Fuzzy Set Induced Time Series Prediction

A Thesis

submitted in partial fulfillment of the requirement for

the Degree of

Master of Technology in Intelligent Automation

and Robotics

offered by Electronics and Tele-Communication Engineering

Jadavpur University

May 2024

By

Anup Dutta

Registration No: 160236 of 2021-2022

Examination Roll No: M6IAR24018

Under the Guidance of

Dr. Pratyusha Rakshit

Department of Electronics & Telecommunication Engineering

Jadavpur University, Kolkata-700032. India

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE

This is to certify that the dissertation entitled **"Fuzzy Set Induced Time Series Prediction"** has been carried out by **ANUP DUTTA** (University Registration No.:**160236** of 2021-2022) under my guidance and supervision and be accepted in partial fulfillment of the requirement for the Degree of **Master of Technology in Intelligent Automation and Robotics**. The research results presented in the thesis have not been included in any other paper submitted for the award of any degree to any other University or Institute.

Dr. Pratyusha Rakshit

Assistant Professor, Electronics and

Telecommunication Engineering Department

Jadavpur University

Dr.Manotosh Biswas

Head of department Electronics &

Telecommunication

Prof. Dipak Laha

Dean,Faculty Council of Engineering and Technology

FACULTY OF ENGINEERING AND TECHNOLOGY
JADAVPUR UNIVERSITY

CERTIFICATE OF APPROVAL*

The forgoing thesis is hereby approved as a creditable study of an engineering subject and presented in a manner satisfactory to warrant acceptance as prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn there in but approve the thesis only for which it is submitted.

Committee on final examination

for the evaluation of the thesis

Signature of the Examiner

Signature of the Supervisor

***Only in the case the thesis is approved**

JADAVPUR UNIVERSITY

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC THESIS

I hereby declare that the thesis entitled **"Fuzzy Set Induced Time Series Prediction"** contains literature Survey and original research work by the undersigned candidate,as Part of his Degree of Master of Technology in Intelligent Automation & Robotics.

All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name:- Anup Dutta

Examination Roll No:- M6IAR24018

Thesis Title:- Fuzzy Set Induced Time Series Prediction

Signature of Candidate

ACKNOWLEDGEMENT

This thesis entitled "**Fuzzy Set Induced Time Series Prediction**" is the result of the work whereby I have been accompanied and supported by many people, including my guide, my friends, and lab seniors. It is a pleasant aspect that now I can express my gratitude to all of them.

First and foremost, I would like to express my sincere gratitude to my thesis supervisor **Dr. Pratyusha Rakshit,** Assistant of the Department of Electronics and Telecommunication Engineering, at Jadavpur University for her valuable guidance, insightful suggestions, and support while conducting this thesis work as well as in the writing of this thesis. I have been very fortunate to have a guide like her. Her positivity, confidence, and ideas helped me to complete my thesis work successfully and she guided me as a guardian. I am also very grateful to our Head, **Prof.Amit Konar** for his continuous help and support and for letting me use all the available resources for my thesis work.

I want to express my gratitude to my parents and family also, as, without their sacrifices, I can't do anything. And their invaluable love, encouragement, and support make me, whatever I am today.

Anup Dutta

JADAVPUR UNIVERSITY

Kolkata-700032

CONTENTS

Title	Pages
ABSTRACT	1
CHAPTER 1: INTRODUCTION TO TIME SERIES	2
1.1 Time Series	3
1.2 Models of Time Series	4
1.3 Time series forecasting opposed to time series analysis	4
1.4 Tests and validation for a time series model	4
1.5 Time series forecasting examples	5
1.6 FUZZY Logic in Time Series Prediction	7
1.7 Application of FUZZY Sets in Time Series Prediction	7
1.8 Objective of Thesis	8
1.9 Organization of Thesis	9
1.10 REFERENCES	10
CHAPTER 2: Fuzzy Set Induced Time-Series Prediction	11
2.1 Introduction	12
2.2 Preliminaries	14
2.3 TIME-SERIES SEGMENTATION USING EDGE PROFILE (EPTS)	16
2.4 FUZZY SEGMENT MATCHING	19
2.5 PREDICTOR GRAPHS	22
2.6 EXPERIMENTS AND RESULTS	26
2.7 CONCLUSION	29
2.8 REFERENCES	31
CHAPTER 3: Conclusion & Future Scope	33
3.1 Conclusion	34
3.2 Future Scope	34
3.3 References	36

List of Figures

Figures	Pages
Fig:1.1 Forecasting Storage	6
Fig:1.2 Machine Learning at Playtech	6
Fig:1.3 Fuzzy logic in TS prediction	7
Fig 2.1: A schematic block diagram illustrating the proposed TAIEX	13
Fig: 2.2 illustrates a Time Series partitioned into 7 equal-space partitions	15
Fig:2.3 4-Patterns of Fuzzy matching scheme	19
Fig:2.4 Membership Functions	21
Fig:2.5(a) A synthetic time series is divided into temporal segments and each one is categorized.	23
Fig:2.5(b) Prediction graphs made from the time series that have been analysed and displayed in (a)	24
Fig:2.6 Segmented TAIEX close-price time-series yielded predictor graphs for 1991	28

List of Abbreviations

Abbreviations	Name
FLR	**FUZZY LOGIC RELATIONS**
PSO	**PARTICLE SWARM OPTIMIZATION**
ACO	**ANT COLONY OPTIMIZATION**
EPTS	**EDGE PROFILE TIME SERIES SEGMENTATION**
RMSE	**ROOT MEAN SQUARE ERROR**

ABSTRACT

This research introduces an innovative multi-stage methodology meticulously crafted for comprehensive knowledge extraction from time-series data. At its core, the proposed approach conceptualizes a given time-series as a dynamic sequence intricately woven with fundamental primitive patterns, each playing a pivotal role in shaping the underlying structure. The primary objective is to unravel the intricate web of first-order probabilistic transition rules essential for accurate predictions. In its initial phase, the model embarks on a strategic journey of time-series segmentation, meticulously dividing the temporal continuum into distinct blocks, each characterized by its unique and discernible dynamic slope. This segmentation process lays the foundation for subsequent analysis, facilitating a granular understanding of the temporal dynamics embedded within the data. Following this, the segmented blocks undergo a meticulous normalization process to ensure consistency and reliability across different segments. The heart of the methodology lies in the intricate matching process, where each normalized segment is meticulously paired with one of four well-established primitive patterns. This matching endeavor is guided by a sophisticated fuzzy matching algorithm, designed to discern subtle nuances and patterns within the data. Once matched, these segments serve as building blocks for the construction of directed graphs, each intricately woven with vertices representing horizontal partitions of the time-series. Within these graph structures, directed edges paint a vivid picture of transitions between partitions, driven by the occurrence of temporal segments. Each edge carries weight, capturing the essence of probabilistic transitions and encapsulating the dynamic interplay between different segments. During the testing phase, these graph representations emerge as powerful predictive tools, offering a glimpse into the future trajectory of the time-series with remarkable precision. Experimental validation, conducted on the TAIEX close-price time-series, serves as a litmus test for the model's predictive prowess. The results underscore its exceptional accuracy, cementing its position as a reliable and indispensable tool for real-world forecasting applications.

Keywords: Time-Series Analysis, Knowledge Extraction, Fuzzy Matching Algorithm, Primitive Patterns, TAIEX Close-Price, Forecasting Applications.

CHAPTER 1

INTRODUCTION
Of
Time-Series Prediction

1.1 *Time Series*

The ability to display how variables change over time through analysis distinguishes time series data from other types of data. Stated differently, time is an important variable since it reveals how the data changes over the course of the data points and the end outcomes. It offers an extra information source and a predetermined hierarchy of dependencies among the data.

For time series analysis to be reliable and consistent, a lot of data points are usually needed. Having a large data collection guarantees that your analysis can sift through erratic data and that your sample size is representative. Additionally, it guarantees that any patterns or trends found are not anomalies and can take seasonal variation into consideration. Time series data can also be utilized for forecasting, which is the process of making predictions about the future based on the past.

Time series forecasting is one of the data science techniques most commonly applied in business, finance, manufacturing, supply chain management, and inventory planning. Many prediction problems with a temporal component need time series forecasting, sometimes called time series extrapolation.

Another crucial field of machine learning (ML) is time series forecasting, which is a supervised learning issue. It can be subjected to machine learning techniques as XG Boost, Random Forests, Support Vector Machines, Regression, and Neural Networks. Using models that have been fitted to historical data, forecasting entails making predictions about future observations.

Astronomy
Planning a business
Engineering controls
Forecasting earthquakes
Using econometrics
Finance in mathematics
Identification of patterns
Allocation of Resources
Signal handling
Figures
predicting the weather

A historical time series is the first step in time series forecasting. Examining the historical data, analysts look for trends, seasonal patterns, cyclic patterns, and uniformity in the temporal decomposition patterns. Time series forecasting is used in many organizational departments, such as marketing, finance, and sales, to assess expected technical costs and customer demand. Time series data models can take many different forms and depict many stochastic processes.

1.2 Models of Time series

On the basis of validated historical data, time series models are used to forecast occurrences. Moving average, smooth-based, and ARIMA are examples of common types. It's important to identify the model that performs best depending on each time series since different models will produce different outcomes for the same dataset.

The goal must be clear to you when forecasting. Asking the following questions can help you focus on the details of your predictive modelling problem:

A larger amount of data is usually more beneficial since it presents additional chances for exploratory data analysis, model testing and fine-tuning, and model fidelity.

Required time horizon for predictions: forecasts with shorter time horizons can frequently be made more confidently than those with larger time horizons.

Frequency of forecast updates: Forecasts may need to be made once and then remain static, or they may need to be updated regularly over time as new information becomes available. More accurate predictions are typically obtained when forecasts are updated.

Forecast temporal frequency: Forecasts can frequently be made at different frequencies, allowing for the utilisation of data down sampling and up sampling (which can be advantageous for modelling purposes).

1.3 Time series forecasting opposed to time series analysis

Forecasting is concerned about making predictions about the dataset, but time series analysis is all about comprehending it. Time series analysis refers to techniques for examining time series data in order to derive important statistics and additional features from the data.
The following are predictive modelling three facets:

Data Sampling: the information we gather that illustrates our issue with established correlations between inputs and outputs.

Model Learning: the process by which we build a model that we can utilize repeatedly using sample data.

Predicting: applying our developed model to fresh data whose results we are unsure of

1.4 Tests and validation for a time series model

Time series forecasting is difficult due to a number of reasons, including:

Time dependency of a time series: In this situation, the fundamental tenet of a linear regression model—that the data are independent—is broken. Owing to the temporal dependencies included in time series data, standard validation procedures are not appropriate for time series forecasting. Training data sets should include observations that happened before those in validation sets in order to prevent biased judgments. After selecting the best model, we can fit it over the whole training set and assess how well it performs over time on a different test set.

Seasonality in a Time Series: Most time series exhibit seasonal trends, or changes unique to a given time period, in addition to a growing or declining trend.

On a given dataset, time series models can perform better than others; a model that excels on one kind of dataset might not perform as well on another.

1.5 Time series forecasting examples

Forecasting customer demand for a certain product over time, the cost of fuel sources for home heating, hotel occupancy rates, hospital inpatient care, fraud detection, and stock prices are a few examples of time series forecasting. Forecasting can be done using machine learning models or storage models.

Let's examine some forecasting examples with the open-source time series data source InfluxDB.

Forecasting storage

This is a screenshot from a storage forecasting use case from Veritas Technologies that you can see below:

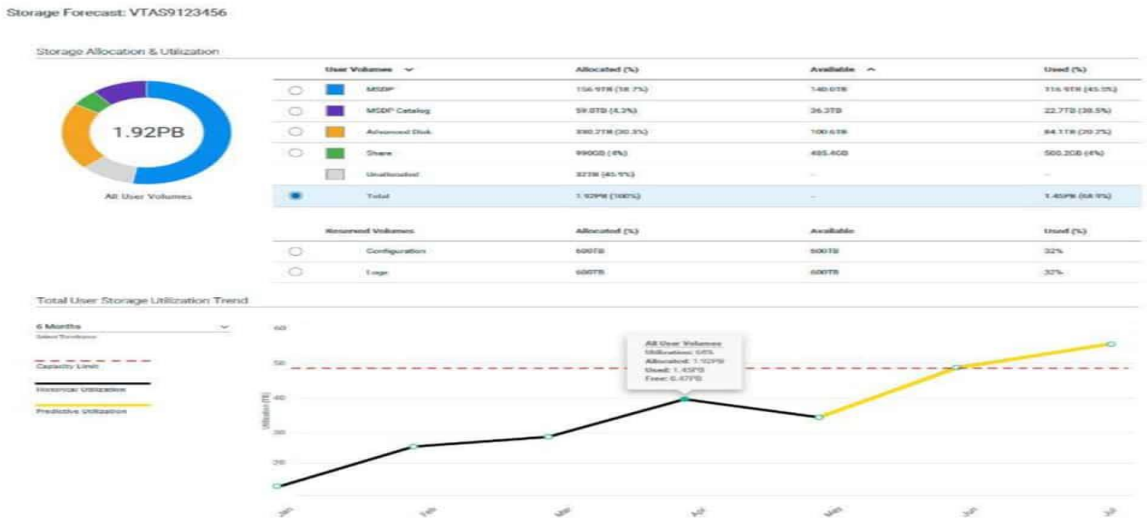


Fig:1.1 Forecasting Storage

Machine learning

This is a screenshot from a machine learning use case at Playtech that you can see below:

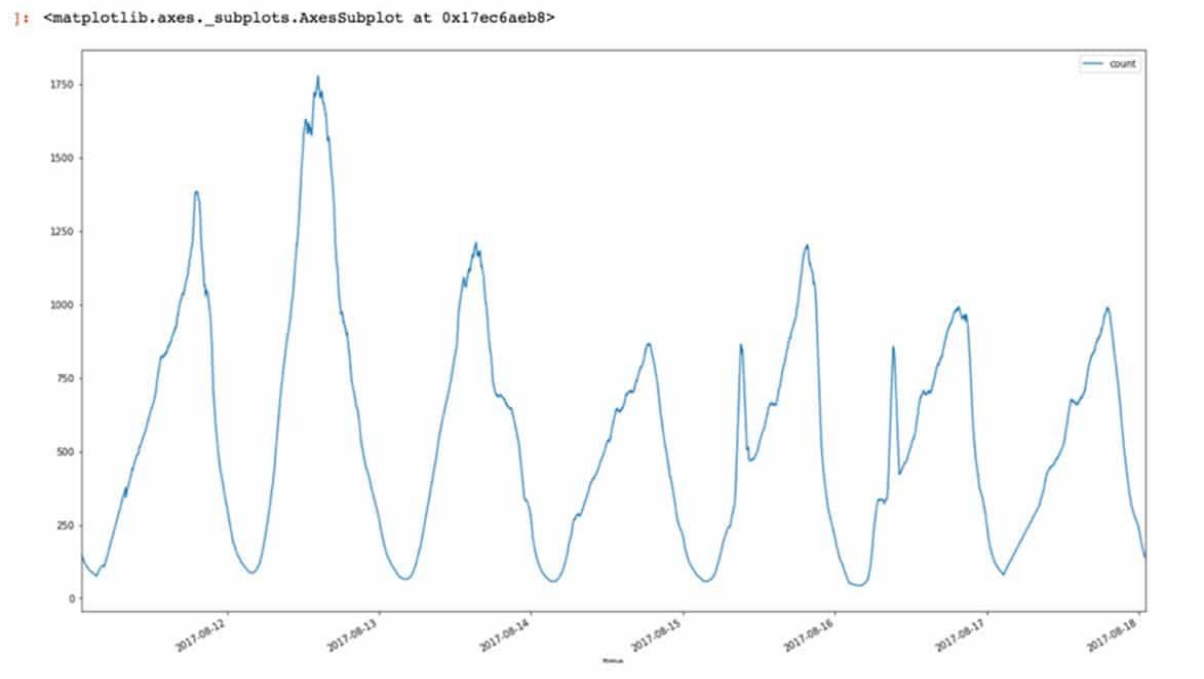


Fig:1.2 Machine Learning at Playtech

1.6 FUZZY Logic in Time Series Prediction

Fuzzy logic does not classify all statements as entirely true or untrue; rather, it is a type of para consistent logic, which treats contradictions weakly. The main difference from classical logic is this .Since it is impossible to abstract everything into a binary system in the real world, fuzzy logic offers an intriguing solution to choice problems .Let's take a look at a dataset, for instance, that tracks people's heights to identify those who are tall or short.

The height of 1.70 meters can be used as a dividing line; people who are taller than 1.70 meters are regarded as such, and those who are shorter than 1.70 meters are seen as short. By conventional reasoning, all the facts should go into one of these categories, but is it really accurate to label someone who is 1.69 m tall or 1.71 m tall?

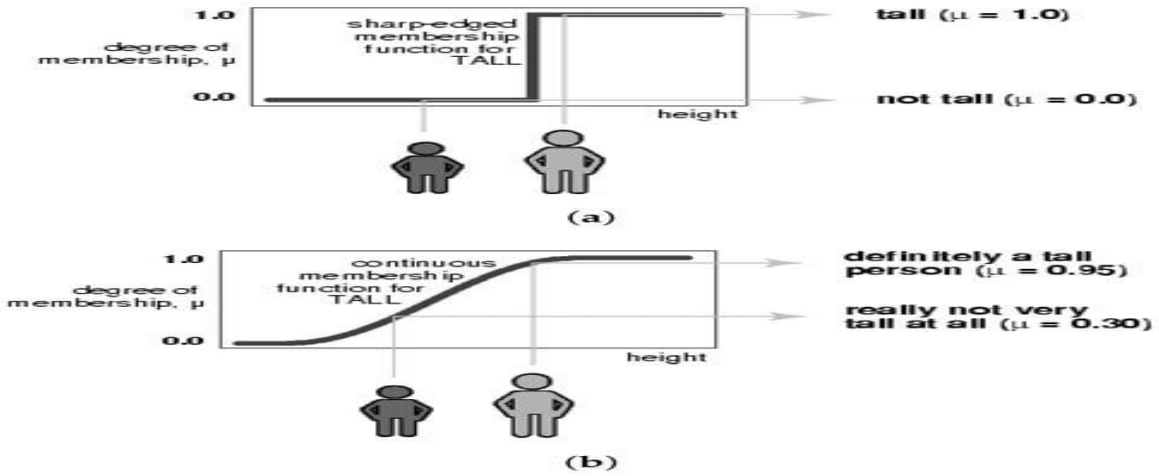


Fig:1.3 Fuzzy logic in TS prediction

1.7 Application of FUZZY Sets in Time Series Prediction

Time series prediction has seen a number of uses for fuzzy sets as a means of addressing the uncertainty and imprecision present in many real-world datasets. Fuzzy sets are used in time series prediction in the following ways:

1. Fuzzy Time Series Models: Fuzzy time series models use fuzzy logic to express linguistic words and uncertainty in time series data. Examples of these models are the Fuzzy Time-Invariant Model (FTIM), the Fuzzy Time-Varying Model (FTVM), and the Fuzzy Time Series Moving Average (FTSMA). Fuzzy rule-based systems are frequently used in these models to identify patterns and relationships in the data.

2. Fuzzy C-Means Clustering: Time series data can be divided into clusters using fuzzy clustering methods such as Fuzzy C-Means (FCM), which makes it possible to identify various patterns or regimes within the data. This can increase forecast accuracy and provide insight into the time series' underlying structure.

3. Fuzzy Neural Networks (FNNs): By incorporating fuzzy logic into neural network architectures, fuzzy neural networks can be produced. Fuzzy Logic Networks (FNNs) integrate the learning powers of neural networks with fuzzy logic's capacity to deal with

imprecision and uncertainty. These networks can increase prediction accuracy and efficiently model intricate relationships found in time series data.

4. Fuzzy Logical Relationship Groups (FLRGs) and Fuzzy Time Series Moving Average (FTSMA): are two examples of fuzzy time series forecasting techniques that use fuzzy logic to predict the patterns and trends found in time series data. These techniques frequently entail the creation of hazy logical connections and the application of hazy inference to forecast outcomes based on past evidence.

5. Interval-Based Fuzzy Time Series: By taking into account time intervals rather than discrete time points, interval-based fuzzy time series models build upon conventional fuzzy time series techniques. In situations when the precise timing of occurrences is uncertain, this makes it possible to reflect ambiguity in the timing of events and may result in more reliable forecasts.

In general, the use of fuzzy sets in time series prediction offers a framework for handling data imprecision, uncertainty, and non-linearity, resulting in more reliable and accurate predictions across a range of fields, including environmental science, finance, and economics.

1.8 Objective of Thesis

This thesis proposes a novel multi-stage method for extracting knowledge from a time-series. The goal of modelling a given time series as a set of well-known primitive patterns is to find first-order probabilistic transition rules for prediction. A time-series is divided into structurally separate temporal blocks of non uniform length in the first stage of the suggested model, with each block having a comparatively small variance in its dynamic slope. The resulting temporal segments are then normalized and matched using a fuzzy matching algorithm with one of four widely recognized primitive patterns in the second stage.

Lastly, a set of four directed graphs that fit the four primitive patterns are created from the sequence of matched segments to represent the time-series. Every vertex on the graph denotes a division on horizontal of TimeSeries, whereas every edge which are directed signifies a shift between two such divisions due to the presence of one or more temporal segments. Graphs are used to forecast potential future time-series values during the testing process. Studies demonstrating the model's applicability to real-world forecasting scenarios are consistent with its high prediction accuracy on the TAIEX close-price time series.

1.9 Organization of Thesis

In this Thesis,

Chapter 1 includes the introduction to the Time-Series prediction,literature review on time-series prediction.It proposes the application of fuzzy sets in Time-Series prediction.

Atlast it contains the objective and organization of thesis work.

Chapter 2 includes the Edge-profile Segmentation and its algorithms.It also contains the predictor graphs and experimental results including codes required.

It conatins the conclusion and references.

Chapter 3 contains the future scope for the thesis work and final conclusion statement.

1.10 REFERENCES

1. Lin, Jessica; Keogh, Eamonn; Lonardi, Stefano; Chiu, Bill (2003). "A symbolic representation of time series, with implications for streaming algorithms". Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. New York: ACM Press. pp. 2–11. CiteSeerX 10.1.1.14.5597. doi:10.1145/882082.882086. ISBN 9781450374224. S2CID 6084733.
2. ^ Warren Liao, T. (November 2005). "Clustering of time series data—a survey". Pattern Recognition. **38** (11): 1857–1874. Bibcode:2005PatRe..38.1857W. doi:10.1016/j.patcog.2005.01.025. S2CID 8973749.
3. ^ Aghabozorgi, Saeed; Seyed Shirkhorshidi, Ali; Ying Wah, Teh (October 2015). "Time-series clustering – A decade review". Information Systems. **53**: 16–38. doi:10.1016/j.is.2015.04.007. S2CID 158707.
4. ^ Keogh, Eamonn; Kasetty, Shruti (2002). "On the need for time series data mining benchmarks: A survey and empirical demonstration". Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 102–111. doi:10.1145/775047.775062. ISBN 1-58113-567-X.
5. ^ Agrawal, Rakesh; Faloutsos, Christos; Swami, Arun (1993). "Efficient similarity search in sequence databases". Foundations of Data Organization and Algorithms. Lecture Notes in Computer Science. Vol. 730. pp. 69–84. doi:10.1007/3-540-57301-1_5. ISBN 978-3-540-57301-2. S2CID 16748451.
6. ^ Chen, Cathy W. S.; Chiu, L. M. (4 September 2021). "Ordinal Time Series Forecasting of the Air Quality Index". Entropy. **23** (9): 1167. Bibcode:2021Entrp..23.1167C. doi:10.3390/e23091167. PMC 8469594. PMID 34573792.
7. ^ Sarkar, Advait; Spott, Martin; Blackwell, Alan F.; Jamnik, Mateja (2016). "Visual discovery and model-driven explanation of time series patterns". 2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). pp. 78–86. doi:10.1109/vlhcc.2016.7739668. ISBN 978-1-5090-0252-8. S2CID 9787931.
8. ^ Bloomfield, Peter (1976). Fourier Analysis of Time Series: An Introduction. Wiley. ISBN 978-0-471-08256-9.[page needed]
9. ^ Shumway, Robert H. (1988). Applied Statistical Time Series Analysis. Prentice-Hall. ISBN 978-0-13-041500-4.[page needed]
10. ^ Arlinghaus, Sandra (1994). Practical Handbook of Curve Fitting. CRC Press. ISBN 978-0-8493-0143-8.[page needed]
11. ^ Kolb, William M. (1984). Curve Fitting for Programmable Calculators. SYNTEC. ISBN 978-0-943494-02-9.[page needed]
12. ^ Halli, S. S.; Rao, K. V. (1992). Advanced Techniques of Population Analysis. Springer Science & Business Media. p. 165. ISBN 978-0-306-43997-1. functions are fulfilled if we have a good to moderate fit for the observed data.
13. ^ The Signal and the Noise: Why So Many Predictions Fail-but Some Don't. By Nate Silver
14. ^ Pyle, Dorian (1999). Data Preparation for Data Mining. Morgan Kaufmann. ISBN 978-1-55860-529-9.[page needed]

CHAPTER 2

Fuzzy Set Induced Time-Series Prediction

2.1 INTRODUCTION

Time-series analysis pertains to the examination of discrete sequences of real-valued observations, acquired through the periodic sampling of a measurable phenomenon over time. The precise forecasting of forthcoming values within such series has long captivated the interest of researchers. However, a time-series's behavior usually depends on a wide range of variables, many of which are still unpredictable and erratic. As a result, the most common approach in time-series analysis is to build mathematical models that capture the properties of the series and make predictions easier. Numerous approaches have surfaced to tackle this endeavor, all with the goal of improving the comprehension and prognostic potential of time-series data.

One notable approach in time-series analysis involves the application of fuzzy set theory. Fuzzy time-series methodology, first introduced by researchers, presents a dynamic framework for analysis. The general procedure entails treating the time-series' dynamic range as universe of discourse and establishing fuzzy sets across disjoint contiguous intervals, referred to as partitions within this universe. After that, every data point in the TimeSeries is converted into the fuzzy set whose membership value is greatest for it. Fuzzification is the procedure that makes it possible to derive first- and higher-order fuzzy logical relations (FLRs), which are then used to anticipate possible future values inside the time-series.

A lot of improvements have been made to improve the original Song and Chissom model. A paper describes one such breakthrough that takes into account the recurrence of fuzzy logical relations (FLRs) by weighting them according to the order in which they occur. Moreover, as suggested in later studies and developed for further improvement, researchers have looked at incorporating secondary and multifactor heuristics into fuzzy time-series modelling to increase prediction accuracy. Notable advancements in this field also include the use of optimization methods like ant-colony optimization and particle swarm optimization (PSO) and the investigation of effective time-series partitioning and fuzzification algorithms. Furthermore, novel techniques such as fuzzy variation groups have been studied to improve the efficiency of fuzzy time-series analysis.

In the current thesis, a novel method is proposed for modelling time-series data by identifying sequences of primitive patterns within contiguous time intervals. The approach comprises three main stages. Firstly, the focus is on semantic segmentation of the time-series into distinct temporal blocks of varying lengths, with each block efficiently categorized into one of the

recognized primitive patterns. This is achieved through an implemented segmentation algorithm that identifies temporal regions exhibiting significant variations in the dynamic slope of the time-series, marking these regions as segment boundaries to signify structural changes. The segmentation algorithm is characterized by its simplicity and low computational overhead.

The next step involves normalizing the time-segments that were extracted from the segmentation procedure and representing them as 10-point sequences to guarantee length consistency. A fuzzy matching method is then used to match these segments with one of 4 deserving patterns: linearly rising, linearly falling, Gaussian_bell curve, and Inverse_Gaussian curve. This method makes it possible to classify segments approximatively in real time without the need to train computer models.

In the third stage, the time-series is represented as a set of weighted directed graphs, each of which corresponds to a primitive pattern, using the information that was obtained from the previous phases. These graphs show the horizontal divisions in the time-series as nodes, and the likelihood of transitions between related partitions due to one or more classed time-segments as weighted directed edges. The time-series data may be intelligently analyzed by using these graphs, also known as predictor graphs, to generate probabilistic transition rules and make predictions based on them.

Analysis performed on the TAIEX economic close-price time-series spanning from 1990 to 1999 reveals a noteworthy achievement: an average Root Mean Square Error (RMSE) of 88.48, demonstrating superior performance over competitive models. Additionally, a schematic block diagram illustrating the proposed model is presented in Figure 1.

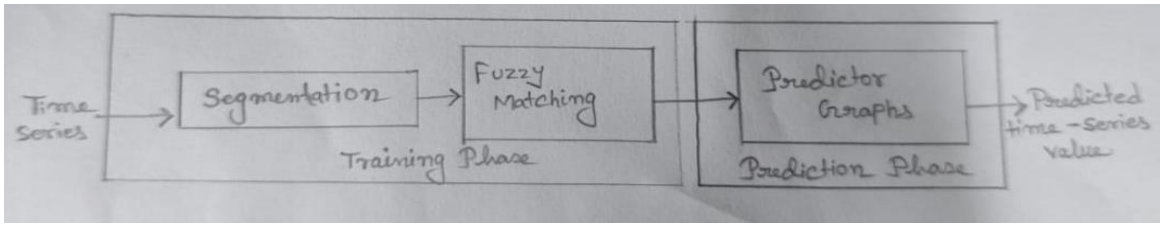


Fig 2.1: A schematic block diagram illustrating the proposed TAIEX

2.2 PRELIMINARIES

This section furnishes an introduction to several key concepts essential for comprehending the subsequent discussions. These concepts encompass a concise explanation of fuzzy sets and fuzzy membership functions, along with an elucidation of principles associated with time-series partitioning.

A. Fuzzy Sets

2.2.1: outlines the concept of a set of fuzzy Y, defined on a universe of discourse J, as a collection of 2-tuples. Each tuple comprises an element x from the universe J and its corresponding membership value, denoted as $\mu_Y(x)$, which ranges between 0 and 1. Essentially, a fuzzy set extends the notion of a traditional set, allowing for elements to possess varying degrees of membership within the set, rather than solely being categorized as either fully belonging (membership value of 1) or not belonging (membership value of 0).

$$Y = \{(x,\mu\ Y\ (x)) \mid x \in J\} \quad (1)$$

2.2.2: Describes the idea of a fuzzy membership function for a fuzzy set Y that is defined on a discourse universe U. Every element x in U is mapped by this function to its associated membership value μY(x), which is a number between 0 and 1. Such a function is required when the universe of discourse is continuous or infinite, as it is impractical to determine a finite number of 2-tuples for the fuzzy set.

Fuzzy sets are frequently used to characterize sets of items that don't have exact definitions. Fuzzy sets can be used to define sets such as "very tall people" or "numbers very close to 0".

The subjective nature of some linguistic phrases, such "very tall" and "very close to 0," makes standard sets unsuitable for precisely characterizing such collections.

B. Time-Series Partitioning

2.2.3: If z is a measurable entity and z(t) is the sampled value obtained at time instant t, then the time-series A of length n can be formalized as a vector, where n is the total number of samples obtained and T is the sampling interval. A time-series is defined as a sequence of discrete values or measurements obtained by means of sampling of a measurable phenomenon at successive time points.

$$A=[z(T),z(2T),...,z(nT)]=[a_1,a_2,...,a_n] \quad (2)$$

2.2.4: Let min and max denote the minimum and maximum values of the time-series C, respectively, defining the dynamic range of the series as (C)=[min, max] This range $R(C)$ is divided into k non-overlapping, contiguous intervals of the form

$$P_1 = [p_1^-, p_1^+],\ P_2 = [p_2^-, p_2^+],\ \ldots,\ P_k = [p_k^-, p_k^+] \text{ where}$$
$$p_i^- > p_{i-1}^+,\ \forall i \in \{2,3,\ldots,k\} \qquad (3)$$

This process is termed partitioning, where each interval Pj is referred to as a partition. It is evident that $P1$ begins at the minimum value min of the series, and ends at the maximum value max. Figure II illustrates a time-series partitioned into 7 equally-spaced partitions.

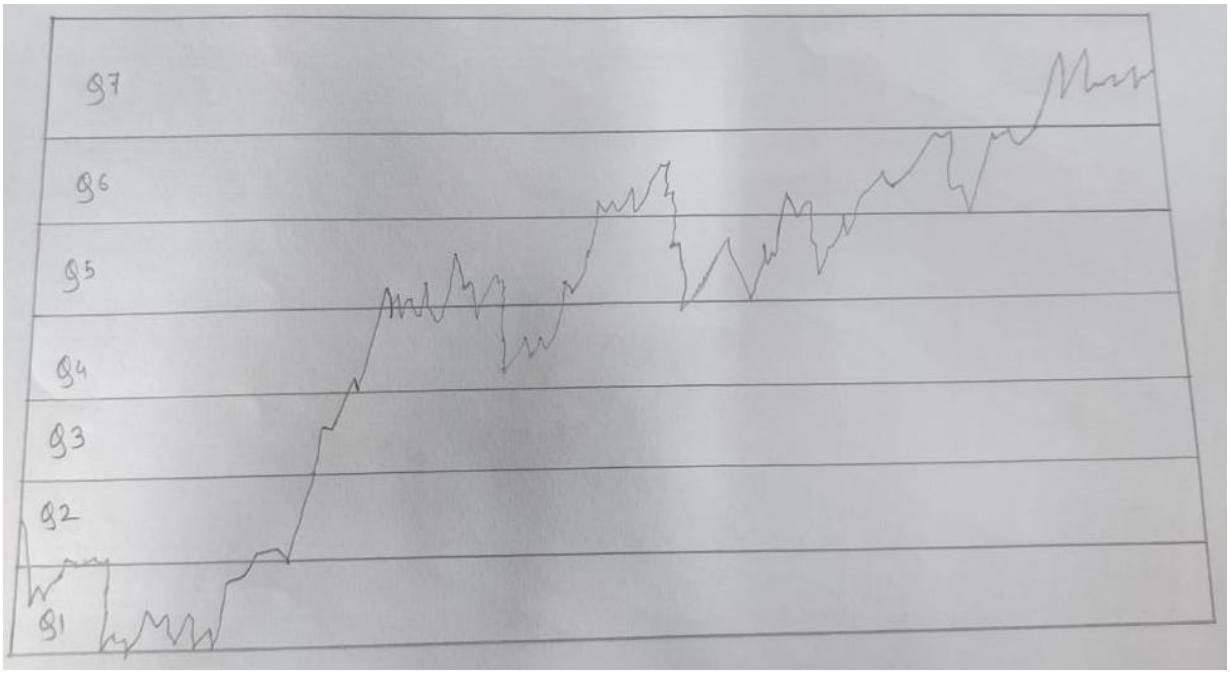


Fig: 2.2 illustrates a TimeSeries partitioned into 7 equal-space partitions

2.3 TIME-SERIES SEGMENTATION USING EDGE PROFILE (EPTS)

Here, we present an effective technique for segmenting time-series data using the dynamic slope of the series as the principal segmentation criterion. Semantic segmentation of a time-series is based on finding times when the series shows notable slope changes; the time-series is viewed as a piecewise linear curve, with each successive pair of data points connected by a straight line, called an edge.

Imagine the perfect time series that is free of low-level noisy fluctuations. In these situations, the locations where the series experiences behavioral pattern changes are the only ones where the slopes of the edges joining successive data points show significant fluctuation. Our method uses this intuition to find these temporal zones and labels them as time-series segment borders. Temporal segments, also known as time blocks, are made up of the data points that fall between two successive segment borders. We call our method Edge Profile Time-Series Segmentation (EPTS), and the segmentation criteria we use is the relative variance of the edge slopes in the series.

Our algorithm's main benefit over other time-series segmentation techniques is its low computational complexity and ease of use. The suggested method is formalized in the definitions that follow.

A. Definitions

2.3.1: Consider two consecutive data points, denoted as z_i and z_{i+1}, extracted from a given time-series C sampled at time instances jT and (j+1) respectively. The straight line, referred to as edge m, connecting these points is defined by the coordinates (jT,z_j) and (($j+1$)T,z_{j+1})) The slope of this edge, denoted as mm, can be calculated using the formula:

$$slope\ (x_1, x_2) = \frac{(z_{i+1} - z_i)}{((i+1)T - iT} = \frac{(z_{i+1} - z_i)}{T} \quad (4)$$

Example 1 Let Z=[z1,z2,z3,z4,z5]=[3,4,6,2,1] be a time-series in which the values are sampled at time instants 1, 2, 3, 4, and 5, respectively. Plotting the time-series data points as ordinates and the time-instants as abscissas results in the following coordinates in two-dimensional space: q1=(1,3), q2=(2,4), q3=(3,6), q4=(4,2), and q5=(5,1). The straight line equation between points q1 and q2, as given in equation (5), defines the edge l created between two successive data points, z1 and z2.

$$l_{1,2} \equiv \frac{y-3}{x-1} = \frac{4-3}{2-1} \rightarrow x - y + 2 = 0 \quad (5)$$

The slope of the edge m formed between point's q1and q2 is clearly determined as the slope between point's q1 and q2, resulting in a value of +1. Similarly, the slopes for the remaining edges $m_{2,3}$, $m_{3,4}$, and $m_{4,5}$ of the time-series are calculated as follows: the slope between points q2 and q3 yields +2, the slope between points q3 and q4 results in −4, and finally, the slope between points q4 and q5 is −1.

2.3.2: In the context of a time-series analysis, a finite sequence of consecutive edges is termed an edge-window. Let x represent the chosen width of each edge-window. The edge-window originating from the i^{th} time-series data point is defined as follows:

$$X_i=\{m_j,m_{j+1},m_{j+2},...,m_{j+x-1}\} \quad (6)$$

Similarly, for a given edge-window, we define a slope-window containing the slope of every edge within the edge-window. For the window described in equation (6), the corresponding slope-window is articulated as:

$$T_i=\{(slope(q_i,q_{i+1})),(slope(q_{i+1},q_{i+2})),...,(slope(q_{i+x-1},q_{i+x}))\} \quad (7)$$

The degree of variation observed in the slopes encompassed within a slope window offers valuable insights into the behaviour of the time-series within the associated temporal region. Higher variations in slopes suggest a greater likelihood of the time-series exhibiting shifts in its behavioural patterns. Consequently, the variation of slopes within a slope-window serves as the basis for deriving the segmentation criterion in the proposed algorithm.

Example 2: In this illustration, we exemplify the construction of edge-windows and slope-windows using the time-series provided in Example 1. Assuming a window width of 3, the edge-windows derived from the time-series are represented as:

$X_1=[m_{1,2}, m_{2,3}, m_{3,4}]$

$X_2=[m_{2,3}, m_{3,4}, m_{4,5}]$

Correspondingly, the slope-windows associated with the time-series are computed as follows:

$$T_1=[slope(q_1,q_2),slope(q_2,q_3),slope(q_3,q_4)]=[+1,+2,-4] \quad (8)$$

$$T_2=[slope(q_2,q_3),slope(q_3,q_4),slope(q_4,q_5)]=[+2,-4,-1] \quad (9)$$

Segmentation criterion: Within the framework of the proposed algorithm, the segmentation criterion is predicated on the detection of slope-windows exhibiting high variance, indicative of potential alterations in the structural composition of the time-series within corresponding temporal domains. This pivotal determination relies upon assessing the variability inherent in the slopes encapsulated within a given slope-window, denoted as (V). If this variability surpasses a pre-established threshold hT, the slope-window V emerges as a candidate for delineating a segment boundary.

Consider an edge denoted as $m_{j,+1}$, positioned at the midpoint of the identified slope-window. Flanking this edge are two consecutive data points d_j and d_{j+1}. While, for the purpose of this paper, we elect to position the segment boundary at d_{j+1}, it's imperative to acknowledge the inherent subjectivity of this decision-making process. Alternative options for locating segment boundaries include anchoring them at d or employing the mean value between c_i and d_{i+1}.

B. Proposed Segmentation Algorithm

The proposed segmentation technique unfolds through three primary stages, each contributing to the effective partitioning of the time-series. Firstly, in the pre-processing stage, efforts are directed towards mitigating the influence of low-level noisy fluctuations on the segmentation algorithm's decision-making process. Various methods exist for achieving this, with this paper opting for the widely recognized Gaussian smoothing technique. By smoothing the time-series, the aim is to attenuate disruptive noise components.

Subsequently, the second stage employs the segmentation algorithm to delineate segment boundaries within the time-series. However, there's a possibility of closely spaced segment boundaries emerging if the high variance condition persists across multiple consecutive slope windows. To address this, two strategies are proposed. The first involves shifting the slope window by more than one unit in the direction of increasing time whenever a segment boundary is placed, thereby reducing the likelihood of consecutive slope windows satisfying the high variance segmentation criterion. Alternatively, the second approach entails reprocessing the segmented time-series to consolidate closely placed segment boundaries into a single boundary. In this paper, the latter option is embraced as the third stage, ensuring the consistent shifting of the slope window throughout the entirety of the time-series.

2.4 FUZZY SEGMENT MATCHING

In this part, we describe a procedure intended to assign one of 4 basic primitive patterns—linear rise, linear fall, Gaussian bell curve, or inverted Gaussian bell curve—to every temporal segment obtained from the previous stage, as shown in Figure 4. There are several ways to go about the process of categorizing a temporal segment into one of these predetermined forms. Conventionally, a typical approach is to train a pattern classifier to identify the class or form that most closely resembles a particular temporal segment. But this method requires a large dataset to be created in order to train the classifier, which takes a lot of time and processing power.

Alternatively, we provide a matching approach that finds the representative pattern for a given temporal segment in real-time without requiring training. The goal of this technique is to find the pattern that most closely matches the temporal segment's properties without having to incur the costs of classifier training.

To facilitate the matching process, each temporal segment is initially represented as a vector of length 10, thereby ensuring uniformity in temporal length across segments. Subsequently, these segments undergo mean normalization to mitigate the impact of differing ranges among segments, which could otherwise yield inaccurate matching results. After normalization, a fuzzy matching approach is applied to determine a basic pattern's similarity metric to a particular temporal segment. Because each fundamental pattern and the segment have different degrees of similarity, this fuzzy matching technique enables a more sophisticated assessment of similarity.

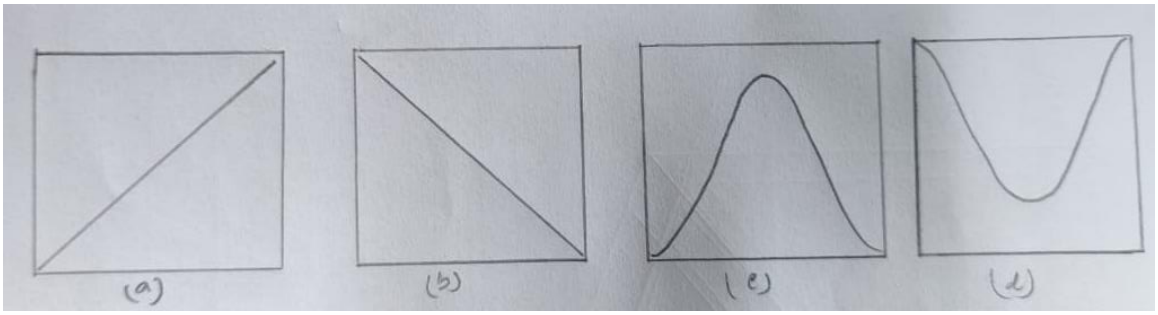


Fig:2.3 4-Patterns of Fuzzy matching scheme

A. Similarity Metric of FUZZY

A set of ten consecutive dots, indicated as, represents a basic pattern R.

As follows:

R=[r1,r2,...,r10].

Let max(r) and min(r) stand for the sequence *P*'s maximum and minimum values, respectively. The universe of discourse for portraying the primitive pattern is the range U=[min(r)−k, max(r)+k], where k is an experimentally selected constant. Ten fuzzy sets in a row, each representing a pattern point, can fit inside this range. Each fuzzy set represents the linguistic expression "CLOSE TO rj" and is specified for a point i in the pattern. The Gaussian and triangle membership functions, which both reach a maximum value of 1 at point pi, are appropriate membership functions for characterizing the fuzzy set. The triangle membership function for point i in this context is as follows:

$$\left.\begin{aligned} \mu_i(x) &= \left(\frac{x - p_i + r}{r}\right) \text{ if } (p_i - r) \le x \le p_i \\ &= \left(\frac{p_i + r - x}{r}\right) \text{ if } p_i \le x \le (p_i + r) \\ &= 0 \text{ otherwise} \end{aligned}\right\} \quad (11)$$

In this case, $\boldsymbol{U}$ denotes a certain proportion of the range max (p−p), where p and p stand for the sequence's maximum and minimum values, respectively. The sum of the individual membership values μi(t) of each point si in the appropriate fuzzy sets for the pattern R defines the similarity between a mean-normalized temporal segment T=[t1,t2,...,t10] and pattern R. Equation (12) represents this computation, which is shown in Figure 4.

$$similarity(T, R) = \sum_{i=0}^{10} \mu_i(t_i) \quad (12)$$

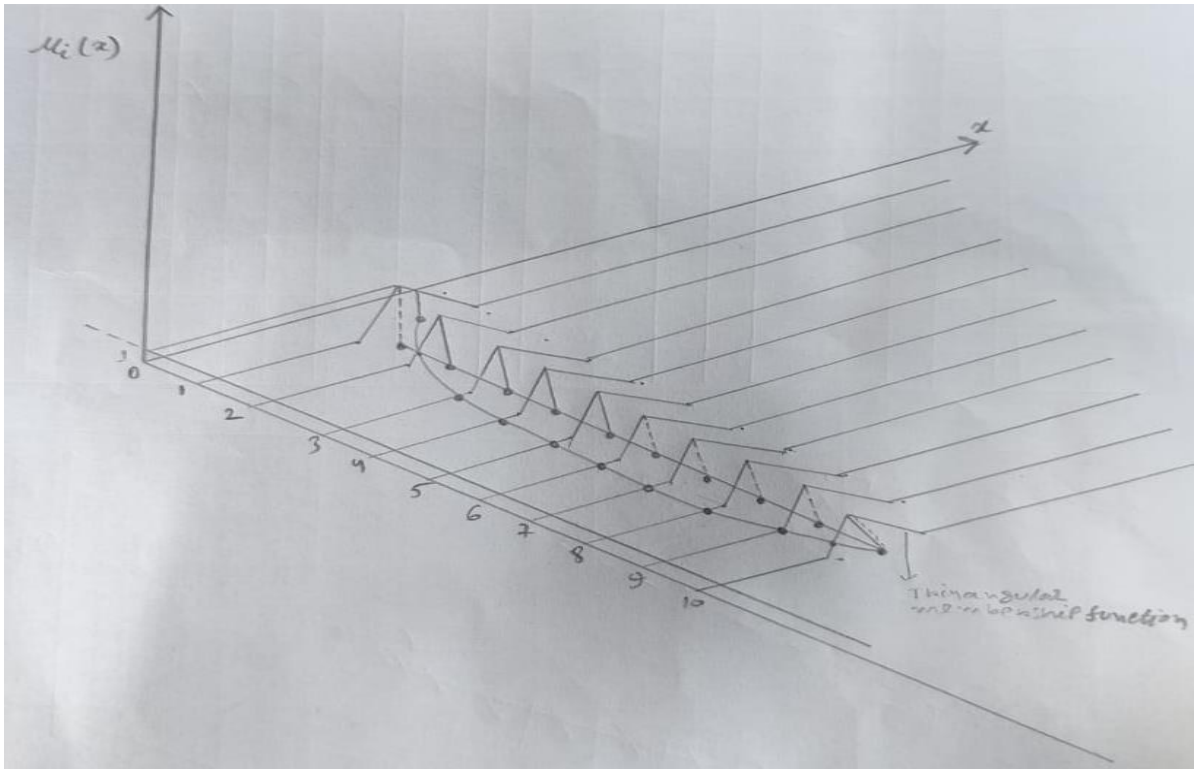


Fig:2.4 Membership Functions

B. Fuzzy Matching Algorithm

Upon establishing the fuzzy similarity metric in the previous phase, the subsequent matching procedure unfolds in a straightforward manner. Every standardized segment of temporal X, its fuzzy resemblance is meticulously computed against each of the four fundamental primitive patterns. This entails a comprehensive assessment of how closely X aligns with each pattern, gauged by their respective fuzzy similarity scores. Subsequently, X is allocated to pattern Y_i if its similarity with Y_i surpasses that with any other pattern, as specified in Equation (13).

$$similarity\ (X, Y_i) \geq similarity\ (X, Y_i)\ \forall j \in \{1, 2, 3, 4\}. \quad (13)$$

It's imperative to acknowledge the existence of temporal segments that may not conform adequately to any of the predefined primitive patterns. To account for such instances, a distinct classification is reserved for these outlier (or anomalous) segments. Thus, while a segment X may be assigned to a specific pattern Y_j following Equation (13), it is flagged as an outlier if its similarity with Y_j falls below a predetermined rejection threshold T, as delineated in Equation (14). In essence, if a segment fails to exhibit a substantial degree of similarity to any of the established patterns, it is categorized as an outlier.

$$similarity\ (X, Y_i) < T_t \qquad (14)$$

The formal presentation of the fuzzy matching algorithm is encapsulated in Pseudo Code 2, providing a comprehensive guideline for its implementation and execution.

2.5 PREDICTOR GRAPHS

In this section, we delve into the intricate process of translating the insights gleaned from the preceding steps into a comprehensible framework. We construct a set of four weighted directed graphs that serve as repositories for the acquired knowledge. Within each graph, nodes meticulously delineate horizontal partitions of the time-series landscape, while the weighted directed edges meticulously capture the transitions between these partitions. These transitions are instigated by the occurrence of one or more temporal segments, which effectively serve as the catalysts for temporal evolution.

As we transition into the prediction phase, the significance of these graphs becomes increasingly apparent. From a designated starting partition, each graph undertakes the task of prognosticating a potential future value of the time-series. Yet, what sets this prediction scheme apart is its nuanced approach. By meticulously dissecting the effects of different primitives and independently incorporating them into the prediction process, the model gains the dexterity to anticipate the outcomes of individual primitives on the time-series dynamics.

This approach not only enhances the predictive accuracy of the model but also underscores its adaptability to the myriad complexities inherent in time-series data. By embracing a granular understanding of the underlying dynamics, the model emerges as a robust tool capable of navigating the intricacies of forecasting with precision and insight.

A. Construction of predictor graphs

To delineate the construction of the predictor graphs, denoted as G_1, G_2, G_3, G_4, a systematic approach is employed, consisting of the following elaborated steps:

Initially, the time-series undergoes partitioning into q equi-spaced intervals or partitions, ensuring a structured segmentation of the data. For each of these partitions, a corresponding vertex is established within each graph G_x, where x ranges from 1 to 4, representing the different primitive patterns identified in the time-series. This results in the creation of q vertices within each graph, each serving to demarcate specific segments of the time-series.

Subsequently, for every temporal segment characterized by pattern , which triggers a transition from one partition to another, an intricate analysis ensues. The algorithm assesses whether d' of these segments lead to a transition to a specific partition, denoted as j. Upon confirming such transitions, directed arcs are meticulously inserted from the vertex corresponding to the initial partition to the vertex representing the destination partition within the graph G_x. These arcs symbolize the directional flow of transitions induced by the identified pattern . Notably, the weight of each arc is meticulously computed using Equation (15), serving to approximate the probability of transition between partitions attributed to the occurrence of pattern P_x. It's crucial to ensure that the sum of weights associated with outgoing edges from any vertex is normalized to 1, preserving the integrity of the probabilistic model.

$$\delta(v_i, v_j) = \left(\frac{d'}{d}\right) \qquad (15)$$

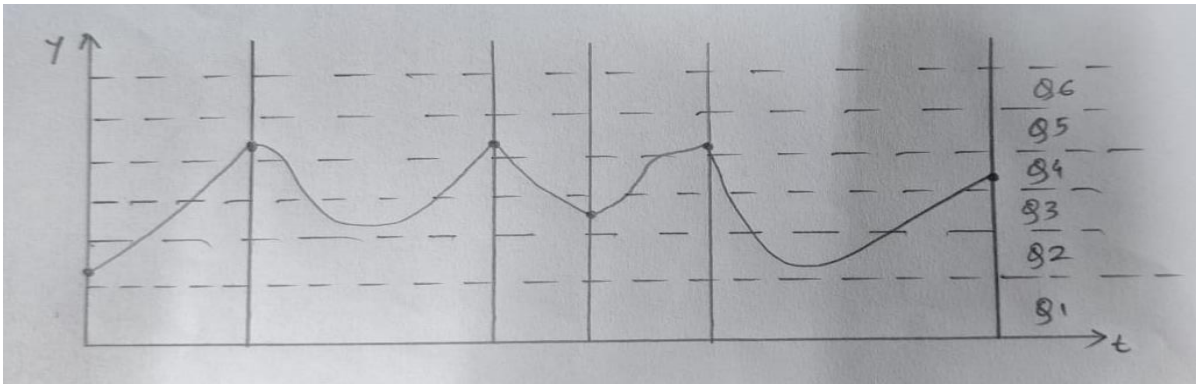


Fig:2.5(a) A synthetic time series is divided into temporal segments and each one is categorized.

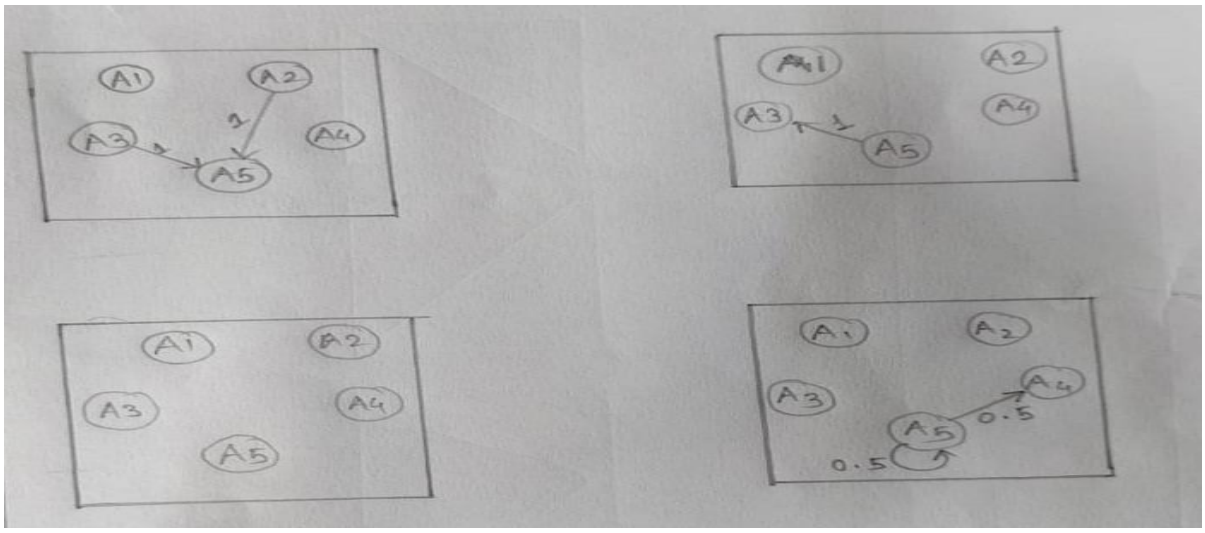


Fig:2.5(b) Prediction graphs made from the time series that have been analyzed and displayed in (a)

To offer a tangible illustration of this intricate construction process, Figure 6 provides visual representations, depicting the sequential formation of the predictor graphs from a hypothetical synthetic time-series. This illustration showcases the intricate interplay between segmentation, classification, and graph construction, elucidating the underlying dynamics of the proposed methodology.

B. Prediction of future time-series values

Once the predictor graphs have been meticulously assembled, the model is poised to undertake prognostications regarding potential future values within the time-series dataset. Assuming a delineated partition aligns with the present day, denoted as cur, the predictive process unfolds across each graph, indexed from 1 to 4.

Within each graph , where p iterates through the set {1, 2, 3, 4}, the predictive algorithm comes into play. It calculates the forecasted value by meticulously considering the following factors: the current partition corresponding to the present day, and the potential transitions delineated within the predictor graph.

This procedure is a careful evaluation of possible partition transitions inside the time-series, which is contained in a weighted summation mechanism. The midpoint of each partition is carefully weighted, taking into account the possibility of transitions caused by temporal segments.

$$Pr_p = \sum_{i=1}^{N} (\delta(v_{cur}, v_i) \times mid(p_i)) \quad (16)$$

Ultimately, the model synthesizes these insights across all predictor graphs to provide a comprehensive and cohesive forecast for the time-series data. This aggregation ensures that the predictive outcomes are robust and reflective of the intricate dynamics encapsulated within the dataset.

In essence, the prediction mechanism intricately weighs each partition's mid-point in light of potential transitions facilitated by temporal segments. This involves meticulously considering the edge-weight of the directed edge originating from the current partition vertex (cur_v) to the respective partition vertex (i_v). The edge-weight ((cur_i,)v_v plays a crucial role in determining the significance of the transition.

Additionally, the mid-point of each partition serves as a pivotal factor in this predictive calculus. It represents a focal point within the time-series continuum, contributing to the nuanced understanding of potential future trends.

$$Pr = \frac{1}{4} \sum_{p=1}^{4} Pr_p \quad (17)$$

Summarily, the predictive process culminates in a weighted summation of partition mid-points, reflecting the probabilities of transitions induced by temporal segments. This meticulous synthesis ensures a comprehensive outlook on the potential trajectories within the time-series.

Furthermore, to derive the overall prediction of the model, these individual predictions are amalgamated through a process of averaging. This aggregation strategy ensures that the predictive outcomes are robust and reflective of the intricate dynamics encapsulated within the dataset.

2.6 EXPERIMENTS AND RESULTS

In this section, we unveil the comprehensive findings derived from our extensive experimentation on the TAIEX economic close-price time-series dataset spanning the substantial timeframe from 1990 to 1999. Our analytical approach meticulously involves segmenting the time-series into seven distinct partitions for each annual cycle. These partitions are strategically designed to encapsulate a broad spectrum of market behaviors, ranging from conditions categorized as EXTREMELY LOW to those denoted as EXTREMELY HIGH.

Each partition serves as a crucial delineation, capturing the dynamic range of price movements and economic indicators prevalent during the specified period. It's imperative to emphasize our deliberate choice of partition count, carefully balancing granularity with practicality. A higher number of partitions could potentially clutter the predictor graph with numerous disconnected nodes, rendering the interpretation and analysis challenging.

By opting for this specific partition scheme, we ensure that each transition observed in the graph is substantively significant, driven by temporal segments comprising consecutive data points. This meticulous strategy underscores the depth and precision of our analytical framework, facilitating a nuanced exploration of market dynamics and trends within the TAIEX dataset.

Throughout the years from 1990 to 1999, our model undergoes a rigorous training regimen, meticulously honing its predictive capabilities. This training spans across the initial ten months of each year, from January to October. Following this intensive training phase, the model's predictive prowess is put to the test during the subsequent months of November and December, serving as the evaluation period.

It's worth noting that our model's predictions are not merely isolated guesses but rather calculated projections based on the expected duration of an average segment length. Thus, each prediction is rooted in a thorough understanding of the underlying time-series dynamics.

To offer a visual representation of our training methodology and predictive outcomes, we present Figure 7, which meticulously showcases the segmented and partitioned time-series. Additionally, this figure provides insight into the predictor graphs derived during the intensive training phase, illustrating the model's ability to discern patterns and anticipate future trends.

Furthermore, to assess the effectiveness of our proposed model, we conduct a comprehensive comparative analysis with other existing models documented in the literature. Leveraging the

widely recognized RMSE error metric, we meticulously evaluate the predictive performance across various models. Through this thorough evaluation process, we aim to demonstrate the superior predictive capabilities of our proposed model compared to existing approaches.

During the experimentation phase, the computational tasks were conducted using MATLAB, a widely-used software platform renowned for its versatility in numerical computing and data analysis. These operations were performed within the Windows 11 operating system environment. The hardware configuration employed for the experiments featured an Ryzen 3 processor, operating at a clock speed of 4.50 GHz, coupled with a substantial RAM capacity of 16 GB.

Throughout the extensive experimental period spanning from 1990 to 1999, the average duration required for model training was meticulously recorded. The computation time, essential for evaluating the efficiency and performance of the model, was meticulously monitored and meticulously logged. The culmination of these efforts yielded an average training time of 2.6832 seconds, providing valuable insights into the computational demands and capabilities of the proposed methodology.

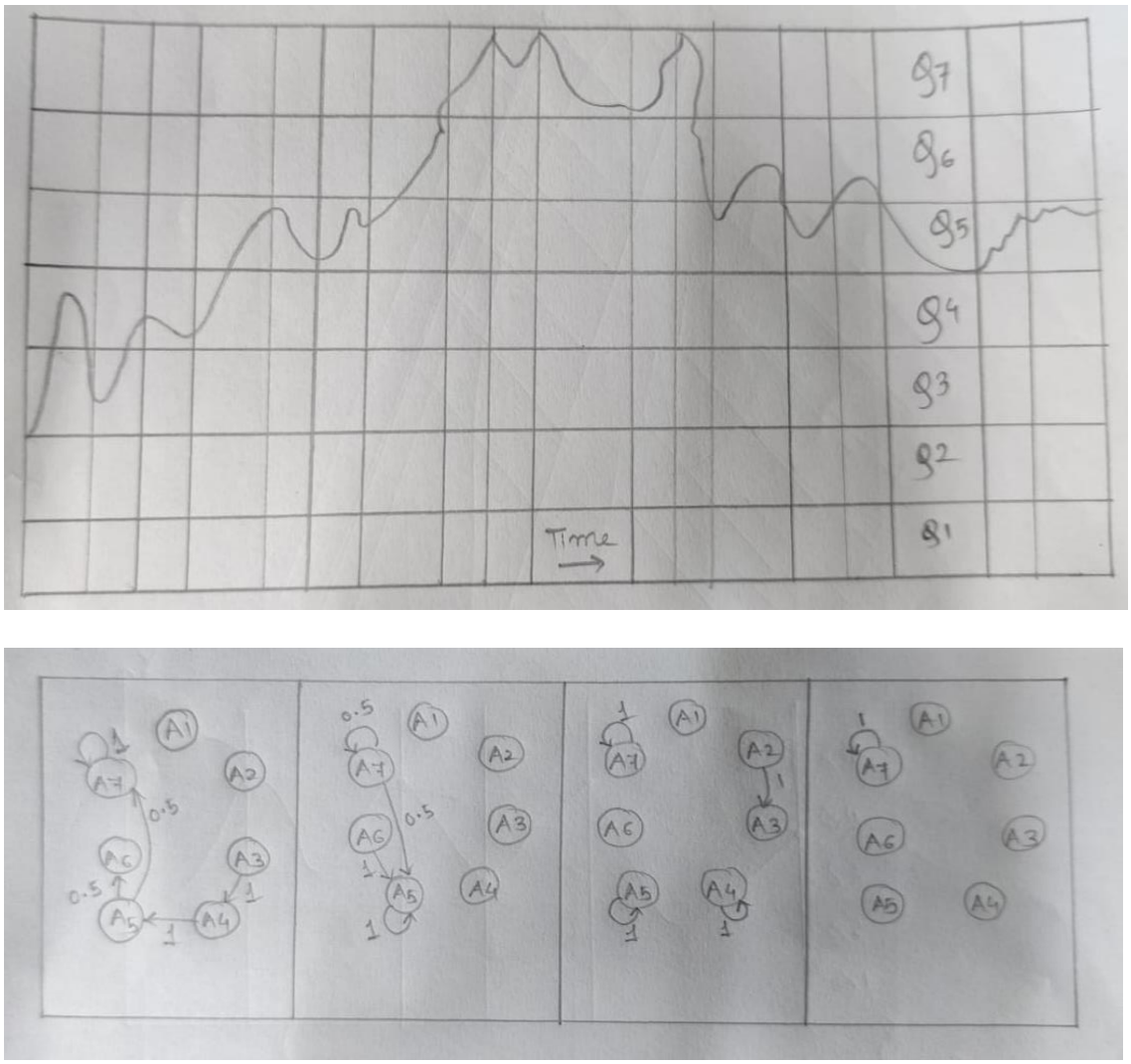


Fig:2.6 Segmented TAIEX close-price time-series yielded predictor graphs for 1991

It's essential to highlight that empirical parameter selection plays a crucial role in the performance of each algorithm. In the case of the EPST segmentation algorithm, meticulous experimentation has revealed that employing a slope-window width of 10 data points yields optimal results. This involves a meticulous process of normalizing slope variances across all windows to ensure consistency within the [0, 1] range. Subsequently, a threshold value of 0.8 is judiciously chosen to guide the segmentation process.

Similarly, in the fuzzy matching algorithm, the scope of fuzzy similarity values spans from a minimum of 0 to a maximum of 10. Through rigorous experimentation, it's established that setting the rejection threshold parameter at 2 produces the most favorable outcomes, ensuring accurate classification of temporal segments.

A comprehensive examination of Table 1 unmistakably showcases the remarkable superiority of the proposed model. With an average RMSE error of 88.48, it decisively outperforms its closest competitor by a margin of 1.0733%. This unequivocal success underscores the model's robustness and positions it as a highly effective tool for real-world applications in stock market trading and investment realms.

TABLE I
COMPARISON OF PREDICTION PERFORMANCE OF PROPOSED MODEL WITH EXISTING MODELS USING THE RMSE METRIC

Years / Models	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	Average
1.Conventional models [5]	220	80	60	110	112	79	54	148	167	149	117.9
2.Weighted models [5]	227	61	67	105	135	70	54	133	151	142	114.5
3.Chen and Chen [13]	172.89	72.87	43.44	103.21	78.63	66.66	59.75	139.68	124.44	115.47	97.70
4. Chen *et. al* [8]	174.62	43.22	42.66	104.17	94.6	**54.24**	50.5	138.51	117.87	101.33	92.17
5. Chen and Kao [11]	**156.47**	56.50	36.45	126.45	**62.57**	105.52	51.50	125.33	104.12	**87.63**	91.25
6.Cai *et. al* [12]	187.10	39.58	39.37	**101.80**	76.32	56.05	**49.45**	123.98	118.41	102.34	89.44
7.Proposed model	184.65	**34.42**	**35.21**	109.63	67.96	78.21	52.34	**115.81**	**101.69**	104.83	**88.48**

2.7 CONCLUSION

In conclusion, this chapter has presented a comprehensive methodology tailored for the effective modeling and forecasting of time-series data. Through a meticulously designed framework encompassing three distinct stages, we have demonstrated the ability to extract valuable insights and leverage them towards informed predictions.

The initial stage of semantic segmentation, facilitated by the innovative Edge Profile Time-Series Segmentation (EPTS) algorithm, has proven instrumental in partitioning the time-series data into coherent temporal segments characterized by distinct dynamic slopes. This foundational step sets the stage for subsequent analysis and interpretation.

Building upon this groundwork, the integration of a sophisticated fuzzy matching technique in the second stage has enabled the correlation of temporal segments with elemental primitive patterns. Notably, the real-time classification capabilities of this fuzzy matching scheme offer a significant advantage over traditional pattern classifiers, enhancing efficiency and reducing computational overhead.

In the final phase, the culmination of insights is encapsulated within a set of four weighted directed graphs, each intricately linked to a specific primitive pattern. These graphs serve as powerful tools for predictive analytics, offering both granular insights into the dynamics of individual primitive patterns and the derivation of weighted probabilistic transition rules of varying orders.

Empirical validation, particularly on the TAIEX close-price economic time-series, has underscored the efficacy and robustness of the proposed methodology. The attainment of a notably low Root Mean Square Error (RMSE) compared to existing fuzzy time-series models serves as a testament to the reliability and predictive accuracy of our approach.

In summary, the methodology outlined in this paper represents a significant advancement in the realm of time-series analysis and forecasting. Its versatility, efficiency, and ability to deliver accurate predictions make it a valuable asset across a diverse array of domains and applications. Moving forward, further exploration and refinement of this methodology hold promise for even greater advancements in time-series modeling and forecasting capabilities.

2.8 REFERENCES

[1] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338-353, 1965.

[2] Q. Song and B.S. Chissom, "Fuzzy Time Series and Its Model," *Elsevier, Fuzzy Sets Syst.*, vol. 54, no. 3, pp. 269–277, 1993.

[3] Q. Song and B.S. Chissom, "Forecasting Enrollments With Fuzzy Time Series—Part I," *Elsevier, Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, 1993.

[4] Q. Song and B.S. Chissom, "Forecasting Enrollments With Fuzzy Time Series—Part II," *Elsevier, Fuzzy Sets Syst.*, vol. 62, no. 1, pp. 1–8, 1994.

[5] H. K. Yu, "Weighted Fuzzy Time Series Models for TAIEX Forecasting," *Elsevier, Physica A*, vol. 349, no. 3-4, pp. 609-624, Apr. 2005.

[6] K. Huarng, "Heuristic Models of Fuzzy Time Series for Forecasting," *Elsevier, Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 369-386, Nov. 2001.

[7] K. Huarng, T. H. K. Yu and Y. W. Hsu, "A Multivariate Heuristic Model for Fuzzy Time-Series Forecasting," *IEEE Trans. Sys. Man Cybern. Part B,* vol. 37, no. 4, pp. 836-846, Aug 2007.

[8] S.M. Chen, H.P. Chu and T.W. Sheu, "TAIEX Forecasting Using Fuzzy Time Series and Automatically Generated Weights of Multiple Factors," *IEEE Trans. Sys. Man Cybern. Part A*, vol. 42, no. 6, pp. 1485-1495, 2012.

[9] S.T. Li, Y.C. Cheng and S.Y. Lin, "A FCM-Based Deterministic Forecasting Model for Fuzzy Time Series," *Elsevier, Computers and Mathematics with Applications*, vol. 56, no. 12, pp. 3052-3063, Dec. 2008.

[10] S.S. Gangwar and S. Kumar, "Partitions Based Computational Method for High-Order Fuzzy Time Series Forecasting," *Elsevier, Expert Systems with Applications,* vol. 39, no. 15, pp. 12158-12164, Nov. 2012.

[11] S.M. Chen and P.Y. Kao, "TAIEX Forecasting Based on Fuzzy Time Series, Particle Swarm Optimization Techniques and Support Vector Machines," *Elsevier, Information Sciences,* vol. 247, pp. 62-71, Oct. 2013.

[12] Q. Cai, D. Zhang, W. Zheng and S.C.H. Leung, “A New Fuzzy Time Series Forecasting Model Combined with Ant Colony Optimization and Auto-Regression,” *Elsevier, Knowledge Based Systems,* vol. 74, pp. 61- 68, 2015.

[13] S.M. Chen and C.D. Chen, “TAIEX Forecasting Based on Fuzzy Time Series and Fuzzy Variation Groups,” *IEEE Trans. Fuzzy Syst.,* vol. 19, no. 1, pp. 1-12, 2011.

[14] TAIEX [Online]. Available: http://www.twse.com.tw/en/products/indices/tsec/taiex.php

[15] L.G. Shapiro and G.C. Stockman, “Computer Vision,” pp. 137-150, Prentice Hall, 2001.

[16] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*,4th ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 2007.

[17] S. Picek, M. Golub, and D. Jakobovic, “Evaluation of crossover operator performance in genetic algorithms with binary representation,” in *Proc. 7th Int. Conf. Intell. Comput. Bio-Inspired Comput. Appl.*, 2011, pp. 223–230.

[18] Zar, J.H., Biostatistical Analysis. Prentice Hall. 1999.

CHAPTER 3

Conclusion

&

Future Scope

3.1 Conclusion

By splitting data of the TIME-SERIES to discrete segments of temporal of non-uniform length, each with constant dynamic slope, the suggested EP TimeSeries method effectively finds ideal segment boundaries. The initial stage of this approach is computationally light, which is advantageous in real-world applications. These segments are categorized into one of four primitive patterns in the second stage using a fuzzy matching technique: linear rise, linear fall, Gaussian bell curve, and inverted Gaussian curve. Compared to conventional pattern classifiers that require a training phase, fuzzy matching offers a substantial advantage because it does not require training and can classify data in real-time.

Using a collection of four weighted directed graphs, each representing a basic pattern, the information is represented in the third step. These graphs can be used directly to make predictions or to create weighted probabilistic transition rules, either higher or lower order. Furthermore, each of the separate graphs offers insightful information on how each fundamental pattern affects the time-series data.

The suggested approach outperforms other fuzzy time-series models in terms of RMSE error value, according to experimental findings utilizing the TAIEX close-price economic time-series. As a result, time-series forecasting and prediction applications find great success with this approach.

3.2 Future Scopes

Fuzzy time series prediction has a bright future ahead of it, with continuous study and possible uses in many fields. The following are some directions for further research and use:

Integration with Deep Learning: More potent and precise time series prediction models may result from the integration of fuzzy logic with deep learning methods like neural networks and recurrent neural networks (RNNs). Whereas fuzzy logic can accommodate imprecision and uncertainty and produce predictions that are more resilient, deep learning systems are capable of discovering intricate patterns and relationships within the data.

Hybrid Models: Creating hybrid models that combine econometric models or machine learning algorithms with fuzzy time series approaches may increase the precision and dependability of forecasts. Hybrid models are able to overcome the shortcomings of individual techniques by utilizing the advantages of hybrid models, which in turn offer more complete forecasting solutions.

Applications in IoT and Cyber-Physical Systems: As Internet of Things (IoT) devices and cyber-physical systems proliferate, the demand for predictive analytics to assist in making decisions in dynamic, real-time contexts is increasing. To foresee future trends and enhance system performance, fuzzy time series prediction can be used in fields including energy management, smart manufacturing, and predictive maintenance.

Dynamic Fuzzy Time Series Models: It's crucial to conduct research on creating dynamic fuzzy time series models that can adjust over time to shifts in the underlying data patterns. In order to maintain accuracy and relevance of predictions in changing contexts, these models have to have the ability to constantly change their parameters and rules in response to new data.

Improving the interpretability and explainability of fuzzy time series prediction: Models is essential to their acceptance in practical settings, particularly those where decision-makers must comprehend the rationale behind forecasts. Developing transparent fuzzy models and visualization strategies to assist users in understanding and relying on the model's outputs might be the main emphasis of this field of study.

Applications in Healthcare and Biomedicine: Personalized medicine, illness prognosis, and patient monitoring are just a few of the uses for fuzzy time series prediction in the healthcare and biomedicine fields. Fuzzy models can assist physicians in identifying high-risk patients, anticipating disease outbreaks, and customizing treatment approaches based on the needs of each patient by evaluating past patient data and medical records.

Online learning and adaptive systems: Future research should focus on creating online learning algorithms and adaptive systems that are able to continually update and enhance fuzzy time series models in real-time. Predictions should be accurate and dependable over time because to these systems' ability to adjust to variations in the environment and data distribution.

To enable their broad implementation in real-world decision-making situations, the future of fuzzy time series prediction mostly consists of improving the state-of-the-art models, investigating new application areas, and resolving practical obstacles.

3.3 REFERENCES

1. *Abarbanel, Henry (Nov 25, 1997). Analysis of Observed Chaotic Data. New York: Springer. ISBN 978-0387983721.*
2. ^ Boashash, B. (ed.), (2003) *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Elsevier Science, Oxford, 2003 ISBN 0-08-044335-4
3. ^ *Nikolić, Danko; Mureşan, Raul C.; Feng, Weijia; Singer, Wolf (March 2012). "Scaled correlation analysis: a better way to compute a cross-correlogram". European Journal of Neuroscience. **35** (5): 742–762. doi:10.1111/j.1460-9568.2011.07987.x. PMID 22324876. S2CID 4694570.*
4. ^ Jump up to: **a** **b** *Sakoe, H.; Chiba, S. (February 1978). "Dynamic programming algorithm optimization for spoken word recognition". IEEE Transactions on Acoustics, Speech, and Signal Processing. **26** (1): 43–49. doi:10.1109/TASSP.1978.1163055. S2CID 17900407.*
5. ^ *Mormann, Florian; Andrzejak, Ralph G.; Elger, Christian E.; Lehnertz, Klaus (2007). "Seizure prediction: the long and winding road". Brain. **130** (2): 314–333. doi:10.1093/brain/awl241. PMID 17008335.*
6. ^ *Land, Bruce; Elias, Damian. "Measuring the 'Complexity' of a time series".*
7. ^ *Chevyrev, Ilya; Kormilitzin, Andrey (2016). "A Primer on the Signature Method in Machine Learning". arXiv:1603.03788.*
8. ^ *Ropella, G.E.P.; Nag, D.A.; Hunt, C.A. (2003). "Similarity measures for automated comparison of in silico and in vitro experimental results". Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439). pp. 2933–2936. doi:10.1109/IEMBS.2003.1280532. ISBN 978-0-7803-7789-9. S2CID 17798157.*
9. ^ *Tominski, Christian; Aigner, Wolfgang. "The TimeViz Browser:A Visual Survey of Visualization Techniques for Time-Oriented Data". Retrieved 1 June 2014.*