Understanding the Trend of Internet of Things Data Prediction

Lu Zhang[1], Lejie Li[2(✉)], Benjie Dong[2], Yanwei Ma[2], and Yongchao Liu[3]

[1] School of EE, University of Jinan, Jinan 250022, China
[2] Jinan Lingsheng Info Tech. Co. Ltd., Huaiyin District, Jinan 250000, China
lejie.li@dmml.stream
[3] Jinan Yiling Electronic Tech. Co. Ltd., Huaiyin District, Jinan 250000, China

Abstract. With the advancement of science and technology in recent years, the Internet of Things has become another technology hotspot after the Internet. It is widely used in various fields under its intelligent processing and reliability of transmission. However, rapid development also brings certain opportunities and challenges. The most prominent is the massive increase in equipment data, which brings huge challenges to the field of data analysis and prediction. Therefore, how to efficiently process and predict the time series data generated by the Internet of Things has become a research hotspot and difficulty. With the improvement of computer indicators in the past ten years, machine learning has developed to a certain extent. Most scholars will use machine learning methods when researching time-series data processing and forecasting of the Internet of Things. Therefore, we provide a preliminary overview of the history and evolution of machine learning-based IoT time-series data analysis and forecasting from a bibliometric perspective.

Keywords: Internet of Things · Time series data · Data analytics · Data prediction · Machine learning

1 Introduction

With the continuous advancement of technology, the Internet of Things has gradually been applied in various fields and has achieved good results [1]. At the same time, a large amount of device data is also generated. These data are usually time series data, and whether they can be effectively analyzed and predicted has become very important in the fields of transportation [2], healthcare, smart home, smart agriculture, and industrial production. Accurate analysis and prediction can not only improve the overall efficiency of the application field but also have far-reaching significance in promoting social stability and ensuring the safety of people's lives and property. At the same time, the research direction of IoT data analysis and prediction is also closely related to the research content of some other fields.

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 535, pp. 308–318, 2024.
https://doi.org/10.1007/978-3-031-50580-5_27

2 Background

At present, in the related research of Internet of Things data prediction, researchers at home and abroad have made in-depth explorations in various fields. In the field of medicine, some mobile healthcare devices are widely used in diseases such as diabetes, obesity, and heart disease. Device sensors can be used to obtain the physical and mental state of patients on time, and further, make subsequent medical plans. At the same time, our mobile phones or some smart bracelets also have the function of health monitoring [3]. Data prediction is more important in the field of intelligent transportation. Bin Sun et al. [4] propose a data-driven approach to anomaly detection. Yisheng Lv et al. [5] used auto encoders for the first time to extract temporal and spatial characteristics in traffic flow data, and the constructed deep framework model performed well in multiple real data sets. Bichen Wu [6] and others proposed that the SqueezeDet algorithm be applied to the target detection of automatic driving. Bin Sun et al. used machine learning to predict emergency traffic [7]. In the industrial field, Shen Yan et al. used deep learning to propose a machine tool anomaly detection method [8]. Mingzhi Chen et al. improved the GAN and significantly improved the accuracy of the fault diagnosis task [9].

Although there has been a lot of research on the fusion of the Internet of Things and artificial intelligence, this is only the beginning, and the Internet of Things based on artificial intelligence still has a lot of room for development.

3 Related Papers and Sources

We selected 759 articles and citations from the core database of Web Of Science (WOS) on machine learning-based IoT/IoV time-series data prediction. It is mainly selected from 167 periodicals and books. Their publication time is mainly from 2015 to 2022, and a total of 2,758 authors participated in the writing of the articles. The average number of citations per article is more than 13 times. A total of more than 20,000 other documents were cited. We import article information into R library bibliometrix for analysis [10].

From Fig. 1, we can see that the prediction of time-series data of the Internet of Things based on machine learning algorithms has been widely studied by scholars from all over the world in recent years. Especially after 2018, related articles have grown exponentially, which means that it shows that the large amount of data generated by the Internet of Things is increasingly relying on machine learning algorithms for processing and prediction.

Figure 2 shows the most productive sources of our articles according to Bradford's law [11]. We can see that these articles mainly come from IEEE Access, IEEE Internet of Things Journal, etc.

However, a journal with a large number of published articles does not mean that the quality of its published articles is high. We sorted all the sources by influence (G-index), as shown in Fig. 3, we can see that the most authoritative source is IEEE Internet of Things Journal, which published some key points of

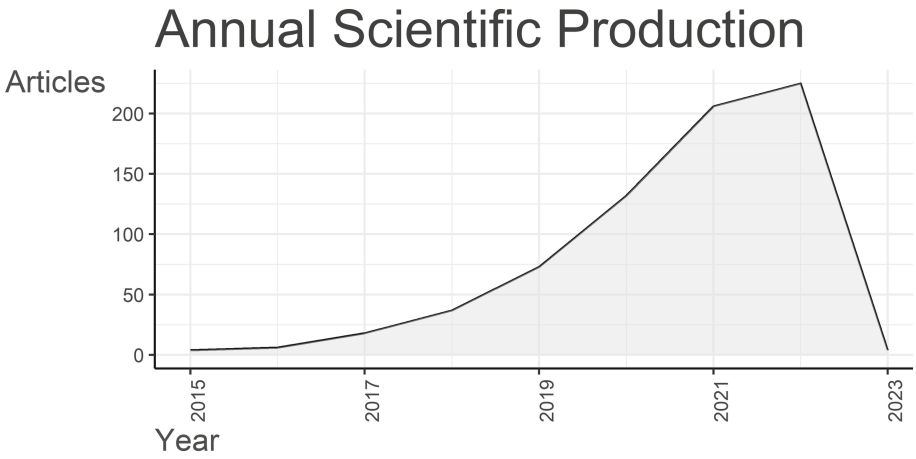

Fig. 1. Annual scientific production

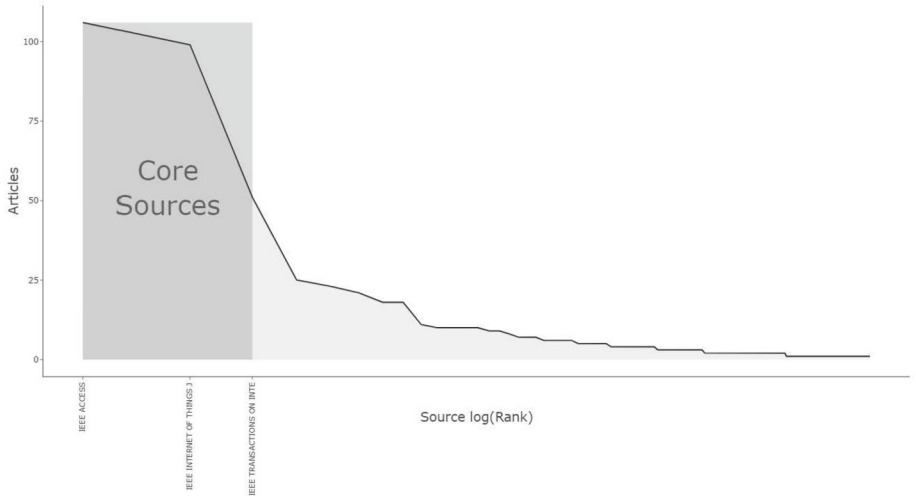

Fig. 2. Analyzing the Literature via Bradford's Law

Tang, FX team paper. A total of 79 papers have been published in the database we collected. For example, An intelligent channel allocation algorithm proposed by Tang, FX [12], which effectively solves the problem of network congestion during data transmission, has been cited 133 times in Web Of Science (WOS). Wang, XF [13] solved the problem of access congestion in the network, which was cited 109 times in WOS.

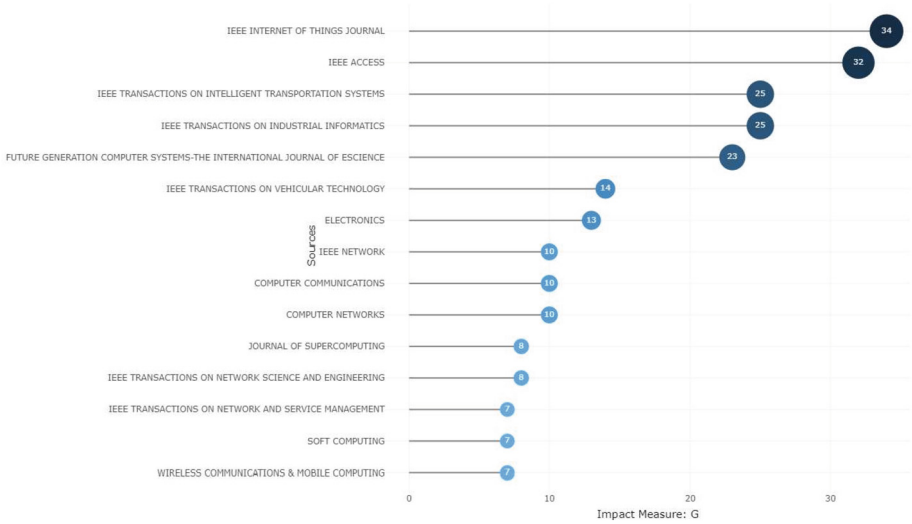

Fig. 3. Source impact (Source G-index)

4 Citations and References

As shown in Fig. 4, this figure shows some of the most cited documents, among which the most cited one is Luong, NC team's survey on the application of reinforcement learning in communication networks [14], cited 585 in WOS Second-rate. The second is from the Mahdavinejad, MS team [15]. They have done some analysis on the application of machine learning in IoT data analysis, which has been cited 375 times in WOS. It is worth mentioning that in the top ten, the team of Tang and FX has two places, and the third paper, which looks forward to the way of artificial intelligence to the future 6G vehicle network [16], has been cited 247 times in WOS. The sixth paper proposes a deep learning-based intelligent channel assignment algorithm [12], which has been cited 133 times in WOS.

The rest of the articles come from different affiliated organizations, and the fourth-ranked article is about the method of information interaction and sharing in the Internet of Vehicles, which has been cited 144 times. The fifth-ranked article is from Munir, and the M team proposed the DeepAnT algorithm to solve common periodic anomalies in data [17], with 142 citations.

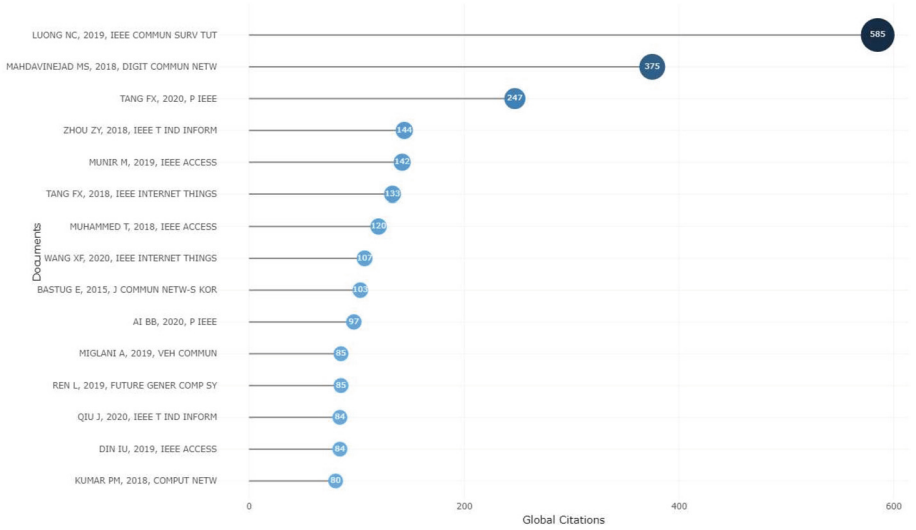

Fig. 4. Most cited documents

According to the literature statistics (Fig. 5), we can see that the more cited papers have a greater impact on the research of other scholars. From this, we can also see that in recent years and the next few years or even decades, the integration of AI and IoT is still a hot topic in the future.

Table 1 analyzes from Lotka's law [18], which reveals the author's scientific productivity and also reflects the quantitative relationship between the author and the paper. We can see that most of the authors have published less than three (including three) papers in the database we collected, and the proportion of these authors is 97.8%. Two authors (Zhang Y and Zhang J) produced 10 and 12 papers respectively. It can be seen that there are not only a large number of authors conducting research in this field but also some scholars have conducted in-depth research.

Figure 6 shows the relationship between the three aspects of papers, authors, and subject keywords. In this three-section diagram, you can see the structural composition of some important papers, authors, and keywords, so that Important research topics have been formed. At the same time, we can see that the Internet, deep learning, predictive models, edge computing, etc. are the most important directions that scholars have studied in recent years.

5 Trending Topics and Theme Evolution

We can learn more from the trend topics. The trend change is mainly caused by the change in topic frequency. The trend of topics and the evolution of topics is shown in Fig. 7 and Fig. 8. In this analysis, we can see that modules, Internet, and performance have received continuous attention in recent years.

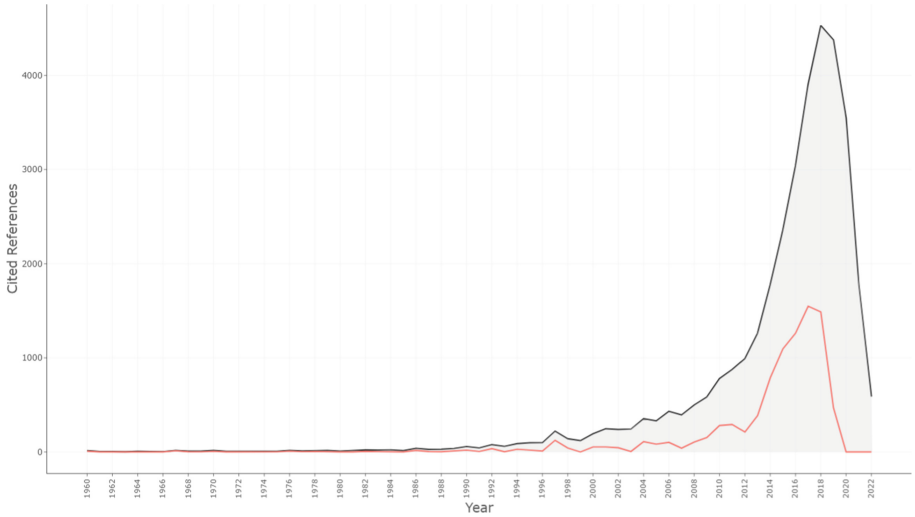

Fig. 5. References spectroscopy (RPYS)

Table 1. Analyzing Author Productivity According to Lotka's Law

Number of articles	Number of authors	Proportion of Authors
1	2357	0.855
2	259	0.094
3	81	0.029
4	26	0.009
5	20	0.007
6	8	0.003
7	2	0.001
8	1	0.000
9	2	0.001
10	1	0.000
12	1	0.000

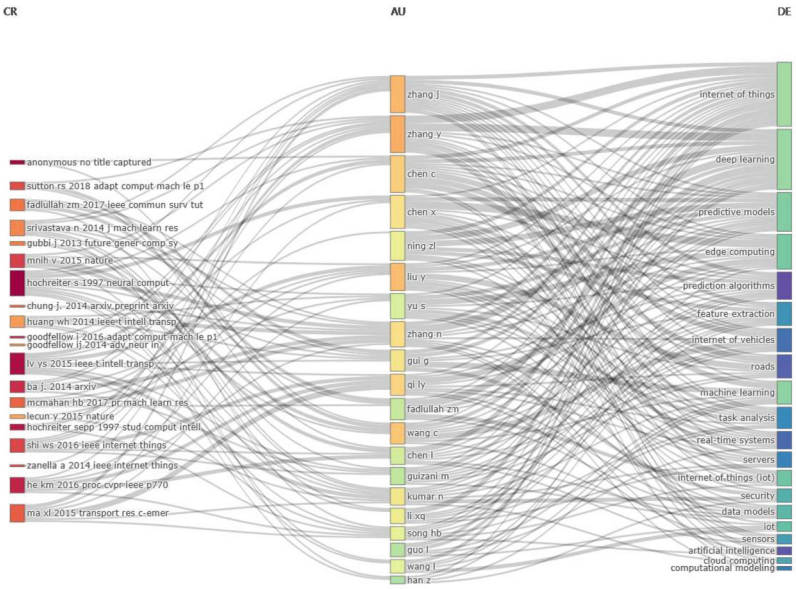

Fig. 6. Three-Field Plot

After 2021, trend topics and topic evolution have undergone major changes, which indicates that the focus of research has shifted to a certain extent in 2021, and some related issues such as wireless sensor networks and resource allocation have been resolved. The Internet and prediction began to gradually shift to neural network. Although neural-network and time-series related issues have been topics of concern in recent years, their frequency of occurrence is among the best.

From another perspective, we divide the trending topics into four different types [19]. In the figure, they are in the first, second, third, and fourth quadrants respectively, as shown in Fig. 9 and Fig. 10. The horizontal axis in the figure represents centrality, and the vertical axis represents density. In the upper right corner of the figure, which is the first quadrant, is Motor Themes. This quadrant contains topics that are not only important but well-developed. The upper left corner of the graph (the second quadrant) is Niche Themes. Although the themes contained in this quadrant have developed well, they have not been applied to the current field. The lower left corner of the graph is Emerging or declining Themes and the themes contained in this quadrant are mainly marginal themes. It may be some themes that are just emerging, or it may be some themes that are about to decline. The last is Basic Themes. The themes in this part are relatively basic but important.

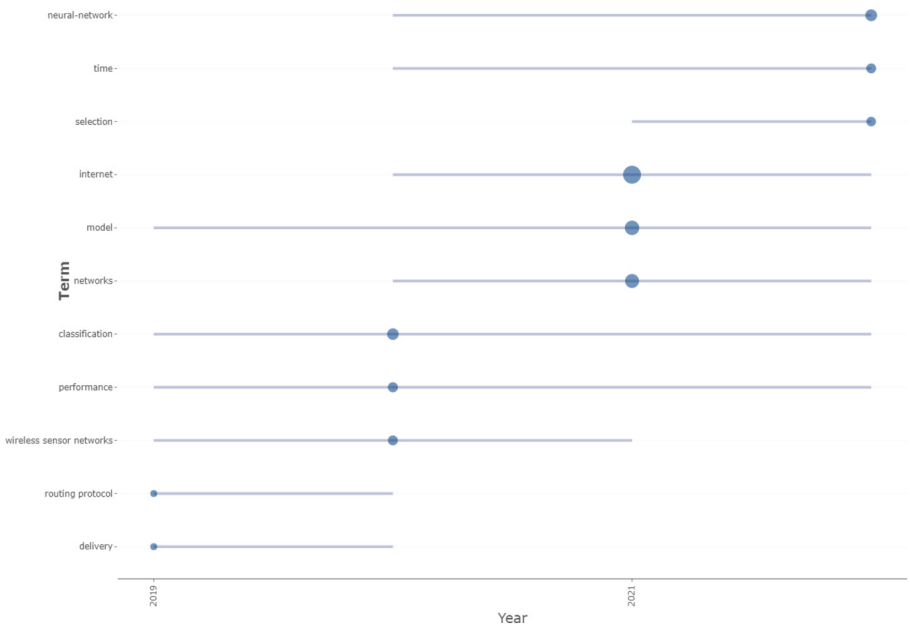

Fig. 7. Topic trend (Frequency)

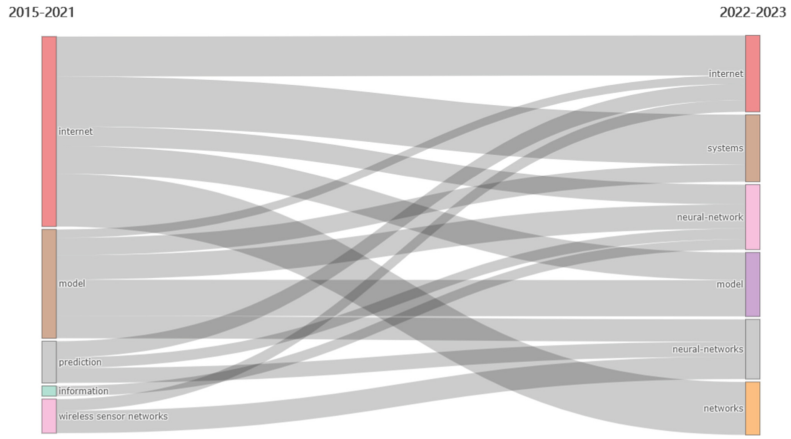

Fig. 8. Thematic evolution

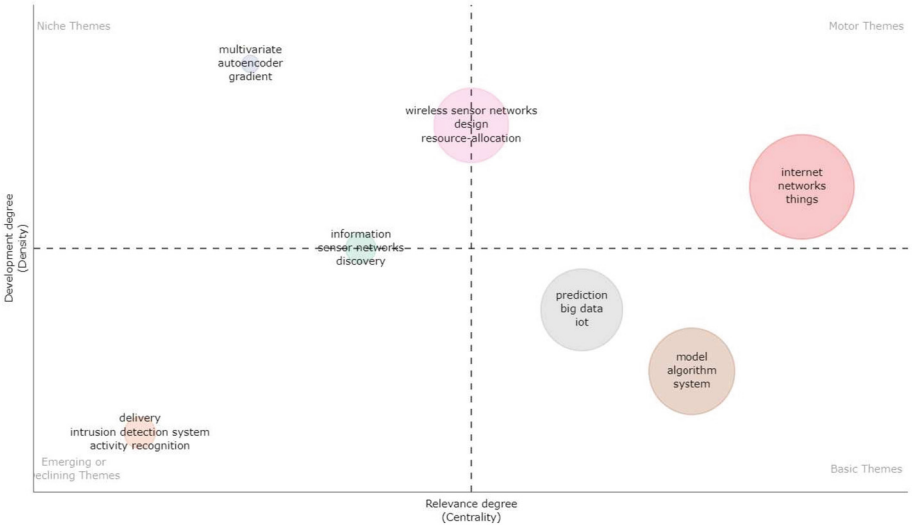

Fig. 9. Thematic map until 2021

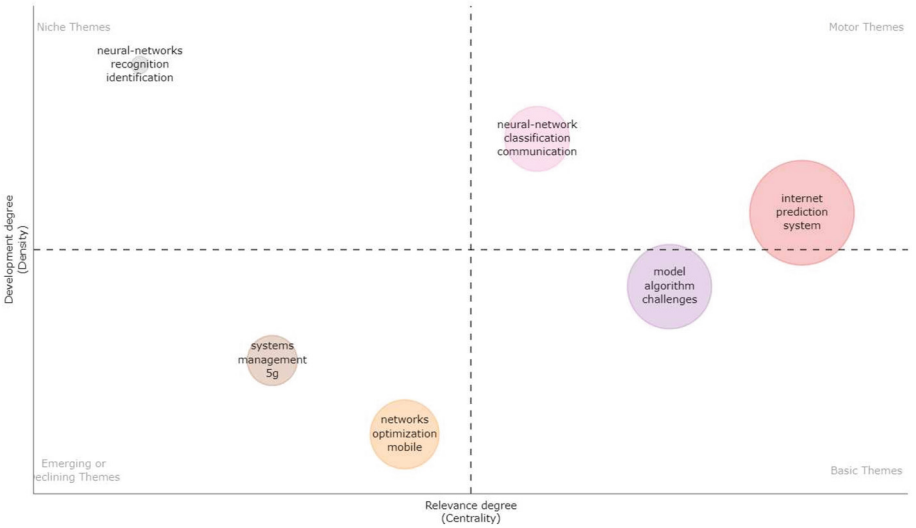

Fig. 10. Thematic map after 2021

From Fig. 9, we can see that from 2015 to 2021, the topics in the first quadrant are mainly composed of the Internet, and the topics in the fourth quadrant mainly include forecasting, big data, algorithms, models, systems, etc. Before 2021, multivariate and autoencoding are the main research work in the second

quadrant. The third quadrant mainly includes delivery and intrusion detection systems, which were undeveloped and marginal research directions at that time.

However, after 2021, the thematic map has undergone major changes, as shown in Fig. 10. The direction of research and the distribution of topics has changed a lot. The most obvious is that in the first quadrant, new neural networks and classifications have emerged, while the original Internet is gradually slipping into the fourth quadrant. The fourth quadrant is mainly composed of algorithms and models, but it is also gradually approaching the first quadrant, which shows that the development of algorithms and models is becoming more and more important, and still faces greater challenges. There is mainly 5G and mobile in the third quadrant, indicating that after 2021, this topic has slowly faded out of the research field.

6 Conclusions and Future Directions

According to the evolution of topics and changes in topic trends, we can see that the integration of the field of artificial intelligence and the Internet of Things is the general trend [20]. Machine learning has become indispensable in predictive research on IoT time-series data. As a topic that has attracted attention in recent years, neural networks, we believe that more and more researchers will devote themselves to the research of neural networks in the future. In particular, the processing of IoT data based on neural networks can be said to be the focus and difficulty of research in the next few years or even decades.

References

1. Sun, B., Geng, R., Zhang, L., Li, S., Shen, T., Ma, L.: Securing 6G-enabled IoT/IoV networks by machine learning and data fusion. EURASIP J. Wirel. Commun. Network. **2022** (2022)
2. Sun, B., Geng, R., Shen, T., Xu, Y., Bi, S.: Dynamic emergency transit forecasting with IoT sequential data. Mob. Netw. Appl. 1–15 (2022)
3. Chen, M., Li, W., Hao, Y., Qian, Y., Humar, I.: Edge cognitive computing based smart healthcare system. Futur. Gener. Comput. Syst. **86**, 403–411 (2018)
4. Sun, B., Ma, L., Shen, T., Geng, R., Zhou, Y., Tian, Y.: A robust data-driven method for multiseasonality and heteroscedasticity in time series preprocessing. Wirel. Commun. Mob. Comput. **2021**, 6692390:1–6692390:11 (2021)
5. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.-Y.: Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intell. Transp. Syst. **16**(2), 865–873 (2015)
6. Wu, B., Wan, A., Iandola, F., Jin, P.H., Keutzer, K.: Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 446–454 (2017)
7. Sun, B., Geng, R., Yuan, X., Shen, T.: Prediction of emergency mobility under diverse IoT availability. EAI Endors. Trans. Pervasive Health Technol. **8**(4), e2 (2022)

8. Yan, S., Shao, H., Xiao, Y., Liu, B., Wan, J.: Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. Robot. Comput. Integr. Manuf. **79**, 102441 (2023)
9. Chen, M., Shao, H., Dou, H., Li, W., Liu, B.: Data augmentation and intelligent fault diagnosis of planetary gearbox using Ilofgan under extremely limited samples. IEEE Trans. Reliab. 1–9 (2022)
10. Aria, M., Cuccurullo, C.: bibliometrix: An r-tool for comprehensive science mapping analysis. J. Informet. **11**(4), 959–975 (2017)
11. Nisonger, T.E.: The "80/20 rule" and core journals. Serials Librarian **55**(1-2), 62–84 (2008)
12. Tang, F., Fadlullah, Z.Md., Mao, B., Kato, N.: An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: a deep learning approach. IEEE Internet Things J. **5**(6), 5141–5154 (2018)
13. Wang, X., Wang, C., Li, X., Leung, V.C.M., Taleb, T.: Federated deep reinforcement learning for internet of things with decentralized cooperative edge caching. IEEE Internet Things J. **7**(10), 9441–9455 (2020)
14. Luong, N.C., et al.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutor. **21**(4), 3133–3174 (2019)
15. Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., Sheth, A.P.: Machine learning for internet of things data analysis: a survey. Digit. Commun. Netw. **4**(3), 161–175 (2018)
16. Tang, F., Kawamot, Y., Kato, N., Liu, J.: Future intelligent and secure vehicular network toward 6g: machine-learning approaches. Proc. IEEE **108**(2), 292–307 (2020)
17. Munir, M., Siddiqui, S.A., Dengel, A., Ahmed, S.: DeepAnt: a deep learning approach for unsupervised anomaly detection in time series. IEEE Access **7**, 1991–2005 (2019)
18. Lotka, A.J.: The frequency distribution of scientific productivity. J. Washington Acad. Sci. (Baltimore) **19**, 317–323 (1926)
19. Sun, B., Ma, L., Geng, R., Xu, Y.: Matrix profile evolution: an initial overview. In: Fu, W., Xu, Y., Wang, S.-H., Zhang, Y. (eds.) ICMTEL 2021. LNICST, vol. 387, pp. 492–501. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-82562-1_48
20. Sun, B., Cheng, W., Bai, G., Goswami, P.: Correcting and complementing freeway traffic accident data using Mahalanobis distance based outlier detection. Tehnicki Vjesnik-Tech. Gazette **24**, 10 (2017)