

Research on the Performance of Text Mining and Processing in Power Grid Networks

Yuzhong Zhou^{1,*}, Zhengping Lin¹, Liang Tu¹, Jiahao Shi¹, and Yuliang Yang¹

¹Research Institute of China Southern Power Grid, Guangzhou, China.

Abstract

This paper employs deep learning technique to perform the research of text mining for power grid networks, focusing on fundamental elements such as loss and activation functions. Through some analysis and formulas, we explain how these functions contribute to deep learning. We also introduce major deep learning training models, including CNN and RNN, and provide visual aids to aid understanding. To demonstrate the impact of various factors on deep learning training, we employ control variable experiments to analyze the influence of factors such as learning rate, batch size, and data noise on model training trends. While the influence of hyperparameters and data noise are covered in this paper, other factors such as CPU and memory frequency, as well as GPU performance, also play a crucial role in deep learning training. Therefore, continuous adjustments to various factors are necessary to achieve optimal training results for deep learning models in power grid networks.

Received on 04 March 2023; accepted on 13 May 2023; published on 01 June 2023

Keywords: Text mining, performance analysis, deep learning

Copyright © 2023 Yuzhong Zhou *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetsis.v10i4.3094

1. Introduction

As a typical form of Internet of Things (IoT) networks [1–3], the power grid network is a critical infrastructure that provides electricity to consumers across various regions. Research on power grid networks focuses on improving the efficiency, reliability, and resilience of the network while also ensuring the safety of the users and the environment. Recent research in this field has explored the integration of renewable energy sources, such as wind and solar, into the power grid network. This requires the development of new technologies to manage and balance the fluctuating supply of renewable energy, which can affect the stability of the grid. Another area of research in power grid networks is the use of smart grid technologies. Smart grids incorporate advanced communication and control systems that enable real-time monitoring and optimization of the power grid network [4–6]. This allows for more efficient energy distribution, reduced energy waste, and improved grid reliability [7, 8]. Research is also being conducted to address the

cybersecurity threats to power grid networks. As power grid networks become more interconnected and reliant on digital technologies, the risk of cyber attacks increases. Developing effective cybersecurity measures is crucial to ensuring the safety and reliability of the power grid network.

Text mining plays an important role in the power grid networks, which is useful for the system operation and running [9–11]. Text mining is a research field that focuses on using computational techniques to extract useful information and knowledge from large collections of textual data. This field encompasses a range of techniques and applications, including natural language processing (NLP), machine learning, and data mining. One area of active research in text mining is sentiment analysis. Sentiment analysis aims to automatically identify the sentiment of a text, such as positive, negative, or neutral. This is important for applications such as opinion mining, customer feedback analysis, and social media monitoring. Researchers are developing new algorithms and techniques to improve the accuracy and effectiveness of sentiment analysis. Another area of research in text mining is topic modeling, to identify latent topics in a collection of documents. This

*Corresponding author. Email: yuzhong_zhou@hotmail.com

is useful for tasks such as information retrieval, document clustering, and trend analysis. Researchers are exploring new methods to improve the efficiency and scalability of topic modeling, such as distributed and parallel algorithms. Named entity recognition (NER) is another area of active research in text mining [12–14]. NER is the process of identifying and classifying named entities, such as people, organizations, and locations, in text. This is important for applications such as information extraction and text summarization. Researchers are developing new methods to improve the accuracy and robustness of NER, such as deep learning-based approaches. Finally, researchers in text mining are exploring new applications of the field, such as fake news detection, biomedical text mining, and social media analysis. These applications are becoming increasingly important in today’s data-driven world, and researchers are developing new techniques to address the unique challenges posed by these domains. In a word, the current research in the text mining is focused on developing new algorithms and techniques to extract useful information and knowledge from large collections of textual data.

Deep learning-based models have already surpassed classical machine learning-based approaches in various text mining fields [15–17]. One of the key steps in traditional machine learning is to extract features, but feature extraction has a disadvantage of relying mainly on human labor and each extraction method is not universal. Deep learning does not rely on human for feature extraction, but is automatically extracted by machines, which gives deep learning excellent expressiveness. Not only that, deep learning is highly dependent on data, the larger the amount of data the better it performs, and sometimes the upper limit can be raised by tuning the parameters, even in some text mining fields it has already exceeded the human performance. It has a wide range of applications, such as the ability to answer questions, handle spam, chat on electronic devices, and more. The most popular model is the transformer model, which is widely used in the field of natural language processing (NLP), such as the famous BERT and GPT-3 models. In recent years, a new algorithm for analyzing written language, named Neural Analysis of Sentiment (NaSent), has been proposed to better understand the emotions that flow between words.

Motivated by the above literature review, this paper employs the deep learning technique to perform the research of text mining for power grid networks, focusing on fundamental elements such as loss and activation functions. Through images and formulas, we explain how these functions contribute to deep learning. We also introduce major deep learning training models, including CNN and RNN, and provide visual aids to aid understanding. To demonstrate the

impact of various factors on deep learning training, we employ control variable experiments to analyze the influence of factors such as learning rate, batch size, and data noise on model training trends. While the influence of hyperparameters and data noise are covered in this paper, other factors such as CPU and memory frequency, as well as GPU performance, also play a crucial role in deep learning training. Therefore, continuous adjustments to various factors are necessary to achieve optimal training results for deep learning models in power grid networks.

2. Basic Principle of Deep Learning

Text mining is theoretically based on statistics and computer linguistics. Deep learning-based text mining is a process that relies on information retrieval techniques to extract meaningful, implicit and useful information from large amounts of text data. The basic principle is to transform text data into sequences that describe the content of the text, and then use techniques such as classification and clustering to classify the text into organised groups and discover new concepts and corresponding relationships based on the structure.

The main steps of deep learning-based text mining are as follows: in the first step, the creation of the dataset is necessary. In the second step, dataset needs to be scaled down to improve the operational efficiency. In the third step, the classification or clustering operation should be performed to form the corresponding knowledge patterns. In the fourth step, the evaluation of the quality of the knowledge patterns should be performed, to analyze modify one of the previous links and then further mining. Here we take word2vec as an example to explain the deep learning based text mining and processing.

2.1. Word2Vec

In Word2Vec, the CBOW (Continuous Bag of Words) and the Skip-Gram model can quickly train word-embeddings and calculate word vectors to operations, where the principle of CBOW and Skip-Gram are shown in Fig. 1 and Fig. 2, respectively. The concept of word-embedding is that the text can be scattered and embedded into another discrete space. Most supervised machine learning models can be summarized as the functional form

$$f(x) = y, \quad (1)$$

and so can word2vec. The ultimate goal of word2vec is to be able to obtain the word vector matrix after the model has been trained. The CBOW model is equivalent to predicting the current word by context, i.e., using $w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}$ to predict w_t . In contrast, the SkipGram model is equivalent to predicting context by

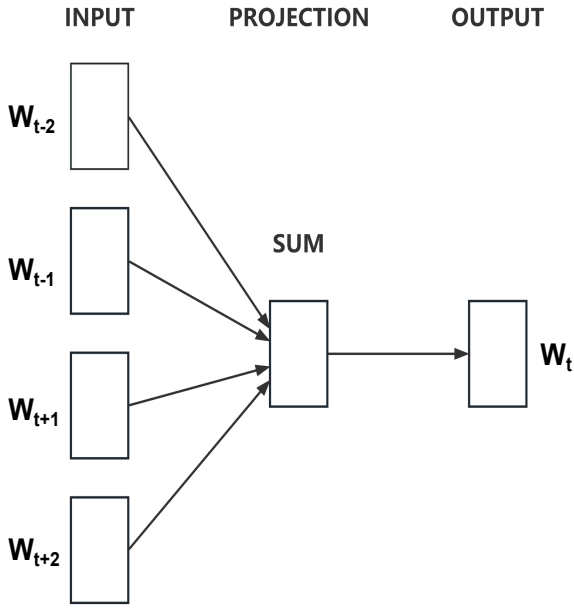


Figure 1. Principle of CBOW

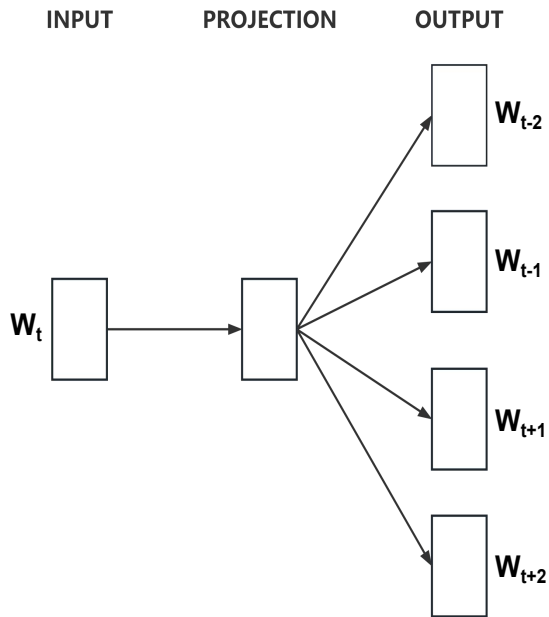


Figure 2. Principle of SkipGram

the current word, i.e., using w_t to predict w_{t-2} , w_{t-1} , w_{t+1} , w_{t+2} .

As shown in Fig. 3, deep learning-based text-mining uses one-hot encoding as the input value to multiply with the weight matrix from the input layer to the

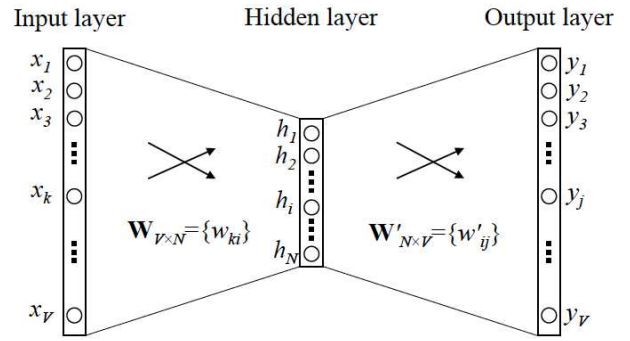


Figure 3. Principle of deep learning based text mining.

hidden layer, which also enables to obtain the N -dimensional word vector of the input words.

$$h = W_T * X = v_{wI}^T \quad (2)$$

The N -dimensional word vector is then multiplied with the weight matrix from the hidden layer to the output layer to obtain a score for the input word, with higher scores representing more accurate predictions.

$$u = W'^T * h. \quad (3)$$

3. ACTIVATION FUCTIONS

Activation functions are mathematical operations applied to the output of a neural network layer to introduce nonlinearity into the network. In deep learning, activation functions are used in almost every neural network layer to help the network learn complex representations of data. There are several activation functions used in deep learning, including:

- Sigmoid function: This function maps any input to a value between 0 and 1, which is useful for binary classification problems, given by [4, 18, 19]

$$sigmoid(x) = 1/(1 + \exp(-x)). \quad (4)$$

- ReLU function: The Rectified Linear Unit (ReLU) function maps any negative input to zero and any positive input to the input value. ReLU is popular because it's simple and computationally efficient, given by

$$ReLU(x) = \max(0, x). \quad (5)$$

- Tanh function: The Hyperbolic Tangent (tanh) function maps any input to a value between -1 and 1, which is useful for multi-class classification problems, given by

$$\tanh(x) = (\exp(x) - \exp(-x))/(\exp(x) + \exp(-x)). \quad (6)$$

- **Softmax function:** The Softmax function maps a vector of input values to a probability distribution over a set of classes, making it ideal for multi-class classification problems, given by

$$\text{softmax}(x_i) = \exp(x_i) / \sum_{n=1}^N (\exp(x_n)). \quad (7)$$

- **Leaky ReLU function:** Leaky ReLU is a variation of ReLU where the negative input values are mapped to a small, non-zero value instead of zero, to prevent dead neurons in the network.
- **ELU function:** Exponential Linear Unit (ELU) function is similar to the Leaky ReLU, but with negative input values, it maps them to a small, non-zero value, which is scaled by an exponential factor.
- **Swish function:** Swish function is a new activation function that is similar to the ReLU function, but has a smoothly varying output, which can result in a better performance in certain types of models.

Choosing the right activation function depends on the type of problem to be solved and the properties of the data. In this work, we choose the ReLU function as the activation function, without loss of generality.

4. LOSS FUNCTIONS

The loss function is also a crucial part for deep learning model training. Generally speaking, the objective function of the deep learning model is to reduce the error between the real value and the predicted value. The smaller the error, the closer the predicted result of the model is to the real value, and the better the model performance will be. The loss function plays an important role because it must faithfully reduce aspects of the model to a single number, and improvements in that number are used to measure improvements in model performance. The choice of loss function often determines whether the model can achieve the performance results we want and whether it can perform the tasks we need. Here are two commonly used loss functions.

One loss function is based on the mean squared error (MSE),

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (8)$$

where y_i is the true label and \hat{y}_i is the predicted label for the i^{th} sample, and n is the number of samples.

Another typical form is based on the binary cross-entropy (BCE),

$$\text{BCE} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (9)$$

where y_i is the true label (either 0 or 1) and \hat{y}_i is the predicted probability that the i^{th} sample belongs to the positive class, and n is the number of samples. Note that BCE loss is commonly used for binary classification problems. In this work, we adopt BCE as the loss function, without loss of generality.

5. Evaluation metrics and Influencing factors of Training results in Deep Learning

In this part, we investigate the training result performance by influencing factors and evaluation metrics. The evaluation metrics are shown as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$F1 - \text{score} = \frac{2\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

$$\text{Acc} = \frac{TP + TN}{N} \quad (13)$$

where TP is denoted as the number of samples which are correctly classified into positive class, TN is the number of the samples that are correctly classified into negative classes, FN is the number of samples that are misclassified into negative samples, FP is the number of the samples that are misclassified into positive samples, N is the number of samples, $F1 - \text{score}$ is the reconciled average of accuracy and recall, Acc describes how many of all predicted positive cases match the actual situation match.

In machine learning, training loss indicates the error between the prediction result of the training set in the model and the true result, and is used to measure the fitting ability of the neural network model trained by the training set. Valid loss denotes the error between the prediction result of the validation set in the model and the true result, and is used to measure the difference between the effect of the neural network model on the test set and the true situation, which can be regarded as the generalization ability. Both the training and valid loss are decreasing indicating that the neural network is still learning, which is the best case; the training loss is decreasing while the valid loss tends to be constant indicating that the neural network is overfitting. When both the training and valid loss tend to be constant, it means that the learning of the neural network has encountered some obstacles and the learning rate needs to be reduced. When both the training and valid loss

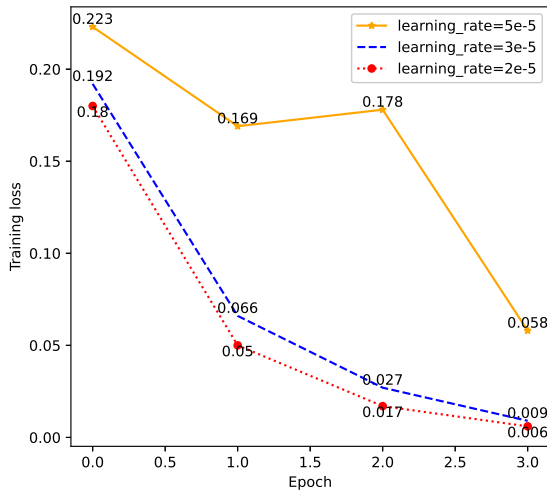


Figure 4. The trend of training loss affected by the learning rate.

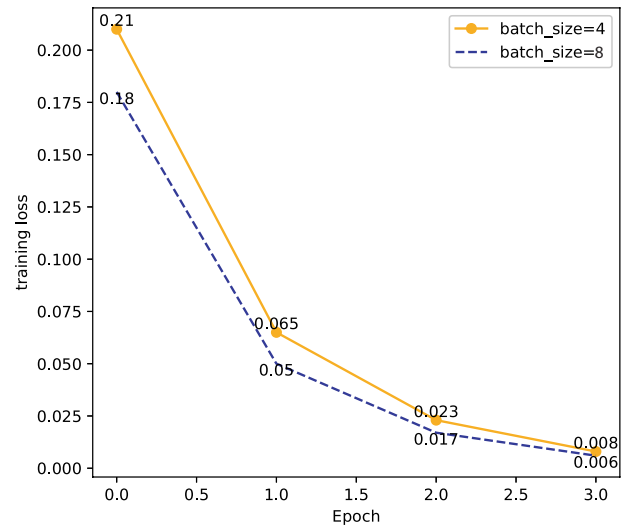


Figure 6. The trend of training loss versus the number of epochs.

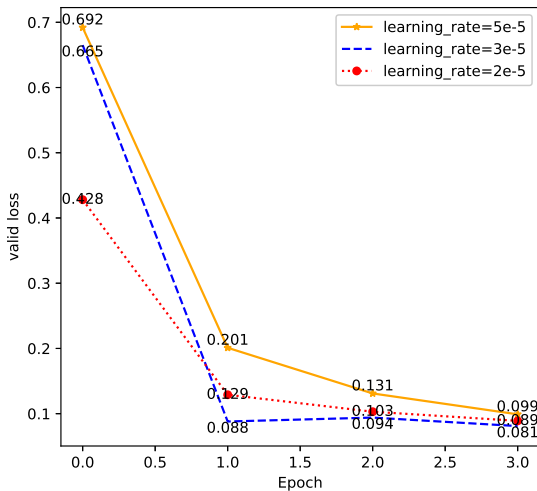


Figure 5. The trend of valid loss affected by the learning rate.

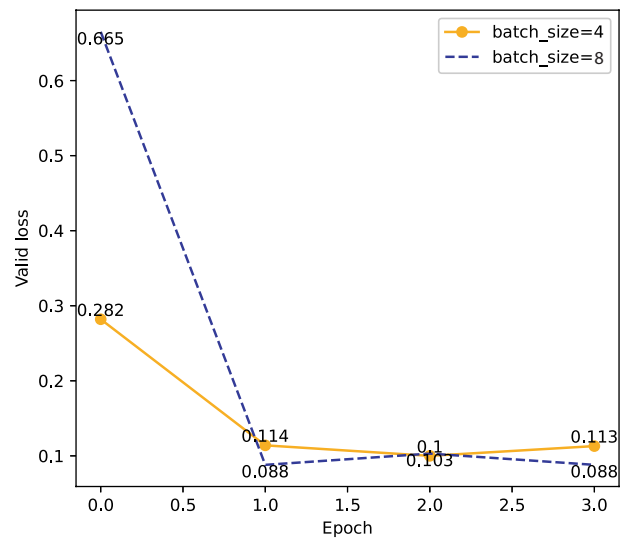


Figure 7. The trend of valid loss versus the number of epochs.

are increasing, it means that there is a problem in the structural design of the neural network and the wrong hyperparameters are used for training.

In our simulations, 38,507 samples were used for the deep learning related competitions, where there are 19222 positive data and 19285 negative data. The experiment were PyTorch with GTX 1060 for acceleration.

Fig. 4 and Fig. 5 show the trend of the training loss and valid loss of the model during training with the different values of learning rate. As seen in these two figures, the training loss and valid loss decrease as the number of training iterations increases at the same learning rate. Note that the training loss and valid loss decrease as the learning rate decrease at the same iteration. The learning rate is one of the very important hyperparameters in deep learning, which determines

the step size of parameter update in each iteration. If the learning rate is set too large, the step size of each parameter update will also become too large, which may lead to unstable oscillation. If the learning rate is set too small, the step size of each parameter update will also become too small, resulting in a slower training speed and an increase in the number of iterations required for the training process, which may even cause the model to fail to reach the optimal point or fall into a local optimum prematurely. A suitable learning rate can ensure that the model converges to the optimal point steadily during the training process, and the training speed can also be guaranteed. In general, the appropriate learning rate needs to be determined according to the specific problem and model structure.

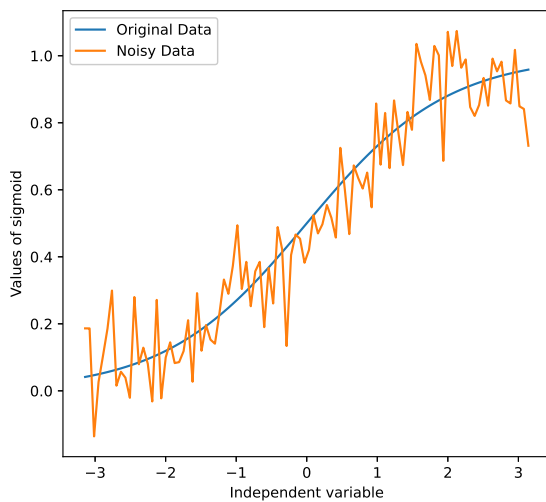


Figure 8. The original data and the data after adding noise.

Fig. 6 shows the trend of training loss versus the number of normalized epochs, where the batch size is set to 4 or 8. Fig. 7 illustrates the trend of valid loss versus the number of normalized epochs, with batch size of 4 or 8. From Fig. 6, we can find that the system training loss decreases with the number of epochs, indicating the efficiency in the training process. In particular, the training loss with 1 epoch is 0.065 or 0.05 when the batch size is 4 or 8; the training loss with 2 epochs is 0.023 or 0.017 when the batch size is 4 or 8; the training loss with 3 epochs is 0.008 or 0.006 when the batch size is 4 or 8. The similar observation can be also found in Fig. 7, where the valid loss also decreases with the number of epochs, indicating the efficiency in the valid process. In particular, the valid loss with 1 epoch is 0.114 or 0.088 when the batch size is 4 or 8; the valid loss with 2 epochs is 0.103 or 0.1 when the batch size is 4 or 8; the valid loss with 3 epochs is 0.113 or 0.088 when the batch size is 4 or 8.

Fig. 8 shows how noise affects the raw data and its impact on the deep learning model training. Note that the noise in the data may make the model training more difficult because the model needs to learn how to distinguish between the signal and noise. In some cases, noise may overfit the model because the model will overfit the noise instead of the real data. As a result, noise can degrade the performance of the model. Although noise may degrade the performance of the model, in some cases it may help improve the generalization ability of the model. When the model is able to learn the commonality of real data from the noise, it will be more robust and able to handle new data better. Understanding the laws affecting data noise and adopting appropriate processing methods can help improve the performance and generalization ability of deep learning models.

6. Conclusion

This paper explored the application of deep learning in text mining for power grid networks, with a focus on important concepts like loss and activation functions. Visual aids and mathematical expressions were used to demonstrate how these functions contributed to deep learning. To investigate the impact of various factors on deep learning training, control variable experiments were conducted, and the influence of factors such as learning rate, batch size, and data noise on model training trends was analyzed. While hyperparameters and data noise were discussed in detail in this paper, it should be noted that other factors such as CPU and memory frequency, as well as GPU performance, also played a critical role in deep learning training. Therefore, constant adjustments to multiple factors are necessary to achieve optimal training results for deep learning models.

6.1. Copyright

The Copyright licensed to EAI.

References

- [1] W. Hong, J. Yin, M. You, H. Wang, J. Cao, J. Li, and M. Liu, "Graph intelligence enhanced bi-channel insider threat detection," in *Network and System Security: 16th International Conference, NSS 2022, Denarau Island, Fiji, December 9–12, 2022, Proceedings*. Springer, 2022, pp. 86–102.
- [2] S. Guo and X. Zhao, "Multi-agent deep reinforcement learning based transmission latency minimization for delay-sensitive cognitive satellite-uav networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 131–144, 2023.
- [3] J. Lu, S. Zhan, and X. Liu, "Intelligent wireless monitoring technology for 10kv overhead lines in smart grid networks," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 2, 2022.
- [4] H. Hui and W. Chen, "Joint scheduling of proactive pushing and on-demand transmission over shared spectrum for profit maximization," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 1, pp. 107–121, 2023.
- [5] J. Yin, M. Tang, J. Cao, M. You, H. Wang, and M. Alazab, "Knowledge-driven cybersecurity intelligence: software vulnerability co-exploitation behaviour discovery," *IEEE Transactions on Industrial Informatics*, 2022.
- [6] Y. Zhou, Z. Lin, Y. La, J. Huang, and X. Wang, "Analysis and design of power system transformer standard based on knowledge graph," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 2, 2022.
- [7] J. Yin, M. Tang, J. Cao, H. Wang, M. You, and Y. Lin, "Vulnerability exploitation time prediction: an integrated framework for dynamic imbalanced learning," *World Wide Web*, pp. 1–23, 2022.
- [8] C. Chaieb, F. Abdelkefi, and W. Ajib, "Deep reinforcement learning for resource allocation in multi-band and hybrid OMA-NOMA wireless networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 187–198, 2023.

- [9] Y. Zhou, Z. Lin, L. Tu, J. Huang, and Z. Zhang, "Analysis and design of standard knowledge service system based on deep learning," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 10, no. 2, 2022.
- [10] F. Xiao, S. Zhang, S. Tang, S. Shen, H. Dong, and Y. Zhong, "Wision: Bolstering MAV 3d indoor state estimation by embracing multipath of wifi," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 253–266, 2023.
- [11] J. Han, J. Zhang, C. He, C. Lv, X. Hou, and Y. Ji, "Distributed finite-time safety consensus control of vehicle platoon with sensor and actuator failures," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 162–175, 2023.
- [12] S. Mosharafian and J. M. Velni, "Cooperative adaptive cruise control in a mixed-autonomy traffic system: A hybrid stochastic predictive approach incorporating lane change," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 136–148, 2023.
- [13] A. Gupta, M. Sellathurai, and T. Ratnarajah, "End-to-end learning-based full-duplex amplify-and-forward relay networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 199–213, 2023.
- [14] Z. Lin, Z. Ni, L. Kuang, C. Jiang, and Z. Huang, "Multi-satellite beam hopping based on load balancing and interference avoidance for NGSO satellite communication systems," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 282–295, 2023.
- [15] D. Orlando, S. Bartoletti, I. Palamà, G. Bianchi, and N. Blefari-Melazzi, "Innovative attack detection solutions for wireless networks with application to location security," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 1, pp. 205–219, 2023.
- [16] M. E. Gonzalez, J. F. Silva, M. Videla, and M. E. Orchard, "Data-driven representations for testing independence: Modeling, analysis and connection with mutual information estimation," *IEEE Trans. Signal Process.*, vol. 70, pp. 158–173, 2022.
- [17] A. T. M. Fisch, I. A. Eckley, and P. Fearnhead, "Innovative and additive outlier robust kalman filtering with a robust particle filter," *IEEE Trans. Signal Process.*, vol. 70, pp. 47–56, 2022.
- [18] Y. Cai, J. Llorca, A. M. Tulino, and A. F. Molisch, "Decentralized control of distributed cloud networks with generalized network flows," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 256–268, 2023.
- [19] X. Fang, W. Feng, Y. Wang, Y. Chen, N. Ge, Z. Ding, and H. Zhu, "Noma-based hybrid satellite-uav-terrestrial networks for 6g maritime coverage," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 1, pp. 138–152, 2023.