EURASIP Journal on
Information Security

RESEARCH Open Access

Research on gray correlation analysis and situation prediction of network information security

Chengqiong Ye, Wenyu Shi and Rui Zhang*

Abstract

In order to further improve the accuracy and efficiency of network information security situation prediction, this study used the dynamic equal-dimensional method based on gray correlation analysis to improve the GM (1, N) model and carried out an experiment on the designed network security situation prediction (NSSP) model in a simulated network environment. It was found that the predicted result of the improved GM (1, N) model was closer to the actual value. Taking the 11th hour as an example, the predicted value of the improved GM (1, N) model was 28.1524, which was only 0.8983 larger than the actual value; compared with neural network and Markov models, the error of the improved GM (1, N) model was smaller: the average error was only 2.3811, which was 67.88% and 70.31% smaller than the other two models. The improved GM (1, N) model had a time complexity that was 49.99% and 39.53% lower than neural network and Markov models; thus, it had high computational efficiency. The experimental results verify the effectiveness of the improved GM (1, N) model in solving the NSSP problem. The improved GM (1, N) model can be further promoted and applied in practice and deployed in the network of schools and enterprises to achieve network information security.

Keywords: Gray correlation analysis, Information security, Situation prediction, Error

1 Introduction

Due to the unique openness of the network, information security issues have become more prominent [1], which brings huge threats to individuals, society, and countries. Technologies such as the traditional firewall [2], intrusion detection [3], and digital encryption [4] have not been able to deal with the existing attacks and threats. An active and reliable security strategy is urgently needed. Network security situation awareness (NSSA) is a process that can comprehensively analyze the network security status [5]. It can obtain the situation elements in a large-scale network and calculate and analyze them to predict the future trend of the network [6]. Network security situation prediction (NSSP) is an important technology in NSSA. With the emergence of machine learning, the computation ability of computers has been further improved [7], which provides more methods for NSSP. The commonly used methods include neural network, time series, and gray correlation analysis [8]. The neural network method has an excellent self-learning ability, but it has randomness and is prone to local convergence. The time series method is based on the periodicity and regularity of the situation, but due to uncertain factors in the actual network, its prediction results are not accurate. The gray correlation analysis method generates new data by accumulating historical data [9], which has the advantages of simple modeling and fast calculation; however, it may cause large errors when the randomness of the network is large. NSSA was first defined by Endsley [10]. Then, with the development of network technology, there are more researches on NSSA. Bode et al. [11] developed a Bayesian network classifier to analyze network traffic and further analyzed

* Correspondence: nzre78u@126.com
Institute of Information Engineering, Anhui Xinhua University, No. 555, Wangjiang West Road, Hefei 230088, Anhui, China

the risk level using the modified risk matrix standard. The experiments on the KDD Cup 99 data set showed that the model was suitable and well developed in network security. Aiming at the deficiency of the gray Verhulst model, Leau et al. [12] designed an adaptive gray Verhulst model with adjustable generation order, tested the model with DARPA 1999 and 2000 benchmark data sets, and found that the model showed good performance in predicting the network. Kim et al. [13] pointed out that the traditional time series analysis could not predict the dynamic network and proposed a hidden Markov model (HMM) to analyze and predict the real-time changes of network traffic. Holsopple et al. [14] designed a FuSIA framework to predict the possible future attacks, which used uncertain observability to determine the current and future impacts of key tasks in the application. Panigrahi et al. [15] proposed a new autoregressive integrated moving average-artificial neural network (ARIMA-ANN) hybrid model for time series prediction, which used a fuzzy filter to decompose the time series into low-volatile and high-volatile components. The low-volatile components were modeled by ARIMA, and the high-volatile components were modeled by ANN. The final prediction was obtained by combining the prediction of ARIMA and ANN models. The experiment found that the hybrid model was superior to ARIMA and ANN models. In the current research, although a lot of achievements have been made, the accuracy and timeliness of predictions need to be further improved. In order to find a more efficient NSSP method to achieve better and faster predictions for the network security situation, this study designed an improved gray relational analysis (GRA)-based NSSP model and performed simulation experiments and analyses on the model. The experimental results verified that the method was effective in predicting the network security situation, which makes some contributions to the further development of network information security.

2 Methods

2.1 GM (1, N) model

In GRA theory, the GM (1, 1) model is a typical model used in the early stage [16]. Suppose the original NSS data sequence is: ${X^{(0)}}_{(t)} = \{{x^{(0)}}_{(1)}, {x^{(0)}}_{(2)}, \cdots, {x^{(0)}}_{(n)}\}$. Let ${X^{(1)}}_{(t)} = \sum_{i=1}^{t} x^{(0)}(i), i = 1, 2, \cdots, n$. After AGO processing, there is: ${X^{(1)}}_{(t)} = \{{x^{(1)}}_{(1)}, {x^{(1)}}_{(2)}, \cdots, {x^{(1)}}_{(n)}\}$. GM (1, 1) model is described by a differential equation, $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = \phi$, where a and ϕ are undetermined parameters. $(a, \phi)^T = (A^T A)^{-1} A^T y$, where $A = \begin{bmatrix} -\frac{[x^{(1)}(t)+x^{(1)}(2)]}{2} & 1 \\ \vdots & \vdots \\ -\frac{[x^{(1)}(n-1)+x^{(1)}(n)]}{2} & 1 \end{bmatrix}$ and $y = [x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n)]$. Then, the cumulative sequence value can be written as: $\hat{x}(t+1) = [x^{(1)}(1) - \frac{\phi}{a}]e^{-at} + \frac{\phi}{a}$. After reduction, the prediction result of GM (1, 1) model can be obtained: $\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t)$.

GM (1, 1) model can only be used in the case of a single change, and the error is uncontrollable; therefore, it is not suitable for solving NSSP problems. GM (1, N) has high accuracy [17], which is more suitable for situation prediction. In GM (1, N) model, the differential equation is written as: $\frac{dx^{(1)}(1)}{dt} + ax^{(1)}(1) = \phi_1 x^{(1)}(2) + \phi_2 x^{(1)}(3) + \cdots + \phi_{N-1} x^{(1)}(N)$. Then, the value of $x^{(1)}(t)$ can be written as: $x^{(1)}(t) = e^{-at}[\sum_{i=2}^{N} \int \phi_{i-1} x^{(1)}(t) e^{at} dt + x^{(1)}(0) - \sum_{i=2}^{N} \int \phi_{i-1} x^{(0)}(t) dt]$. After reducing the cumulative sequence value, there is $\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t)$.

2.2 NSSP model based on improved GRA

Situation value is an ever-changing dynamic value. In order to predict it better, this study improved the GM (1, N) model and combined the dynamic equal dimension method. According to GM (1, N), the original data are accumulated once: $x^{(1)}(t) = \sum_{i=1}^{t} x^{(0)}(k)$. According to the original data, GM (1, N) of different dimensions is established. The approximate time response formula is obtained: $\hat{x}^{(1)}(t+1) = [x_1^{(0)}(1) - \frac{1}{a}\sum_{i=2}^{N} \phi_i x_i^{(1)}(t+1)]e^{-at} + \frac{1}{a}\sum_{i=2}^{N} \phi_i x_i^{(1)}(t+1)$. After reduction, the prediction model is: $\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(1)(t)$. The predicted value $\hat{x}^{(0)}(n+1)$ is substituted into the original sequence to remove the original $x^{(0)}(1)$ and generate a new sequence, i.e., the real-time data obtained by prediction replace the early data. The above steps repeat until the predicted target is obtained.

The method is applied to NSSP. The original security situation sequence is set as: $s^{(0)} = (s^{(0)}(1), s^{(0)}(2), \cdots, s^{(0)}(n))$, where s is the gray correlation factor of the security situation. $s^{(0)}$ can be obtained by performing 1-AGO on $s^{(0)}$. Then, through the adjacent mean generating sequence, $z^{(1)}$ is obtained: $z^{(1)} = 0.5s^{(1)}(t) + 0.5s^{(1)}(t-1)$. Accuracy test was performed on the improved GM (1, N). The error is:

$$\varepsilon(t) = X^{(0)}(t) - Y^{(0)}(t),$$

where $X^{(0)}(t)$ refers to the actual security situation sequence and $Y^{(0)}(t)$ is the sequence predicted by GM (1, N).

3 Results

3.1 Experimental data

Suppose several hosts are included in the network system, providing p kinds of services $S_i(1 \le i \le p)$ and being attacked by A_j attacks, the severity degree of attack is T_{A_j}, the value of attack class is C, $1 \le j \le c$, the time interval of attacks is τ, the importance of time interval is w_τ, the detection value of A_j is N_{A_j}, and the value for dividing time intervals is C_T. Then, the risk index of S_i is written as:

$$I_{S_i} = \sum_{\tau=1}^{C_T} w\tau \sum_{j=1}^{C} 10 T_{A_j} N_{A_j}.$$

Suppose that the total amount of S_i is W_{S_i} in the system. Then, the NSS value of the system can be written as:

$$NSS = \sum_{i=1}^{p} W_{S_i} \sum_{\tau=1}^{C_T} w_\tau \sum_{j=1}^{C} 10 T_{A_j} N_{A_j}.$$

A network system was simulated, including three servers, which provided WWW service, e-mail service, and file transfer protocol (FTP) service, respectively. The computer of the simulated sensor was responsible for linking the small local area network of the attacker and the attacked computer. When the attack was launched, the attack packet was crawled and reported to the data aggregation server. The specific process is as follows. One network card grabbed the attack packet and transmitted it to another network card. The network card analyzed the data and finally reported it to the aggregation server for final processing. The aggregation server collected the state information of the attacked host while receiving the data transmitted by the sensor. The two kinds of data were compared to determine whether the host was attacked. The NSS value of the network was calculated through the calculation formula of the NSS value. The calculation of the NSS value lasted for 20 h, as shown in Table 1. The first 10 h were used for model training, and the last 10 h were used for prediction.

Table 1 NSS values within 20 h

Time	NSS value
1	3.2158
2	4.2514
3	4.4128
4	14.5514
5	15.2541
6	16.2845
7	17.8524
8	18.2987
9	21.5585
10	22.2541
11	27.2541
12	28.9518
13	19.2514
14	16.2517
15	17.2599
16	18.3784
17	19.2865
18	21.2854
19	25.2168
20	26.8518

3.2 Prediction results

GM (1, 1), GM (1, N), and improved GM (1, N) models were used in NSSP. The predicted results were compared with the actual NSS values, and the results were drawn into a line chart (Fig. 1).

It was seen from Fig. 1 that there was a large gap between the results of GM (1, 1) and GM (1, N) models and the actual value. Taking the 11th hour as an example, the actual NSS value was 27.2541, the prediction result of the GM (1, 1) model was 46.1285, which was 18.8744 larger than the actual value; the prediction result of the GM (1, N) model was 37.2651, which was 10.011 larger than the actual value; the predicted value of the improved GM (1, N) model was 28.1524, which was only 0.8983 larger than the actual value. It was found that the result of the improved model was closest to the actual NSS value, and it had a better performance in solving NSSP problems.

In order to further verify the effectiveness of the model, it was compared with the neural network model [18] and the Markov model [19]. The line chart was also used to compare the predicted results between different models (Fig. 2).

As shown in Fig. 2, the predicted value of the improved model was in good agreement with the actual value, and the values were close; the prediction results of the other two models fluctuated greatly, and the differences with the actual values were large. For example, at the 20th hour, the actual NSS value was 26.8518, and the predicted value of the three models was 34.9477, 20.6485, and 24.3196, respectively, and the difference between the predicted value and the actual value was 8.0959, 6.2033, and 2.5322, respectively. The results of the improved model were closest to the actual values.

The errors of the models in Fig. 2 are calculated, and the results are shown in Table 2.

It was seen from Table 2 that the maximum and minimum errors of the neural network model were 8.6667 and 2.6606, respectively, the maximum and minimum

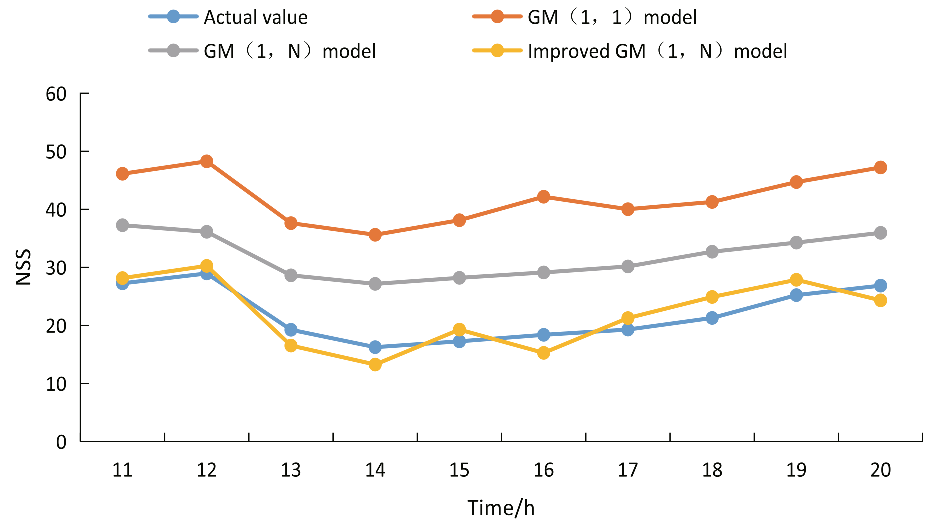

Fig. 1 Prediction results of different gray models

errors of the Markov model were 9.5692 and 5.4318, respectively, and the maximum and minimum errors of the improved model were 3.6167 and 0.8983, respectively, which were significantly smaller than the other two models. The average error of the three models was 7.4138, 8.0211, and 2.3811, respectively; the average error of the improved model was 67.88% and 70.31% smaller than the other two models, which indicated the advantage of the improved model in NSSP.

The time complexity of different models in prediction was compared, and the results are shown in Fig. 3.

Figure 3 shows the time complexity of different models in NSSP. When predicting the NSS value, the neural network model had the largest time complexity, followed by the Markov model and the improved GM (1, N) model. The time the three models needed was 42.67 s, 35.29 s, and 21.34%, respectively. The time complexity of the Markov model was 17.3% lower than that of the neural network model. The time complexity of the improved GM (1, N) model was 49.99% lower than that of the neural network and 39.53% lower than that of the Markov model, which verified the advantage of the improved GM (1, N) model in computation efficiency.

4 Discussion

With the development of the Internet of things, more and more devices have been connected to the network [20], further strengthening the openness of the network [21].

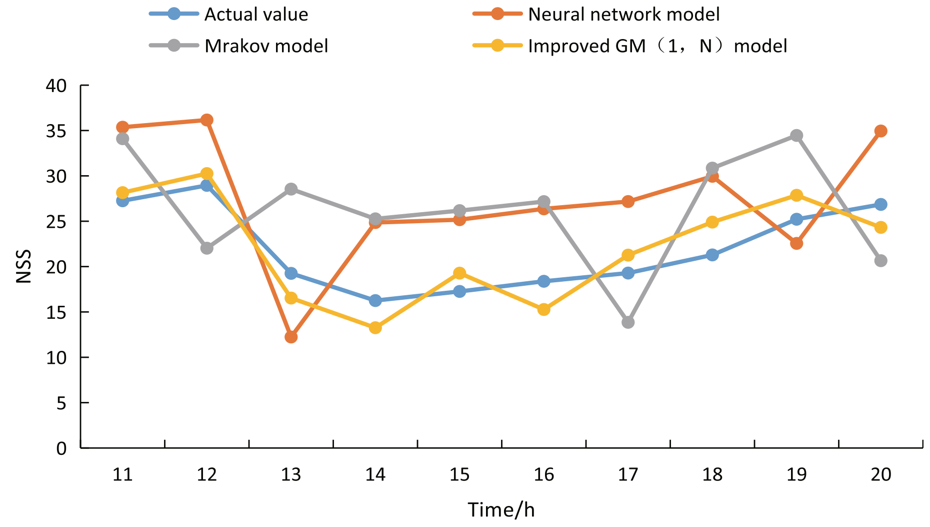

Fig. 2 The results of comparison with the other models

Table 2 Comparison of errors between errors

	Neural network model	Markov model	Improved GM (1, N) model
11	8.1113	6.8424	0.8983
12	7.2079	6.9267	1.303
13	7.0146	9.2898	2.7227
14	8.603	9.0118	2.9933
15	7.9086	8.9099	2.0163
16	7.9864	8.7863	3.1086
17	7.8833	5.4318	1.9773
18	8.6667	9.5692	3.6167
19	2.6606	9.2398	2.6427
20	8.0959	6.2033	2.5322

The network is always faced with a variety of malicious attacks and threats [22], which will damage any operation of the target computer and bring huge reputation and property losses [23]. In order to achieve network security, it is necessary to detect the attacks in the network in advance, take corresponding measures to curb the threats in time, and protect the information security in the network actively. Therefore, network managers need to sense the threat in time, accurately grasp the status of the network, and predict the future development trend. NSSA technology can convert the changes in network traffic and resource occupancy rate into security situation information when attacks happen to provide reliable support for risk assessment and prediction, including data fusion [24], network security situation evaluation [25], and NSSP. This study mainly analyzed NSSP.

The GRA method can find the rule in the sequence and use the rule to predict the sequence, which has a good performance in short-term prediction. Based on the GRA method, this study introduced GM (1, 1) and GM (1, N) models and applied them to the solution of the NSSP problem. In order to obtain better accuracy, the GM (1, N) model was improved by the dynamic equal dimension method. It was found that the GM (1, 1) model and GM (1, N) model both showed large errors in the prediction of situation value, more than ten, and the results of the improved GM (1, N) model were closer to the actual NSS value, which showed that the method had a high prediction accuracy. Then, the comparison with the other methods demonstrated that the neural network and Markov model showed great volatility and large errors in NSS prediction, and the average errors were 7.4138 and 8.0211, respectively. It was seen from Fig. 2 that the results of the improved GM (1, N) model had better similarity with the actual value and the average error was 2.3811, which was significantly smaller than the other two methods. The above results revealed that the improved GM (1, N) model had a better performance in the NSSP problem.

5 Conclusions

Aiming at the NSSP problem, this study analyzed the advantages of the GRA method, improved the GM (1, N)

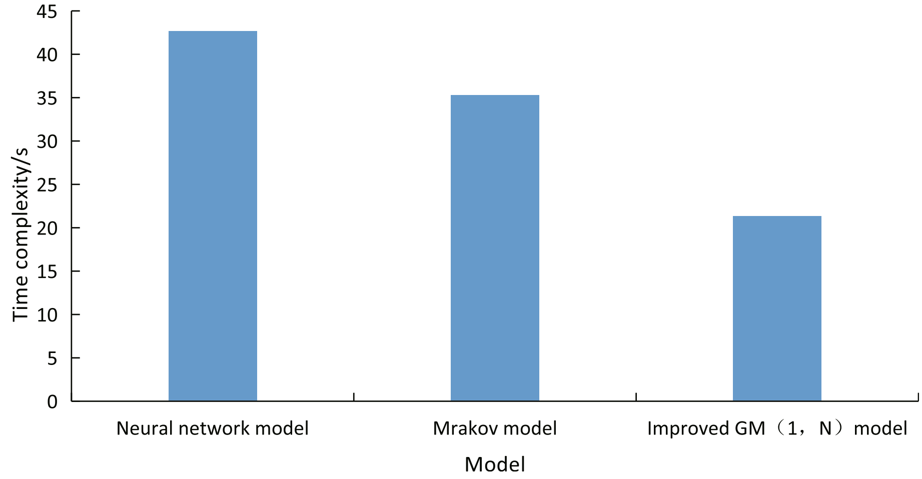

Fig. 3 Comparison of time complexity between models

model, and conducted simulation experiments. The results showed that (1) there were large errors between the prediction results of GM (1, 1) and GM (1, N) models and the actual values; (2) the predicted value obtained by the improved GM (1, N) model was closer to the actual value. Taking the 20th hour as an example, the error between the predicted value and the actual value was only 2.5322; (3) compared with the neural network and Markov model, the prediction accuracy of the improved GM (1, N) model was higher, and the average error was only 2.3811.

The experimental results verify that the improved GM (1, N) model is reliable and can be popularized and applied in practice to accurately predict the situation to realize the network information security. In future research, the accuracy of the GM (1, N) model will be further improved, and experiments will be carried out in a larger network and actual network environment to further verify the performance and practical application ability of the model.

Abbreviations

HMM: Hidden Markov model; ARIMA: Autoregressive integrated moving average; ANN: Artificial neural network; NSSA: Network security situation awareness; NSSP: Network security situation prediction; PSO: Particle swarm optimization; BP: Back propagation; BPA: Basic probability assignment; RNN: Recurrent neural network; GRA: Gray relational analysis; FTP: File transfer protocol

Acknowledgements

Not applicable.

Authors' contributions

CQY conceived of the study. CQY and RZ designed the study. All authors analyzed the data and were involved in writing and revisions of the manuscript. The author(s) read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 19 January 2021 Accepted: 7 April 2021
Published online: 20 April 2021

References

1. P. Bhandari, M. Singh, Semantic web based technique for network security situation awareness status prediction. Int. J. Nat. Eng. Sci. **14**, 2229–6913 (2015)
2. A.D. Brucker, L. Brügger, B. Wolff, Formal firewall conformance testing: an application of test and proof techniques. Softw. Test. Verif. Rel. **25**, 34–71 (2015). https://doi.org/10.1002/stvr.1544
3. Y. Hamid, V.R. Balasaraswathi, L. Journaux, M. Sugumaran, Benchmark Datasets for network intrusion detection: a review. Int. J. Netw. Secur. **20**, 645–654 (2018)
4. R. Yadav, V. Kapoor, A hybrid cryptography technique for improving network security. Int. J. Comput. Appl. **141**, 25–30 (2016)
5. T. Jirsik, P. Celeda, *Toward real-time network-wide cyber situational awareness* (2018), pp. 1–7. https://doi.org/10.1109/NOMS.2018.8406166
6. M. Azhagiri, A. Rajesh, S. Karthik, A multi-perspective and multi-level analysis framework in network security situational awareness. Int. J. Comput. Netw. Commun. Sec. **5**, 71–75 (2017)
7. D.M. Li, L.B. Deng, B.B. Gupta, H.X. Wang, C. Choi, A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. Inform. Sci. **479**, 432–447 (2019). https://doi.org/10.1016/j.ins.2018.02.060
8. M. Husák, J. Komárková, E. Bou-Harb, P. Celeda, Survey of attack projection, prediction, and forecasting in cyber security. IEEE Commun. Surv. Tut. **21**, 640–660 (2019)
9. Z.M. Yunos, W. Idrus, S.M. Shamsuddin, M.S.I. Saadon, S.M. Yusuf, *Enhanced of autism spectrum disorder using grey relational analysis and supervised learning for classification. Paper presented at 2nd Joint Conference on Green Engineering Technology & Applied Computing* (IOP Publishing Ltd, Bangkok, 2020)
10. M.R. Endsley, Design and evaluation for situation awareness enhancement. Proc. Hum. Factors Soc. Ann. Meet. **32**, 97–101 (1988). https://doi.org/10.1177/154193128803200221
11. M.A. Bode, S. Oluwadare, B.K. Alese, A. Thompson, *Risk analysis in cyber situation awareness using Bayesian approach. Paper presented at 2015 International Conference on Cyber Situational Awareness* (Data Analytics and Assessment (CyberSA), London, 2015)
12. Y.B. Leau, S. Manickam, A novel adaptive grey Verhulst model for network security situation prediction. Int. J. Adv. Comput. Sci. Appl **7**, 90–95 (2016)
13. D. Kim, T. Lee, S.D. Jung, H.P. In, H. Lee, *Cyber threat trend analysis model using HMM. Paper presented at Proceedings - IAS 2007 3rd Internationl Symposium on Information Assurance and Security* (IEEE Computer Society, Manchester, 2007)
14. J. Holsopple, S.J. Yang, FuSIA: Future situation and impact awareness. Inform. Fusion, 1–8 (2008). https://doi.org/10.1109/ICIF.2008.4632456
15. S. Panigrahi, H.S. Behera, A. Abraham, *A fuzzy filter based hybrid ARIMA-ANN model for time series forecasting. Paper presented at Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016)* (VIT University, Vellore, 2016), pp. 19–21
16. M. Elsayed, A. Soliman, Prediction of technical reserves based on grey model - GM(1,1): evidence from non-life Egyptian insurance market. J. Busin. Econ **10**, 852–860 (2019). https://doi.org/10.15341/jbe(2155-7950)/09.10.2019/006
17. K.C. Chiu, C.S. Lai, T. Thi, *GM(1,N) Analysis of effect of consumer confidence and transaction security on Vietnamese online shopping intention. Paper presented at ICEMC 2019: Proceedings of the 2019 5th International Conference on E-business and Mobile Commerce* (ICEMC, Taichung, 2019), pp. 25–29. https://doi.org/10.1145/3332324.3332331.
18. J. Schmidhuber, Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015). https://doi.org/10.1016/j.neunet.2014.09.003
19. G.T. Lakshmanan, D. Shamsi, Y.N. Doganata, U. Merve, R. Khalaf, A markov prediction model for data-driven semi-structured business processes. Knowl. Inf. Syst. **42**, 97–126 (2015). https://doi.org/10.1007/s10115-013-0697-8
20. A. Tewari, B.B. Gupta, Cryptanalysis of a novel ultra-lightweight mutual authentication protocol for IoT devices using RFID tags. J. Supercomput. **73**, 1–18 (2017). https://doi.org/10.1007/s11227-016-1849-x
21. A. Tewari, B.B. Gupta, Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework. Future Gener. Comp. Sy. **108**, 909–920 (2018). https://doi.org/10.1016/j.future.2018.04.027
22. V. Adat, B.B. Gupta, Security in Internet of Things: issues, challenges, taxonomy, and architecture. Telecommun. Syst **67**, 423–441 (2018). https://doi.org/10.1007/s11235-017-0345-9
23. A. Dahiya, B.B. Gupta, A reputation score policy and Bayesian game theory based incentivized mechanism for DDoS attacks mitigation and cyber defense. Future Gener. Comp. Sy. **117**, 193–204 (2021). https://doi.org/10.1016/j.future.2020.11.027
24. A.G. Morosan, F. Pop, *OCPP security - neural network for detecting malicious traffic. Proceedings of the International Conference on Research in Adaptive and Convergent Systems* (RACS '17, Krakow, 2017), pp. 190–195. https://doi.org/10.1145/3129676.3129693.
25. I. Kotenko, E. Doynikova, *Security evaluation for cyber situational awareness* (2014), pp. 1197–1204. https://doi.org/10.1109/HPCC.2014.196

Publisher's Note