A Project Report

On

Fuzzy Logic in TimeSeries Data Analysis

BY

Chandrahas Aroori

2016A7PS0100H

Under the supervision of

Dr Jabez Christopher

SUBMITTED IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS OF

CS F367 : LAB ORIENTED PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

SEPTEMBER 2019

ACKNOWLEDGMENTS

I would like to thank Dr. Jabez Christopher for the opportunity to pursue this project. I would also like to thank BITS Pilani Hyderabad Campus introducing the project system.

Birla Institute of Technology and Science-Pilani,

Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Fuzzy Logic in TimeSeries Data Analysis**" submitted by Mr/Ms. Chandrahas Aroori (ID No. 2016A7PS0100H) in fulfillment of the requirements of the course CS F367, Lab Oriented Project Course, embodies the work done by him under my supervision and guidance.

Date:

(Dr Jabez Christopher)

BITS- Pilani, Hyderabad Campus

ABSTRACT

Spatio-Temporal data has now permeated the everyday life of the modern world, from health data to stock market data. Processing this data requires specialized methods to ensure that the underlying component of time is clearly understood by the prediction methods. One of the most common methods associated with spatio-temporal data is forecasting, i.e. the prediction of values based on the previously known data in the time axis.

This project attempts to introduce fuzzy systems into spatio-temporal data driven forecasting. Fuzzy systems are dependent on a branch of mathematics, Fuzzy Logic. In which there are degrees of truth, instead of the 1 and 0 system of Boolean Logic, Fuzzy variables can take values in between 0 and 1. Fuzzy systems work well with missing or sporadic data as they can represent the uncertainty present within the data. We attempt to see the improvement obtained by using a fuzzy inference system and measure the improvement. In the pre-processing absent values are inputted with the forecasts obtained by Holts double exponential smoothening. The obtained values are then put through common regression methods such as Artificial Neural Network, Decision Tree Regression, Random Forest Regression, Support Vector Machine Regression.

Surprisingly, for our data, the Random Forest Regression seems to have done the best job without allowing over-fitting. The Artificial Neural Network seems to have done the poorest even with many iterations, possibly because the amount of data was not enough for the model to learn.

This project wishes to ascertain the improvements of augmenting the current regression systems with classical Fuzzy Systems, both in pre-processing and post-processing.

CONTENTS

1 - Introduction

With the increase in data collection, newer and newer methods are being introduced for data pre-processing and data-mining. Temporal data is very different than other types of data, and requires its own pre-processing and data-mining methods to enable proper forecasting. Temporal data is data which contains a time dimension, i.e. data is collected over a period of time. Usually the period of time remains relatively constant, if the time period is not constant, we say that the data obtained is sporadic.

Data itself within the past couple of years has increased tremendously. Most of the common items in the world has now become computerized. Recent statistics show that by 2020, every person on earth will generate 1.7 megabytes per second. Newer tools are also being generated for these types of data. A few of these tools are, NoSQL databases, Hadoop, and Spark. The sharp rise in data can be attributed to both the rise of the internet and systems like IoT (Information of Things), which is all aout increasing the computerization of everyday objects.

Data itself are of many types. We can split data into many types based on its application source, nature of its correlation, nature of its relation with time and nature of its origin. For our needs we can classify two types of data, static(/persistent) and dynamic data. Dynamic data is information that is periodically &/or non-periodically updated. This data basically changes with the passage of time as we add the current moment's data. Persistent data is data that doesn't change with the passage of time, this can mean that the total amount of data, is not corelated with time, or that in the future there will not be more of this data present.

A commonly associated problem with temporal data is sporadic observation. Most real time data is dependent on a variety of real-time factors and is subject to many delays. Thus, most pre-processing methods for temporal data are related to inputting absent values. This paper wishes to solve this problem by using a fuzzy system for the pre-processing that takes the uncertainty with inputting these absent values into account.

1.1 - Components of Time Series Data

Most time series can be described by three components. These components are level, trend, seasonality and noise. We must also define systematic and unsystematic components.

- **Systematic:** Components of time series data that are recurring in nature and can be modelled. These can be both linear and non-linear.
- **Unsystematic:** Components of time series data that are not recurring in nature and cannot be modelled.

We chose to express a time series data with both systematic components and unsystematic components. Systematic components are, level, trend, seasonality. And the unsystematic component is noise.

The decomposition of the given data into these components can be either, additive, or multiplicative or logarithmic.

Exponential Smoothening is a technique devised in the late 1950's and is one of the most commonly used time series forecasting technique. In exponential smoothening, the forecasts produced are the weighted averages of the past observations. So, more recent observations hold more weight than the older observations, and yet older observations still affect the forecasts. These smoothening techniques work quite well to model ideas of level and trend, and also help in eliminating noise, but lack the power to capture long term seasonality variations.

There are many issues while working with time series data. One of them is, analysing the impact of single events. While forecasting, we may not be sure whether the recent spike was due to an anomaly or part of a long-term seasonal component. This usually stems from 'correlation does not imply causation', and the fact that real life data has many hidden variables.

Another problem is, the dependency of causal patterns. While studying time series data, we tend to look for the effect of one variable on another. Implying the existence of one independent variable and one dependent variable. But in real life there exists no such purely dependent and independent variable.

Other problems lie in the data collection itself. Most models work best with evenly spaced data without missing values, but due to the nature of time series data, such an ideal situation is rarely possible. Most time series data is sporadic and or unevenly spaced and littered with missing values from data loss. To fix these problems can improve the overall output of the system.

1.2.1 - Machine Learning and Time Series Data

With the recent increase in machine learning, data scientists are now retrying to improve older models that once had hit the limits of what could be achieved.

Machine learning itself can be defined as the way of using algorithms and statistical models to enable computer systems to perform tasks without explicitly giving instructions on how the task must be performed, but to build the model by training it on real life data so that the computer system can use this model to make decisions.

Much research is underway in areas of time series data, such as imputation, time series classification, time series forecasting, stock market prediction, traffic prediction, time series clustering, irregular time series forecasting, time series denoising, time-to-event prediction and more.

The major issues with Machine Learning and time series data is that unlike most data that Machine Learning models have been trained on, time series data inherently contains a different set of challenges that must be faced. Time series data tend to have a time-based correlation and exhibit significant autocorrelation. Autocorrelation is the correlation of a signal with a previously delayed copy of itself and can be expressed as a function of itself with a delay parameter introduced. Cross correlation between variables also plays a major role in time series data.

Most machine learning models cannot express the inherent uncertainty present within the variables of time series data. For example, the duration variable, i.e. the distance between measurements of data must have a certain uncertainty present within it, due to missing data values. To solve most of these problems we can use fuzzy systems to allow for these uncertainties.

1.2.2 - Machine Learning Algorithms for Time Series Data

There have been quite a number of machine learning algorithms crafted to learn temporal-based information without hand-crafting features. Most common methods are Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU).

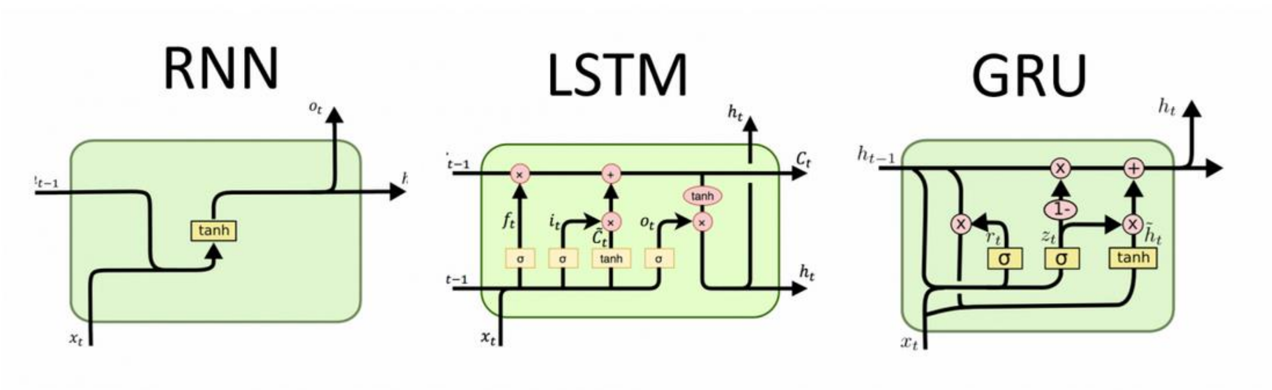

Figure 1: Comparison in the Neurons of an RNN, LSTM, & GRU

RNN, also known as Recurrent Neural Network, perform the same task for each element of the sequence with the output of each neuron being dependent on the previous neuron and its own output in the previous computation. These Recurrent Neural Networks, have been very successful at time series data and natural language processing.

To address the problem of short-term memory present with Recurrent Neural Networks, we try to build in a memory module in Long Short-Term Memory (LSTM's), with gates. These gates have the ability to learn which data is most important or which is irrelevant.

Gated Recurrent Units (GRU) also work in the same manner as Long Short-Term Memory (LSTM's), and is good at processing long sequences.

1.3.1 - Fuzzy Systems

Traditionally most tools for modelling, reasoning and computing work with well-defined and precise inputs. In conventional logic theory, we have a dichotomous, true or false. There can be no in between values present. Problems that arise from this precision are that, real situations are not deterministic and cannot be described precisely. A full description of the system would require more data and information that can possibly be obtained.

To address these problems, a fuzzy set theory has been designed. Fuzzy logic is the way of defining the imprecision present with precision. We use imprecision as the vagueness associated with the variable rather than the absence of knowledge.

The advantages to using such a system are:

a) **Modeling of uncertainty:** Uncertainty of data and or systems can be represented using this technique.
b) **Compactification of data:** Data reduction using insertion into a fuzzy system can lead to the reduction of massive data into clusters that can be easily understood.
c) **Relaxation of duality:** Most methods are sharply defined as a dual system even when there are cases which do not fit into this model.
d) **Preserving Meaning Reasoning:** Using fuzzy logic we can now build inference engines that can embed the inherent uncertainty of inputs and outputs allowing powerful knowledge processing to be available.
e) **Efficient estimation of Approximate Results:** The usage of fuzzy logic to solve approximate solutions of real problems efficiently can allow greater complexity to be embedded within the solution.

We define a fuzzy set through the use of a Membership function (grade of membership). This is the point of which a value x lies within the given set with a value between 0 and 1.

These fuzzy sets now can have operations defined on them, standard operations such as intersection, union, complement. We use the membership function along with the min and max operator to define these operations.

This now leads to the definition of Type II fuzzy systems. Type 2 fuzzy sets are fuzzy sets whose membership functions' values are itself type I fuzzy sets through [0, 1]. Similarly a Type 'm' fuzzy system can be defined as, a fuzzy set in Y whose membership functions' values itself are Type (m-1) fuzzy sets through the values [0, 1].

1.3.2 - Fuzzy Control Systems

The fuzzy logic control systems seek to control complex processes though the use of defined uncertainty. Most control systems are designed by modelling the real world into the computational world. Since building such a model is complex, certain information is naturally simplified. To codify the possible complexity, we use fuzzy logic and its systems to handle this.

Most automatic control system all work with the design of the automatic control system. The control system reaches the desired states through the adjustment of the input values and through the measurement of the output values. Through constant tweaking of the input-ouput system we are able to obtain the required state.

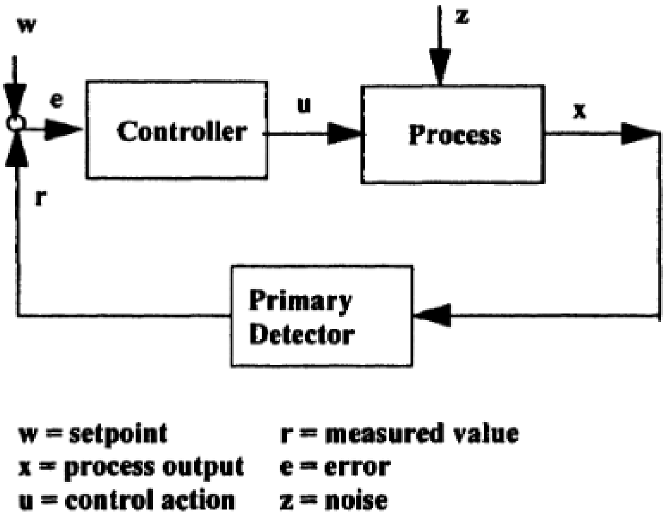

Figure 2: Automatic Control System

Fuzzy control systems use rules to model this 'knowledge'. Instead of using an algorithmic implementation, the designer designs knowledge-based rules for the Fuzzy logic system to use. The rule structure is commonly seen as, IF condition, THEN change_condition. The conditions themselves are expressed using fuzzy logic. This is where the uncertainty is placed in the model.

1.3.3 - Fuzzy Inference Systems

These systems also use 'IF, THEN' based rules along with logical and/or connectors to build a rule set. The output obtained from such a system is always crisp, irrespective of its input which can be either fuzzy or crisp. The blocks that comprise of such a system are:

a) **Rule Base** − The antecedent and the consequent rule system is stored in here.
b) **Database** – The fuzzy set definition along with the corresponding membership functions are placed here. This is different than the rule base as this defines the fuzzy system, and rule base defines the fuzzy rules.
c) **Decision-making Unit** – This computes the decision based on the fuzzy rules.
d) **Fuzzification Interface Unit** − This transforms the non-fuzzy inputs into fuzzy outputs.
e) **Defuzzification Interface Unit** − This transforms the fuzzy inputs into crisp outputs.

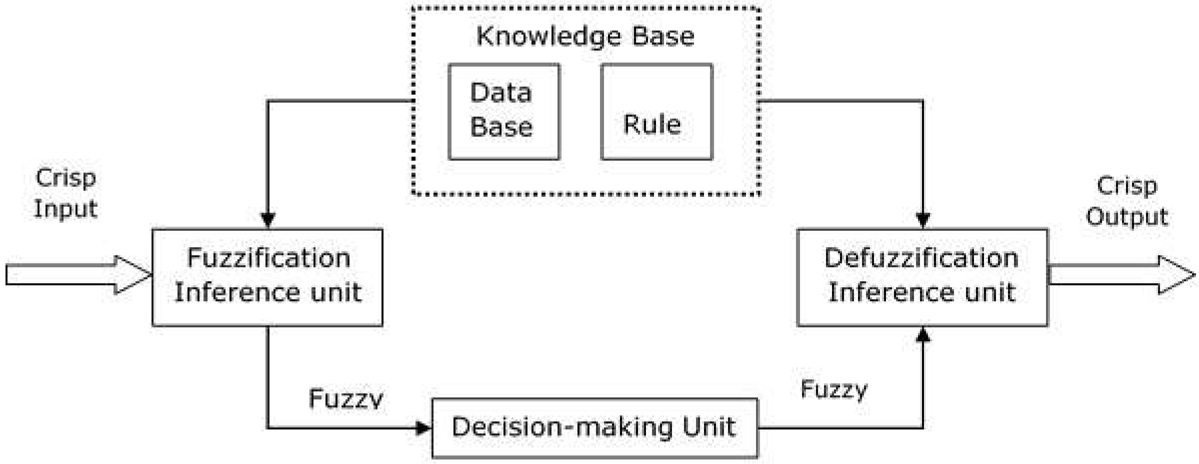

Figure 3: Fuzzy Inference System

1.3.4 - Mamdani Fuzzy Inference System

Mamdani controller describes process states through the use of linguistic variables and then uses these as inputs to control rules. We can then use this to visualize the rule base as a tabulated entry. With given IF conditions vs the existing THEN conditions.

Process for Mamdani Fuzzy Inference System:

1. Determine the set of fuzzy based rules to follow.
2. Fuzzify the given input values based on the defined fuzzy function.
3. Combine the now fuzzy input with the defined fuzzy rules, to get the rule strength.
4. Determine which consequent must be followed by using the rule strength and the output membership function.
5. Combine all the consequents to get the distribution of the output.
6. Defuzzify the output distribution to obtain the final output

1.3.5 - Takagi-Sugeno Fuzzy Inference System

In this model we change the rule statements itself to be more accommodating. The output Z is a crisp output and is not fuzzy. The format followed by Takagi-Sugeno is,

IF m is P and n is Q THEN Z = f(m, n)

Process for Takagi-Sugeno Fuzzy Inference System:

1. Fuzzify the given input values based on the defined fuzzy function.
2. Apply the given fuzzy operator f(.) to the outputs to obtain the result.

2 - Dataset

The given dataset is of a Biogas digester. There are two types of bacteria present in the digester, Thermophilic Bacteria along with Mesophilic Bacteria. Thermophilic Bacteria are a type of extremophile, that survive best at higher temperatures, between 45 and 122 °C. Mesophilic Bacteria live in a common room temperature range. The commonly associated temperature ranges with these bacteria are between 20 and 45 °C.

The given dataset has 4 parameters measured, pH, Alkalinity (mg/L), Biogas (m3/m3/d), TS. pH is a way of calculating the amount of hydrogen ion [H^+] present within a solution. Alkalinity measures the ability of a solution/substance to. It is also a measure of the capacity of the solution to act as a buffer. Buffers are solutions to which even if a little acid is added the concentration of H^+ ions (i.e. pH) are not changed appreciably. Biogas is a gaseous mixture of methane and other hydrocarbons, produced anaerobically (i.e. in the absence/minimal amount of oxygen), by the breakdown of organic matter. Biogas can be produced from any organic material, commonly used raw materials are associated waste products from other processes. This can be human waste, sewage or agricultural waste. Total Solids (TS) is a measure of the suspended and dissolved solids in waste water.

The dataset itself presents itself with a number of missing values and requires imputation to be further analysed.

3 – The System

The project is divided into the modules consisting of:

1. **Holt Double Smoothening & Plotting**

   This module calculates the Holt-Winters Double Exponential Smoothening for the given of data. After applying the smoothening, we save the data in a CSV(Comma Separated Value) file and we plot the graphs of TS vs Days.
2. **FIDES (Fuzzy Inference Double Exponential Smoothening)**

   This module comprises of the methods used to build FIDES. This involves defining a universe for the fuzzy sets, defining the Antecedent and Consequent rules. After which the Fuzzy Membership Functions must be defined, for this project we used all trapezoidal functions. And then the model has the output calculated and imputed via the use of a CSV (Comma Separated Value) file.
3. **Prediction-Module**

   This module is in charge of value prediction through the use of common regression methods. The regression methods we use are, Artificial Neural Network, Decision Trees, SVR (Support Vector Regression), Random Forests. The module loads the data, splits in into 80-20 batches and then runs the common regression method and then calculates the accuracy metrics involved. The accuracy metrics that we calculate for this paper are, R^2 value, Mean Absolute Error (MAE)a, Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

The project itself is built using Python3 and popular libraries like Sklearn, Numpy and Pandas for the data processing. For the fuzzy aspects we used the library SkFuzzy, which allowed easy fuzzy variable definition and rule definition.

4 - Experiment and Results

The objective of the paper is to perform imputation of missing data and regression on the provided data. Commonly, missing data is either deleted or imputed. Although many methods already exist for missing data imputation and sparse sporadic data, we wish to use a Fuzzy Inference System and compare it with traditional smoothening methods.

The dataset was initially processed, and missing input values for the TS, were inserted so that further data analysis could take place. Following Holt's two-parameter model, also known as linear exponential smoothening, further values were predicted with both windows of 1 and 5 days respectively.

$$
\begin{array}{lll}
\text{Forecast} & \hat{y}_{t+h|t} & = \ell_t + hb_t \\
\text{Level} & \ell_t & = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend} & b_t & = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1},
\end{array}
$$

Formulae Set 1: Holts Double Smoothening

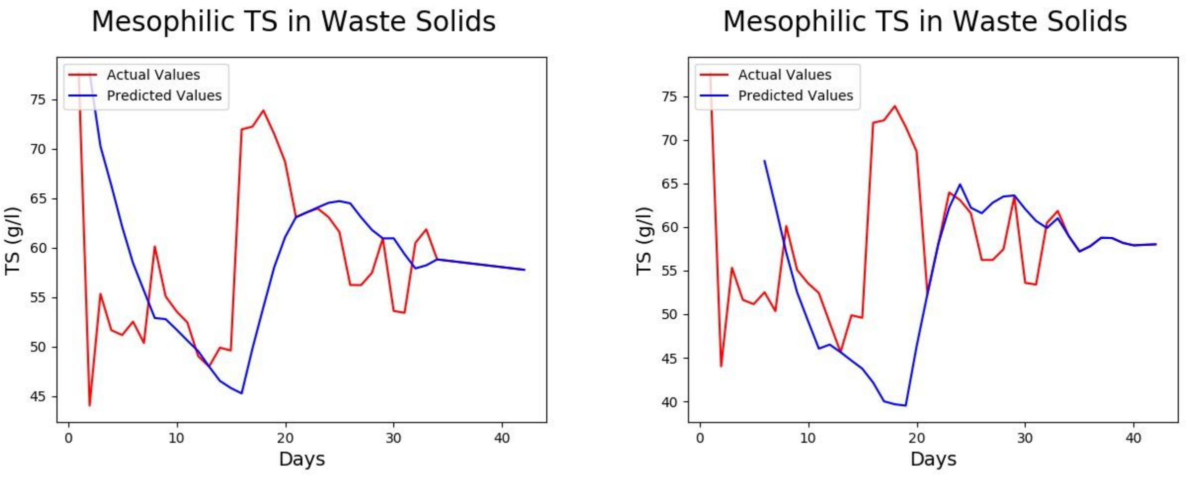

Figure 4: Value Prediction for Mesophilic bacteria over 5 days and 1 day respectively (Left to Right)

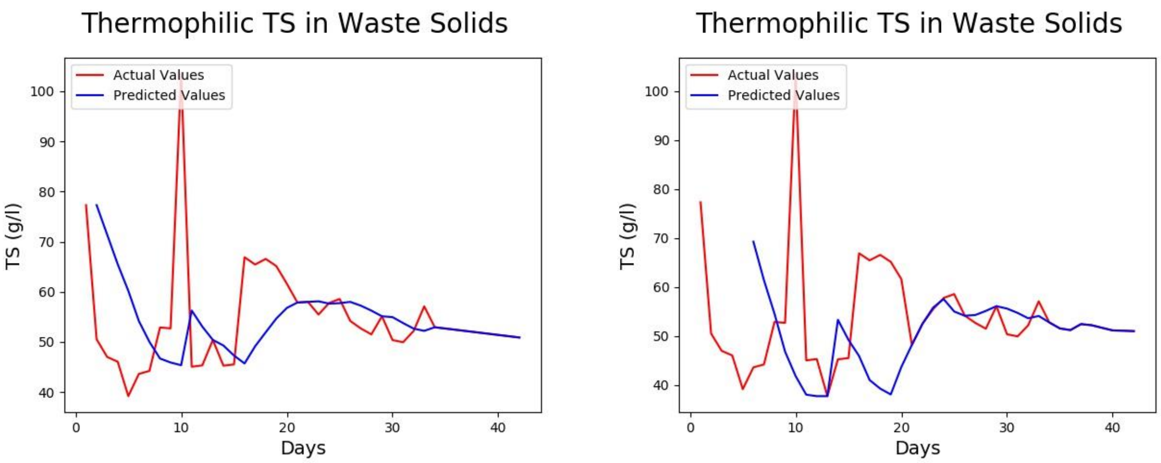

Figure 5: Value Prediction for Thermophilic Bacteria over 5 days and 1 day respectively (Left to Right)

As we have seen the window of 5 days offers too wide a window for our extrapolation purposes.

Using the now processed data, we fed the data into a variety of Machine Learning Algorithms, namely Artificial Neural Network, Decision Trees, Random Forests, Support Vector Machines. The results obtained were as shown in Figure 2

.

Regression Method	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	R^2 Value
Artificial Neural Network	6.91017302939	75.5839740211	8.6939044175	-2.1766182965199
Decision Trees	8.2562098628	388.500322101	19.71041151526	-15.3277632245
Random Forests	3.982460513385	30.85392032855	5.55463053033	-0.2967183732306
Support Vector Machines	6.29788985146	65.9500324568	8.12096253265	-1.771726182320

Figure 6: Values for each regression method

Now we use a Fuzzy Inference System named FIDES(Fuzzy Inference Double Exponential Smoothening) [1] for data imputation. FIDES work by using two separate Fuzzy Inference Systems to dynamically adjust the value of alpha and beta to use in Holt's two-parameter model. Thus, we have a Level FIS and a Trend FIS.

A fuzzy system works through, fuzzification, fuzzy rule base, implication-aggregation and, defuzzification. For our dataset the Fuzzy variables and their respective fuzzy sets are defined as follows.

FIS	Variable	Range	Fuzzy Set	Parameters
Trend	Trend	[50, 100]	*Decrease*	[-2.0, -2.0, -1.4, -0.4]
			Stable	[-0.6, -0.4, 0.4, 0.6]
			Increase	[0.2, 1.0, 2.0, 2.0]
	Duration	[0, 5]	*Minimum*	[0, 0, 0.5, 2]
			Average	[0.75,2.0,2.25,3]
			Maximum	[2.5, 4.5, 5, 5]
	Trend parameter	[0,1]	*Low*	[-2, -2, -1.4, -0.4]
			Medium	[-0.6, -0.4, 0.4, 0.6]
			High	[0.2, 1.0, 2, 2]
Level	Error rate	[0, 100]	*Low*	[0, 0, 4, 13]
			Medium	[10, 26, 34, 46]
			High	[44, 66, 100, 100]
	Duration	[0, 5]	*Minimum*	[0, 0, 0.5, 2]
			Average	[0.75, 2.0, 2.25, 3]
			Maximum	[2.5, 4.5, 5, 5]
	Level parameter	[0,1]	*Low*	[-2, -2, -1.4, -0.4]
			Medium	[-0.6, -0.4, 0.4, 0.6]
			High	[0.2, 1.0, 2, 2]

Figure 7: Parameter Values of Fuzzy Variables

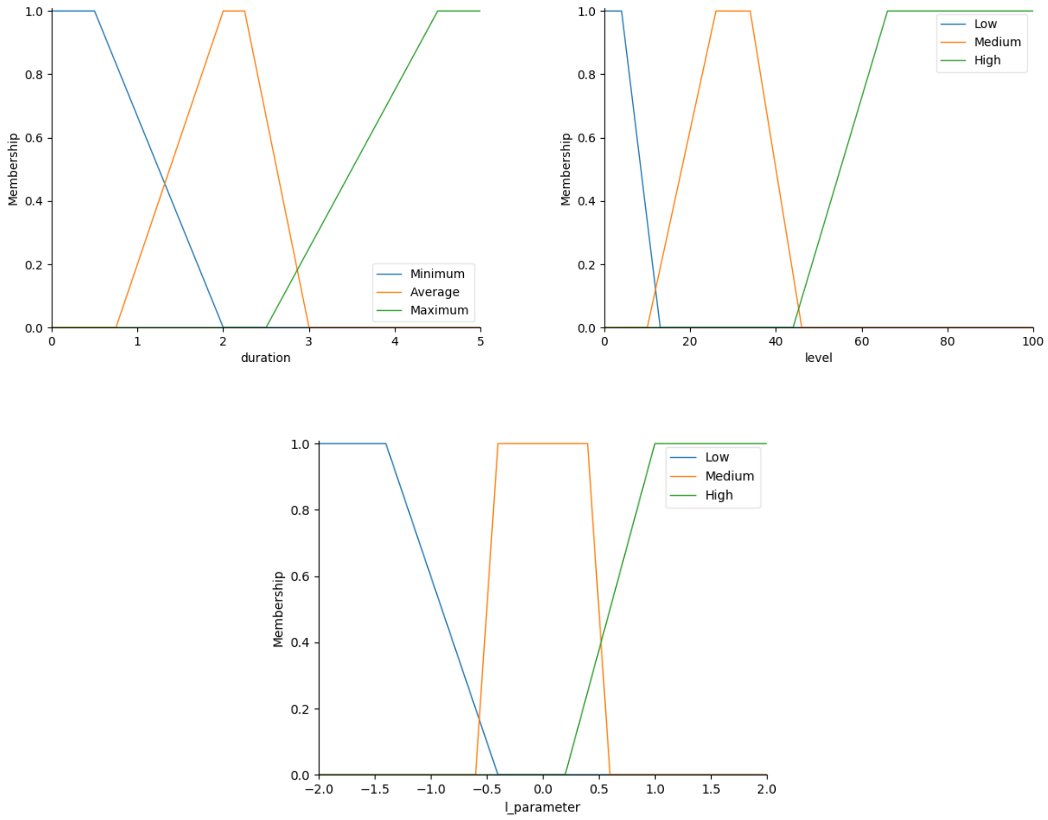

Figure 8: Membership Function Graphs (Level)

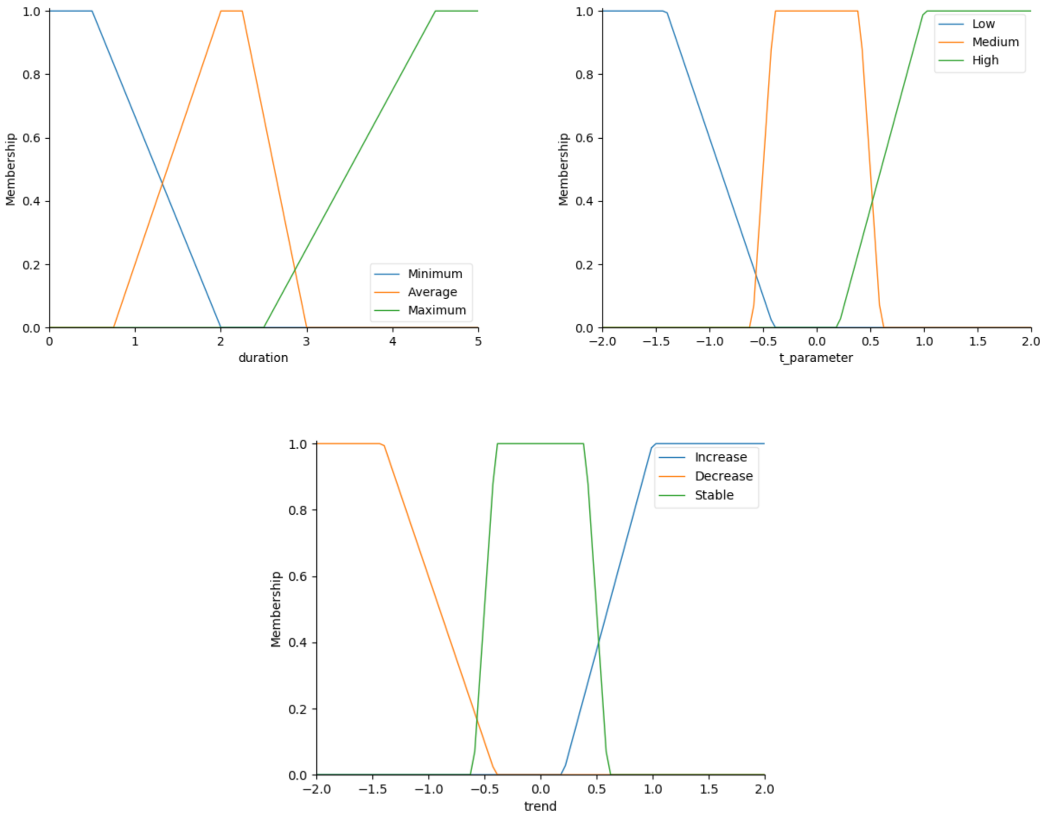

Figure 9: Membership Function Graphs (Trend)

Let A,B,C,D, & E be the complete universe, which represents Error, Duration, Trend, Mean Parameter and Trend Parameter respectively. Let us define the fuzzy sets, U,V,X,Y & Z universes where V= { Decrease, Stable, Increase} , W= {Minimum, Average, Maximum} , X = {Low , medium , high } represents fuzzy sets of C, B and A respectively & Y= {Low, Medium, High}, Z={Low, Medium, High} represents fuzzy sets of D & E.

We can thus define the fuzzy rule set as the following if then statements containing one consequent and two antecedents:

Level FIS: IF (A) is X[i] AND (B) is W[i] THEN (D) is Y[i]

Trend FIS: IF (C) is V[i] AND (B) is W[i] THEN (E) is Z[i]

Formulae Set 2: Antecedents & Concequents

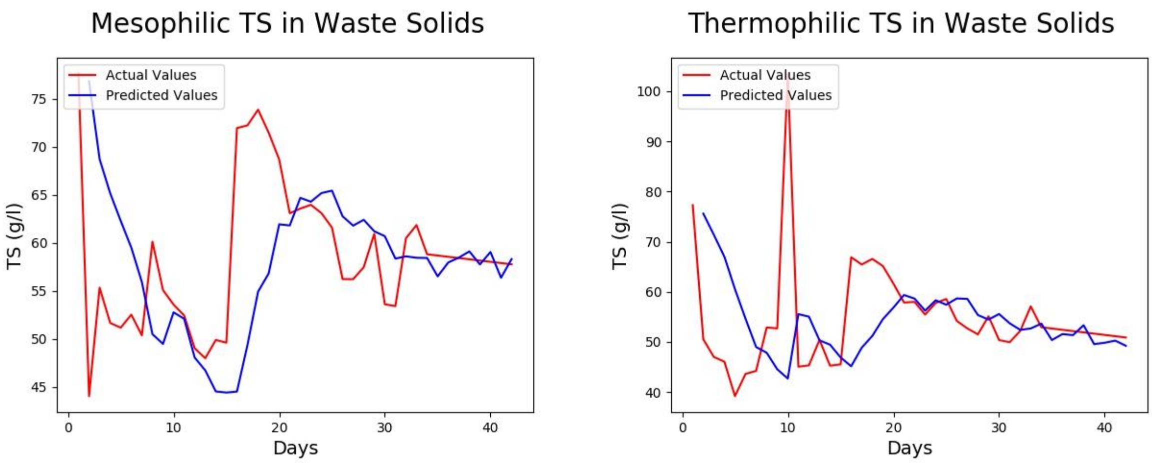

Figure 10: Value Prediction for Mesophilic & Thermophilic bacteria via FIDES

Using the now processed data, we fed the data into a variety of Machine Learning Algorithms, namely Artificial Neural Network, Decision Trees, Random Forests, Support Vector Machines. The results obtained were as shown in Figure 7.

Regression Method	**Mean Absolute Error (MAE)**	**Mean Squared Error (MSE)**	**Root Mean Squared Error (RMSE)**	**R^2 Value**
Artificial Neural Network	6.910173029391	75.58397402119	8.69390441753	-2.176618296519
Decision Trees	8.25620986285	388.5003221011	19.710411515265	-15.32776322457
Random Forests	3.9824605133859	30.853920328559	5.554630530337	- 0.29671837323063
Support Vector Machines	6.297889851464	65.95003245681	8.120962532656	-1.7717261823208

Figure 11: Values for each regression method

5 - Literature Review

Jane Y Nancy(2017) and team have attempted to solve the problem of missing data and/or null values, and sporadic data in clinical time-series data. They proposed a data mining model, named statistical tolerance rough set induced decision tree, also known as STRiD. Their proposed framework, allows temporal pre-processing, and classification. They have also used Fuzzy inference double exponential smoothing (FIDES) system to compensate the weaknesses of Double Exponential Smoothening, this leads to enhanced pre-processing. While attribute selection, the corresponding values are taken from a rough set, which then requires further de-fuzzification.

Jane Y. Nancy(2017) also presented another novel way to input absent values present in unevenly spaced, sporadic clinical temporal data. Classical methods such as IDW, experiences problems such as data point selection and influence factor selection. The proposed method wishes to solve these problems through the use of TR (tolerance rough set) analysis, and then the use of Influence Factor optimization (PSO), to find the influence factor to be optimized so that the weightage of the data point can then be fixed. They have obtained higher reduction of error rates than commonly used EM and KNN.

Doyup Lee(2019) and team have designed a model called Temporal-Guided Network (TGNet) which employs the usage of graph based networks and temporal-guided embedding. This method allows the complex patterns that are present in short-term demand forecasting models, and also exploits the long-term patterns present such as seasonality, and higher order periodicity. The inherent advantage of the graph networks is that it can extract immutable patterns present in the adjacent positioned data, while temporal-guided embedding allows understanding the temporal hidden contexts present within the data.

According to the research by Jiaqiu Wang(2007) and peers, linear addition of various factors is a simplification of the complexity present in the actual data. Through the use of Support Vector Machines, they choose to combine these factors in a non-linear manner so as to introduce the actual complexity that is present in the real-world data. Support Vector Machines, were chosen because of the fact that they adhere to the learning method of Statistical Learning Theory (SLT). By testing the method on average temperature data from weather stations from China, they were able to obtain results better than traditional methods such as ARIMA, RNN, and STIFF.

Wei Xu(2004) and his team analyzed the limitations of the existing forecast methods, and proposes their integrated method, that works on data and method fusing. Through the existing methods, he obtains the temporal, spatial and spatio-temporal auto-regressive forecast, which he then combines to produce the final forecast method, through the use of an integrative spatio-temporal forecast.

Tung Wan Cheng(2006) and team chose the field of forest fire prevention, to implement their model called Spatio Temporal Data Mining and Knowledge Discovering hereby shortened to STDMKD. This field required a modified approach to prediction due to the dynamic nature of forest fires. They proposed a method dubbed as ISTIFF (improved spatio-temporal integrated forecasting framework). The proposed framework makes use of not just trend analysis but also of Association Rule Mining and Sequential Patterns Mining. This allows them to generalize trends such as fire in an area will lead to haze in that area. fire in an area will lead to haze in that area. This along with the designing of defining spatially-correlated siblings allowed them to equitably model the problem.

Ali Mert Ertugrul(2019) and team built a spatio-temporal predictive model for opioid overdose forecasting through the usage of the spatio-temporal patterns of the criminal incidents. Their CASTNet (Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting) , defines each location as an aggregation of static features and dynamic features. Key neural networks used spatial attention blocks to map the correlation between the features and the forecasts.

Jessica Hwang(2019) and team built an ensemble containing two separate non-linear regression based models. One of them was a Local Linear Regression present with a Multitask Feature Selector model that had lags based on the resolution and frequency of the data updates. The other model was an Autoregression k-Nearest Neighbour. The resultant ensemble, was the average of the L2-normalized predicted anomalies of the two models.

Yen-Yu Chang(2018) and team proposed an alternative to the commonly used Recurrent Neural Network (RNN) since such methods lack the ability to analyse the long term periodicity patterns present within the dataset. Inspired by the memory network built for answering questions, the team built a Memory Time-series network (MTNet). The system contains a memory module considerably larger than contemporary methods, present with three distinct separate encoders, along with a builtin autoregressive element. Together the whole system can be more understood when compared to a simple RNN.

Yulia Rubanova(2019), have proposed a novel approach to Time Series data with non-uniform intervals. Such data is quite hard to model with Recurrent Neural Networks (RNN), they generalized that the given data has hidden dynamics that can be defined using Ordinary Differential Equations (ODE). The resultant ODE-RNN, can handle the random time gaps placed by using Poisson Processes. Such models are more suited for interpolation tasks, since the hidden components allow them to work with more scare and sparse data.

Shun-Yao Shih(2019), have proposed a pattern of filters to augment the attention mechanism, since the attention model only can focus on previous time step and cannot capture long term patterns of seasonality, etc. These time-invariant patterns can be then be reintroduced into the RNN to allow the attention model to capture long term data regarding the time series data.

Guokun Lai(2017) and team, explored how Recurrent Neural Network(RNN) can be combined alongside a Convolution Neural Network (CNN) to build a LSTNet (Long and Short Term Timeseries Network). Built-in the convolutional component extracts thee short-term patterns and also local dependencies between the variables. The recurrent component allows the model to retain some form of 'memory' to remember long term patterns. To ensure that the LSTNet is autoregressive, they also placed a linear Bypass. The following model can now capture long and short-term patterns while also being non-linear in nature.

Xiaoyuan Liang(2019) and team have proposed a Spatio Temporal Fuzzy Neural Network, allowing for the complex interactions present in their domain. In their chosen domain of Passenger Demand Prediction, it would be difficult to map all of the temporal relations. They have fused a deep recurrent neural network along with a neural network with embedded fuzzy components, using a specific attention model to fuse both of these architectures. With the real-world dataset they are able to obtain a 10% improvement on the RMSE as compared with the other alternative state-of-the-art methods present currently.

Karthik H. Shankar(2019) and team have explored the possibilities of Fuzzy based memory systems. For accurate predictive models the learner would recall all of the information present, yet to optimise the space-complexity, a fuzzy memory system that sacrifices the accuracy of the information for the more dominant information. Although accuracy is sacrificed, the memory gain by the ability to observe a longer window is a better trade-off.

Francesca Rossi(2006) designed a set of algorithms and a framework to be used in cases of soft temporal constraint problems. These algorithms and the field have been used by NASA for optimal plan generation of the for planetary surface rover exploration. Existing ideas such as Simple Temporal Problems with Preferences, do not take uncertainty into account, using fuzzy frameworks such uncertainty can be embedded within the system.

Rafik Mahdaoui(2011) and team have designed and investigated the usability of Temporal Neuro based Fuzzy Systems (TNFS) for Fault diagnosis along with prediction of failure. The neuro-fuzzy models can inherently tackle uncertainty in data, and unpredictability. Training the system allows not just the neural network to be learnt but also the fuzzy inference rules.

Zhang Lifu(2015) and his team have designed a system to work with four dimensional datasets, (datasets with spatial and temporal aspects), to output the various characteristic parameters present within the dataset. Their work uses high frequency satellite data, and they have also proposed a method to allow them to select specific temporal data from specified regions like a point, or a rectangle or an abstract polygon.

Yves Julien(2015) and team compared the various temporal-spatial series based reconstruction of atmospheric data which contains atmospheric based cloud contamination of the dataset. Usage of a probabilistic model allows them to simulate the real time series of the influence of the cloud contamination.

Yao Zhao(2016) and team worked with remote sensing image data to understand field of land cover detection. Due to surface events the ground objects change with the passage of time. The main difficulty with such data lies in the sharp noise and cloud noise which obscure similarity calculation and clustering. By using the Canberra Distance-Dynamic Time Warping, which is a modification of the classical time-warping distance with the improved addition of Canberra-distance.

Ildar Batyrshin(2013) proposed a novel method for methods of association to calculate the extent of relationships (either inverse or direct) between existing parameters in temporal data. He defines association measures through axiom definitions and also proposes measures for similarity measurement through the usage of Minkowski Distance and alternate data standardization methods. Other partial associative methods that he proposes are the use of Pearson's correlation coefficient and cosine similarity.

Abbas Madraky(2012) and his team propose a novel data model called Hair data model. Most temporal data models have problems in applying analytical operations upon the data and it also faces poor compatibility with real life objects. Most real-life scenarios can be emulated with the introduction of object oriented models, but with relational data models the data present and the possible functions that one can apply on them are disjoint. The authors solve this problem through modelling data similar to human hair.

J.F. Roddick(2005) and team investigate possible modifications for the accommodations of semi-intervals present in temporal interval relationships. The authors build upon the models built by Allen and Freksa. Through the redefinition of the overlap relationships to consider the midpoints of the equally segmented intervals and the accommodation of the linear token sequences.

Apeksha Aggarwal(2018) and her team proposes a hash based temporal associative rule mining algorithm. The advantage of using hash-based systems is that such a system will hasten the memory accessing time allowing for overall higher performance. In addition to this, they have also expressed the granularity in present in the data for higher performance. The algorithm uses the apriori technique mixed with direct access hashing for querying. Such a method would reduce the candidate generation through multiple scans of the database instead of the classical apriori technique which grows exponentially with the amount of data present.

Pierre-Francois Marteau(2017) in his paper applies time elastic centroids for the temporal dataset for the purpose of denoising. In the paper he tries to answer the question of how averaging subsets of the data can be used in the form of Dynamic Time Warping and also using a Dynamic Time Warping Kernel as well. Because exact calculation of the centroid is a NP-complete problem with exponential complexity, a variety of heuristic methods are being made. The introduction of a probabilistic element is through the use of a stochastic automata, on the pairwise kernel alignment.

Edward De Brouwer(2019) and his team proposed a method for sporadic time series based data, by the usage of a GRU-ODE-Bayes model. The author builds upon the Neural Ordinary Differential Equation model. This model allows a continuous representation of neural networks. The main system works through GRU-ODE and the combination of GRU-Bayes, which presents the sporadic observability into the ODE’s understanding of probability distribution. The author achieves this through the placement the ODE between each element of the input steps as well. The advantages of such a method, is that the model is bounded, continuous and has a general numerical integration.

Such a model allows extension into variables that can be either discrete or continuous , and can be extended to multinomial observations as well.

Yaguang Li(2017) and his team built a modified version of a Convolutional Recurrent Neural Network, monikered Diffusion Convolutional Recurrent Neural Network (DCRNN), which includes both spatial dimensions and temporal dimensions present in the data. The modified neural network can capture the spatial dimensions through the use of random walks in the graph allowing it to capture the temporal dependency through the usage of an encoders-decoders. The team was using traffic data and was trying to perform data driven traffic forecasting (instead of knowledge-driven traffic forecasting), which faces problems like, complex dependencies on the position of road networks, non-linear time based dependencies on the shifting prevailing road conditions, along with the difficulty of forecasting over long time periods.

Boris N. Oreshkin(2019) and his team solve point forecasting in univariate time series data using deep neural networks which inject forward and backward residual links along with a large number of fully connected neural network layers. Their network uses ensemble as a regularization technique instead of dropout or the L2-norm. They also designed the network so that the common decomposition techniques of seasonality and etc can be seen in certain parts of the architecture.

Andrew Redd(2019) and his team build ES-RNN which is a composite of the classical spatio-temporal forecasting models and the current Recurrent Neural Networks(RNN). The base of the model is a skip connection-based LSTM layer, this offers it lower computational time, while also allowing the network to remember the hidden information present within. After using a primer estimate for the level and seasonality coefficients, we can train the model to have its own level and seasonality.

Conclusion

We have used a fuzzy inference system along with a conventional inference system to observe the difference between our proposed system and the conventional systems. The addition of FIDES to our system allows us to model the uncertainty present within our system.

REFERENCES

1. Jane Y. Nancy, Nehemiah H. Khanna, Arputharaj Kannan (2017), ”A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals”, ELSEVIER Expert Systems with Applications, 0957-4174
2. Jane Y. Nancy, Nehemiah H. Khanna, Kannan Arputharaj (2017), 'Inputing missing values in unevenly spaced clinical time series data to build an effective temporal classification framework', ELSEVIER Computational Statistics & Data Analysis, 0167-9473
3. Doyup Lee and Suehun Jung and Yeongjae Cheon and Dongil Kim and Seungil You(2019), ‘Demand Forecasting from Spatiotemporal Data with Graph Networks and Temporal-Guided Embedding’, arXiv, 1905.10709
4. J. Wang, T. Cheng and X. Li(2007), "Nonlinear Integration of Spatial and Temporal Forecasting by Support Vector Machines," *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, 10.1109
5. Wei Xu, Houkuan Huang and Yong Qin(2004), "Study and its application of spatio-temporal forecast algorithm," *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP '04. 2004.*, Beijing, 2004, 10.1109/ICOSP.2004.1441646
6. *Cheng, T.W., & Wang, J.W. (2006). APPLICATIONS OF SPATIO-TEMPORAL DATA MINING AND KNOWLEDGE DISCOVERY (STDMKD) FOR FOREST FIRE PREVENTION.*
7. *Ertugrul, A.M., Lin, Y., & Taskaya-Temizel, T. (2019). CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting. ArXiv, abs/1905.04714.*
8. Hwang, Jessica et al (2019). “Improving Subseasonal Forecasting in the Western U.S. with Machine Learning.” Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD ’19
9. Yen-Yu Chang and Fan-Yun Sun and Yueh-Hua Wu and Shou-De Lin(2018), ‘A Memory-Network Based Solution for Multivariate Time-Series Forecasting’, arXiv, 1809.02105
10. Yulia Rubanova and Ricky T. Q. Chen and David Duvenaud(2019), ‘Latent ODEs for Irregularly-Sampled Time Series’, ArXiv, 1907.03907
11. Shih, Shun-Yao, Fan-Keng Sun, and Hung-yi Lee (2019), “Temporal Pattern Attention for Multivariate Time Series Forecasting.” Machine Learning 108.8-9: 1421–1441.
12. Lai, Guokun et al (2018), “Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks.” The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR ’18:
13. Liang, Xiaoyuan et al. “A Deep Spatio-Temporal Fuzzy Neural Network for Passenger Demand Prediction.” 2019 SIAM International Conference on Data Mining (2019): 100–108.
14. Karthik H. Shankar and Marc W. Howard(2012), ‘Optimally fuzzy temporal memory’, arXiv,

15. Rossi, F., K. B. Venable, and N. Yorke-Smith. “Uncertainty in Soft Temporal Constraint Problems: A General Framework and Controllability Algorithms for The Fuzzy Case.” Journal of Artificial Intelligence Research 27 (2006): 617–674.
16. Rafik Mahdaoui and Leila Hayet Mouss and Mohamed Djamel Mouss and Ouahiba Chouhal(2011), ‘A Temporal Neuro-Fuzzy Monitoring System to Manufacturing Systems’, arXiv, 1107.3302
17. Z. Lifu, C. Hao, F. Dongjie, W. Taixia, L. Jia and C. Huang(2015), "Method for time series extraction of characteristic parameters from multidimensional remote sensing datasets," 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, 10.1109/WHISPERS.2015.8075500
18. Y. Julien and J. A. Sobrino(2015), "CloudSim: A fair benchmark for comparison of methods for times series reconstruction from cloud and atmospheric contamination," 2015.7245761
19. Yao Zhao, Lei Lin, Wei Lu and Yu Meng, "Landsat time series clustering under modified Dynamic Time Warping," 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Guangzhou, 2016, 7552767
20. I. Batyrshin, "Constructing Time Series Shape Association Measures: Minkowski Distance and Data Standardization," 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence, Ipojuca, 2013, pp. 204-212
21. A. Madraky, Z. A. Othman and A. R. Hamdan, "Hair data model: A new data model for Spatio-Temporal data mining," 2012 4th Conference on Data Mining and Optimization (DMO), Langkawi, 2012, pp. 18-22. doi: 10.1109/DMO.2012.6329792
22. J. F. Roddick and C. H. Mooney, "Linear temporal sequences and their interpretation using midpoint relationships," in IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 1, pp. 133-135, Jan. 2005. doi: 10.1109/TKDE.2005.12
23. A. Aggarwal and D. Toshniwal, "Spatio-Temporal Frequent Itemset Mining on Web Data," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, Singapore, 2018, pp. 1160-1165. doi: 10.1109/ICDMW.2018.00166
24. Marteau, Pierre-Francois. “Times Series Averaging and Denoising from a Probabilistic Perspective on Time–Elastic Kernels.” International Journal of Applied Mathematics and Computer Science 29.2 (2019): 375–392.
25. Edward De Brouwer and Jaak Simm and Adam Arany and Yves Moreau (2019), “GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series”, arXiv, 1905.12374
26. Yaguang Li and Rose Yu and Cyrus Shahabi and Yan Liu (2017), “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting”, arXiv, 1707.01926
27. Boris N. Oreshkin and Dmitri Carpov and Nicolas Chapados and Yoshua Bengio(2019), “N-BEATS: Neural basis expansion analysis for interpretable time series forecasting”, arXiv, 1905.10437

28. Andrew Redd and Kaung Khin and Aldo Marini(2019), “Fast ES-RNN: A GPU Implementation of the ES-RNN Algorithm”, arXiv, 1907.03329
29. Guokun Lai and Wei-Cheng Chang and Yiming Yang and Hanxiao Liu(2017), “Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks” , arXiv, 1703.07015