FUZZY SETS AND FUZZY LOGIC PROJECT

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Bawana Road, Delhi-110042

PREDICTING COVID-19 INFECTION TREND USING FUZZY TIME SERIES

Submitted By:

Jai Garg (2K18/MC/044)
Jatin Papreja (2K18/MC/049)
Kumar Apurva (2K18/MC/058)
Madhuresh Mayank (2K18/MC/062)

INDEX

INTRODUCTION

TIME SERIES ANALYSIS

Time series are sets of data representing the behavior of one (or more) random variable over time, and its main characteristic is that the successive records of this variable are not independent of the previous records and their analysis must take into account the order in which they were collected. Time series analysis and forecasting methods are indispensable on several fields, for instance on engineering, medicine, economy, meteorology, etc.

There are several methods of analysis and forecasting, from the traditional and consecrated statistical tools (ARMA, ARIMA, SARIMA, Holt-Winters, etc.), to the new computational intelligence tools (recurrent neural networks, LSTM, GRU, etc.). But some key features including Readability, Manageability, Simplicity and Scalability distinguish the Fuzzy Time Series to turn it in an attractive option.

FUZZY TIME SERIES

The use of fuzzy sets for modeling and predicting time series arises almost intuitively, first based on the ability of fuzzy models to approximate functions, but also on the readability of rules using linguistic variables that make them more accessible to experts and non-experts analysis.

The pioneer work on fuzzy time series is Song and Chisson (1993) followed by the evolution in field published by Chen (1996). The idea is to divide the Universe of Discourse from time series in intervals/partitions (the fuzzy sets), and learn how each area behaves (extracting rules through the time series patterns). The rules of these models tell how the partitions relate with themselves over the time, as values jump from one place to another. In other words: let's create a linguistic variable to represent the numerical time series, and these areas will be the linguistic terms of our variable.

When we create a linguistic variable to represent the universe of discourse, we create a "vocabulary," and then the fuzzyfied series is composed of words in that vocabulary. The sequences of these words — the sentences or phrases — are the patterns we need to learn.

BACKGROUND KNOWLEDGE

FUZZY SETS AND FUZZY LOGIC

The logic theory and classical mathematics define a Set as a dichotomy: every element is in OR out of the set. There is no middle point! The membership of an element is a boolean value, that is, a value on set {0, 1}, imposing to each set strong and inflexible boundary. This dichotomous way of thinking is uncomfortable for the human being because innumerable realities are not so. We will have difficulty when we try to classify people into categories with strict limits, for example: weight = {thin, slender, fat}, age = {child, adolescent, young, adult, elderly}, height = {low, medium, high}.

The Fuzzy Logic, proposed by Zadeh (1965), state a duality instead of this dichotomy: a certain element may belong and simultaneously do not belong to the same set at certain levels, such that the membership is a value in the interval [0, 1]. The fuzzy sets have no strict boundaries and they are usually overlapping, a person can be medium high and medium, or 90% medium and 10% high for example.

Given X, a Numerical Variable, such that $X \in \mathbb{R}$ — for instance, a height measure — its Universe of Discourse, abbreviated to U, is the is the range of values that this variable can assume, such that U = [min(X), max(X)].

LINGUISTIC VARIABLE

A linguistic variable A is the transformation of the values of the numerical variable X into a set of words/linguistic terms (what we call fuzzification). Each word/linguistic term is a fuzzy set $\tilde{a} \in \tilde{A}$, and each fuzzy set $\tilde{a}$ is associated to a function μ (mu greek letter), such that $\mu_{\tilde{a}}: X \rightarrow [0,1]$.

FUZZY INFERENCE SYSTEMS

A fuzzy system is a repository of fuzzy expert knowledge that can reason data in vague terms instead of precise Boolean logic. The expert knowledge is a collection of fuzzy membership functions and a set of fuzzy rules, known as the rule-base, having the form:

IF (conditions are fulfilled)
THEN (consequences are inferred)

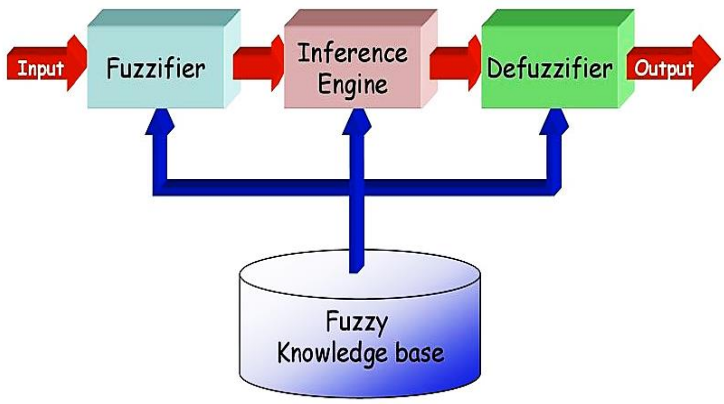

Figure 1: Fuzzy Inference Systems

APPROACH AND IMPLEMENTATION

The approach and implementation of the project can be broadly classified under following sections:

- Dataset Analysis
- Fuzzy Time Series System
- Partitioning Types

DATASET ANALYSIS

In this project, we will use COVID-19 data of all continents from January, 2020 to December, 2020. The overlook of the data is given in Figure 2.

continentExp	Africa	America	Asia	Europe	Oceania	Other
dateRep						
2019-12-31	0	0	27	0	0	0
2020-01-01	0	0	0	0	0	0
2020-01-02	0	0	0	0	0	0
2020-01-03	0	0	17	0	0	0
2020-01-04	0	0	0	0	0	0
...	...	...	...	...	...	...
2020-12-10	18892	325238	76203	255762	19	0
2020-12-11	19111	325486	75131	273397	227	0
2020-12-12	18909	339515	71851	226696	169	0
2020-12-13	18517	304854	72723	230310	17	0
2020-12-14	16142	250922	64995	208571	29	0

Figure 2: Daily Covid-19 Cases in different Continents

Figure 3 shows the trend of rise and fall of Covid-19 cases across the continents. Europe and America are the most affected continents during this time period with daily cases rising over 3 lacs. We have constrained to Covid-19 trend of continent of ‘Asia’ as the scope of this project.

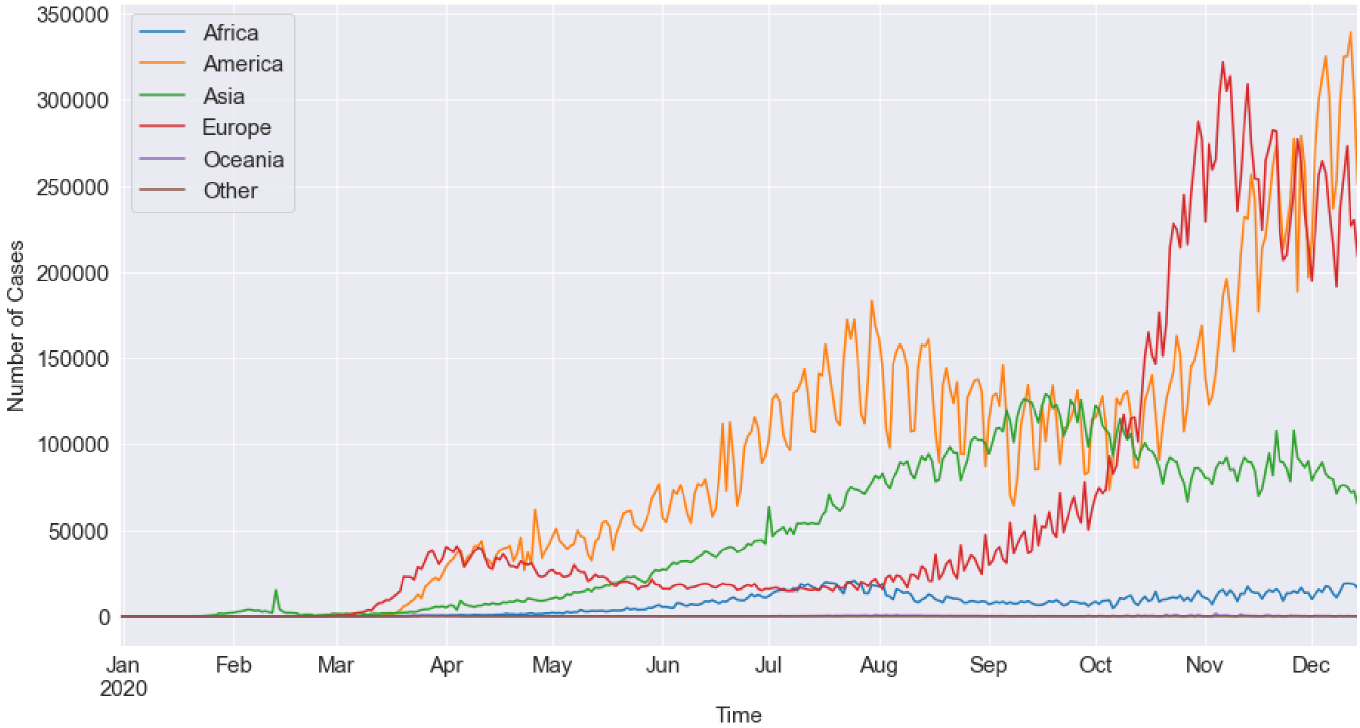

Figure 3: Covid-19 Trend across Continents

FUZZY TIME SERIES SYSTEM

We will explain the Fuzzy Time Series model using a basic example for the purpose of improving understanding. Following subsections present a detailed look of the system.

1. Definition of the Universe of Discourse U

First, we need to know the universe of discourse U from the training data, such as U = [min(X), max(X)]. Usually we extrapolate the upper and lower bounds by 20%, as a security margin.

2. Create the Linguistic Variable Ã (Universe of Discourse Partitioning)

Now we need to split U on several overlapping intervals (partitions) and create a fuzzy set for each one of them. The number of intervals is one of the most important parameters on Fuzzy Time Series and it will directly imply on model's accuracy. Beyond the number of partitions, the way we split U also have great impact on accuracy.

As an example, will create 10 partitions based on grid partitioning scheme where all partitions have the same length and format, such that the linguistic variable be Ã = {A0, A1, ..., A9}. Each linguistic variable represents a triangular fuzzy number as shown in Figure 4.

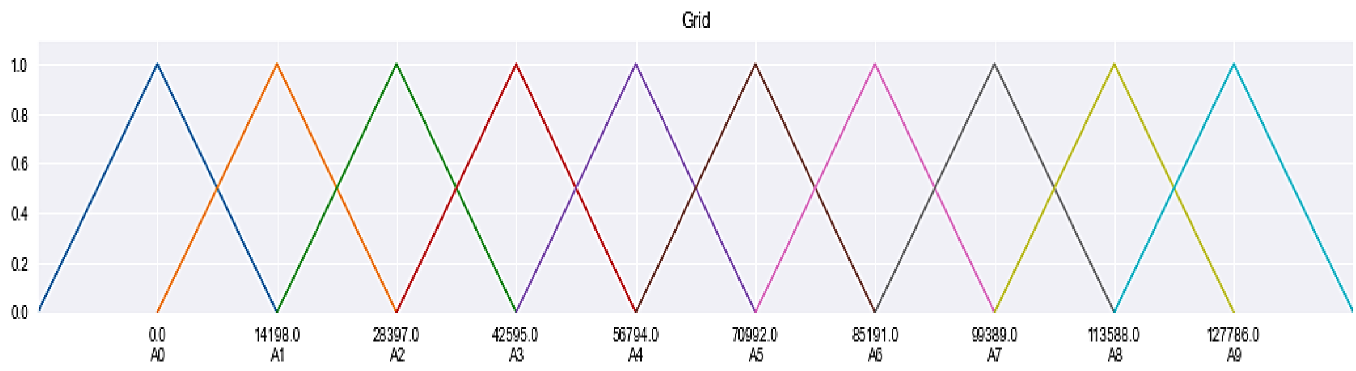

Figure 4: Linguistic Variables formed using Grid Partitioning

3. Fuzzification

Now we can convert the numerical values of X(t) into fuzzy values of the linguistic variable Ã, giving rise to the fuzzy time series F(t). It is always good to remind that the fuzzy sets on Ã are overlapped, so for each x ∈ X(t) it is possible that it belongs to more than one fuzzy set A_i ∈ Ã. Using Chen method, just the maximum membership fuzzy set is chosen.

4. Creating the temporal patterns

A temporal pattern indicates two fuzzy sets that appear sequentially on fuzzy time series F(t) and have the format Precedent → Consequent, where the precedent indicates a fuzzy set on time t and the consequent the fuzzy set that appears soon after on time t+1.

The generated Temporal Patterns for example considered will be:
{'A0 -> A0', 'A0 -> A1', 'A1 -> A0', 'A1 -> A1', 'A1 -> A2', 'A2 -> A2', 'A2 -> A1', 'A2 -> A3', 'A3 -> A3', 'A3 -> A2', 'A3 -> A4', 'A4 -> A3', 'A4 -> A4', 'A4 -> A5', 'A5 -> A5', 'A5 -> A4', 'A5 -> A6', 'A6 -> A6', 'A6 -> A5', 'A6 -> A7', 'A7 -> A6', 'A7 -> A7', 'A7 -> A8', 'A8 -> A8', 'A8 -> A7', 'A8 -> A9', 'A9 -> A9', 'A9 -> A8', 'A6 -> A8', 'A8 -> A6'}.

5. Creating the rules

The model rules also have the format Precedent → Consequent. Given the previously generated temporal patterns we will group them by its precedents. The model will contain a rule for each distinct precedent found, and the consequent of each rule will be the union of all consequents of each temporal pattern with the same precedent.

Thus, generated rules will be:

```
A0 -> A0,A1          A5 -> A4,A5,A6
A1 -> A0,A1,A2       A6 -> A5,A6,A7,A8
A2 -> A1,A2,A3       A7 -> A6,A7,A8
A3 -> A2,A3,A4       A8 -> A6,A7,A8,A9
A4 -> A3,A4,A5       A9 -> A8,A9
```

The rule set is, in fact, the FTS model. They describe how time series behaves and, if it was stationary enough (well behaved), we can use this model to forecast the next values on the time.

A simple and readable model like this has other advantages:

- It is easy to parallelize/distribute, what makes it very attractive for big data.
- It is easy to update, what makes it very attractive for frequently changing data.

Since we know the numerical value for time t, $x(t) \in X(t)$, we want to predict the next instant, $x(t + 1)$.

1. Input value fuzzification

The input value x(t) will be converted into fuzzy values of the linguistic variable $\tilde{A}$, generating the value f(t). As in the training process only the most pertinent set is chosen. Considering the example, for at any time t, if the value of x(t) = 18876, then Fuzzification of x(t) sets of most pertinent value to A1, so f(t) = A1.

2. Find the compatible rules

Find the rule whose precedent is equal to f(t). The consequence of the rule will be the fuzzy forecast for t + 1, i.e., f(t + 1). For f(t) = A1 we have the rules A1 → A0, A1, A2.

3. Defuzzification

Now we need to convert f(t + 1) to a numeric value. For this we use the center-of-mass method, where the numerical value is equal to the mean of the centers of the fuzzy sets of f(t + 1), that is $x(t+1) = \Sigma A_i/n$, for i = 0, 1, 2, …., n-1 and n equal to the number of sets in f(t). Therefore, x(t+1) = (A0 + A1 + A2)/3 = 14198.47.

Figure 5 shows the complete Fuzzified forecast for the considered example.

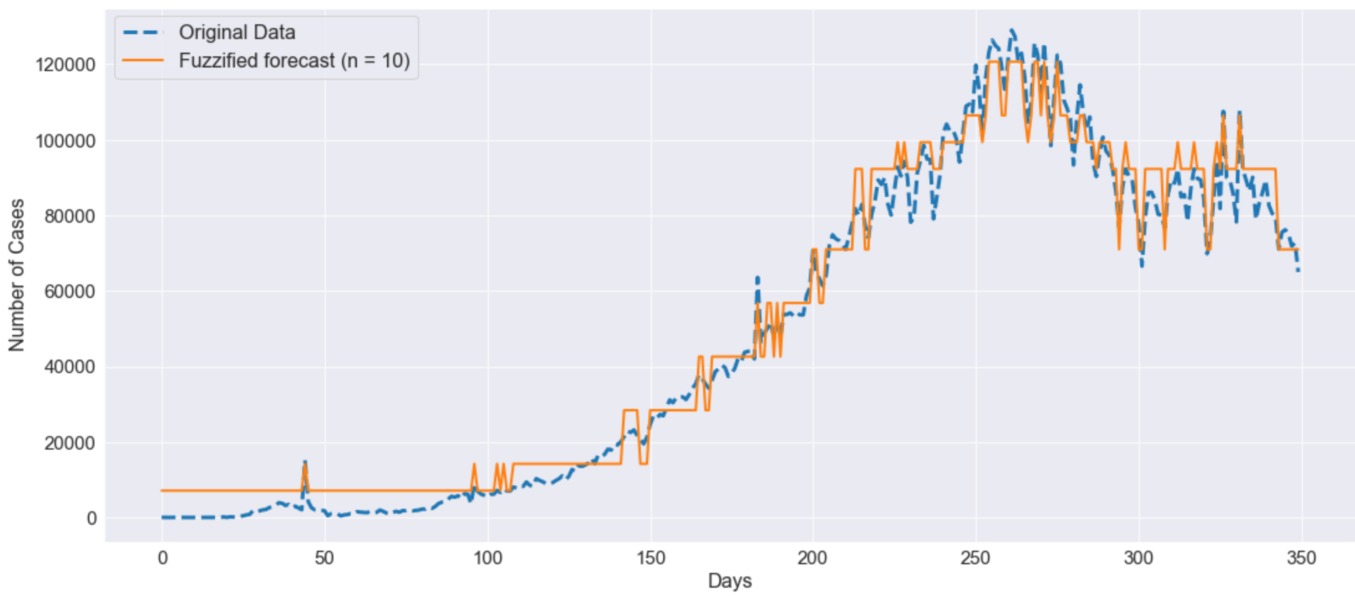

Figure 5: Fuzzified Forecast for 10 partitions

RESULTS

The results achieved using the proposed approach can be broadly classified under following sections:

- Influence of Number of Partitions
- Influence of Partition Types
- Comparison of FTS with Moving Average
- Comparison of Improved Weighted FTS with ARIMA

INFLUENCE OF NUMBER OF PARTITIONS

This is simply the most influential parameter in the model accuracy. The precision in the capture of the characteristics of the time series increases with increase in number of partitions but with some traps:

- Too few fuzzy sets generate underfitting, due to signal over simplification.
- Too much fuzzy sets generates overfitting, making the model to start learning noise on data.

For analysis, we have taken an example of Fuzzy Time Series model with different number of partitions (5, 10, 20) using Grid Partitioning method as shown in Figure 6.

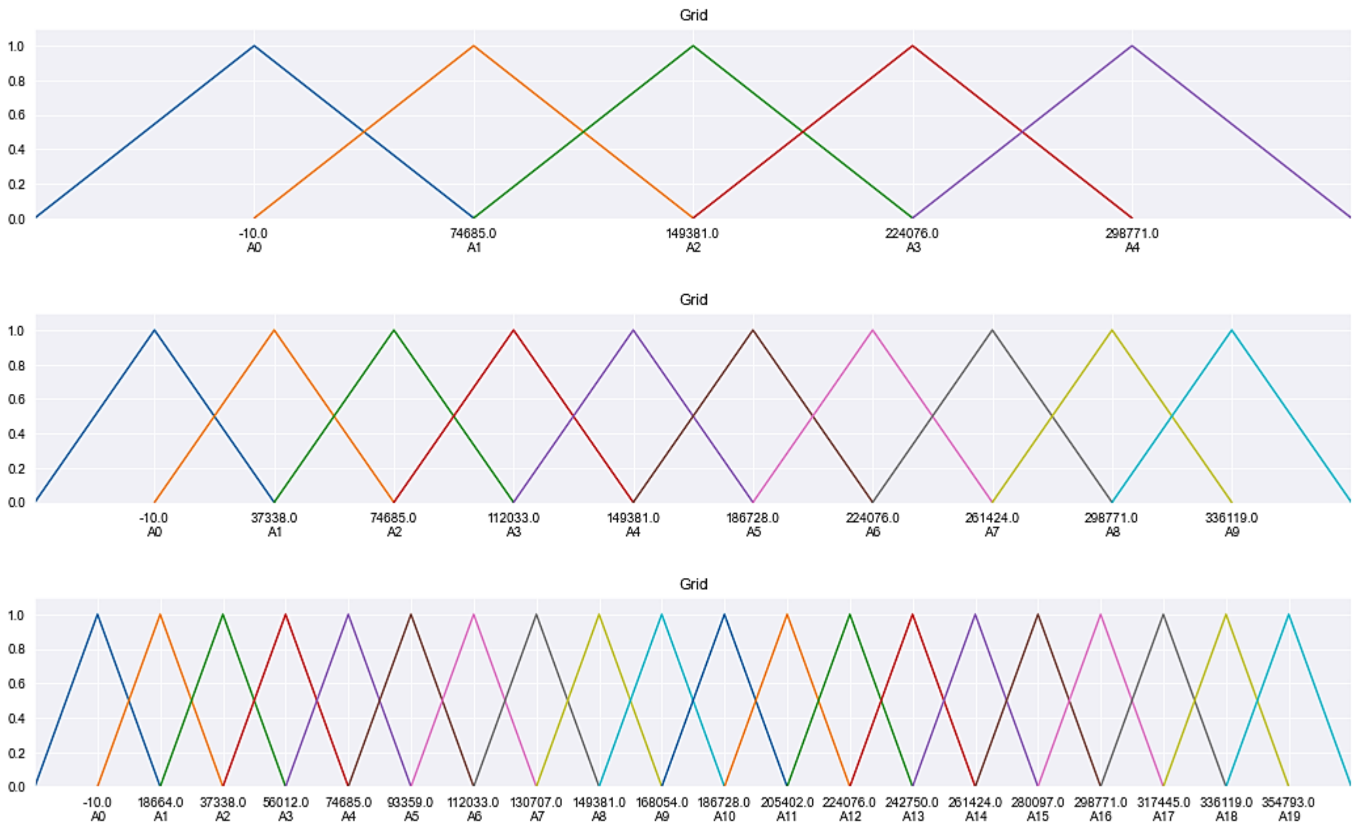

Figure 6: Linguistic Variables with different number of partitions

Figure 7 shows the Fuzzified forecasts obtained using different number of partitions for a section of data. We can observe that as the number of partitions increases the precision of the model also increases.

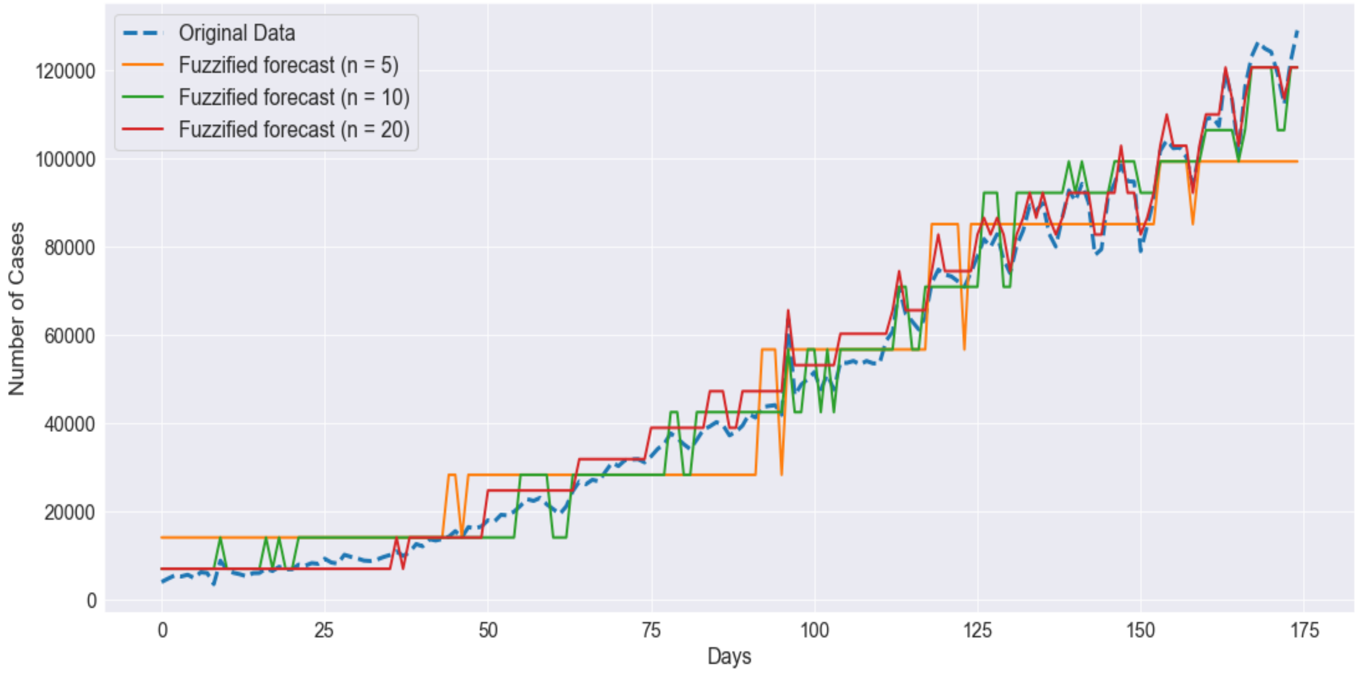

Figure 7: Fuzzified Forecast with different number of partitions

INFLUENCE OF NUMBER OF PARTITIONS

There are many types of partitioning, from Grid partitioning where all sets are evenly distributed and have the same format, going through partitioners where sets have distinct sizes such as entropy-based and cluster-based partitioners. Figure 8 shows the different types of partitioning schemes (Grid, C-Means, Entropy) used for FTS model using triangular fuzzy function as membership function.

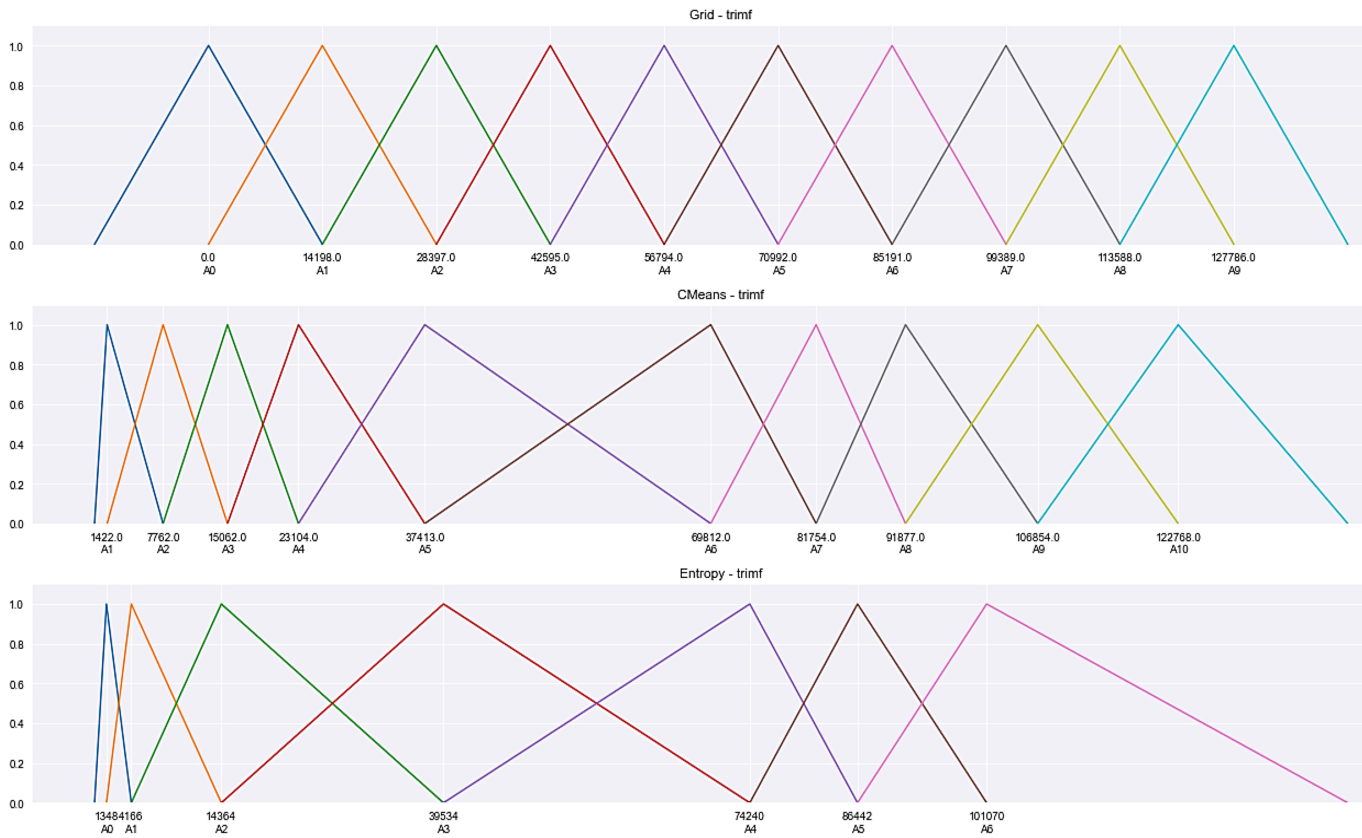

Figure 8: Linguistic Variables formed using different partitioning schemes

Figure 9 shows the Fuzzified forecasts obtained using different types of partitioning. We can observe that Grid Partitioning performs the best for the data considered since it closely mimics the trend of Covid-19 cases over the time period considered. Entropy Based Partitioning performs poorly in this case.

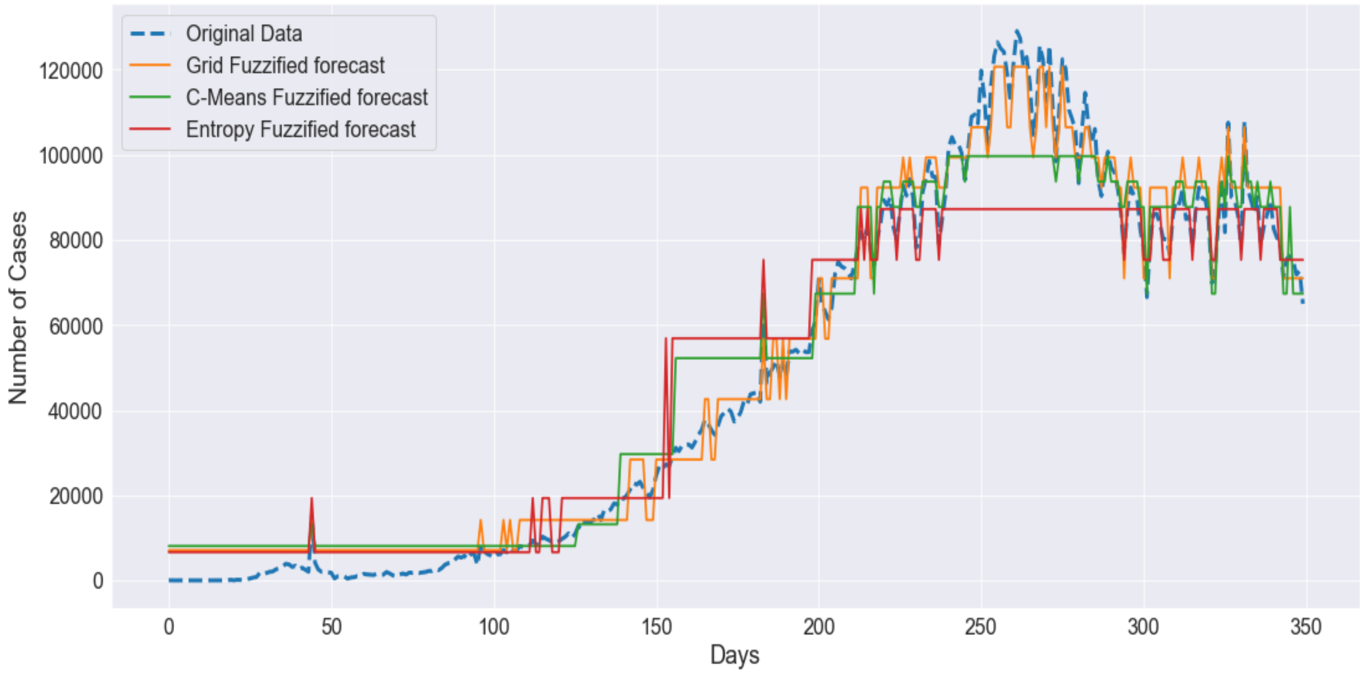

Figure 9: Fuzzified Forecast formed using different partitioning schemes

COMPARISON OF FTS WITH MOVING AVERAGE

In order to emphasize the usefulness of Fuzzy Time Series (FTS) model, we compare it with Moving Average Time Series (MATS) Model with a sliding window of 6 lags. The FTS model used is based on 35 partitions using Grid Partitioning scheme. Figure 10 shows the results of both models. It can be observed that FTS model is much closer in terms of resembling the trend of Covid-19 cases in Asia.

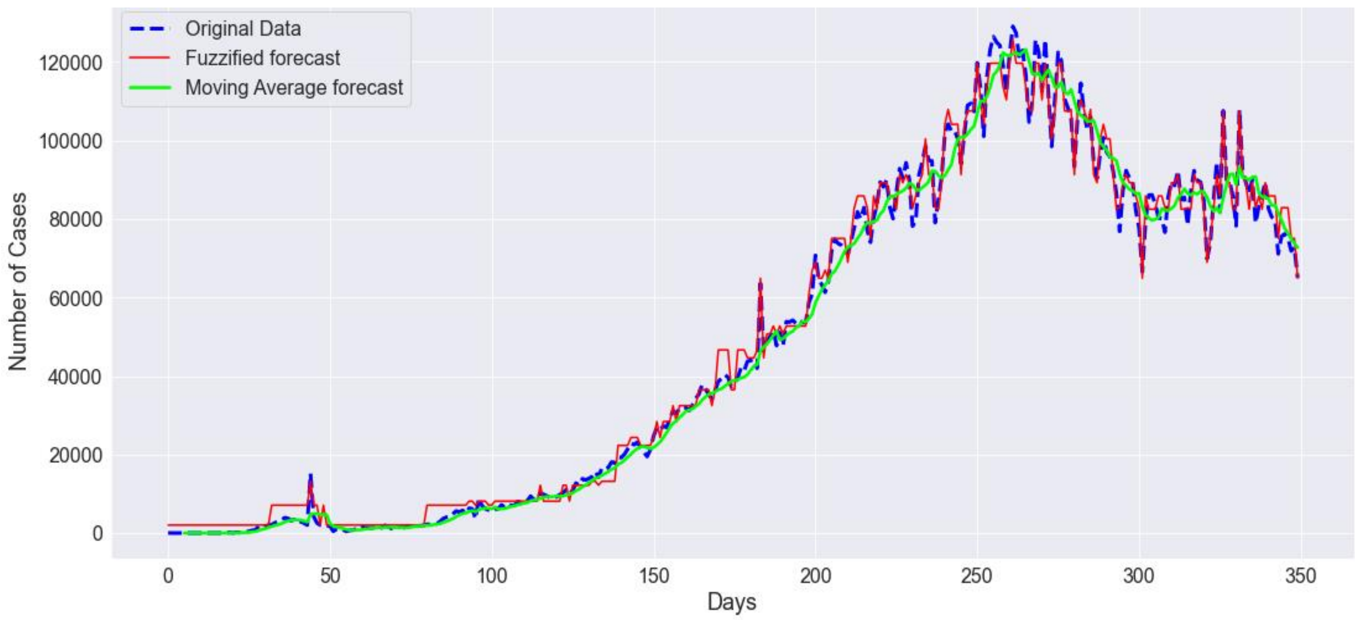

Figure 10: Comparison of FTS and MATS

This is further confirmed by looking at the error metrics which comes out as follows:

RMSE (FTS): 2890.12 RMSE (MATS): 4746.06
MSE (FTS): 8352802.45 MSE (MATS): 22525069.86

COMPARISON OF IMPROVED WEIGHTED FTS WITH ARIMA

We can further improve the FTS model by generating weighted rules. For this purpose, we have considered an improved Weighted Fuzzy Time Series (WFTS) model involving Grid Partitioning Scheme and compared it with an Autoregressive Integrated Moving Average (ARIMA) model which is a well-known statistical time series model.
The following rules are generated using Weighted FTS model:

```
A3 -> A0(1.0)
A4 -> A0(1.0)
A2 -> A0(0.583),A1(0.333),A2(0.083)
A1 -> A0(0.676),A1(0.243),A2(0.081)
A0 -> A0(0.882),A1(0.081),A2(0.027),A3(0.003),A4(0.007)
```

Figure 11 shows the results obtained using both models. Both WFTS and ARIMA perform really well, for the data considered.

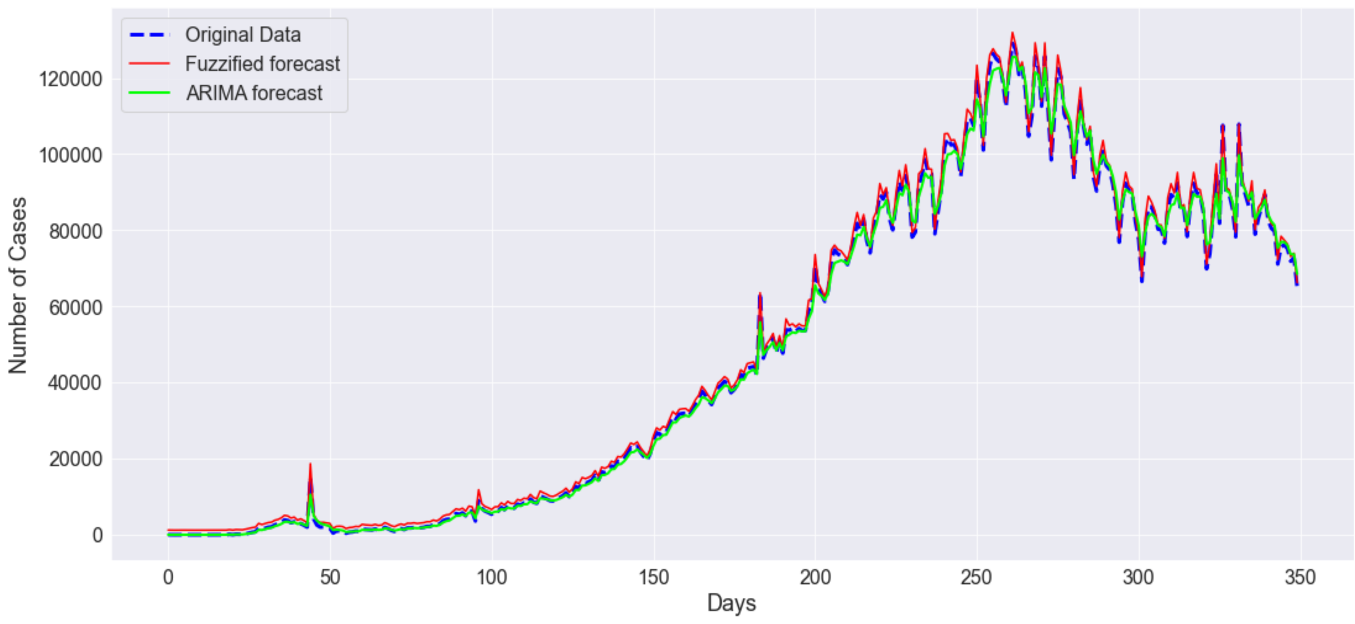

Figure 11: Comparison of WFTS and ARIMA

Looking at the error metrics which comes out as follows:

RMSE (WFTS): 1604.05 RMSE (ARIMA): 1964.64
MSE (WFTS): 2572968.65 MSE (ARIMA): 3859825.33

We can confirm that WFTS works better than the ARIMA model for this scenario. It offers about 18% improvement over the well-known ARIMA model. Thus, it is a really useful model for analysis of time series datasets.

APPLICATIONS

1. **Aerospace:** Fuzzy logic proportional-integral-differential (PID) controllers are developed to perform both stability augmentation and automatic flight control functions. It operates for controlling both longitudinal and lateral-directional motions for an example aircraft, the X-29.

2. **Business:** Fuzzy logic plays very important roles, especially in business, because it helps reduce costs. It differs from conventional (hard) computing in that it is tolerant of imprecision, uncertainty, partial truth, and approximation. In effect, the role model for fuzzy logic is the human mind.

3. **Defense:** The future of warfare will be defined by technological innovation rather than pure destructive power. While development is ongoing, fuzzy logic based artificial intelligence systems are capable of operating in complex battlefield environments and producing better results than existing control systems. Fuzzy controllers in military applications such as mobility, cybersecurity, target tracking, interoperability and unmanned combat aerial vehicle control have shown encouraging results when introduced to uncertainties and complexities of the battlefield.

4. **Electronics:** Fuzzy electronics is an electronic technology that uses fuzzy logic, instead of the two-state Boolean logic more commonly used in digital electronics. Fuzzy electronics is fuzzy logic implemented on dedicated hardware. This is to be compared with fuzzy logic implemented in software running on a conventional processor.

5. **Finance:** Fuzzy logic has been successfully applied in the field of finance due to its ability to address imprecise, incomplete and vague data. This methodology has also been used in the field of banking, although to a lesser extent, with particular relevance to areas such as risk management and credit scoring.

6. **Manufacturing:** Fuzzy logic control is presented as a technique to implement real-time control algorithms that can be embedded in workstations in the framework of intelligent hierarchical control of manufacturing systems.

7. **Healthcare:** Fuzzy logic is a suitable way to provide the physician with the support he needs in handling linguistic concepts and get rid of the loss of precision. Fuzzy logic technologies are applied to each area of medicine, and they have been proven to be successful.

CONCLUSION

In this project, we have studied **Fuzzy Time Series** to tackle the problem of forecasting COVID-19 cases.

Fuzzy Time Series algorithm mainly focuses on the following steps:

- **Training Procedure**: It consists of defining of universe of discourse, creating linguistic variables and thus fuzzification of values to their corresponding linguistic variable. Temporal patterns are then created which then decides the rule set. This rule set acts as the model for prediction of COVID-19 cases.
- **Forecasting Procedure**: as For a given input value, we predict the future value of COVID-19 cases. This input value is first fuzzyfied. Then, we find rule for which the precedent value equals the fuzzyfied input value. The consequence of this rule is then considered to fuzzy forecast. Defuzzification is then done in order to find the fuzzy forecast.

We have also studied the influence of parameters such as number of partitions and partitioning schemes on the performance of FTS models. Further, comparisons with Moving Averages Time Series model and Autoregressive Integrated Moving Average (ARIMA) are performed to highlight the real-life usefulness of the FTS models.

Results shows that Fuzzy Time Series forecasting model performs better than ARIMA time series forecasting model. The error value of the fuzzy time series is less than the ARIMA time series error value. So it is concluded that the fuzzy time series can gives us better results than other time series models.

REFERENCES

- https://towardsdatascience.com/a-short-tutorial-on-fuzzy-time-series-dcc6d4eb1b15
- https://www.sciencedirect.com/science/article/abs/pii/S0888613X18306376#:~:text=Fuzzy%20Time%20Series%20Forecasting%20is,been%20observed%20in%20this%20direction.
- https://www.sciencedirect.com/science/article/abs/pii/0165011495002200
- https://pyfts.github.io/pyFTS/build/html/index.html
- https://www.sciencedirect.com/science/article/abs/pii/016501149390372O