DOI: 10.37943/13EOTU7482

Gulnara Bektemyssova
Master, Senior Lecturer, Department of Radioengineering, Electronics, and Telecommunications
g.bektemisova@gmail.com, orcid.org/0000-0002-0850-0558
International Information Technology University, Almaty, Kazakhstan

Zhanar Ibraeva
Master, Senior Lecturer, Department of Radioengineering, Electronics, and Telecommunications
zhanarbazar@gmail.com, orcid.org/0000-0003-3196-696X
International Information Technology University, Almaty, Kazakhstan

Abd Rahim Ahmad
PhD, Associate Professor, Department of Systems and Networking
abdrahim.ahmad@gmail.com, orcid.org/0000-0001-7923-0105
College of Computer Science and IT (CCSIT)
Universiti Tenaga Nasional (UNITEN), Malaysia

FUZZY MODEL FOR TIME SERIES FORECASTING

Abstract: In 2007, in Kazakhstan, there was a transition of TDM (Time Division Multiplexing) circuit-switched technologies to IP (Internet Protocol) packet technology, which created a modern infrastructure for the ICT (information communication technologies) sphere. The advent of the IoT (Internet of Things) concept has led to the growth of a functioning network at a faster rate. It is currently developing in the direction of a cognitive infocommunication network. Its evolutionary development is characterized by a change in the volume of transmitted information, types of its presentation, methods of transmission and storage, the number of sources and consumers, distribution among users, and requirements for timeliness and reliability (quality) [1]. Types of traffic and their structure are changing; therefore, data processing becomes more complicated. For this reason, the tasks of analyzing and predicting network traffic remain relevant.

In this work, the prediction of the measured traffic on a real network is performed. The series under study shows the totality of packets transmitted over the backbone network for each second. Forecasting of a one-dimensional time series is carried out on the basis of fuzzy logic methods. This class of models is well suited for modeling nonlinear systems and time series forecasting. The use of fuzzy sets is based on the ability of fuzzy models to approximate functions, as well as on the readability of rules using linguistic variables. The results of the software algorithm of fuzzy inference models were obtained using the Python environment. Membership functions and predictive graphs were built, and their evaluation was carried out. The numerical values of the root mean square error (MSE) are calculated. As a result, it was found that the Cheng fuzzy prediction model has higher forecast accuracy than the Chen forecasting method.

Keywords: network traffic, time series, fuzzy logic, data analysis, forecasting.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

Introduction

The evolution of the telecommunications network based on packet-switched infocommunication networks has led to a huge increase in the amount of data associated with information flows. Today, the current trends in the Republic of Kazakhstan (RK) are data center management, cloud and cognitive technologies, IT security, etc. [1].

Management of network devices of a functioning multiservice network provides an opportunity to respond to an ever-increasing amount of transmitted information and quickly allocate the necessary resource. As users generate ever-increasing data, predicting network traffic remains an urgent task. Predictive data provide the necessary information to solve the problem of managing information flows in the network. Time series modeling is one of the ways to predict them [2].

The series under study in this work shows the totality of packets transmitted over the backbone network for each second received within 5 hours.

Literature review and problem statement

The term "fuzzy set" belongs to Professor Lotfi Zadeh; he owns many techniques for describing algorithms in the theory of fuzzy sets created in 1965.

In 1993, foreign scientists Song and Chissom proposed the concept of fuzzy time series (FTS). They were the first to propose the FTS (Fuzzy Time Series) methodology. In 1996, the scientist Chen carried out an extension of the method based on simplified arithmetic operations. The idea of the scientist Chen was based on fuzzy logical tables of group relations to reduce the computational complexity of the previously created model [3]. As for the algorithm of the scientist Cheng, he, in turn, expanded the model developed by scientist Chen and introduced the trend-weighted FTS model. At the same time, for forecasting, he assigned appropriate weights to individual fuzzy relations [4]. In 2001, Hwang introduced a heuristic model by integrating the Chen model. In 2002, Chen introduced the high-order FTS. In 2006, Hwang started working with non-stationary FTS. The general idea is to divide the universe of time series discourse into intervals - partitions in order to know how each area behaves with the extraction of rules from time series patterns [5]. The rules of these models show how sections are related to each other over time as values move from one place to another.

In [6], it is described that for a wide class of dynamic processes, it is possible to predict time series based on their fuzzy values. 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 initial data was obtained using the Wireshark sniffer program. 278,557 packets were tracked in 5 hours: including 25,733 MPEG (Moving Picture Experts Group).

5 0.008388 192.168.172.30 239.2.4.4 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
6 0.008391 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
7 0.010750 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
8 0.010753 192.168.172.30 239.2.4.4 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
9 0.012718 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
10 0.013497 192.168.172.30 239.2.4.4 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
11 0.014810 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
12 0.016488 192.168.172.30 239.2.4.4 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
13 0.016931 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
14 0.018195 192.168.172.30 239.2.4.4 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
15 0.019525 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes
16 0.020687 DTS_14321.272788888 PTS_14321.312788888 MPEG_PES_video-stream 1358 bytes
17 0.021821 192.168.172.20 239.2.2.40 UDP Source_port:_bre Destination_port:_cisco-sccp 1358 bytes

Figure 1. Fragment of measured network traffic

The graph of the measured data (MPEG) in this work is shown in Figure 2. The number of packets is displayed vertically, the time (in seconds) is displayed horizontally. It is visually observed that the series has an uneven distribution. The practice has shown that the measured time series are non-stationary. In such processes, fuzzy models have the best non-linear processing capabilities; for this reason, many researchers use them for forecasting.

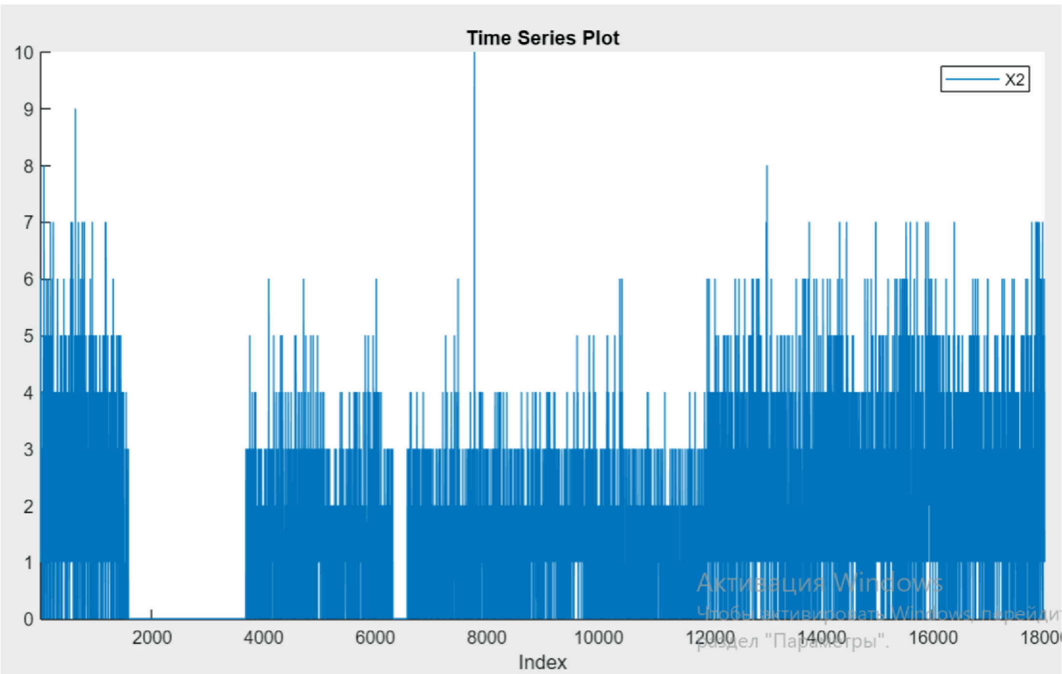

Figure 2. Packet intensity time series

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

Purpose and objectives of the study

The purpose of this work is to predict network traffic for controlling information flows in the network in order to avoid overloads and losses. The paper considers a one-dimensional time series of the intensity of MPEG (Moving Picture Experts Group) protocol packets, measured on the backbone network of the city for five hours at constant time intervals. It was visually established that the intensity of the studied series is heterogeneous, and the structure is complex. Fuzzy logic algorithms are one of the reliable methods for complex process control systems. For the software implementation of forecasting a one-dimensional time series using fuzzy logic, by authors the following tasks were set:

- making a selection from the original data series (1800 points);
- splitting the data series into non-overlapping intervals;
- creation of a fuzzy set for each interval;
- construction of graphs of membership functions for intervals used for fuzzification;
- building a forecast of future values of the time series;
- MSE mean square error estimation.

Methods and materials

Machine learning and deep learning algorithms [7] are increasingly used for time series forecasting problems. And statistical methods are based on strict obedience to the restrictive assumptions of a linear, normal distribution, stationarity of the series, or transformation of a non-stationary series into a stationary series [8].

Non-linear identification with prediction is also possible using neural network prediction. Neural networks and deep learning show very reliable and adequate results when working with similar tasks. The capabilities of neural networks can be enhanced by technologies of information processing based on fuzzy sets and fuzzy inference. In this paper, the fuzzy logic algorithms of Chen and Cheng [9] are chosen as forecasting tools.

With the help of fuzzy logic, an arbitrary relationship can be approximated exactly. Fuzzy time series is a concept that can be used for forecasting when historical data is formed in linguistic terms. At the same time, it seems that the time series has a common property, and its elements have this property to varying degrees and belong to the given set of the series with varying degrees. The algorithms are implemented using the Python programming language.

The actually measured series (MPEG protocol packet intensity) in this work shows the totality of packets transmitted over the backbone network for each second. Total data - 18000 points. Data sampling (1800 points) from the original data series (18000 points) was carried out for use in software forecasting algorithms to solve the first task. For this, the authors used the Matlab numerical simulation environment. Thus, the authors obtained the number of network packets for every second. Further, these data will be used for forecasting. The sample program code is shown in fig. 3.

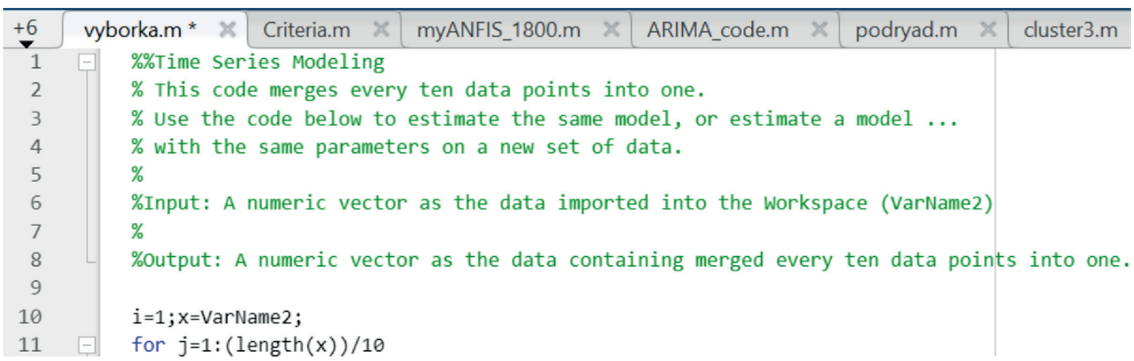

Figure 3. A sample from the original data series

DOI: 10.37943/13EOTU7482

Results

The library contains built-in modules that provide access to system functions, which provide standardized solutions to many programming problems. The “pyFTS.partition” module in Python provides functionality to split the original time series into non-overlapping intervals (Fig. 4). The pyFTS: Fuzzy Time Series for Python library is developed on MINDS – Machine Intelligence and Data Science of Federal University of Minas Gerais (UFMG) in Brazil, and is intended for researchers and data scientists to exploit the Fuzzy Time Series methods.

To solve the second and third tasks of the research in this paper, the package pyFTS was used. Several data transformations can be used for the pre-processing and/or post-processing data, which directly impacts the partitioning of the universe of discourse.

```
# importing libraries into the current environment

import dill
from scipy.spatial import KDTree

# The "pyFTS.partition" module to split the time series into non-overlapping intervals
from pyFTS.data import Enrollments
from pyFTS.partitions import Grid

from pyFTS.models import chen, cheng
import matplotlib.pylab as plt
from openpyxl import load_workbook

wb = load_workbook('C:/Users/Жанар/Desktop/Докторантура/МША/РядMPEG.xlsx')
```

Figure 4. Code snippet

The time series is divided into 10 “GridPartitioner” intervals (Fig. 5), and a fuzzy set is created for each interval. 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) [10]. The rules of these models tell how the partitions relate with themselves over time, as values jump from one place to another.

```
count_intervals = 10

    #Universe of Discourse Partitioner
    #A class containing all the entities referred to in a discourse or argument.
partitioner = Grid.GridPartitioner(data=train,npart=count_intervals)
```

Figure 5. Partitioning into “GridPartitioner” intervals

To solve the fourth problem in this research - construction of graphs of membership functions for intervals used for fuzzification, some definitions are needed. A fuzzy set is an extension of a classical set. If X is the universe of discourse and its elements are denoted by x, then a fuzzy set A in X is defined as a set of ordered pairs.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

$$A\ \{x, \mu_A(x)|\ x \in X\}$$
$$A = \{x, \mu_A(x)|\ x \in X\} \quad (1)$$

$\mu_A(x)$ is called the membership function (or MF) of x in A. The membership function maps each element of X to a membership value between 0 and 1. Figure 6 shows a graph of membership functions for 10 intervals used for fuzzification. Centroid defuzzification returns the center of gravity of the fuzzy set along the x-axis. The centroid is computed using the following formula, where $\mu(x_i)$ is the membership value for point x_i in the universe of discourse.

$$\text{xCentroid}=\frac{\sum_i \mu(x_i)(x_i)}{\sum_i \mu(x_i)} \quad (2)$$

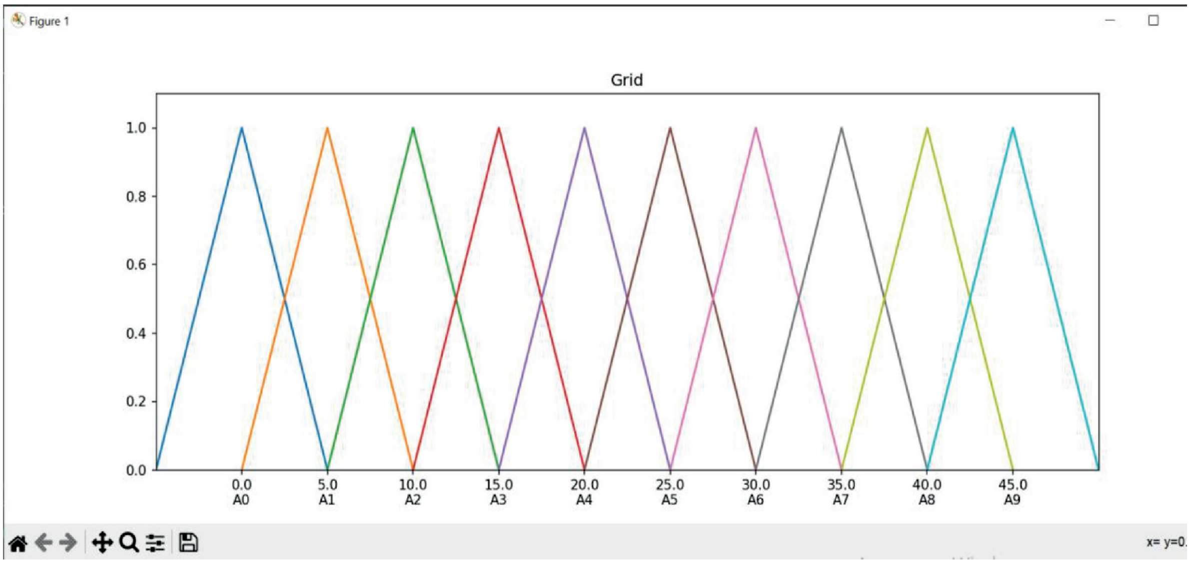

Figure 6. Graph of membership functions for 10 intervals, used for fuzzification

The fifth task in this research was building a forecast of future values of the time series. For solving this task, the neuro-fuzzy algorithms of Chen and Cheng were used, and the software implementation was carried out in the Python environment. Figure 7 shows a fragment of the program code of using Chen's algorithm and output results.

DOI: 10.37943/13EOTU7482

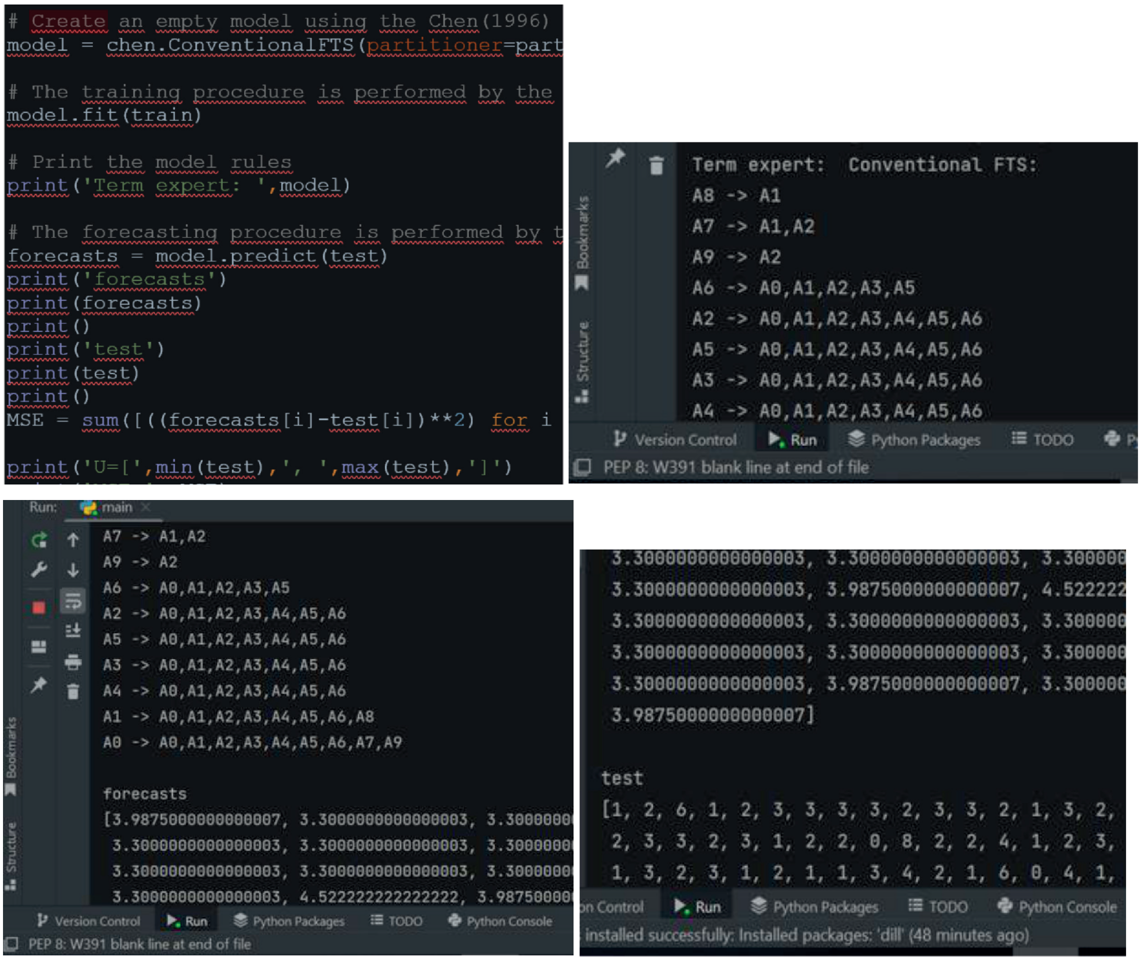

Figure 7. Chen’s Algorithm Numeric Output

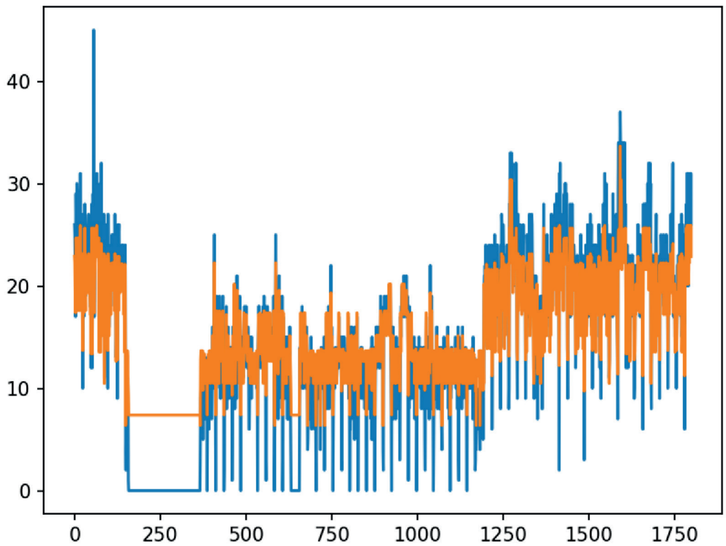

Figure 8. Combined chart of initial and forecast data

Figure 8 demonstrates a combined graph consisting of initial data and forecast data. The graph shows that Chen’s method has low accuracy rates, the original series is barely covered by predicted data points.

Figure 9 shows a fragment of the program code of using Cheng's algorithm and output results.

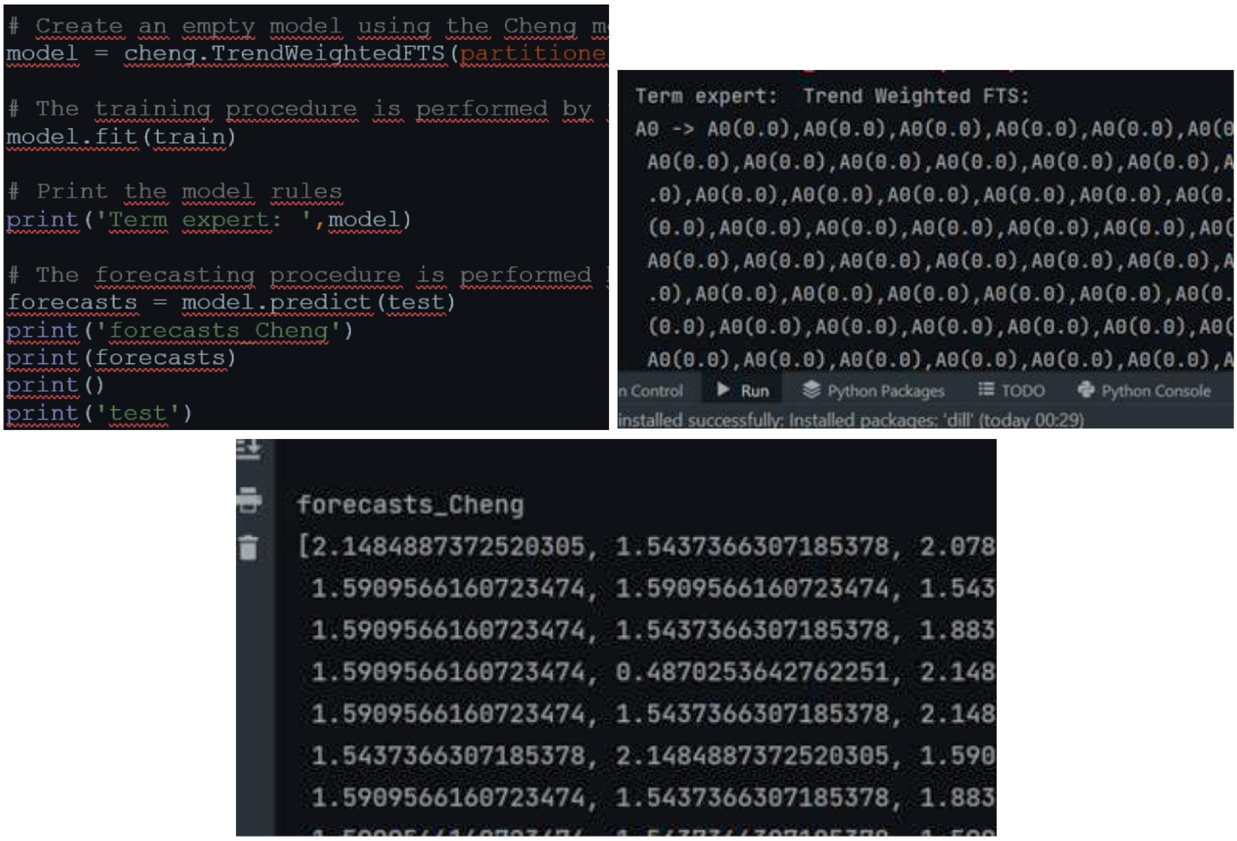

Figure 9. Cheng's Algorithm Numeric Output

Figure 10 demonstrated a combined graph consisting of initial data and forecast data built using the Cheng's algorithm. The original series is almost completely covered by the predicted data points, which indicates the adequacy of the forecast.

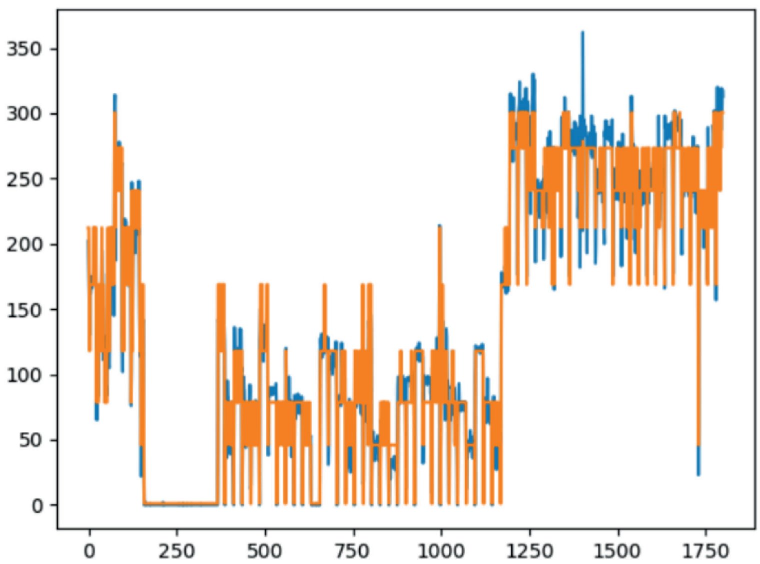

Figure 10. Combined chart of initial and forecast data

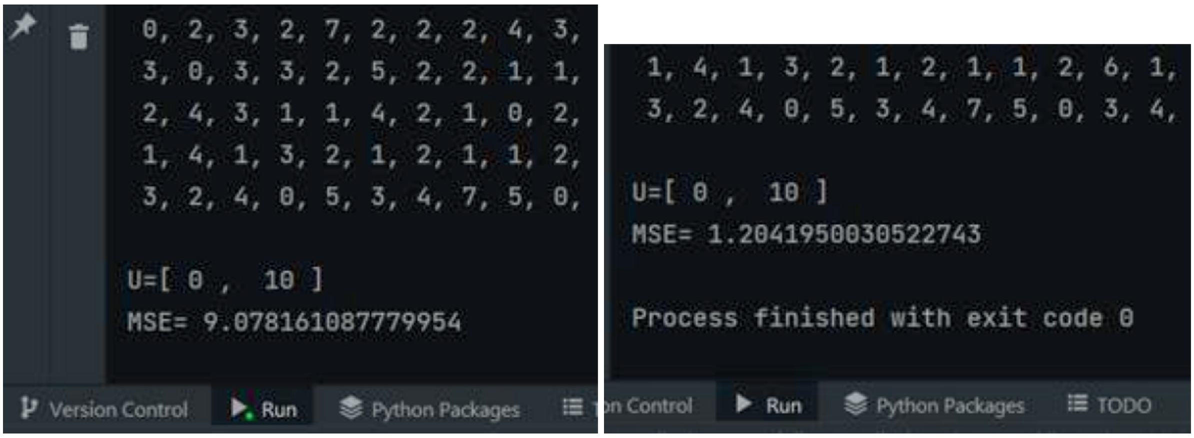

Figure 11. Output values of MSE

Table 1 shows the output values of the program using Chen's and Cheng's algorithms. Cheng algorithm shows that this fuzzy forecasting method has higher forecast accuracy indicators, as evidenced by a lower numerical value of the MSE indicator (Fig.11.).

Table 1. Comparison output values of Chen's and Cheng's algorithms

	Chen's algorithm	Cheng's algorithm
number of intervals, used for fuzzification	10	10
MSE (mean square error)	9.078161087779954	1,2041950030522743

Conclusion

The modern heterogeneous network grows every second, and network traffic changes. The more complex the system, the more it depends on a large number of parameters. Therefore, they should be processed by non-linear approaches. As a result of the study, the time series graph (Fig. 2) showed that it has an uneven packet intensity, which allows for the non-stationarity of the series. Numerous studies have shown that fuzzy logic methods give good results for non-linear processes.

The authors have done a great deal of work on the study of this problem, and the software implementation of neuro-fuzzy inference algorithms has been carried out. The described research methods clearly demonstrate the advantages of using neuro-fuzzy algorithms (in particular, those implemented in the universal programming language Python) for modeling the characteristics of complex systems. One of the most popular in the world and, perhaps, the universal programming language Python has all the necessary sets of modules for organizing the research methods indicated for the purpose of this article and for their further development and adaptation for use in almost any field of science and technology.

Chen's algorithm has some drawbacks: a decrease in the accounting of repetition with an increase in the duration of observation and the absence of weighting factors. Cheng's method has higher forecast accuracy than Chen's forecast method. The differences in these methods are in the stages of formation of fuzzy sets, and in the presence of weights in each group of fuzzy relations.

A combined graph consisting of initial data and forecast data was built using Cheng algorithm (fig.10.), which shows that this fuzzy forecasting method has higher forecast accuracy indicators, as evidenced by a lower numerical value of the MSE indicator [11]. This result indicates that Cheng's fuzzy time series method is good enough to be used in time series forecasting.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

The proposed method of fuzzy forecasting is quite universal and will increase the level of validity of predictive solutions in the problems of designing complex systems and managing their development.

Reference

1. Serikov, T., Zhetpisbayeva, A., Mirzakulova, S., Zhetpisbayev, K., Ibrayeva, Z., Tolegenova, A., ... & Zhumazhanov, B. (2021). Application of the NARX neural network for predicting a one-dimensional time series. *Eastern-European Journal of Enterprise Technologies*, *5*(4), 113. https://doi.org/10.15587/1729-4061.2021.242442
2. Bektemyssova, G., Ahmad, A. R., Mirzakulova, S., & Ibraeva, Z. (2022). TIME SERIES FORECASTING BY THE ARIMA METHOD. *Scientific Journal of Astana IT University*, 14-23. https://doi.org/10.37943/HFCH4395
3. Al-Saati, N.H., Omran, I.I., Salman, A.A., Al-Saati, Z., & Hashim, K.S. (2021). Statistical modeling of monthly streamflow using time series and artificial neural network models: Hindiya Barrage as a case study. *Water Practice and Technology*, *16*(2), 681-691. https://doi.org/10.2166/wpt.2021.012
4. Chen, S. M. (2002). Forecasting enrollments based on high-order fuzzy time series. *Cybernetics and Systems*, *33*(1), 1-16. https://doi.org/10.1080/019697202753306479
5. Afanaseva, T.V., Namestnikov, A.M., Perfileva, I.G., & Romanov, A.A. (2014). Prognozirovanie vremennykh ryadov: nechetkie modeli [Forecasting time series: fuzzy models]. *Ulyanovsk, UlGTU Publ*. http://lib.ulstu.ru/venec/disk/2015/192.pdf
6. Yu, H. K. (2005). Weighted fuzzy time series models for TAIEX forecasting. *Physica A: Statistical Mechanics and its Applications*, *349*(3-4), 609-624. https://doi.org/10.1016/j.physa.2004.11.006
7. Manenti, F., Rossi, F., Goryunov, A.G., Dyadik, V.F., Kozin, K.A., Nadezhdin, I.S., & Mikhalevich, S.S. (2015). Fuzzy adaptive control system of a non-stationary plant with closed-loop passive identifier. *Resource-Efficient Technologies*, *1*(1), 10-18. https://doi.org/10.1016/j.reffit.2015.07.001
8. El-Bardini, M., & El-Nagar, A. M. (2014). Interval type-2 fuzzy PID controller for uncertain nonlinear inverted pendulum system. *ISA transactions*, *53*(3), 732-743. https://doi.org/10.1016/j.isatra.2014.02.007
9. F. Manenti. (2011). Considerations on nonlinear model predictive control techniques. *Comput. Chem. Eng*, *35*(11), 2491–2509. https://doi.org/10.1016/j.compchemeng.2011.04.009
10. Dalvandi, A., Gurusamy, M., & Chua, K. C. (2015). Power-efficient resource-guaranteed VM placement and routing for time-aware data center applications. *Computer Networks*, *88*, 249-268. https://doi.org/10.1016/j.comnet.2015.06.017
11. Tu, R., & Wang, L. (2014). Real-time coseismic wave retrieving by integrated Kalman filter with observations of GPS, Glonass and strong-motion sensor. *Advances in Space Research*, *53*(1), 130-137. https://doi.org/10.1016/j.asr.2013.10.011