sets. Therefore, the linguistic value of "too many" and "too many many" are labelled for the fuzzy set A_6 and A_7. Let $Y(t)$ and $F(t)$ two time series data at year t, where $Y(t)$ is actual enrollment and $F(t)$ is the fuzzy set of $Y(t)$. So, all the elements of $Y(t)$ are integer representing actual enrolment and all the elements of $F(t)$ are linguistic value (i.e. fuzzy set) with respect to the corresponding element of $Y(t)$. Table 3 represents the results of fuzzification on enrollments of the University of Alabama. Actual enrollment $Y(t)$ is fuzzified to a corresponding linguistic value of the fuzzy set $F(t)$. For instance, $Y(1988) = 18{,}150$ is converted to $F(1988) = A_6$ with the linguistic value "too many"; $Y(1991) = 19{,}337$ is converted to $F(1991) = A_7$ with the linguistic value "too many many", and so on.

Step 4: Generate all fuzzy relationships

After the creation of fuzzy time series $F(t)$, the fuzzy relationship under different orders can be constructed easily. All linguistic values of the current state and the next state used as training data. The first order fuzzy relationship can be constructed using the pattern as $F(t-1) \rightarrow F(t)$ based on fuzzy time series definition 2, where $F(t-1)$ is the current state and $F(t)$ is the next state, respectively. Fuzzy sets of $F(t-1)$ and $F(t)$ can be found from the corresponding historical enrolment of $Y(t-1)$ and $Y(t)$, i.e. $F(t-1) = A_i$ and $F(t) = A_j$ where $i \leqslant j$ and $1 \leqslant i,\ j \leqslant 7$. Then a fuzzy relationship $A_i \rightarrow A_j$ can be created by replacing the $F(t-1)$ and $F(t)$ with the corresponding linguistic values. For example, $F(1971) \rightarrow F(1972)$ be a fuzzy time series relationship with the fuzzy sets as $F(1971)$ and $F(1972)$. $F(1971) = A_1$ and $F(1972) = A_1$ can be observed according to table 3.3 and a fuzzy relationship $A_1 \rightarrow A_1$ is obtained by replacing the $F(1971)$ and $F(1972)$ with linguistic values of A_1 and A_1, respectively. The first order fuzzy relationships of the historical enrollments from the year 1971 to 1992 are listed in column 4 of table 3.3. To find all λ ($\lambda \geq 2$) order fuzzy relationships, λ consecutive fuzzy sets are necessary in the training phase, these are $F(t-\lambda), F(t-\lambda+1),.., F(t-2), F(t-1) \rightarrow F(t)$, where the pattern "$F(t-\lambda), F(t-\lambda+1), \ldots, F(t-2), F(t-1)$" is called the current state and $F(t)$ is called the next state. Then the λ order fuzzy relationships can be found by replacing the corresponding linguistic values by fuzzy set. For instance, a third-order fuzzy relationship $A_1, A_1, A_1 \rightarrow A_2$, has got as $F(1971), F(1972), F(1973) \rightarrow F(1974)$. From table 3.3, it also obtained that $F(1971) = A_1$, $F(1972) = A_1$, $F(1973) = A_1$ and $F(1974) = A_2$ and by replacing fuzzy sets $F(1971)$, $F(1972)$, $F(1973)$ and $F(1974)$ with linguistic values A_1, A_1, A_1 and A_2 the fuzzy relationships $A_1, A_1, A_1 \rightarrow A_2$ is created. The linguistic value of $F(1993)$ does not exist within the historical data, the symbol '#' is used to denote the unknown next state. As the fuzzy relationships are untrained pattern, it can be used for testing purpose. For example, a three-order relationship is $F(1990), F(1991), F(1992) \rightarrow F(1993)$ where the linguistic values are $F(1990) = A_7$, $F(1991) = A_7$, $F(1992) = A_6$ and $F(1993) = unknown$. Therefore, the fuzzy relationship is expressed as $A_7, A_7, A_6 \rightarrow \#$.

Table 3.3 First order and Third-order fuzzy relationships for enrollment

Year	Actual Enrolments	Fuzzy sets	First Order	Third Order
1971	13 055	A_1		
1972	13 563	A_1	$A_1 \rightarrow A_1$	
1973	13 867	A_1	$A_1 \rightarrow A_1$	
1974	14 696	A_2	$A_1 \rightarrow A_2$	$A_1, A_1, A_1 \rightarrow A_2$
1975	15 460	A_3	$A_2 \rightarrow A_3$	$A_1, A_1, A_2 \rightarrow A_3$
1976	15 311	A_3	$A_3 \rightarrow A_3$	$A_1, A_2, A_3 \rightarrow A_3$
1977	15 603	A_3	$A_3 \rightarrow A_3$	$A_2, A_3, A_3 \rightarrow A_3$
1978	15 861	A_4	$A_3 \rightarrow A_4$	$A_3, A_3, A_3 \rightarrow A_4$
1979	16 807	A_5	$A_4 \rightarrow A_5$	$A_3, A_3, A_4 \rightarrow A_5$
1980	16 919	A_5	$A_5 \rightarrow A_5$	$A_3, A_4, A_5 \rightarrow A_5$
1981	16 388	A_5	$A_5 \rightarrow A_5$	$A_4, A_5, A_5 \rightarrow A_5$
1982	15 433	A_3	$A_5 \rightarrow A_3$	$A_5, A_5, A_5 \rightarrow A_3$
1983	15 497	A_3	$A_3 \rightarrow A_3$	$A_5, A_5, A_3 \rightarrow A_3$
1984	15 145	A_2	$A_3 \rightarrow A_2$	$A_5, A_3, A_3 \rightarrow A_2$
1985	15 163	A_3	$A_2 \rightarrow A_3$	$A_3, A_3, A_2 \rightarrow A_3$
1986	15 984	A_4	$A_3 \rightarrow A_4$	$A_3, A_2, A_3 \rightarrow A_4$
1987	16 859	A_5	$A_4 \rightarrow A_5$	$A_2, A_3, A_4 \rightarrow A_5$
1988	18 150	A_6	$A_5 \rightarrow A_6$	$A_3, A_4, A_5 \rightarrow A_6$
1989	18 970	A_7	$A_6 \rightarrow A_7$	$A_4, A_5, A_6 \rightarrow A_7$
1990	19 328	A_7	$A_7 \rightarrow A_7$	$A_5, A_6, A_7 \rightarrow A_7$
1991	19 337	A_7	$A_7 \rightarrow A_7$	$A_6, A_7, A_7 \rightarrow A_7$
1992	18 876	A_6	$A_7 \rightarrow A_6$	$A_7, A_7, A_7 \rightarrow A_6$
1993		#	$A_6 \rightarrow \#$	$A_7, A_7, A_6 \rightarrow \#$

Step 5: Set all fuzzy relationship groups

Once the fuzzy relationships of time series are identified, all fuzzy relationships with the same current state can be found to form fuzzy relationship groups. To find out all first order and higher order relationship groups, two suitable examples have been considered. Based on the table 3.4, a first order fuzzy relationship group G_1 with the current state A_1, consists of fuzzy relationships listed as follows:

$$G_1: A_1 \rightarrow A_1,\ A_2.$$

Here $A_1 \rightarrow A_1$, $A_1 \rightarrow A_1$ and $A_1 \rightarrow A_2$ at years 1972, 1973 and 1974, respectively. Based on table 3.5, a third-order fuzzy relationships group G_7 with the current state "A_3, A_4, A_5", consists of fuzzy relationships listed as follows:

$$G_7: A_3,\ A_4,\ A_5 \rightarrow A_5,\ A_6.$$

Here A_3, A_4, $A_5 \rightarrow A_5$ and A_3, A_4, $A_5 \rightarrow A_6$ at years 1980 and 1988, respectively. Table 3.5 shows the three order fuzzy relationships with 19 groups in the training phase and group G_{19} is represented as A_7, A_7, $A_6 \rightarrow$ # which contains the unknown linguistic value of the next state at year 1993. The forecasted value of the year 1993 is decided in testing phase.

Table 3.4 First-order fuzzy relationship groups for enrollment

Group label	**Fuzzy relationships**		
1	$A_1 \rightarrow A_1$	$A_1 \rightarrow A_2$	
2	$A_2 \rightarrow A_3$		
3	$A_3 \rightarrow A_3$	$A_3 \rightarrow A_2$	$A_3 \rightarrow A_4$
4	$A_4 \rightarrow A_5$		
5	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_6$	$A_5 \rightarrow A_3$
6	$A_6 \rightarrow A_7$		
7	$A_7 \rightarrow A_7$	$A_7 \rightarrow A_6$	

Table 3.5 Third-order fuzzy relationship groups for enrollment

Group label	**Fuzzy relationships**
G_1	$A_1, A_1, A_1 \rightarrow A_2$
G_2	$A_1, A_1, A_2 \rightarrow A_3$
G_3	$A_1, A_2, A_3 \rightarrow A_3$
G_4	$A_2, A_3, A_3 \rightarrow A_3$
G_5	$A_3, A_3, A_3 \rightarrow A_3,$
G_6	$A_3, A_3, A_4 \rightarrow A_5$
G_7	$A_3, A_4, A_5 \rightarrow A_5, A_6$
G_8	$A_4, A_5, A_5 \rightarrow A_5$
G_9	$A_5, A_5, A_5 \rightarrow A_3$
G_{10}	$A_5, A_5, A_3 \rightarrow A_3$
G_{11}	$A_5, A_3, A_3 \rightarrow A_2$
G_{12}	$A_3, A_3, A_2 \rightarrow A_3$
G_{13}	$A_3, A_2, A_3 \rightarrow A_4$
G_{14}	$A_2, A_3, A_4 \rightarrow A_5$
G_{15}	$A_4, A_5, A_6 \rightarrow A_7$
G_{16}	$A_5, A_6, A_7 \rightarrow A_7$

G_{17}	$A_6, A_7, A_7 \rightarrow A_7$
G_{18}	$A_7, A_7, A_7 \rightarrow A_6$
G_{19}	$A_7, A_7, A_6 \rightarrow \#$

Step 6: Calculate the forecasting values

Forecasting accuracy can be improved by introducing two new terms like, global information of fuzzy relationships and local information of current fuzzy fluctuation (LFF). The global information of fuzzy relationships with the local information of current fuzzy fluctuation combined to calculate the predicted value. Eq. (3.8) represented the forecasted value of enrollments with two weighted parts *Glob_info* and *Local_info*, respectively, for each of the groups of the forecasted value where w_1 and w_2 are adaptive weights for global information of the fuzzy relationships and local information of the LFF. The forecasted value of enrollments can be represented as:

$$Forecasted\_value = w_1 \times Glob\_info + w_2 \times Local\_info \quad (3.8)$$

where $w_1 + w_2 = 1$ and assume that w_1 and w_2 are equally weighted as $0 \leqslant w_1, w_2 \leqslant 1$. On the basis of the (Chen and Chung, 2006) defuzzification method, the defuzzified value for each fuzzy relationship can be calculated through the midpoint of the next state. In Eq. (3.8), the *Glob_info* represents the global information decided by the fuzzy groups created in Step 5. The midpoint m_t of each interval u_t can be calculated by applying seven intervals in Step 2 and can be represented as follows: $m_t = (bs_t - be_t)/2$, where $1 \leqslant t \leqslant 7$ and u_t is bounded within $(bs_t, be_t]$. Therefore, the midpoints are

m_1 = 13,610, m_2 = 14,655.5, m_3 = 15,374, m_4 = 15,843, m_5 = 16,504,
m_6 = 17,879.5 and m_7 = 18,672.5.

For more than one fuzzy relationship exists in a fuzzy relationship group, the value of Glob_info is the average of the respective midpoints of all intervals with respect to all linguistic values in the next states of all fuzzy relations. Assuming a first-order fuzzy relationship group is $A_{t-1} \rightarrow A_{t1}, A_{t2}, \ldots, A_{tk}$, and the midpoints of linguistic values $A_{t1}, A_{t2}, \ldots, A_{tk}$, are $m_{t1}, m_{t2}, \ldots, m_{tk}$, respectively. Then, the value of *Glob_info* is calculated as follows:

$$Gobl\_info = \frac{m_{t1} + m_{t2+} \ldots\ldots + m_{tk}}{k} \quad (3.9)$$

In Eq. (3.8) the *Local_info* represents the local information derived by the LFF scheme. The LFF scheme is determined by the next state and the latest past in the current state. Suppose a λ-order fuzzy relationship is $A_{t-\lambda}, A_{t-\lambda+1}, \ldots, A_{t-2}, A_{t-1} \rightarrow A_t$, where $\lambda \geqslant 1$ and $t \geqslant 2$. Latest past in the current state and the next state are represented by A_{t-1} and A_t, respectively. Here, m_{t-1} and m_t are midpoints of the fuzzy intervals u_{t-1} and u_t with respect to A_{t-1} and A_t, where $u_{t-1} = (bs_{t-1}, be_{t-1}]$ and $u_t = (bs_t, be_t]$. The fuzzy difference between A_{t-1} and A_t using m_{t-1} and m_t can be computed by the LFF scheme. Then the fuzzy difference should be normalized by dividing $m_{t-1} + m_t$ and the LFF scheme can be expressed as follows:

$$Local\_info = (bs_t + \frac{be_t - bs_t}{2} \times \frac{m_t - m_{t-1}}{m_t + m_{t-1}}) \qquad (3.10)$$

To find out the forecasted enrollment of the year 1975, the linguistic enrollment of current state at 1974 is A_2. From table 4 it can be found that a fuzzy relationship $A_2 \rightarrow A_3$ in the group G2 appears the same linguistic value of the current state A_2. The fuzzy sets A_2 and A_3 maximum membership values occur at intervals u_2 and u_3, respectively, where $u_2 = (bs_2, be_2]$ and $u_3 = (bs_3, be_3]$.

The obtained values from step 2 are as follows, bs_2 = 14,166, be_2 = 15,145, bs_3 = 15,145 and be_3 = 15,603. The midpoints of the intervals u_2 and u_3 are m_2 = 14,655.5 and m_3 = 15,374, respectively, where m_2 = ½(14,166 + 15,145) and m_3 = ½(15,145 + 15,603). The global information of year 1975 is equal to m_3, that is $Glob\_info$ = 15,374. According to Eq. (3.10), by setting $bs_t = bs_3$, $be_t = be_3$, $m_{t-1} = m_2$, $m_t = m_3$, $u_{t-1} = u_2$ and $u_t = u_3$, the value of the *Local_info* on the enrollment of the year 1975 can be calculated as follows:

$$Local\_info = (bs_3 + \frac{be_3 - bs_3}{2} \times \frac{m_3 - m_2}{m_3 + m_2})$$

$$= 15{,}145 + \frac{15{,}603 - 15{,}145}{2} \times \frac{15{,}374 - 14{,}655.5}{15{,}374 + 14{,}655.5}$$

$$= 15{,}150.4$$

The forecasted value of the year 1975 can be computed, which is 15,217.9 (i.e. 0.5 × 15,374 + 0.5 × 15,150.4) from the obtained values of the *Glob_info* and the *Local_info*. For the forecasted enrollment of year 1982, the current state of the enrollment at year 1981 is A_5 in Table 3 need to be considered. From Table 4, it can be found that three-order fuzzy relationships $A_5 \rightarrow A_5$, $A_5 \rightarrow A_6$ and $A_5 \rightarrow A_3$ in group G_5 appear the same current state A_5. Based on Eq. (3.5), (3.6) and (3.7), the forecasted enrollment of the year 1982 can be calculated as follows:

$$Forecasted\_Value = w_1 \times Global\_info + w_2 \times Local\_info$$

$$= 0.5 \times \frac{m_5 + m_6 + m_3}{3} + 0.5 \times (bs_3 + \frac{be_3 - bs_3}{2} \times \frac{m_3 - m_5}{m_3 + m_5})$$

$= 0.5 \times \frac{16{,}504 + 17{,}897.5 + 15{,}374}{3} + 0.5 \times (15{,}145 + \frac{15{,}603 - 15{,}145}{2} \times \frac{15{,}374 - 16{,}504}{15{,}374 + 16{,}504})$

$= 15864.35$

Now, the higher order fuzzy relationship group and the corresponding midpoint of the linguistic values need to be considered. Suppose a λ-order fuzzy relationship group is $A_{t-\pi}, A_{t-\pi+1}, \ldots, A_{t-1} \rightarrow \#$, and the midpoints of the linguistic values are $m_{t-\pi}, m_{t-\pi+1}, \ldots, m_{t-1}$, respectively. Two voting schemes are introduced by Kao and Chen (Kuo et al., 2010) to deal with untrained testing data. Kuo (Kuo et al., 2010) proposed a master voting (MV) which gives latest past and other past linguistic values in the current state the highest votes and one vote respectively. The MV scheme calculates the forecasted value by using the following formula in Eq. (3.11) where the w_{high} denotes the highest votes that predefined by the user.

$$Forecasted\_value = \frac{m_{t-1} \times w_{high} + m_{t2} + \cdots + m_{t-\lambda+1} + m_{t-\lambda}}{w_{high} + \lambda - 1} \quad (3.11)$$

Decreasing voting scheme (called DV) was proposed by Chen et al. (2008) that considered different votes decreasingly for all linguistic values in the current state. The DV scheme calculates the forecasted value by using the following formula:

$$Forecasted\_value = \frac{m_{t-1} \times \lambda + m_{t2} \times (\lambda-1) + \ldots + m_{t-\lambda+1} + m_{t-\lambda} \times 1}{\lambda + (\lambda-1) + \ldots + 2 + 1} \quad (3.12)$$

To deal with untrained data in the testing phase the voting schemes have been simplified. Assume A_{t-2} and A_{t-1} denote two latest past linguistic values before unknown next state of time t, where m_{t-2} and m_{t-1} are two midpoints of the fuzzy intervals u_{t-2} and u_{t-1} with respect to linguistic values A_{t-2} and A_{t-1}. LFF scheme calculates the fuzzy difference between two consecutive linguistic values of A_{t-2} and A_{t-1} using $m_{t-2} - m_{t-1}$ to obtain the local information of untrained data. Then the fuzzy difference should be normalized by dividing $m_{t-2} + m_{t-1}$. The intervals for the linguistic values A_{t-2} and A_{t-1} are u_{t-2} and u_{t-1}, respectively, where $u_{t-2} = (bs_{t-2}, be_{t-2}]$ and $u_{t-1} = (bs_{t-1}, be_{t-1}]$. The global information and complete LFF scheme for untrained data are formulated as follows:

$$Glob\_info = \mathrm{m}_{\mathrm{t-1}} \quad (3.13)$$

$$Local\_info = (\mathrm{bs}_{\mathrm{t-1}} + \frac{\mathrm{be}_{\mathrm{t-1}} - \mathrm{bs}_{\mathrm{t-2}}}{2} \times \frac{m_{t-1} - m_{t-2}}{m_{t-1} + m_{t-2}}) \quad (3.14)$$

To forecast the enrollment of the year 1993 by using three-order fuzzy relationship, the current state is composed of three linguistic values of years 1990, 1991 and 1992, which are A_7, A_7 and A_6, respectively. After searching the current state in Table 3.5, a fuzzy relationship group G_{19} can be obtained with the unknown next state in the last row, i.e., $A_7, A_7, A_6 \rightarrow \#$. Two latest past linguistic values before the next state are A_7 and A_6 corresponding to A_{t-2} and A_{t-1}, respectively from the table 3.3. The maximum membership values of A_7 and A_6 occur at intervals u_7 and u_6, respectively, where $u_7 = (bs_7, be_7]$ and $u_6 = (bs_6, be_6]$. It also observed that bs_7 = 18,876, be_7 = 19,337, bs_6 = 16,919 and be_6 = 18,876. The midpoints of the intervals u_7 and u_6 are m_7 = 18,672.5 and m_6 = 17,897.5, where m_7 = ½(18,876 + 19,337) and m_6 = ½(16,919 + 18,876). From Eq. (3.13) and Eq. (3.14), the *Glob_info* of year 1993 is equal to 17,897.5 and $m_{t-2} = m_7$, $m_{t-1} = m_6$, $u_{t-2} = u_7$ and $u_{t-1} = u_6$, the value of the *Local_info* on enrollment of the year 1993 can be calculated as follows

$$Local\_info = (\mathrm{bs}_6 + \frac{\mathrm{be}_6 - \mathrm{bs}_6}{2} \times \frac{m_6 - m_7}{m_6 + m_7})$$

$$= 16{,}919 + \frac{18{,}876 - 16{,}919}{2} \times \frac{17{,}897.5 - 18{,}672.5}{17{,}897.5 + 18{,}672.5}$$

$$= 16{,}898.3$$

Now, the forecasted value of year 1993 is 18,244 (i.e., 0.5 × 18,500 + 0.5 × 17,987). The forecasted enrollments of the first-order fuzzy relationships are listed in Table 3.6

Table 3.6 Forecasted enrollment of the first-order fuzzy relationships

Year	**Fuzzy sets**	***Glob_ info***	***Local_ info***	**Forecasted values**
1971	A_1			
1972	A_1	14,133	13,055	13,594
1973	A_1	14,133	13,055	13,594
1974	A_2	14,133	14,184.1	14,158.55
1975	A_3	15,374	15,150.4	15,262.2
1976	A_3	15,290.8	15,145	15,217.9
1977	A_3	15,290.8	15,145	15,217.9
1978	A_3	15,290.8	15,606	15,448.4
1979	A_4	16,504	16,097.4	16,300.8
1980	A_4	16,591.8	16,504	16,547.9
1981	A_4	16,591.8	16,504	16,547.9
1982	A_3	16,591.8	15,137	15,864.3
1983	A_3	16,591.8	14,137	15,364.3

1984	A_3	16,591.8	14,154	15,372.9
1985	A_3	15,374	15,150.4	15,262.2
1986	A_3	16,591.8	15,606	16,098.9
1987	A_4	16,504	16,097.4	16,300.7
1988	A_6	16,591.8	16,958.6	16,775.65
1989	A_6	18672.5	18880.8	18776.6
1990	A_7	18285	19337	18811
1991	A_7	18285	19337	18811
1992	A_6	18285	16,898.1	17,591.55
1993	A_7	18672.5	18880.8	18776.65

Step 7: Fuzzy forecasting rules creation

To find out the fuzzy forecasting rule, the fuzzy relationship groups and relative forecasting values mentioned above has been considered. The basic format for the fuzzy forecast rule represents by the *if-then* statements. The first order fuzzy forecasting rules to forecast the enrollments $Y(t)$ using fuzzy group, just simply find out the corresponding linguistic value of $F(t-1)$ with respect to the data $Y(t-1)$, and then a forecasted value from the forecasting part of the matched forecast rule can be obtained. the fuzzy forecasting rule R_1 as:

$$\textit{if } F(t-1) = A_1 \textit{ then } Y(t) = Glob\_info + Local\_info. \quad (3.15)$$

As mentioned earlier, the *Glob_info* value is determined by fuzzy groups and the *Local_info* value is determined by LFF scheme. The first order fuzzy relationship for enrollment is mentioned in Table 3.7

Table 3.7 First-order fuzzy relationship rules for enrollment

Rules	**Antecedent**	**Consequent**
1	*if* $F(t-1) = A_1$	*then* $Y(t) = 14{,}133 + Local\_info$
2	*if* $F(t-1) = A_2$	*then* $Y(t) = 15{,}374 + Local\_info$
3	*if* $F(t-1) = A_3$	*then* $Y(t) = 15{,}290.5 + Local\_info$
4	*if* $F(t-1) = A_4$	*then* $Y(t) = 16{,}504 + Local\_info$
5	*if* $F(t-1) = A_5$	*then* $Y(t) = 16{,}591.8 + Local\_info$
6	*if* $F(t-1) = A_6$	*then* $Y(t) = 18{,}672.5 + Local\_info$
7	*if* $F(t-1) = A_7$	*then* $Y(t) = 18{,}285 + Local\_info$

Step 8: Forecasted accuracy estimation using MSE values

Forecasted performance of fuzzy time series can be measured by several evaluation criterions like, MSE, SE, RMSE, MPE, MAE etc. The mean square error (MSE) is an effective one to represent the forecasted accuracy. The MSE value is calculated by the following formula:

$$MSE = \frac{\sum_{i=1}^{N} (FD_i - TD_i)^2}{N} \qquad (3.16)$$

where the number of historical data in time series is denoted by N, fv_i and av_i denotes the forecasted value and actual value at time i. The MSE value of the forecasted enrollment from year 1972 to year 1992 is calculated by using the Eq. (3.16) as follows:

$$MSE = \frac{\sum_{i=1}^{N} (FD_i - TD_i)^2}{N} = \frac{\sum_{i=1}^{21} (FD_i - TD_i)^2}{21}$$

$$= \frac{(13594 - 13055)^2 + (13594 - 13563)^2 + \cdots \ldots\ldots + (18776.6 - 18876)^2}{21}$$

$$= 313{,}626$$

3.6 Particle Swarm Optimization in forecasting model

In this article, two essential factors have been addressed which have an influence on fuzzy time series forecasting accuracy; these are the contents of forecasting rules and the effective lengths of intervals. A hybrid forecasting model (MFPSO) has proposed by using fuzzy time series, automatic clustering algorithm and particle swarm optimization (PSO) to adjust the length of the interval in the training phase and minimize the MSE value. After all the training data has been well trained by the PSO method based on fuzzy forecast rules and LFF scheme, the intervals with minimum MSE value has used to forecast in the testing phase. Assume the number of the intervals be n, the lower bound of historical data $Y(t)$ be b_0 and the upper bound of historical data $Y(t)$ be b_n. A vector b *consisting of n-1 elements, i.e.* $b = \{b_1, b_2, \ldots, b_i, \ldots, b_{n-1}\}$ is used in each particle, where $b_1 \leqslant b_i \leqslant b_{n-1}$ and $b_i \leqslant b_{i+1}$. The universe of discourse cut by the vector b into n intervals which are $u_1 = (b_0, b_1]$, $u_2 = (b_1, b_2]$, … , $u_i = (b_{i-1}, b_i]$, … , $u_{n-1} = (b_{n-2}, b_{n-1}]$ and $u_n = (b_{n-1}, b_n]$, respectively. If a particle moves to another position, the elements b_i $(1 \leqslant i \leqslant n-1)$ of the corresponding vector b must be sorted in ascending order.

Each particle in the MFPSO model uses the intervals to create an independent group of fuzzy forecast rules to get the forecasted accuracy for each particle depending on all historical training data. To denote forecasted accuracy of a particle the mean square error (MSE) value defined in Eq. (3.16) is used and if the MSE value of the particle is lower the better the forecasted accuracy is. The MFPSO model moves all the particles to a new position in the training phase according to Eq. (3.6) and (3.7). To evaluate the forecasted accuracy of all the particles the steps mentioned above will be repeated until the predefined stop condition is satisfied or the optimal solution is found. If the stop condition is satisfied, then all fuzzy forecast rules trained by the best one of all personal best positions of all particles are chosen to be the end result. The MFPSO model uses all the trained fuzzy forecast rules to forecast the new testing data in the testing phase. The whole process is mentioned in the following algorithm.

Algorithm MFPSO

1. randomly initialize all particles positions and velocity
2. **while** the ending condition (the optimal solution is found, or the maximal moving steps are reached) is not fulfilled **do**
3.   **for** every particle id **do**
4.   Partition universe of discourse into new intervals by automatic clustering algorithm
5.   fuzzify all historical training data according to all intervals
6.   establish all fuzzy relationships of different order according to all fuzzified training data
7.   create all fuzzy forecast rules depending on all high order fuzzy relationship
8.   calculate forecasting values by step 6
9.   forecast all historical training data according to all fuzzy forecast rules
10.   calculate the MSE value for particle *id*
11.   update the local best position and the global best position according to according to the MSE value
12.   **end for**
13.   **for** all particle *id* **do**
14.   move particle *id* to another position according to velocity (V_{id}) and current position (X_{id})
15.   **end for**
16. **end while**

The hybrid forecasting model uses the PSO to train all fuzzy forecast rules for all historical training data $Y(t)$ i.e. ($1971 \leqslant t \leqslant 1992$), where the lower bound and upper bound of the universe of discourse represented by the symbol b_0 and b_7, by letting the values 13055 and 19337, respectively. The universe of discourse of $Y(t) = (13055, 19337]$. Let the number of particles and the number of intervals is 5 and 7, respectively. From Eq. (3.6) and (3.7), let X_{id} be limited to (13055, 19337], V_{id} be limited to [−100, 100], both C_1 and C_2 be 2, and ω be 1.4 (ω linearly decreases its value to the lower bound, 0.4, through the whole procedure), respectively. The initial positions and the initial velocities of all the particles listed in a table below are selected on a random basis. Each particle defines an independent group of seven intervals represented in Table 3.8, which are $u_1 = (b_0, b_1]$, $u_2 = (b_1, b_2]$, $u_3 = (b_2, b_3]$, $u_4 = (b_3, b_4]$, $u_5 = (b_4, b_5]$, $u_6 = (b_5, b_6]$, $u_7 = (b_6, b_7]$, respectively. So, the intervals of the initial position of particle 1 can be represented as follows:

u_1 = (13 055, 14 166], u_2 = (14 166, 15 145], u_3 = (15 145, 15 603], u_4 = (15 603, 16 089],

u_5 = (16 089, 16 919], u_6 = (16 919, 18 876] and u_7 = (18 876, 19 337].

In particle 1 seven intervals are considered that are identical like the previously used forecasting example in section 2. The randomized initial positions of all particles are represented in Table 3.8

Table 3.8 The randomized initial position of all particles

	b_1	b_2	b_3	b_4	b_5	b_6	**MSE**
Particle 1	14 166	15 145	15 603	16 089	16 919	18 876	313 626
Particle 2	13 582	14 843	14 785	15 920	16 756	18 589	332 127
Particle 3	13 357	14 225	15 010	15 746	16 412	17 982	201 426
Particle 4	13 829	15 124	14 979	15 895	16 843	19 456	642 364
Particle 5	14 063	14 768	15 357	16 054	16 937	18 687	221 833

The forecasting procedure described in Section 2 needs to be followed with respect to the MFPSO Algorithm mentioned above and from the forecasted results in Table 3.6. By considering the formula from Eq. (16), the MSE value for particle 1 is calculated where FD_i ($1 \leqslant i \leqslant 21$) denotes the forecasted data on $Y(1972 + i)$ and TD_i denotes the corresponding historical training data (i.e. $Y(1972 + i)$). The randomized initial velocities of all particles are listed in Table 3.9

Table 3.9 The randomized initial velocities of all particles

	v_1	v_2	v_3	v_4	v_5	v_6
Particle 1	77.692	33.421	22.462	19.653	92.123	15.247
Particle 2	91.453	68.036	75.685	35.942	42.433	76.998
Particle 3	82.129	54.761	16.224	48.534	81.554	4.567
Particle 4	63.875	71.352	39.743	16.765	35.941	35.223
Particle 5	55.748	83.145	6.864	3.975	73.565	41.457

After getting the MSE value for all the particles, each particle needs to update its own personal best position. Initial personal best positions are set as the initial positions of all particles. The initial personal best positions of all particles are represented in Table 3.10.

Then all particles move to the second position by the PSO model according to Eq. (3.6) and (3.7). The second position and the corresponding new MSE values of all the particles are listed in Table 3.11.

Table 3.10 The initial personal best position of all particles

	b_1	b_2	b_3	b_4	b_5	b_6	**MSE**
Particle 1	14 166	15 145	15 603	16 089	16 919	18 876	313 626
Particle 2	13 582	14 843	14 785	15 920	16 756	18 589	332 127
Particle 3	13 357	14 225	15 010	15 746	16 412	17 982	201 426
Particle 4	13 829	15 124	14 979	15 895	16 843	19 456	642 364
Particle 5	14 063	14 768	15 357	16 054	16 937	18 687	221 833

The global best position is created by particle 3 as its MSE is the least among all particles.

Table 3.11 The second position of all particles

	b_1	b_2	b_3	b_4	b_5	b_6	**MSE**
Particle 1	14 066	15 045	15 503	15 989	16 919	18 776	209 471
Particle 2	13 582	14 843	14 785	15 920	16 756	18 589	351 642
Particle 3	13 357	14 225	15 010	15 746	16 412	17 982	189 158
Particle 4	13 829	15 124	14 979	15 895	16 843	19 456	417 951
Particle 5	14 063	14 768	15 357	16 054	16 937	18 687	235 748

Table 3.12 The personal best position of all particles

	b_1	b_2	b_3	b_4	b_5	b_6	**MSE**
Particle 1	14 066	15 045	15 503	15 989	16 919	18 776	209 471
Particle 2	13 482	14 943	14 885	15 820	16 856	18 489	332 127
Particle 3	13 257	14 325	15 090	15 791	16 512	18 054	189 158
Particle 4	13 729	15 024	14 979	15 895	16 743	19 356	417 951
Particle 5	14 063	14 868	15 257	16 154	16 837	18 787	221 833

By considering the datasets from Table 3.10 and Table 4.11 listed above and comparing the MSE values, it is obvious that particle 1, particle 3 and particle 4 reached a better position than their own personal best position so far. In Table 3.12, the three particles update their own personal best positions. The MSE value of the particle 3 represents the least, so the new global best position is created by particle 1.

3.7 Experimental results in training phase

All historical enrollments from year 1971 to 1992 are used as training data set and the experimental results for MFPSO model are compared with the existing models. The MFPSO model is executed 100 runs. The best result of all runs is taken to be the result. The MFPSO model has the following parameters. The number of particles is 30, the maximal movement for each particle is 100, the inertial weight (i.e. ω) value is linearly decreased from 1.4 to 0.4, the self-confidence coefficient (i.e. C_1) and the social-confidence coefficient (i.e. C_2) both are 2, the velocity V_{id} is limited to [−100, 100]. MSE value is considered for evaluating the performance of forecasted accuracy.

To compare the forecasted accuracy of the proposed model under different order and different number of intervals, three hybrid fuzzy time series models are considered. The models are, CC06F (Bruce et al., 2006) model, HPSO (Kuo et al., 2009) model, AFPSO (Huang et al., 2011) model under a different number of intervals and listed in Table 3.13. The MSE value of the proposed model is smaller, comparison to any other model mentioned above. All the models use the Chen's (Bruce et al., 2006) method to create the first order fuzzy forecast rules to forecast the training data. The key difference between CC06F model and the MFPSO model is that the former uses the genetic algorithm, but latter uses particle swarm optimization to get the appropriate intervals, respectively.

From Table 3.13, it is obvious that the PSO algorithm is more powerful than the genetic algorithm in terms of efficiently searching virtual problem space. The difference between HPSO method and MFPSO method is that LFF scheme used in MFPSO can provide better forecasted accuracy than HPSO, though each model uses the PSO method. For the AFPSO model and the MFPSO model, both utilize the PSO method and LFF scheme, however, MFPSO model can provide much better forecasted accuracy as the intervals from the universe of discourse are not in fixed length.

To compare the forecasted accuracy of the proposed model with those of the existing high order models like HCL98 (Hwang et al., 1998), CC06H model (Bruce et al., 2006), HPSO (Kuo et al., 2009) model, AFPSO (Huang et al., 2011) is selected for comparison and listed in Table 5.2. The experimental results illustrate that the proposed model achieves the lowest MSE value and is more precise than any other existing model.

Table 3.13 Forecasted accuracy comparisons among CC06F model, HPSO model and AFPSO model with different number of intervals.

Models		*Number*	*of*	*intervals*			
	8	***9***	***10***	***11***	***12***	***13***	***14***
CC06F (Chen & Chung, 2006a)	132,963	96,244	85,486	55,742	54,248	42,497	35,324
HPSO (Kuo et al. 2009)	119,962	90,527	60,722	49,257	34,709	24,687	22,965
FLK-means (Tinh et al. 2016)	78,950	42,689	37,265	35,647	33,834	21,308	18,770
AFPSO	27,435	24,860	19,698	19,040	16,995	11,589	8224
MFPSO	23,128	19,470	17,356	16,173	14,405	9447	6819

Table 3.14 Forecasted accuracy with different high order models with different intervals where number of intervals=7

Order	**C02 (Chen, 2002)**	**CC06b (Chen 2006a)**	**HPSO(Kuo et al., 2009)**	**FRPSO (Tinh et al., 2017)**	**FRH(Tinh et al., 2019)**	**AFPSO**	**MFPSO**
2	89,093	67,834	67,123	67,104.9	42,650	19,594	19,243
3	86,694	31,123	31,644	31,641	56,65.5	31,189	29,687
4	89,376	32,009	23,271	23,27.8	55,13.8	20,155	17,589
5	94,539	24,948	23,534	23,533.8	36,71.8	20,366	17, 942
6	98,215	26,980	23,671	23,662	31,47.7	22,276	18,765
7	104,056	26,969	20,651	20,645	N/A	18,482	15,836
8	102,179	22,387	17,106	17,090.6	N/A	14,778	12,920
9	102,789	18,734	17,971	17,962	N/A	15,251	13,534
Avg. MSE	95,868	31,373	28,121	28,113.8	N/A	20,261	16,453

The performance of the forecasted enrollments of the proposed model under a different number of intervals with those of the existing first order models like the SC93b model (Bruce et al., 2006), the C96 model (Bruce et al., 2006), the H01H model (Bruce et al., 2006), CC06a model (Bruce et al., 2006) with different number of intervals are listed in Table 3.15. From the proposed method the smallest value obtained 6819**.** The experimental results show that the proposed model performs more precise than existing model in terms of first order fuzzy time series.

Table 3.15 Comparison of the forecasted results of the proposed model with the existing model with first order of the time series under different number of intervals

Year	Actual data	H01H	CC06a	HPSO	FRPSO	FIPSO (2019)	AFPSO	MFPSO
1971	13,055							
1972	13,563	14,000	13,714	13,555	13,715.6	13,469	13,579	13,618
1973	13,867	14,000	13,714	13,994	13,715.6	13,952	13,812	13,784
1974	14,696	14,000	14,880	14,711	14,768.4	14,596	14,565	14,352
1975	15,460	15,500	15,467	15,344	15,330.4	15,439	15,422	15,516
1976	15,311	15,500	15,172	15,411	15,437.1	15,241	15,307	15,255
1977	15,603	16,000	15,467	15,411	15,437.1	15,925	15,618	15,675
1978	15,861	16,000	15,861	15,411	15,437.1	15,880	15,660	15,791
1979	16,807	16,000	16,831	16,816	16,806.4	16,810	16,794	16,722
1980	16,919	17,500	17,106	17,140	16,918.1	17,009	17,032	17,013
1981	16,388	16,000	16,380	16,464	16,416.8	16,260	16,390	16,420
1982	15,433	16,000	15,464	15,505	15,502.8	15,435	15,504	15,480
1983	15,497	16,000	15,172	15,411	15,437.1	15,212	15,431	15,471
1984	15,145	15,500	15,172	15,411	15,437.1	15,282	15,077	15,018
1985	15,163	16,000	15,467	15,344	15,330.4	15,344	15,297	15,145
1986	15,984	16,000	15,467	16,018	16,040	15,714	15,848	15,254
1987	16,859	16,000	16,831	16,816	16,806.4	16,833	16,835	16,902
1988	18,150	17,500	18,055	18,060	18,148.8	18,016	18,145	18,227
1989	18,970	19,000	18,998	19,014	18,943	18,937	18,880	18,794
1990	19,328	19,000	19,300	19,340	19,304.9	19,345	19,418	19,375
1991	19,337	19,500	19,149	19,340	19,304.9	19,147	19,260	18,943
1992	18,876	19,000	19,149	19,014	18,943	19,152	19,031	19,182
	MSE	226,611	35,324	22,965	20,318.3	23,710	8224	6819

To evaluate the performance of the proposed model based on different order and distinct intervals with those of the existing model (i.e. SC94 model (Bruce et al., 2006), HCL98 model (Bruce et al., 2006), S07S model (Bruce et al., 2006), C02 model (Bruce et al., 2006), CC06b (Bruce et al., 2006) model , HPSO (Kuo et al., 2009) model, AFPSO (Huang et al., 2011) are represented in Table 3.16 where the HPSO model, AFPSO model and MFPSO model use 9 order fuzzy relationships and 14 intervals to train the forecasting enrollments. The experimental results show that MFPSO model is more accurate than other existing forecasting model under different number of intervals.

Table 3.16 Comparison of the forecasted results of the proposed model with the existing model with high order models of the time series under different order and different number of intervals

Year	Actual data	HCL98	S07S	C02	CC06b	HPSO	FRH (2019)	AFPSO	MFPSO
1971	13,055								
1972	13,563								
1973	13,867								
1974	14,696	14,500							
1975	15,460	15,361	15,500						
1976	15,311	16,260	15,468	15,500					
1977	15,603	15,511	15,512	15,500					
1978	15,861	16,003	15,582	15,500			15,877		
1979	16,807	16,261	16,500	16,500	16,846		16,836		
1980	16,919	17,407	16,361	16,500	16,846	16,890	16,910	16,920	16,960
1981	16,388	17,119	16,362	16,500	16,420	16,395	16,385	16,388	16,362
1982	15,433	16,188	15,744	15,500	15,462	15,434	15,442	15,467	15,475
1983	15,497	14,833	15,560	15,500	15,462	15,505	15,482	15,472	15,398
1984	15,145	15,497	15,498	15,500	15,153	15,153	15,153	15,158	15,185
1985	15,163	14,745	15,306	15,500	15,153	15,153	15,153	15,159	15,235
1986	15,984	15,163	15,442	15,500	15,977	15,971	15,970	15,976	15,994
1987	16,859	16,384	16,558	16,500	16,846	16,890	16,836	16,858	16,772
1988	18,150	17,659	17,187	18,500	18,133	18,124	18,151	18,142	18,253
1989	18,970	19,150	18,475	18,500	18,910	18,971	18,957	18,974	18,998
1990	19,328	19,770	19,382	19,500	19,334	19,337	19,328	19,338	19,387
1991	19,337	19,928	19,487	19,500	19,334	19,337	19,328	19,335	19,318
1992	18,876	19,537	18,744	18,500	18,910	18,882	18,885	18,882	18,794
	MSE	321,418	133,700	86,694	1101	234	169	173	112

3.8 Experimental results in testing phase

To evaluate the future enrollments forecasted accuracy, the historical enrollments data set can be divided into two separate parts i.e. training part and testing part. In this paper, the historical enrollments data from 1971 to 1989 is used as the training data set and the historical enrollments data of year 1990, 1991, and 1992 are used as the testing data set. To forecast the new enrollment of the next year, historical enrollment data set from the previous year are used. For instance, to forecast the new enrollment of the year 1991, the past years' historical enrollment data set from 1971 to 1990 is used. From Table 3.17 it can be found that the proposed model has a smaller MSE value compared to the model like C96 model (Bruce et al., 2006), HPSO (Kuo et al., 2009) model, AFPSO (Huang et al., 2011) under first order fuzzy time series.

HPSO model employs a master voting scheme (MV) and the highest votes for the MV scheme are assigned 15. Better forecasted accuracy can get by a good voting scheme no matter what the order of the fuzzy time series is. In HPSO voting scheme, the number of votes given to the latest past year directly affects the forecasted accuracy when deals with the untrained trained fuzzy forecast rules for the testing phase. According to high order time series comparison, smaller MSE values can be got by the proposed model from order 2 to 5. From table 3.17, the best MSE value for HPSO model is 98,607 is achieved from 4-order fuzzy time series with seven intervals. Moreover, AFPSO model revealed an MSE value of 90,538 in the same order and the same number of intervals. By considering same order the proposed method obtains the MSE value of 83,586. The proposed method obtained the lowest three MSE values which are 86,674, 83,586 and 88,764 for 3-order, 4-order, and 5-order fuzzy time series, respectively, at the same intervals. In the testing phase, two models (HPSO, AFPSO) have been compared with the proposed model and it is observed that the proposed method produces the smallest MSE value which is 83,586 under 4-order fuzzy time series and seven intervals. To recapitulate, it can be said that, the proposed model performs tremendously well compared to some other models for enrollments.

Table 3.17 Comparison of the forecasted results of C96 model, HPSO(MV) model, AFPSO model with the proposed model for the testing phase (highest vote for the MV scheme =15, the number of intervals=7)

Year	**Actual data**	**Order = 1**			**Order = 2**			**Order = 3**		
		HPSO	AFPSO	MFPSO	HPSO	AFPSO	MFPSO	HPSO	AFPSO	MFPSO
1990	19,328	18,685	18,970	18,852	18,599	18,983	18,892	18,988	18,975	18,869
1991	19,337	19,138	19,433	19,365	19,246	19,142	19,058	19,167	19,156	19,135
1992	18,876	19,176	19,473	19,752	19,246	19,471	19,534	19,265	19,214	19,182
	MSE	181,017	164,596	159,845	230,089	170,358	146,984	98,607	90,538	86,674

Year	**Actual data**	**Order = 4**			**Order = 5**		
		HPSO	AFPSO	MFPSO	HPSO	AFPSO	MFPSO
1990	19,328	18,821	18,982	18,986	18,593	18,971	18,987

1991	19,337	19,040	19,148	19,234	18,886	19,141	19,373
1992	18,876	19,192	19,212	19,248	19,076	19,210	19,248
	MSE	148,371	89,444	83,586	261,209	92,474	88,764

3.9 Analysis Discussion

The proposed hybrid forecasting model (MFPSO) for the historical enrollments of the university of Alabama based on two advanced methods, fuzzy time series and particle swarm optimization. To improve the forecasting accuracy of the proposed model in comparison with HPSO (Kuo et al., 2009) model, AFPSO (Huang et al., 2011), an automatic clustering algorithm is considered for the interval calculation from the universe of discourse and combined the global information of fuzzy relationships with the local information of latest fuzzy fluctuation to get the defuzzified forecasting value. In addition to that, particle swarm optimization is used to adjust the length of each interval in the universe of discourse.

The experimental results of forecasting enrollments of students of the University of Alabama represent that the proposed model obtained higher forecasting accuracy compared to any other existing models. It also performs best for fuzzy time series with various orders in training and testing phases, respectively. In the training phase the minimal MSE value for the proposed model is 112, which is the lowest forecasting error mentioned in Table 3.16. In the testing phase, the minimal MSE value for the proposed model is 83,586, which is the smallest forecasting error mentioned in Table 3.17.

Finally, the proposed forecasting model was tested for the forecasting enrollment problem and the model is effective enough for others practical domains. Figures 3.2 and 3.3 represented the actual and forecasted enrollments of students of the University of Alabama. First-order and high order forecasting of students of the University of Alabama are represented in figures 3.3 and 3.4 that compared the graph of the proposed model to other state-of-the-art forecasting model.

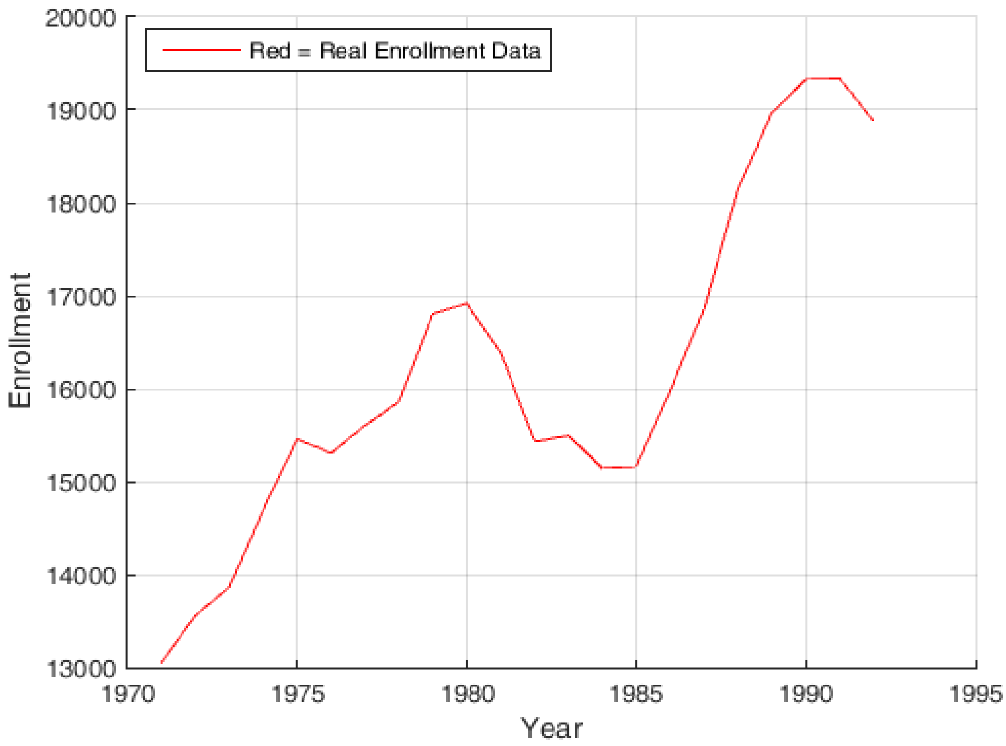

Figure 3.2 Actual Student Enrollment Dataset

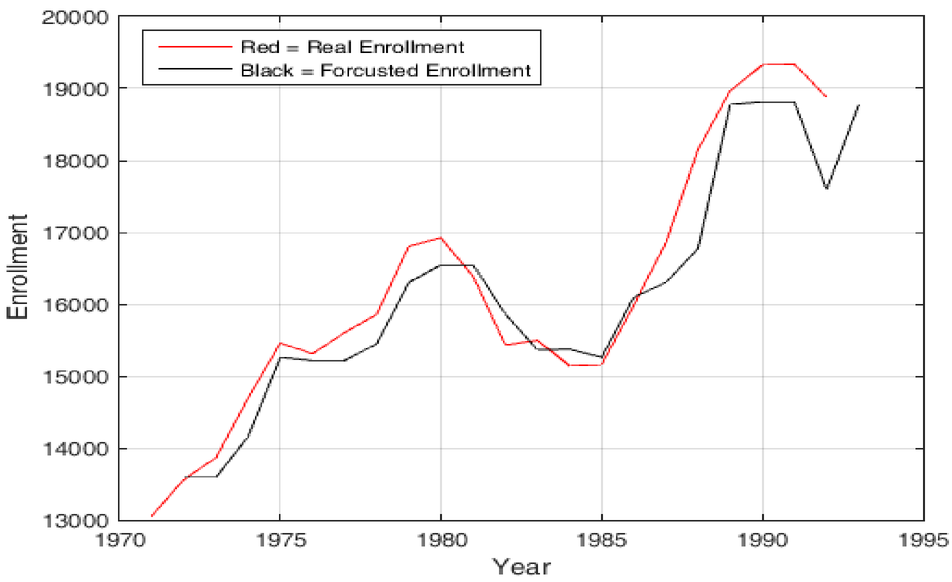

Figure 3.3 Forecasting Student Enrollment Dataset

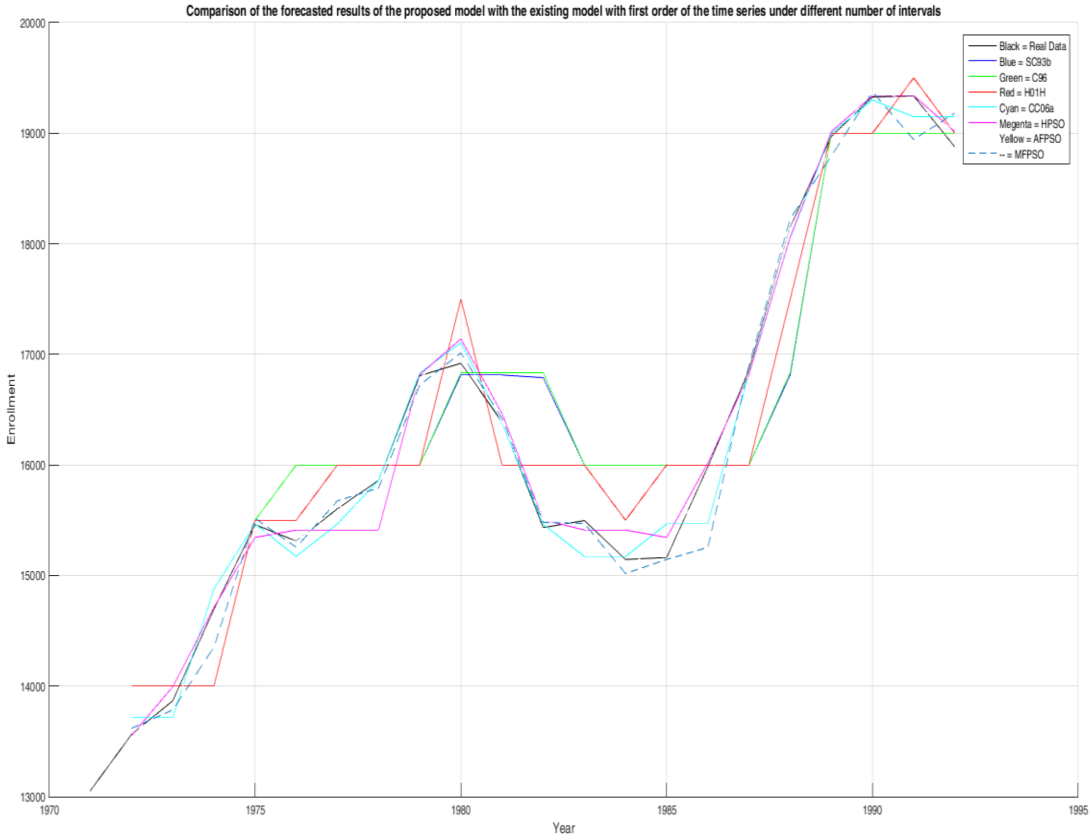

Figure 3.4 Frist-Order Forecasting of Student Enrollment Dataset (MSE= 6819)

Figure 3.5: High Order Forecasting of Student Enrollment Dataset (MSE=112)

3.10 Summary

The main purpose of the research was to explore the state-of-the-art fuzzy time series forecasting methods and to propose a new hybrid forecasting technique. The main findings being very general and evidence frequently suggesting that hybrid forecasting methods based on fuzzy time series can perform better than individual ones. Empirical assessment of forecasting model was performed and characterized using the student enrollment datasets of university of Alabama. The main objective here was to ponder a hybrid forecasting model where an automatic clustering algorithm was utilized to analyze the datasets interval in a more efficient manner. Conventional forecasting practices have some shortcomings due to not dealing with specific forecasting problems where the historical data are symbolized by linguistic values. Fuzzy time series forecasting is employed to overcome that weakness. Moreover, the hybrid forecasting model has been investigated through particle swarm optimization method to attain improved forecasting outcome.

Chapter 4
Combination Forecast

4 Combination Forecasting Model

Based on literature review and information collected from the existing forecasting model, a linear and nonlinear combination forecasting model was proposed in this chapter, and the approaches are described below with a relevant pictorial diagram.

4.1 Forecast combinations

Forecasting method is appropriate for roughly all the circumstances. Research shows the significant impact of the individual forecast can be found by the combination of the models that can produce substantial gain in forecasting accuracy. There is also evidence that adding up quantitative forecasts to qualitative forecasts reduces forecast accuracy. Research has not yet revealed the conditions or methods for the finest possible combinations of forecasts. Judgmental forecasting usually entails combining forecasts from more than one source. Informed forecasting begins with a set of key assumptions and retains a combination of historical data and expert opinions. Moreover, involved forecasting search for the views of all those directly affected by the forecast (e.g., the sales force would be included in the forecasting process). These methods normally produce better quality forecasts than can be attained from a single source.

Forecast combination lead a way to compensate for insufficiencies in a forecasting technique. The effectively selection of the complementary methods, the shortcomings of one technique can be offset by the advantages of another. Since the publication of the seminal paper on forecast combination by Bates and Granger in 1969, research in this area has been active. In general, four key reasons for the prospective advantages of forecast combinations have been discovered:

- The situation appears doubtful to be able to precisely model a real data generation method based on only one model. The single forecasting models are presumably be interpretations of a significantly more intricate reality. Therefore, numerous models might be complementary to each other to be able to estimate the actual method better.
- Since a single finest model is available, lots of professional knowledge is essential to discover the suitable functions and parameters. Forecast combinations assist to attain excellent results without any depth knowledge about the application. Moreover, the time-consuming, computationally complex fine-tuning processes of a single model need to be care about.

- It is not always feasible to consider all the evidence an individual forecast and to establish a superior model, because the information may be private, unobserved, or provided by a closed source.
- In the data generation process, individual models may have different velocities to acclimate the changes. Those changes are difficult to detect in real-time. Therefore, a combination of forecasts with distinct capabilities might perform well.

4.2 Linear Forecast Combination

The linear forecast combination handles a combined forecast $\hat{y}^c$ as the weighted sum of m individual forecasts $\hat{y}_1$,$\hat{y}_2$,.., $\hat{y}_m$ as shown below:

$$\hat{y}^c = \sum_{i=1}^{m} \omega_i \hat{y}_i \qquad (4.1)$$

In various ways, the weights can be estimated and calculated. The easiest and robust example is the simple average combination with identical weights. A variance-based approach first mentioned by Bates and Granger in (1969) and further extended by Newbold and Granger in 1974 uses the average of the sum of the past squared forecast errors (MSE) over a certain period. Granger and Ramanathan (1984) propose the regression method and treat individual forecasts as regressors in an ordinary least squares' regression including a constant. In a rank-based approach, according to Bunn (1975), each combination weight is expressed as the likelihood that the corresponding forecast is going to outperform the others, based on the number of times where it performed best in the past. Gupta and Wilton (1987) additionally consider the relative performance of other models using a matrix with pairwise odd ratios. The elements of the matrix exemplify the probability of the model of the subsequent field, will surpass the model on the subsequent column.

4.3 Proposed Linear Forecast Combination Model

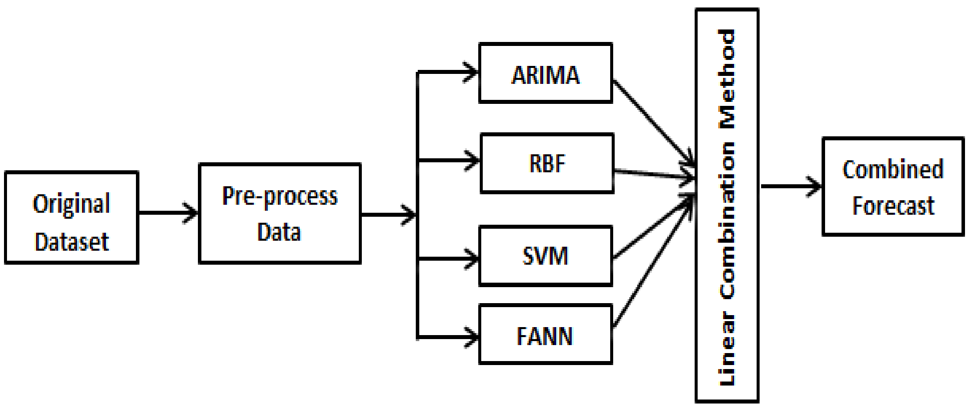

Figure 4.1 Flow chart of the proposed linear combination forecasting method

In figure 4.1, the pictorial diagram represented a linear forecast combination method. The original dataset is pre-processed and several individual forecasting models like ARIMA, RBF. SVM, and FANN (fast artificial neural network) are combined and each combination weight is expressed as the likelihood based on a linear combination technique that outperforms state-of-the-art combination forecasting model.

4.4 Nonlinear Forecast Combination

Theoretically, a linear forecast combination never considers the nonlinear relationships among the forecasts, delivering the key claim regarding the usage of nonlinear combination methods. Backpropagation feedforward neural networks be the most examined nonlinear methods for forecast combination that considers individual forecasts are input data and the combined forecast obtained as the output. This method was first mentioned by (Shi et al., 1999). Fuzzy systems for forecast combination can be found following two different paradigms. First, fuzzy systems can be observed as a kind of regime model where two or more different forecasting models can be active at one time. Second, the resulting fuzzy system almost always outperforms or draws level with the individual forecasts and linear forecast combination methods. In 2002, Xu presents a self-organizing algorithm based on the Group Method of Data Handling (GMDH) technique that was proposed by Ivakhnenko in the 1970s.

In combination algorithm, the individual forecasts are carried as an input variable, different transfer functions, usually polynomials, then create intermediate model candidates for the first layer. The best models are selected iteratively with an external criterion and applied as input variables for the next layer, producing more complex model candidates until the best model is found. Several authors favour the approach of pooling forecasts before combining them. By grouping similar forecasts and subsequently combining the pooled forecasts, several issues like increased weight estimation errors because of a high number of forecasts to combine can be addressed. Research in this area recently started with clustering forecasts based on their recent past’s error variance in and continued with investigations by Riedel and Gabrys (2005) on how to extend and modify the clustering criteria in the context of a big pool of individual forecasts that have been diversified by different methods. The treelike structures of these multi-level and multi-step forecast combinations can be evolved with genetic programming, using the quality of the combined predictions on the validation data as the fitness function to optimize.

4.5 Proposed nonlinear forecast combination method

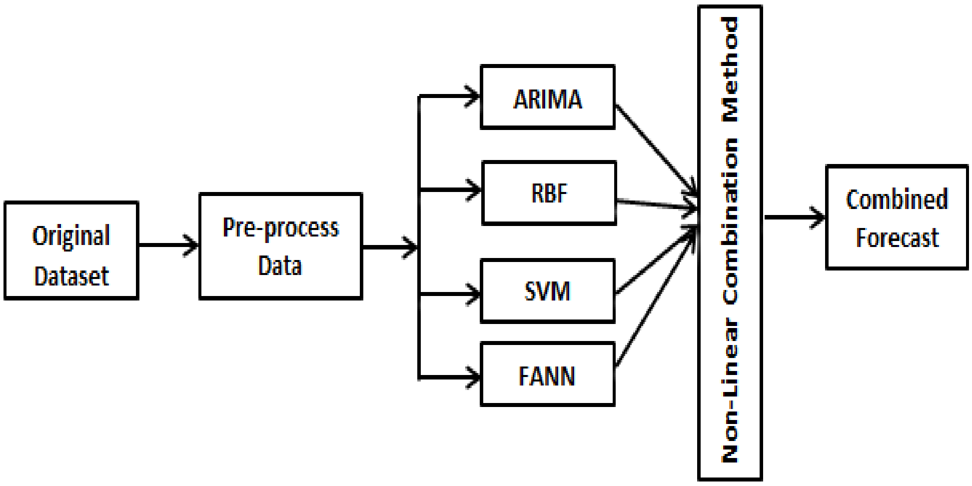

Figure 4.2 Flow chart of the proposed nonlinear combination forecasting method

In figure 4.2, the pictorial diagram represented a non-linear forecast combination method. The original dataset is pre-processed and several individual forecasting models like ARIMA, RBF. SVM, and FANN (fast artificial neural network) are combined and each combination weight is expressed as the likelihood based on a non-linear combination technique that outperforms state-of-the-art combination forecasting model.

4.6 Datasets

For empirical verification of forecasting performances of our proposed ensemble technique, three real-world time series are used in this paper. These are the Canadian lynx, Wolf's sunspots, and the monthly international airline passenger's series. All three series are available in the well-known Time Series Data Library (TSDL). The description of these three-time series is presented in Table 4.1 and their corresponding time plots are shown in figure 4.3 to 4.6. Table 4.1 represented the time series dataset used for evaluating the performance of combination forecast model. Different categories of the time series dataset along with the type, total training, and testing size are represented below.

Table 4.1 Description of the Time Series Datasets

Series	**Type**		**Total Size**	**Testing Size**
Lynx	Stationary, noseasonal	Number of lynx trapped per year in the Mackenzie River district of Northern Canada (1821–1934).	114	14
Sunspots	Stationary,	The annual number of observed	288	67

	noseasonal	sunspots (1700–1987).		
Airline	Monthly seasonal	Monthly number of international airline passengers (in thousands) (January 1949–December 1960).	144	12
River flow	Stationary, noseasonal	River flow of Idaho (1830–1930).	600	100

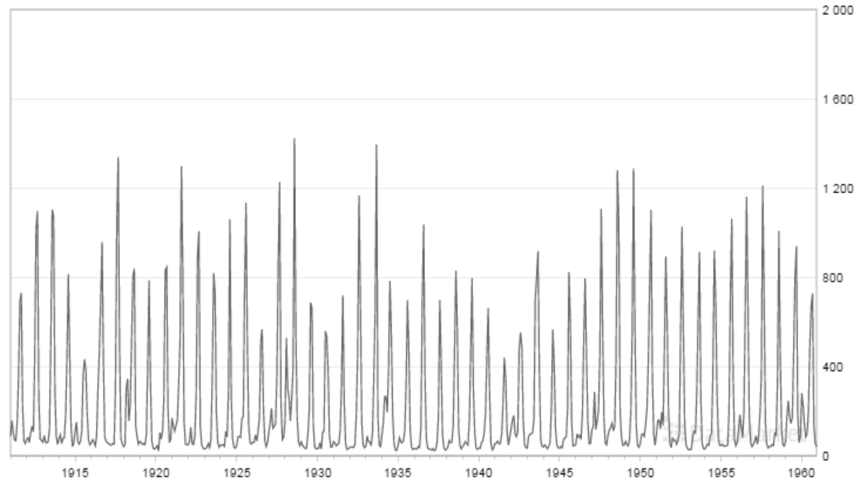

Figure 4.3 Time Plots (Lynx)

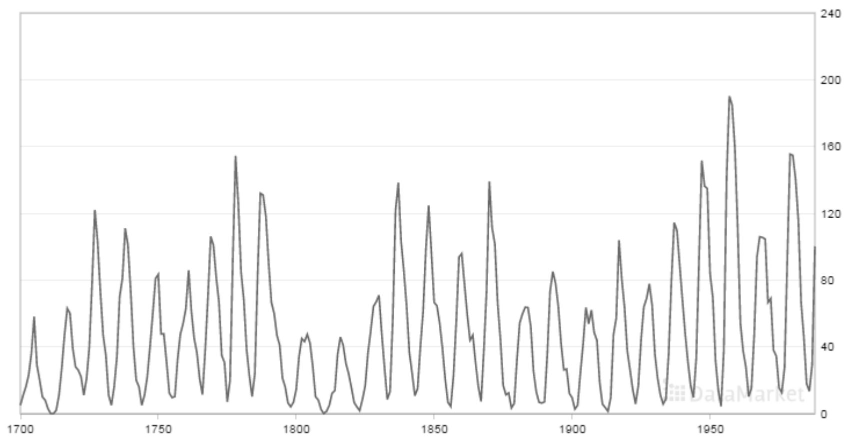

Figure 4.4 Time Plots (Sunspots)

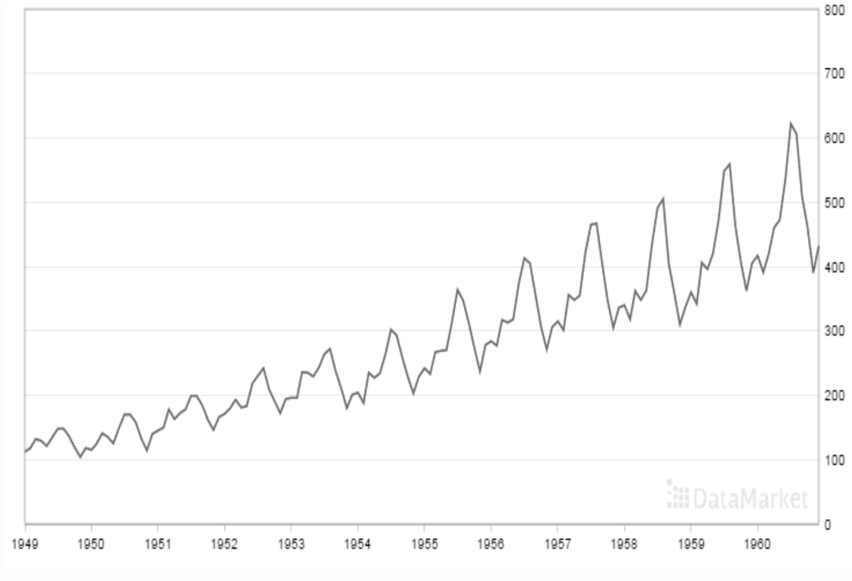

Figure 4.5 Time Plots (Airline)

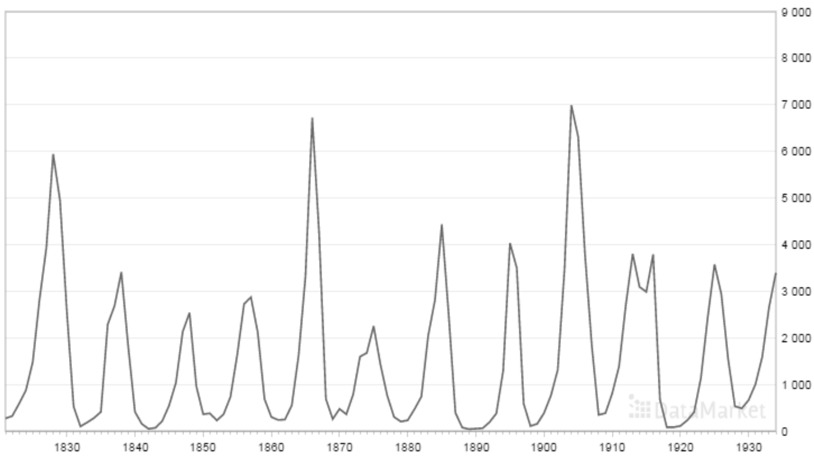

Figure 4.6 Time Plots (River Flow)

4.7 Implementation of Individual Models

Due to lack of individual forecasting model, the forecast combination concept come into action and performed well with greater accuracy result in a real-life dataset. Figure 4.7 to 4.9 depicted the individual forecasting model prediction with the actual dataset based on the model named RBF, ANN, and BPNN. The individual model is combined to get significant improvement in combination forecast model.

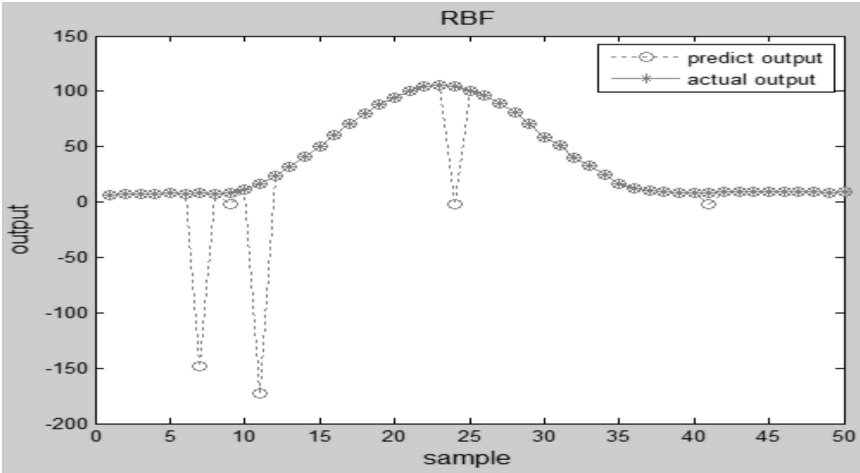

Figure 4.7 Forecasting (RBF)

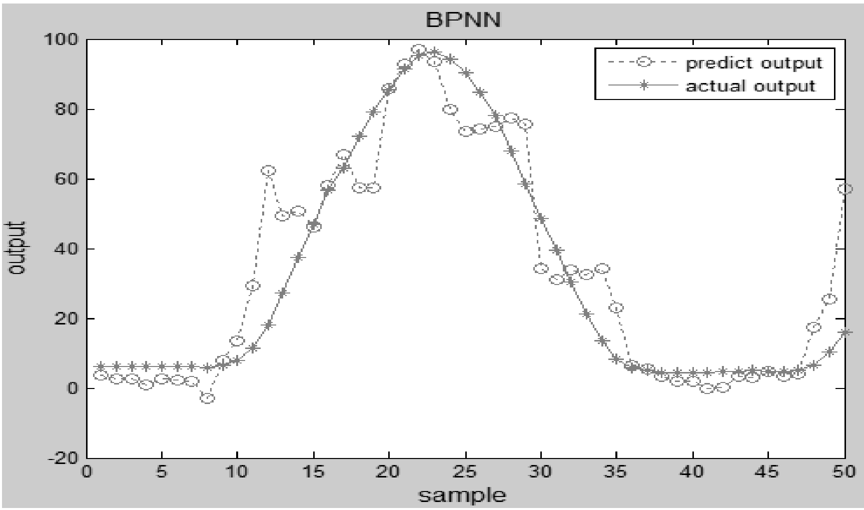

Figure 4.8 Forecasting (BPNN)

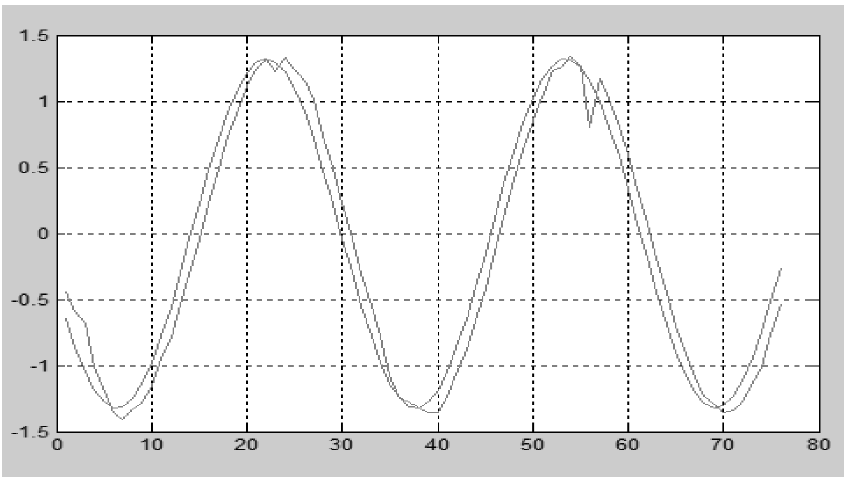

Figure 4.9 Forecasting (ANN)

Table 4.2 represented the comparison results of the various individual forecasting models as well as corresponding forecast combination method. Different forecast combination techniques like average, mean, median, out performance method have applied to give a clear comparison of the models.

Table 4.2 The obtained Forecasting results

Series	Type	Individual Models				Combination Models			
		ARIMA	RBF	SVM	FANN	Avg.	Median	EB	Out_Pf.
Lynx	MAE	0.103	0.173	0.173	0.154	0.112	0.133	0.097	0.107
	MSE	0.015	0.053	0.053	0.032	0.018	0.026	0.013	0.015
Sunspots	MAE	14.91	20.06	20.06	20.06	15.96	14.98	13.78	13.75
	MSE	348.5	630.3	630.3	630.3	371.8	428.4	352.3	375.4
Airline	MAE	12.49	10.49	10.49	16.17	11.63	11.73	10.85	10.85
	MSE	291	176.9	176.9	378	176.5	176.5	157.8	152.5
River Flow	MAE	1.26	0.687	0.687	0.66	0.751	0.676	0.712	0.665
	MSE	2.606	1.172	1.172	1.217	1.158	1.138	1.197	1.068

4.8 Summary

The main purpose of the research was to explore the state-of-the-art individual forecasting methods and to propose a new forecast combination method. The main findings being very general and evidence frequently suggesting that forecast combination methods based on several individual forecasting method can perform better than individual ones. Empirical assessment of forecasting model was performed and characterized using the real-life Canadian datasets with different categories.

Chapter 5
Conclusions
&
Future Work

5 Conclusions

With the profusion of time series forecasting algorithms available, exclusively concentrating on expanding new methods, enhancing current techniques, and conducting countless empirical experiments on diverse datasets does not seem reasonable. The research investigated time series forecasting from a distinct point of view and produced contributions to deliver a hybrid and the combination forecasting model with various datasets that behaved well and provided reasonable predictions in certain circumstances. This chapter presented a synopsis of the thesis; its findings, conclusions, and original contributions, associated to the research objective presented in the introductory chapter. A discussion of openings for future research will round up this chapter and the thesis.

5.1 Summary of the Research

The primary objective of the research was to investigate and explore the time series forecasting methods to determine motivations to design a new forecasting technique. The requirement for a deeper understanding of the fuzzy time series forecasting, a hybrid forecasting model and the combination forecasting methods and their usage was motivated in the introduction. At the same time, forecasting of student enrollment of the university of Alabama was introduced as a practical application that utilized in the forecasting model to attain greater accuracy by comparison with other individual and hybrid forecasting model. The literature review conducted at the beginning of this research project revealed the background of time series forecasting, qualitative and quantitative forecasting, hybrid forecasting and more general behaviour of fuzzy time series and forecast combination.

Chapter 3 started by critically observing the studies published in the literature. Therefore, the main findings being very general and evidence frequently suggesting that hybrid forecasting methods based on fuzzy time series can perform better than individual ones. Empirical examination of forecasting model was conducted and described using the student enrollment datasets of the university of Alabama. The main objective here was to consider a hybrid forecasting model where an automatic clustering algorithm was used to calculate the datasets interval in a more effective manner. Traditional forecasting methods have some drawbacks that it cannot deal with any forecasting problems where the historical data are represented by linguistic values. Fuzzy time series forecasting is used to overcome that drawback. Moreover, the hybrid forecasting model has been investigated through particle swarm optimization method to obtain better forecasting result.

In Chapter 4, a combination forecasting method is proposed by following the literature published in a different article. Linear and non-linear forecast combination techniques were mentioned. The combination forecast model is developed by using a few individual forecasting models and the combination results found based on statistical methods. Therefore, the main findings being very general and evidence frequently suggesting that the combination forecast methods based on several individual methods can perform better than individual ones. Empirical examination of forecasting model was conducted and described using the Canadian lynx, Wolf's sunspots, and the monthly international airline passenger's series datasets.

5.2 Future Work

In any empirical study, there can always be a broader range of forecasting methods, more parametrisations, more attributes, and diverse datasets. According to Ord (2001) the future of time series forecasting must lie in obtaining a good understanding of the performance of existing forecasting methods in diverse scenarios rather than increasing the number of empirical studies. Therefore, the experiments on a hybrid forecasting model presented in chapter 3 seem to be particularly promising for future research. Combination forecast model certainly has potential for further research. Other topics suggested by Ord (2001) include the need to develop model selection procedures that make effective use of both data and prior knowledge, and the need to specify the objectives for forecasts and develop forecasting systems that address those objectives. These areas are still in need of consideration and the future research will contribute tools to solve these problems. Furthermore, Big Data analysis and forecasting in a distributed system might be a key issue where the forecasting models can be used to get better performance. More research is to be expected in this context to provide a certain contribution. The key issues that need to be addressed to contribute more to this research are mentioned below.

Analysis of Hybrid forecasting model

In this analysis, the following points are the key considerations for hybrid forecasting model-

1. More real-life datasets should be introduced to evaluate the model.
2. Comparison with other hybrid forecasting models where ant colony optimization, bee colony optimization, and genetic algorithm might be used.

Analysis and design of combination forecasting model

For combination forecasting model following points are the main considerations in this analysis-

1. Analysis of certain combination forecasting models using ARIMA, SARIMA, ANN, RBF, Naïve Bias, etc.
2. Comparing existing combination models and evaluates their performance and accuracy in terms of forecasting.
3. Model selection approaches while combining forecasting model.
4. Linear and nonlinear forecast combination methods should be considered to get an appropriate combination.

Performance evaluation of forecasting model in a distributed system using Hadoop/MapReduce

For performance evaluation of forecasting model in a distributed environment using Hadoop MapReduce setting, the following points are the major considerations in this analysis-

1. Analysis of the machine learning applications using large scale parallelism and small-scale parallelism.
2. In most existing machine learning applications, the researchers just apply single learning algorithm or technique to deal with practical problems, but it is important to realize that each approach has strengths and weaknesses. Therefore, the idea of hybrid learning should be further considered at present in the big data background.
3. The simplest and most common technique of parallel processing is to simply run the same learning algorithm with different parameters on different processors. The approach needs to be carefully considered to identify whether it speed up the individual run of a learning algorithm.
4. Hadoop/MapReduce workflow should be considered for effectively process the large dataset. Many of the Big Data tools in this domain (non-linear supervised learning) are clunky, slow, memory-inefficient and buggy (affecting predictive accuracy).

References-

AGRAFIOTIS, D. K. & XU, H. 2002. A self-organizing principle for learning nonlinear manifolds. *Proceedings of the National Academy of Sciences,* 99**,** 15869-15872.

AL-ALAWI, S. & AL-HINAI, H. 1998. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. *Renewable Energy,* 14**,** 199-204.

BALDWIN, M. E. & KAIN, J. S. 2006. Sensitivity of several performance measures to displacement error, bias, and event frequency. *Weather and Forecasting,* 21**,** 636-648.

BATES, J. & GRANGER, C. 1969. Combined forecasting. *Journal of Operational Research,* 20**,** 451-468.

BIDLOT, J.-R., HOLMES, D. J., WITTMANN, P. A., LALBEHARRY, R. & CHEN, H. S. 2002. Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Weather and Forecasting,* 17**,** 287-310.

BONISSONE, P. P. A fuzzy sets based linguistic approach: theory and applications. Proceedings of the 12th conference on Winter simulation, 1980. IEEE Press, 99-111.

BOWERMAN, B. L. & O'CONNELL, R. T. 1993. Forecasting and time series: An applied approach. 3rd.

BRUCE, J., SHYU, H., SIEH, K., CHEN, Y.-G. & CHUNG, L.-H. 2006. Geomorphic analysis of the Central Range fault, the second major active structure of the Longitudinal Valley suture, eastern Taiwan. *Geological Society of America Bulletin,* 118**,** 1447-1462.

CASSAR, G. 2014. Industry and startup experience on entrepreneur forecast performance in new firms. *Journal of Business Venturing,* 29**,** 137-151.

CHAPELLE, O. & VAPNIK, V. Model Selection for Support Vector Machines. NIPS, 1999. 230-236.

CHATFIELD, C. 2000. *Time-series forecasting*, CRC Press.

CHEN, C.-Y., FENG, H.-M. & YE, F. 2007. Hybrid recursive particle swarm optimization learning algorithm in the design of radial basis function networks. *Journal of Marine Science and Technology,* 15**,** 31-40.

CHEN, M.-C., CHEN, L.-S., HSU, C.-C. & ZENG, W.-R. 2008. An information granulation based data mining approach for classifying imbalanced data. *Information Sciences,* 178**,** 3214-3227.

CHEN, M., MAO, S. & LIU, Y. 2014. Big data: A survey. *Mobile Networks and Applications,* 19**,** 171-209.

CHEN, S.-M. 1996. Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems,* 81**,** 311-319.

CHEN, S.-M. 2002. Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems,* 33**,** 1-16.

CHEN, S. M. & CHUNG, N. Y. 2006. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of Intelligent Systems,* 21**,** 485-501.

CHEN, S. M. & ZOU, X. Y. 2019. Fuzzy time series forecasting based on proportions of intervals and particle swarm optimization techniques. *Information Sciences,* 500**,** 127-139.

CHEN, Y., ALSPAUGH, S. & KATZ, R. 2012. Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads. *Proceedings of the VLDB Endowment,* 5**,** 1802-1813.

COSHALL, J. T. & CHARLESWORTH, R. 2011. A management orientated approach to combination forecasting of tourism demand. *Tourism Management,* 32**,** 759-769.
DE GOOIJER, J. G. & HYNDMAN, R. J. 2006. 25 years of time series forecasting. *International journal of forecasting,* 22**,** 443-473.
DIEBOLD, F. X. & LOPEZ, J. A. 1996. 8 Forecast evaluation and combination. *Handbook of statistics,* 14**,** 241-268.
DUBOIS, D. J. 1980. *Fuzzy sets and systems: theory and applications*, Academic press.
EBERHART, R. C. & SHI, Y. Tracking and optimizing dynamic systems with particle swarms. Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, 2001. IEEE, 94-100.
FENG, H.-M., CHEN, C.-Y. & YE, F. 2006. Adaptive hyper-fuzzy partition particle swarm optimization clustering algorithm. *Cybernetics and Systems: An International Journal,* 37**,** 463-479.
GRANGER, C. W. & PESARAN, M. H. 2004. Economic and statistical measures of forecast accuracy.
HERODOTOU, H., LIM, H., LUO, G., BORISOV, N., DONG, L., CETIN, F. B. & BABU, S. Starfish: A Self-tuning System for Big Data Analytics. CIDR, 2011. 261-272.
HUA, X., NI, Y., KO, J. & WONG, K. 2007. Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *Journal of Computing in Civil Engineering,* 21**,** 122-135.
HUANG, K. Y. & JANE, C.-J. 2009. A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. *Expert systems with applications,* 36**,** 5387-5392.
HUANG, Y.-L., HORNG, S.-J., HE, M., FAN, P., KAO, T.-W., KHAN, M. K., LAI, J.-L. & KUO, I.-H. 2011. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization. *Expert systems with applications,* 38**,** 8014-8023.
HYNDMAN, R. J. & KOEHLER, A. B. 2006. Another look at measures of forecast accuracy. *International journal of forecasting,* 22**,** 679-688.
KAYACAN, E., ULUTAS, B. & KAYNAK, O. 2010. Grey system theory-based models in time series prediction. *Expert systems with applications,* 37**,** 1784-1789.
KENNEDY, J., KENNEDY, J. F., EBERHART, R. C. & SHI, Y. 2001. *Swarm intelligence*, Morgan Kaufmann.
KNOTT'S, J. 2010. Big Data, Industry 4.0 and Engineering Structural Integrity.
KRASKA, T. 2013. Finding the needle in the big data systems haystack. *IEEE Internet Computing***,** 84-86.
KUO, I.-H., HORNG, S.-J., CHEN, Y.-H., RUN, R.-S., KAO, T.-W., CHEN, R.-J., LAI, J.-L. & LIN, T.-L. 2010. Forecasting TAIFEX based on fuzzy time series and particle swarm optimization. *Expert systems with applications,* 37**,** 1494-1502.
KUO, I.-H., HORNG, S.-J., KAO, T.-W., LIN, T.-L., LEE, C.-L. & PAN, Y. 2009. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert systems with applications,* 36**,** 6108-6117.
LABRINIDIS, A. & JAGADISH, H. 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment,* 5**,** 2032-2033.
LUH, W. M. & GUO, J. H. 1999. A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *British Journal of Mathematical and Statistical Psychology,* 52**,** 303-320.
LUXHØJ, J. T., RIIS, J. O. & STENSBALLE, B. 1996. A hybrid econometric—neural network modeling approach for sales forecasting. *International Journal of Production Economics,* 43**,** 175-192.
MADDEN, S. 2012. From databases to big data. *IEEE Internet Computing***,** 4-6.

MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C. & BYERS, A. H. 2011. Big data: The next frontier for innovation, competition, and productivity.

MARTENS, M., CHANG, Y. C. & TAYLOR, S. J. 2002. A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research,* 25**,** 283-299.

MARX, V. 2013. Biology: The big challenges of big data. *Nature,* 498**,** 255-260.

MCAFEE, A., BRYNJOLFSSON, E., DAVENPORT, T. H., PATIL, D. & BARTON, D. 2012. Big data. *The management revolution. Harvard Bus Rev,* 90**,** 61-67.

MEADE, N. & ISLAM, T. 1998. Technological forecasting—Model selection, model stability, and combining models. *Management Science,* 44**,** 1115-1130.

PALIT, A. K. & POPOVIC, D. 2006. Computational intelligence in time series forecasting.

POLI, R., KENNEDY, J. & BLACKWELL, T. 2007. Particle swarm optimization. *Swarm intelligence,* 1**,** 33-57.

ORTIZ-ARROYO, D. & POULSEN, J. 2018. A weighted fuzzy time series forcasting model. *International Journal of Science and Technology,* 11(27)**,** 0974-6846.

SALERNO, F., CAMMÀ, C., ENEA, M., RÖSSLE, M. & WONG, F. 2007. Transjugular intrahepatic portosystemic shunt for refractory ascites: a meta-analysis of individual patient data. *Gastroenterology,* 133**,** 825-834.

SHETE, S., BEASLEY, T. M., ETZEL, C. J., FERNÁNDEZ, J. R., CHEN, J., ALLISON, D. B. & AMOS, C. I. 2004. Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection. *Behavior genetics,* 34**,** 153-159.

SHI, S. M., DA XU, L. & LIU, B. 1999. Improving the accuracy of nonlinear combined forecasting using neural networks. *Expert systems with applications,* 16**,** 49-54.

SHI, Y. & EBERHART, R. C. Fuzzy adaptive particle swarm optimization. Evolutionary Computation, 2001. Proceedings of the 2001 Congress on, 2001. IEEE, 101-106.

SKAMAROCK, W. C. & KLEMP, J. B. 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics,* 227**,** 3465-3485.

SONG, Q. & CHISSOM, B. S. 1993a. Forecasting enrollments with fuzzy time series—part I. *Fuzzy sets and systems,* 54**,** 1-9.

SONG, Q. & CHISSOM, B. S. 1993b. Fuzzy time series and its models. *Fuzzy sets and systems,* 54**,** 269-277.

STOCK, J. H. & WATSON, M. W. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association,* 97**,** 1167-1179.

STOCK, J. H. & WATSON, M. W. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting,* 23**,** 405-430.

STOCK, J. H. & WATSON, M. W. 2006. Forecasting with many predictors. *Handbook of economic forecasting,* 1**,** 515-554.

TIMMERMANN, A. 2006. Forecast combinations. *Handbook of economic forecasting,* 1**,** 135-196.

TINH, N. & DIEU, N. 2018. Improving the forecasted accuracy model based on Fuzzy Time Series and K-Means clustering. *Journal of Science and Technology, Vol 3 No 2*.

TINH, N., VU, V., & Linh, T. 2016. A new method for forecasting enrolments combining Time-Variant fuzzy logical relationship groups and K-means clusering. *Inernational Rsearch Journal of Engineering and Technology, Vol 3 No 3*, *2395-0072.*

TINH, N., VU, V., & Linh, T. 2019. A new hybrid fuzzy time series forecasting model based on combining fuzzy c-means clustering and particle swarm optimization. *Journal of .Computer Science and Cybernetics, Vol 35 No 3*, *267-292.*

TINH, N. & DIEU, N. 2016. A novel enrollments forecasting model based on automatic clustering techniques and time-variant fuzzy logical relationship groups. *International Journal of Engineering and Appied Sciences, 2394-3661.*

TINH, N. 2017. Handling forecasting problems based on combining fuzzy logical relationship and particle swarm optimization. *Journal of Computer Science and Engineering, Vol 19 No 6, 26-35.*

TSENG, F.-M. & TZENG, G.-H. 2002. A fuzzy seasonal ARIMA model for forecasting. *Fuzzy sets and systems,* 126**,** 367-376.

WONG, K. K., SONG, H., WITT, S. F. & WU, D. C. 2007. Tourism forecasting: to combine or not to combine? *Tourism Management,* 28**,** 1068-1078.

WU, X., ZHU, X., WU, G.-Q. & DING, W. 2014. Data mining with big data. *Knowledge and Data Engineering, IEEE Transactions on,* 26**,** 97-107.

WU, H., LONG, H., & JIANG, J. 2019. Handling forecasting probles based on fuzzy time series model and model error learning*, Applied Soft Computing Journal,* 78**,** 109-118.

XIAN, S., XIAO, Y. & PANG, J., & JIANG, J. 2017. A novel fuzzy time series forecasting method based on improved artificial fish swarm optimizatoin algorithm*, Soft Computing Journal,* 22**,** 3907-3917.

YE, Y., XIAO, L., YEN, I. & BASTANI, F. Cloud storage design based on hybrid of replication and data partitioning. Parallel and Distributed Systems (ICPADS), 2010 IEEE 16th International Conference on, 2010. IEEE, 415-422.

ZADEH, L. A. 1976. A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man-machine studies,* 8**,** 249-291.

ZHANG, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* 50**,** 159-175.