Journal of Autonomous Intelligence (2023) Volume 6 Issue 3
doi: 10.32629/jai.v6i3.882

ORIGINAL RESEARCH ARTICLE

Cybersecurity threat perception technology based on knowledge graph

A. Sali[1,*], Abdulmajeed Al-Jumaily[1,2,*], Víctor P. Gil Jiménez[2], Dhiya Al-Jumeily[3]

[1] *Wireless and Photonic Networks Research Centre of Excellence (WiPNET), Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia*

[2] *Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Madrid, Spain*

[3] *School of Computer Science & Mathematics, Liverpool John Moores University, Liverpool L3 3AF, UK*

*** Corresponding author:** A. Sali, aduwati@upm.edu.my; Abdulmajeed Al-Jumaily, abdulmajeed@tsc.uc3m.es

ABSTRACT

The issue of complex sources, difficult to understand and share security threat intelligence, this paper realizes deep learning of threat intelligence features based on Restricted Boltzmann Machine, which graphs the original threat intelligence features from high dimensional space to low dimensional space layer by layer, and constructs the cyberspace security threat knowledge graphs. The deep learning used to build a multi-level and structured knowledge graph of cyberspace security threats can reflect the structural characteristics of the knowledge graph, making the graph have a lower dimension and a higher level of abstraction. The experiment verifies the feasibility of constructing the cyberspace security threat knowledge graph, and verifies the security threat perception method based on the knowledge graph is more suitable for the perception of high-intensity security threats by comparing with traditional threat detection methods.

Keywords: knowledge graphs; threat intelligence; Restricted Boltzmann Machine; security threat perception; threat detection

ARTICLE INFO

Received: 4 July 2023
Accepted: 17 July 2023
Available online: 26 September 2023

1. Introduction

Cyberspace security threat intelligence data comes from a wide range of sources, resulting in serious data redundancy and even errors, and the relationship between different security entities is complex and hidden. Therefore, it is necessary to carry out security knowledge fusion. That is to build a security knowledge specification, form a unified and complete basic knowledge expression, and on this basis, perform operations such as debug, deduplication, association, verification, and update on multi-source data, and finally form a high-quality and understandable cyberspace security threat knowledge graph. under the new situation, cyberspace security threats are developing towards intelligence, automation, and scale. The types and harms of security threats are also growing rapidly, and their impact is getting bigger and wider.

After a long-term systematic and targeted development, traditional security protection technologies have become mature, such as access control, attack detection, malicious code prevention, which have played an important role in network security protection, greatly reducing network damage caused by the attack. However, under the new situation of cyberspace attack and defense confrontation, security threats are changing with each passing day, and the development of

traditional security protection methods has not kept up with the update of attack technology. Traditional network security protection focuses on passive defense, which belongs to static protection. Pure passive and static protection can no longer meet the needs of cyberspace security. It is urgent to innovate security concepts, and develop dynamic active security defense technologies combined with knowledge graphs, big data, and security threat intelligence.

Knowledge graph technology can clearly display the logical relationship between various information subjects in cyberspace, such as the relationship between different attacks, the relationship between attacks and vulnerabilities, the relationship between vulnerabilities and vulnerabilities, and the relationship between different security threats. Knowledge graph is essentially a semantic network, which consists of nodes and edges, where nodes represent entities/concepts, and edges represent semantic relationships between entities/concepts. Using the knowledge graph, through the analysis of various security subjects, attributes of security subjects and the relationship between security subjects, the deeper relationship between security subjects can be judged and reasoned.

2. Related work

2.1. Knowledge graph

Knowledge graph is one of the important research contents of artificial intelligence technology. The knowledge base built on it has efficient and open semantic processing capabilities, and has been widely used in scenarios such as intelligent recommendation and intelligent question answering[1]. Furthermore, the follows authors are Huang et al.[2] and Li et al.[3]. Carried out research on human-computer interaction model and recommendation model based on knowledge graphs, evaluating the probability of user-entity interaction, and recommending interested parties to those involved in the interaction content. In the field of network security, knowledge graphs have also been widely used. Foreign researchers were the first to put forward and continuously improve the network security ontology, defining the ontology types such as goals, methods, results, vulnerabilities, threats, products, services, and processes[4,5], and clarified assets, risks, threats, attacks, Ontology definitions such as defense and influence[6], the attributes of the ontology are extended, and the relationship between the ontology is analysed and fused[7].

Domestically, Yan et al.[8] conducted in-depth research, proposed methods and deduction rules for building a network security knowledge graph, and used machine learning and Stanford NER to build a network security knowledge base[9–13], applied knowledge graphs to content security, flow rule evolution and Application, security situation prediction, network security data organization, DDoS attack source detection, etc., have achieved relevant results in network security monitoring and situation awareness, network security knowledge learning and data analysis technology[14,15] proposed a network security knowledge graph model, which represented the complexity of the knowledge graph through information entropy, and proposed a knowledge graph selection technology based on fuzzy sets. In the field of threat intelligence, Dong et al.[14] and Wang et al.[16] carried out related research, designed a knowledge graph construction framework for threat intelligence, and proposed a deep learning model for entities and entity relationships for threat intelligence knowledge graphs, and visualized using a graph database. Security threat intelligence plays an increasingly important role in the network security defense system, but there are current sources of security threat intelligence complex, difficult to understand, difficult to share and other issues. In response to these problems, this paper implements deep learning of threat intelligence features based on Restricted Boltzmann Machine (RBM), graphs the original threat intelligence features from high-dimensional space to low-dimensional space layer by layer, and builds cyberspace security threat knowledge graph, depicting the relationship between cyberspace security threat characteristics and security threat intelligence; then using the cyberspace security threat knowledge graph, combined with the current context, based on event stream processing, security threat

path evolution and traceability, accurate perception of cyberspace security threats.

2.2. Structure RBM network

RBM is a two-layer neural network mode[17–21], including visible layer and hidden layer, which can effectively complete the encoding from high-dimensional space to low-dimensional space. In view of the fact that the construction of cyberspace security threat knowledge graph needs to graph the original threat intelligence features from high-dimensional space to low-dimensional space, and to achieve multi-layer graphs, this paper uses multiple stacked RBMs to implement deep learning models to obtain cyberspace security, threat knowledge graph. The schematic diagram of the RBM network structure is shown in **Figure 1**.

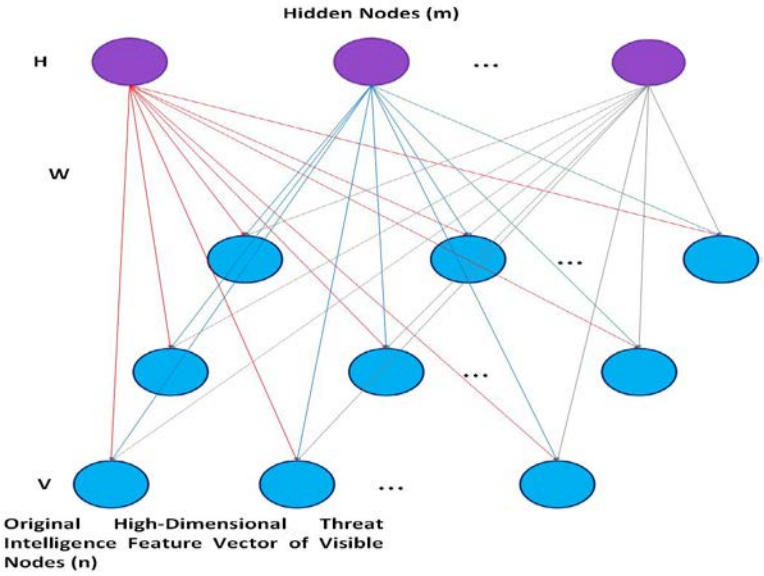

Figure 1. Schematic diagram of RBM network structure.

There are n visible nodes and m hidden nodes ($m < n$), and each visible node is only related to m hidden nodes. The RBM network includes the following parameters: the weight matrix $W_n \times m$ between the visible layer and the hidden layer, the offset set $V = [v_1, v_2, \ldots, v_n]$ of the visible nodes, the offset set $H = [h_1, h_1, \ldots, h_m]$. These parameters determine what kind of m dimensional sample the RBM network encodes an n-dimensional sample into. Multiple layered RBMs are used to reduce the dimension of high-dimensional threat intelligence features, and the threat intelligence feature output of each layer of RBM is used as the input of the next layer of RBM. Specifically, the first layer of RBM is trained, and the uncelebrated threat intelligence data is input. The visual layer of this layer has n_1 nodes, and m_1 nodes are generated after processing. If $m_1 < n_1$, then enter the second layer of RBM training, otherwise end the training; when training the second layer of RBM, the output of the first layer of RBM is used as the input, the visual layer of this layer has n_2 nodes ($n_2 = m_1$), after processing, generate m_2. If $m_2 < n_2$, enter the third layer of RBM training, otherwise end the training. By analogy, the input dimension of threat intelligence features is reduced through continuous training. Finally, the output of the $k-1$ layer of RBM is used as the input of the k layer of RBM to train the k layer. The visual layer of this layer has nk nodes ($n_k = m_{k-1}$), and m_k nodes are generated after processing $m_k < n_k$. From this, the parameters of each layer are obtained.

3. Proposed method

The construction process of cyberspace security threat knowledge graph is shown in **Figure 2**, which mainly includes the following two processes:

1) Security knowledge extraction. Knowledge elements such as security entities, relationships between

security entities, and security entity attributes are extracted from semi-structured and unstructured data of cyberspace security threat intelligence.

2) Security knowledge integration. Including security data integration, security entity alignment, security knowledge reasoning, security ontology construction, security ontology quality assessment and other steps, eliminate the ambiguity between security entities, relationships between security entities, security entity attributes and other elements and actual objects, and finally form a high-quality. The knowledge graph of cyberspace security threats.

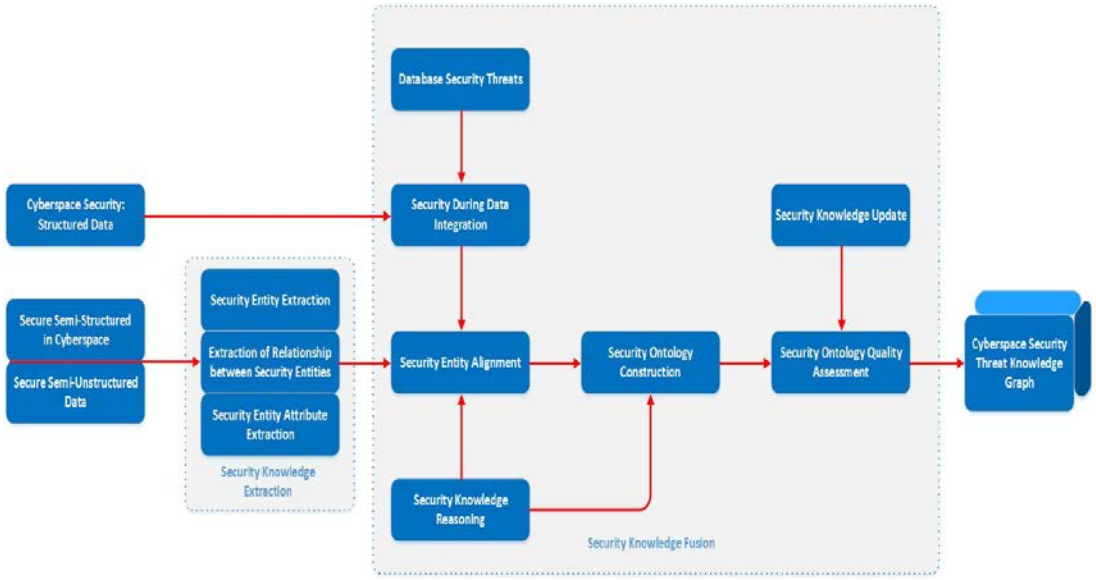

Figure 2. Proposed construction model.

The construction of cyberspace security threat knowledge graph is an unsupervised automatic feature learning process, and its deep learning model is shown in **Figure 3**. The model contains multiple hidden layers, from the first hidden layer to the kth hidden layer is trained individually layer by layer, and the number of nodes decreases layer by layer. There is a connection relationship between nodes in adjacent layers, and there is no connection between nodes within and across layers, and the connection strength is represented by connection weight.

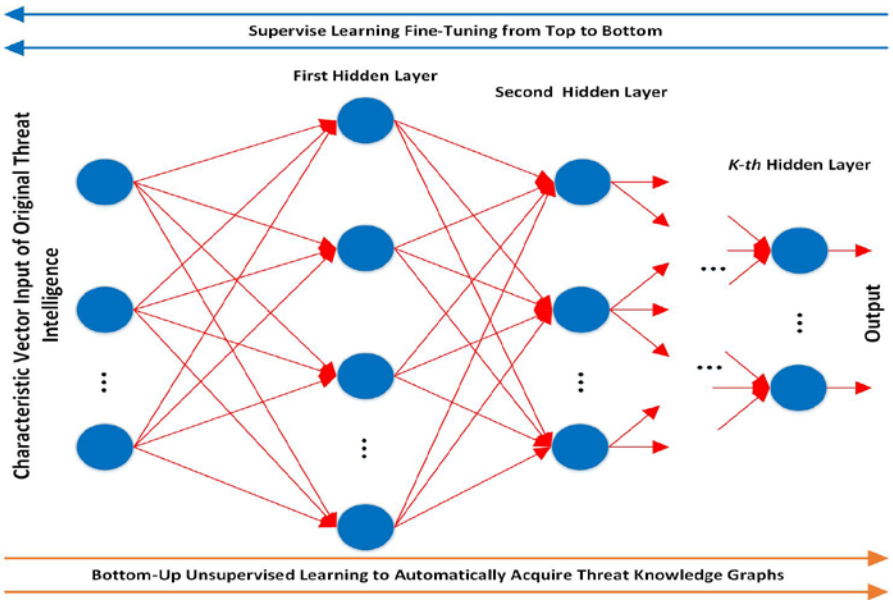

Figure 3. Deep learning model of cyberspace security threat knowledge graph.

4. Experimental result

4.1. Network simulation environment

Build a network simulation environment, and deploy some business systems and security equipment/systems in the network. Based on the big data basic platform, the global security logs, terminal logs, audit logs, etc., in the network are collected, normalized and stored, and the knowledge graph of cyberspace security threats is constructed, and security threats are analysed, predicted and visualized. Deploy network attack tools in a simulated environment, simulate network attack threats, compare the accuracy of the security threat perception method based on knowledge graph and traditional threat detection methods proposed in this paper, and present it to users in a visual way. The network simulation environment is shown in **Figure 4**. The list of equipment and systems involved in the simulation environment is shown in **Table 1**.

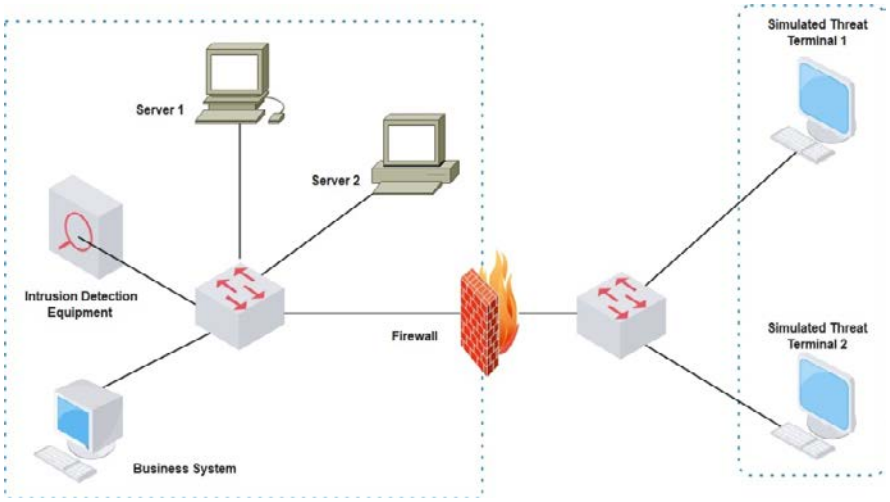

Figure 4. Network simulation environment.

Table 1. List of simulated environment equipment/systems.

Serial number	Device/system name	Function	Configure	Quantity
1	Server 1	Build and store cyberspace security threat knowledge graph.	P920-2 × Intel Xeon4214R 2.4 GHz, main frequency; 12-core processor; 128 GB DDR4. RAM; 64 TB HDD; RTX3090 GPU graphics card.	1
2	Server 2	Deploy a knowledge graph-based security threat perception method.	P920-2 × Intel Xeon4214R 2.4 GHz, main frequency; 12-core processor; 128 GB DDR4. RAM; 64 TB HDD; RTX3090 GPU, graphics card.	1
3	Simulate a threat endpoint	Store various threat samples and simulate security threats.	-	2
4	Intrusion detection equipment	Implementing traditional threat detection methods.	-	1
5	Firewall	Isolate the analog network and the service network to prevent the service network from being polluted.	-	1
6	Business system	Simulate typical B/S architecture business.	-	1

4.2. Construction a knowledge graph of cyberspace security threats

Select the Internet open source threat intelligence dataset malware traffic-analysis as the data source. The dataset contains all threat intelligence from 2015 to 2022, a total of 1837 textual threat intelligence collections and more than 30,000 threat intelligences. The structured and unstructured intelligence in malware-traffic-analysis is cleaned and processed, which takes 13 min/32 s to form a completely independent security threat knowledge graph 1137. It covers all 1837 textual threat intelligence collections and more than 30,000 pieces of threat intelligence, and mines 3564 new security threat features through association rules. This paper focuses on the detection accuracy under high intensity and high speed, that is, the continuous replay the threat sample set. Replay CICIDS 2017 multiple times at different rates. They are normal rate, 2X, 4X, 8X, 16X and 32X respectively. The security threat analysis method based on event stream processing constructed in this paper and the traditional threat detection method are respectively used for detection, and the detection accuracy is compared, as shown in **Figure 5**.

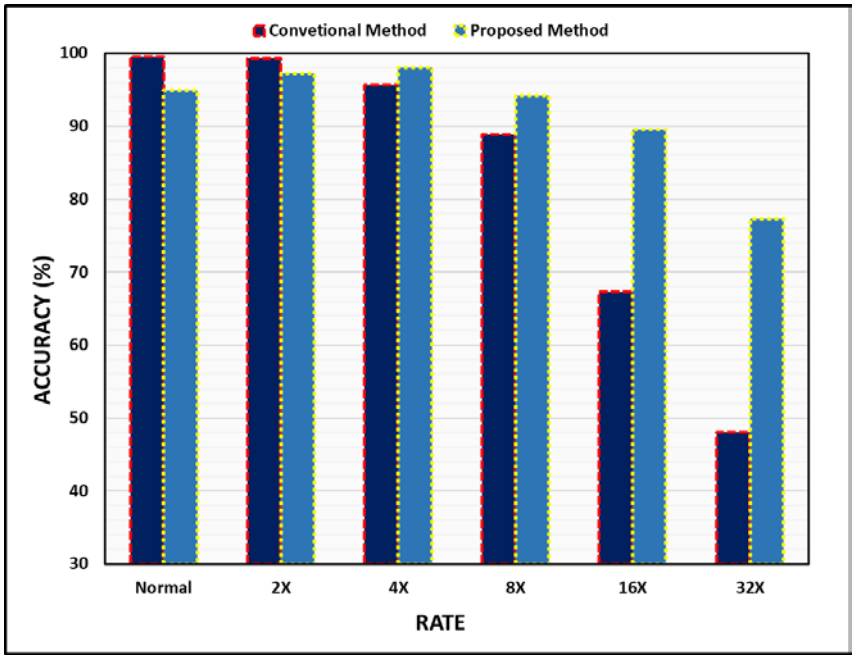

Figure 5. Comparison of detection accuracy.

From **Figure 5**, it can be seen that the detection accuracy of the method in this paper is slightly lower than that of the traditional method in the case of threat simulation under 4 times the rate. In the case of threat simulation with a rate of more than 4 times, the traditional method is based on linear rule matching, and the detection accuracy declines seriously, less than 70% at 16 times the rate, and even less than 50% at 32 times the rate. The method in this paper is close to 85% at a rate of 16 times, and still higher than 70% at a rate of 32 times, indicating that the matching based on the knowledge graph of security threats can meet the perception requirements under high-intensity security threats.

5. Conclusion

The cyberspace security threat perception technology based on knowledge graph can graph the original threat intelligence features from high-dimensional space to low-dimensional space layer by layer, realizing efficient and accurate perception of security threats, and the perception results are highly understandable, which provides a new idea for security threat detection under the new situation of cyberspace confrontation. The security knowledge extraction is mainly for semi-structured and unstructured data of security threat intelligence, using machine learning and other technologies to extract available elements. Future works will

explore the combination of evidence theory and security threat knowledge graph construction algorithm, and further expand hardware resources to improve the accuracy of threat perception in the actual environment and perceive higher-intensity security threats.

Author contributions

Conceptualization, AS and AAJ; methodology, AAJ; software, AAJ; validation, AS, VPGJ and DAJ; formal analysis, AAJ; investigation, AS; resources, AAJ; data curation, AS and AAJ; writing—original draft preparation, AS and AAJ; writing—review and editing, AS; visualization, VPGJ; supervision, DAJ; project administration, AS; funding acquisition, AS. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Chen X, Jia S, Xiang Y. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications* 2020; 141(6): 112948. doi: 10.1016/j.eswa.2019.112948
2. Huang H, Liao Q, Hu M, et al. Human-computer interaction model based on knowledge graph ripple network. *Journal of Electronics & Information Technology* 2022; 44(1): 221–229. doi: 10.11999/JEIT200817
3. Li S, Zhang Y, Liu J, et al. Recommendation model based on public neighbor sorting and sampling of knowledge graph. *Journal of Electronics & Information Technology* 2021; 43(12): 3522–3529. doi: 10.11999/JEIT200735
4. More S, Matthews M, Joshi A, Finin T. A knowledge-based approach to intrusion detection modeling. In: Proceedings of 2012 IEEE Symposium on Security and Privacy Workshops; 24–25 May 2012; San Francisco, CA USA. pp. 75–81.
5. Joshi A, Lal R, Finin T, Joshi A. Extracting cybersecurity related linked data from text. In: Proceedings of 2013 IEEE Seventh International Conference on Semantic Computing; 16–18 September 2013; Irvine, CA, USA. pp. 252–259.
6. Syed Z, Padia A, Finin T, et al. UCO: A unified cybersecurity ontology. In: Proceedings of AAAI Workshop on Artificial Intelligence for Cyber Security; February 2016; Phoenix, Arizona, USA.
7. Atighetchi M, Simidchieva BI, Yaman F, et al. Using ontologies to quantify attack surfaces. In: Proceedings of Semantic Technology for Intelligence, Defense, and Security (STIDS); November 2016; Fairfax, VA, USA. pp. 10–18.
8. Yan J, Yulu Q, Huaijun S, et al. A practical method of constructing network security knowledge map. *Engineering* 2018; 4(1): 117–133.
9. Pingle A, Piplai A, Mittal S, et al. Relext: Relation extraction using deep learning approaches for cybersecurity knowledge graph improvement. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 27–30 August 2019; Vancouver British, Columbia, Canada. pp. 879–886.
10. Chowdhary A, Alshamrani A, Huang D, Liang H. MTD analysis and evaluation framework in software defined network (MASON). In: Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization; 19–21 March 2018; Tempe, AZ, USA. pp. 43–48.
11. Liu D. Prediction of network security based on DS evidence theory. *ETRI Journal* 2020; 42(5): 799–804. doi: 10.4218/etrij.2019-0147
12. Guo W, Tang X, Cheng J, et al. DDoS attack situation information fusion method based on dempster-shafer evidence theory. In: Proceedings of 5th International Conference on Artificial Intelligence and Security; 26–28 July 2019; New York, NY, USA. pp. 396–407.
13. Jiang Y, Li C, Yu L, Bao B. On network security situation prediction based on RBF neural network. In: Proceedings of 2017 36th Chinese Control Conference (CCC); 26–28 July 2017; China. pp. 4060–4063.
14. Dong C, Jiang B, Lu Z, et al. Knowledge graph for cyberspace security intelligence: A survey. *Journal of Cyber Security* 2020; 5(5): 56–76. doi: 10.19363/J.cnki.cn10-1380/tn.2020.09.05
15. Al-Jumaily A, Sali A, Jiménez VPG, et al. Evaluation of 5G and fixed-satellite service earth station (FSS-ES) downlink interference based on artificial neural network learning models (ANN-LMS). *Sensors* 2023; 23(13): 6175. doi: 10.3390/s23136175
16. Wang T, Ai Z, Zhang X. Knowledge graph construction of threat intelligence based on deep learning. *Computer and Modernization* 2018; 12: 21–26.
17. Zhang CX, Ji NN, Wang GW. Restricted Boltzmann machines. *Chinese Journal of Engineering Mathematics* 2015; 32(2): 59–173.

18. Zhang N, Ding S, Zhang J, Xue Y. An overview on restricted Boltzmann machines. *Neurocomputing* 2018; 275: 1186–1199. doi: 10.1016/j.neucom.2017.09.065
19. Nomura Y. Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry. *Journal of Physics: Condensed Matter* 2021; 33(17): 174003. doi: 10.1088/1361-648X/abe268
20. Al-Jumaily A, Sali A, Riyadh M, et al. Machine learning modeling for radiofrequency electromagnetic fields (RF-EMF) signals from mmwave 5G signals. *IEEE Access* 2023; 11: 79648–79658. doi: 10.1109/ACCESS.2023.3265723
21. Al-Jumaily AFM, Al-Jumaily A, Al-Jumaili SJ. Original research article prediction method of business process remaining time based on atten-tion bidirectional recurrent neural network. *Journal of Autonomous Intelligence* 2023; 6(1): 639. doi: 10.32629/jai.v6i1.639