Wireless Network Topology Discovery Based on Spectrum Data by Convolutional Neural Network

Xinfeng Deng[1], Zhihui Xie[2], and Li Zhou[1](✉)

[1] College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China
zhouli2035@nudt.edu.cn

[2] Songjiang Power Supply Company of State Grid Shanghai Electric Power Company, Shanghai, China

Abstract. Wireless network topology can reflect the communication relationships among network node. Since there are significant challenges and difficulties in deciphering the communication contents, spectrum data is adopted to discover communication relationships and network topology of a wireless network. In this paper, we propose a wireless network topology discovery method based on spectrum data to determine the communication relationships of nodes. Since the spectrum data features of nodes are correlated during the communication process, we construct the wireless network topology by mining the communication behaviors of nodes from the spectrum data features based on maximum similarity and hierarchical clustering. Simulation results demonstrate that the proposed method can achieve a better performance of hierarchical clustering than the existing methods.

Keywords: Spectrum Data · Communication Relationship · Network Topology

1 Introduction

With the development of spectrum monitoring and cognitive radio networks, the analysis and mining of spectrum data are receiving more and more attentions [1, 2]. As a reflection of user activities, the study of spectrum data can provide many essential pieces of information and intelligence, especially in anti-terrorism, military communication, and network security fields. Currently, the applications of spectrum data are mainly focused on spectrum situation awareness [3], signal classification [4], and signal feature extraction [5]. The physical characteristics of spectrum data and the statistical laws presented by these features also reflect the through-connection relationships and relevant information. However, mining the connections between the massive amount of spectrum data and the communication relationships among network nodes have yet to be further investigated.

There is a large amount of literature on communication relationship discovery or topology discovery. On one hand, the research of communication relationship discovery can be deciphered with the contents of spectrum data [6–8]. On the other hand,

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 398–410, 2024.
https://doi.org/10.1007/978-3-031-65126-7_34

there are numerous works that investigate network topology by mining the statistical laws of spectrum data. The research based on physical characteristics of spectrum data [9–12] needs to obtain a large amount of physical information data before clustering, such as power and frequency. In recent years, machine learning is increasingly being applied to analyze communication relationships with spectrum data. Wu et al. [14] proposed a method to identify different automatic link establishment (ALE) behaviors with an improved DenseNet. The method requires a highly labelled sample size, and the ALE signal strength significantly impacts network performance. Cheng et al. [15] presented a data-enhanced communication behavior recognition scheme to cope with insufficient spectrum data samples. However, the communication scenarios constructed by this method have strict hierarchical relationships and weak generalization ability. Cheng et al. [16] studied squeeze-excitation based communication relationship, but there is an apparent hierarchical relationship between the two communication nodes. Zhang et al. [17, 18] investigated the identification of communication relationship based on convolutional neural networks, but did not perform the network topology. Instead of relying on the communication contents and many physical characteristics, this paper extracts the spectrum data feature vectors combined with hierarchical clustering to mine communication relationships and network topology.

In this paper, by pre-processing the spectrum data and extracting the spectrum data features, we first determine the location of nodes relative to the monitoring station based on the power and orientation information of the spectrum data and then use the image feature vector extracted by convolutional neural network (CNN) to group the ones with high similarity into a category based on the similarity of the features using hierarchical clustering method. Since the spectrum data characteristics of the two nodes are very similar in fixed-frequency communication, it is possible to determine the through-connection relationship between nodes based on spectrum characteristics. Finally, the entire wireless network topology is formed according to the through-linkage association of all nodes. The experimental results prove that the method is simple to implement, the code is concise and has good clustering and adaptability to the spectrum data. The contributions of this paper are as follows:

- We transform the problem of communication relationship into spectrum data image classification problem to recognize communication behavior.
- We study the wireless network topology without relying on the communication content and a large number of physical characteristics, which are obtained by mining the spectrum data features and using the spectrum features similarity to get the association relationships.
- We use a CNN model VGG16 to get feature vectors and then use an unsupervised classification method (clustering) for classification. Our clustering method can achieve better performance than other clustering methods in the Mutual Information (MI) and Normalized Mutual Information (NMI) evaluation index.

The rest of this paper is organized as follows. Section 2 describes the wireless communication network scenario, and also introduces the overall framework for wireless network topology discovery. Section 3 introduces data pre-processing including feature selection and feature representation. Section 4 uses VGG16 model and the hierarchical clustering to mine the communication relationship and network topology. Section 5 introduces the simulation platform and discuss the experiment results. Finally, Section 6 summarizes the work of the full paper.

2 System Model

Wireless communication always happens in more than one node, and communication relationship can reflect the wireless network topology. As shown in Fig. 1, it is a scenario of wireless network topology discovery based on spectrum data. The work we do is to obtain the communication relationship between communication nodes and further mine the network topology.

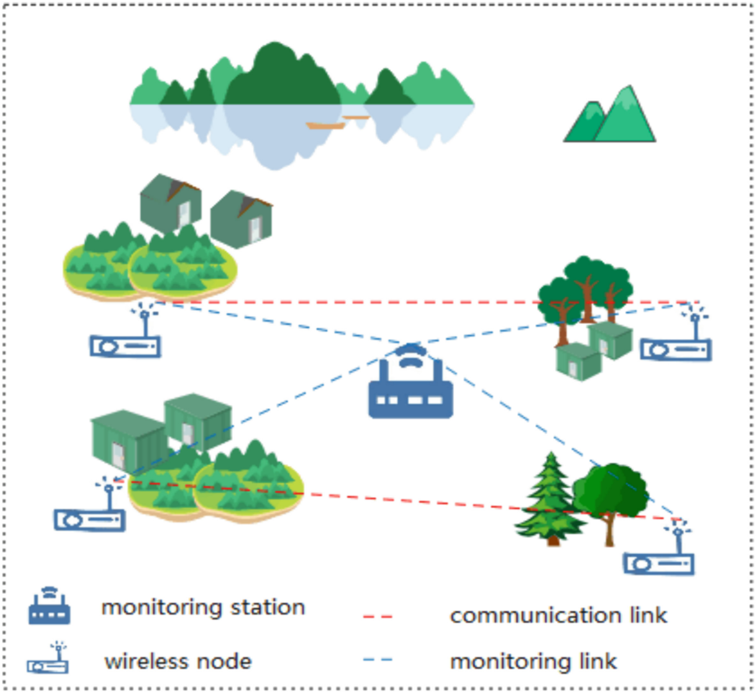

Fig. 1. Scenario of wireless network topology discovery based on spectrum data.

To achieve this goal, we first preprocess the spectrum data and use two variables, i.e., distance and orientation, to train the prediction of node locations. Then, CNN is used to extract image feature vectors, and hierarchical clustering is adopted for classification based on similarity to obtain communication relationships. Finally, the wireless network topology is mined by combining node locations and communication relationships. The overall framework is shown in Fig. 2.

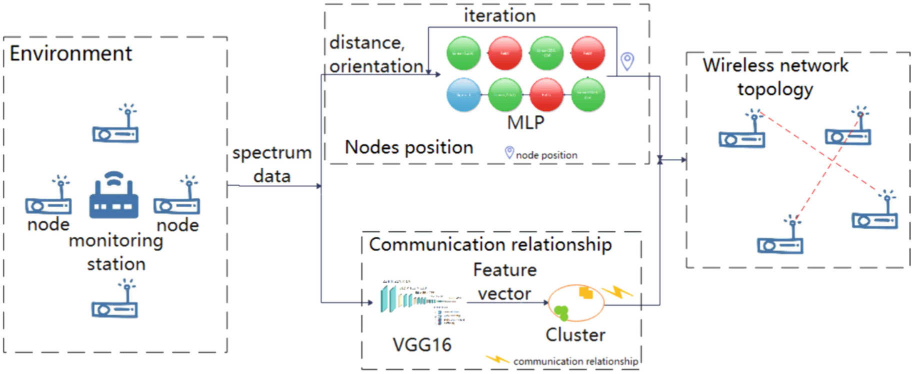

Fig. 2. Overall framework of wireless network topology discovery.

3 Data Pre-processing

3.1 Feature Selection

To mine the communication relationship and topology in wireless nodes from spectrum data, we need to first select the spectrum data features. In fixed-frequency communication, the communication relationship between different nodes can be distinguished by carrier frequency. Then, clustering is then performed based on the obtained spectrum data features and different clustering sets can be obtained. Consequently, each clustering set can represent a communication association and communication relationships are determined according to the carrier frequency characteristics. Here are the features that we select for further feature representation.

- **Signal power:** It is a unit of measurement of signal energy in the communication process. Due to the effect of large-scale signal propagation and fading, smaller power is measured with the increase of distance, and the signal strength is also stable if the node's location keeps unchanged. If a node moves within a short time, we assume that the received signal power remains constant.
- **Carrier frequency:** In long-distance transmission, the signal is not transmitted directly but moved to a fixed high frequency for transmission to improve the propagation distance. This high-frequency signal is called the carrier, and the frequency is called the carrier frequency, also known as the fundamental frequency.
- **Signal orientation:** The node's orientation information is related to the spectrum monitoring station. When the node's location is fixed, the orientation of the received signal is also stable. If the node moves in a relative short time, we consider that the orientation data remains the same.

The features are extracted from the spectrum data by CNN to predict the relative location of the nodes (relative to the spectrum monitoring station), and by using the carrier frequency, the communication association relationships can be uniquely determined.

3.2 Feature Representation

Assuming that $\boldsymbol{N} = \{1, 2, ..., N\}$ represents the index of network nodes. According to the spectrum data from the monitoring station, the spectrum data at time slot t is $\boldsymbol{U}^t = \left\{\boldsymbol{u}_1^t, \boldsymbol{u}_2^t, ...\boldsymbol{u}_i^t, ..., \boldsymbol{u}_N^t\right\}$, where $\boldsymbol{u}_i^t = \left\{p_i^t, \theta_i^t, t_i^t, f_i^t\right\}$, $i \in \boldsymbol{N}$ represents the set of signal power, signal orientation, monitoring time, and carrier frequency of the node i at time slot t. The prediction of the node's location relative to the monitoring station is based on the spectrum data features. The deviation is found to be small when compared with the actual location. This provides the location of a specific node for network topology mining.

4 Communication Relationship and Network Topology Discovery

4.1 Communication Relationship Mining

The key point of communication relationship discovery is to extract the features of the spectrum data and classify them according to the features. The location of a node is predicted by the power and orientation information as the communication network nodes. After that, according to the communication relationship between the classifications, each network node is connected to form a communication network to complete the mining of the network topology.

To determine the communication relationship, we need to obtain the location of each node first. The data set $\boldsymbol{Z} = \{z_1, z_2, ..., z_i, ..., z_N\}$ represents the orientation and power information in the spectrum data, where $z_i = \{p_i, \theta_i\}$, $\boldsymbol{p}_i = \{p_r, p_t\}$, $i \in \boldsymbol{N}$. We assume that the signal propagates in free space, and the signal power decreases gradually with increasing distance. The node distance from the monitoring station is d, the transmit power is p_r, and the received power is p_t. Therefore, the ratio of the transmit power and received power is defined as

$$\frac{p_t}{p_r} = \left[\frac{\sqrt{G_l}\gamma}{4\pi d}\right]^2, \tag{1}$$

where $\sqrt{G_l}$ represents the product of the transmitting antenna gain and the receiving antenna gain. The location of each node can be uniquely determined by combining its orientation and distance, as shown in Fig. 3.

Communication relationships may vary from time to time. In order to study the evolving communication relationships, we can divide the data set by time slots. In this way, we can further intuitively explore the communication sequences, network connectivity path, and communication directions and lay the foundation for the subsequent construction of the network topology. The predicted location of each node is denoted as (d_i, θ_i).

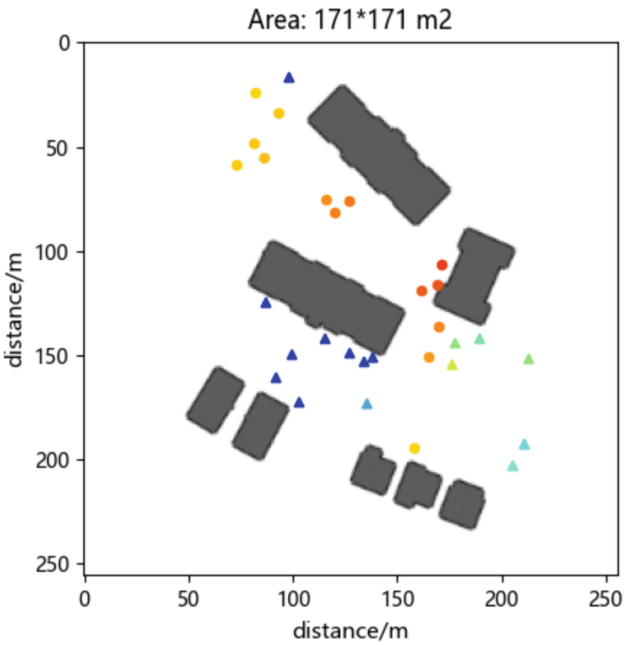

Fig. 3. Node location map.

4.2 VGG16 Model

VGG16 is a well-known CNN model whose name is derived from the initials of the Visual Geometry Group, Oxford University. The model is greatly adapted to the classification task. VGG16 uses small convolutional kernels and small pooling kernels. The model architecture is concise because the convolutional kernels focus on expanding the channel count. The pooling kernels focus on reducing the width and height so that the model is more profound and broader while the computation increases more slowly. At the same time, the fully connected layer is replaced with three convolutional layers in the training phase of the network. The test of the convolutional network has no limitation of full connection, so the input can be an image of any size. The structure of VGG16 model is illustrated in Fig. 4.

The communication relationship can be obtained as there is similarity in spectrum data features. To obtain the communication relationship, we first use CNN to extract the feature vectors of the spectrum data. We do not rely on the physical characteristics of the spectrum data, but the features exhibited by the image of spectrum.

4.3 Hierarchical Clustering

Hierarchical clustering methods put the nearest sample points into one class by calculating the distance between samples and then merges the nearest classes into a larger class. As a branch, split hierarchical clustering method initially put all samples in the same class. By continuously excluding dissimilarities, the samples are finally grouped

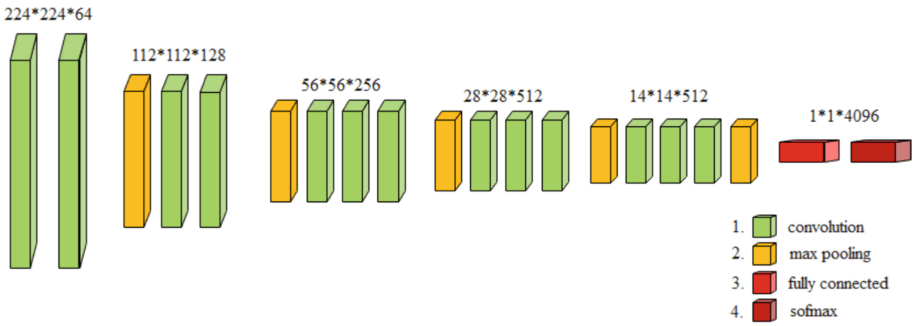

Fig. 4. The structure of VGG16 model.

into different data clusters. On the contrary, cohesive hierarchical clustering treats each individual as a cluster and then keeps merging these clusters until the clusters of different classes are obtained under a particular condition.

We use the feature vectors obtained by VGG16, then clusters the feature vectors based on the maximum similarity. The number of clusters is decided based on the similarity of the clustering tree. As a result, different clusters represent different communication relationships.

4.4 Network Topology Discovery

We obtain network topology according to the PageRank algorithm in the literature [13]. Therefore, if a network node is connected to other nodes, this node behaves as member node in the communication network, i.e., the node has a high PageRank value. Assuming that a node has a high PageRank value, the PageRank value of the nodes connected to it will increase accordingly. Therefore, the laws of edges and nodes should be considered when analyzing the network topology.

Suppose the wireless network topology mined from the spectrum data is $\boldsymbol{C} = \{\boldsymbol{N}, \boldsymbol{R}\}$, where $\boldsymbol{R} = \{(e_1, p_1), (e_2, p_2), \ldots, (e_i, p_i), \ldots, (e_M, p_M)\}$ stands for the pass-through relationship between nodes, p_i represents the number of connections, and e_i defines the edges. Besides, we record the number of time slots that the node acts as a sender or receiver and it participates in a communication. This is used to analyze the number of network nodes and wireless network topology. Based on the above description, the procedure of the whole network topology discovery can be presented in Algorithm 1.

Algorithm 1. Network topology discovery

1: Initialize communication scenarios
2: Acquisition of spectrum data
3: Determine distance and orientation information
4: **for** epochs = 1, M **do**
5: MLP prediction node location
6: **end for**
7: VGG16 extracts spectrum data feature vectors
8: Hierarchical clustering for classification
9: Obtain communication relationships
10: Obtain network topology
11: **end**

In military scenarios, information is usually transmitted step by step due to a strict hierarchy, i.e., information is transmitted in steps. Although network nodes are interconnected, communication devices across levels are subject to certain restrictions, i.e., each communication device has a different communication range and authority, and the flow of information from lower to higher levels often requires a step-by-step delivery. Therefore, the communication behavior of nodes at different levels differs. When we analyse network topology, the path contains the direction of information transmission, network hierarchy, and communication order. Determining the network key nodes and sub-networks can also be done by analyzing the network topology.

5 Simulation Experiments

5.1 Scene Setting and Data Collection

We adopt the dataset which was built based on the real map scenarios[1]. We consider 30 nodes are randomly distributed in a square area of 171 m × 171 m. The nodes communicate in the frequency range of 885–909 MHz and 930–954 MHz, and the monitoring station has a scan bandwidth of 20 MHz and a scan rate of 80 GHz/s [5].

Based on the nodes and monitoring station setup in Fig. 3, the communications between nodes are simulated, the spectrum data is monitored using the monitoring station, and the spectrum data characteristics are analyzed, from which the communication relationship and network topology are mined. In the simulation experiment, node terminals communicate with each other and generate feedback. The wireless network topology is determined based on the location of nodes and simulated communication.

The experiment parameters are presented in Table 1, which contains bandwidth, number of pickup carriers, and transmitter and receiver antenna wavelength spacing.

5.2 Analysis of Experimental Results

According to the VGG16 network and hierarchical clustering method, features are extracted from the spectrum data to mine the wireless network topology. Figure 3 displays

[1] https://www.mobileai-dataset.com/html/default/zhongwen/shujuji/1592719963402108929.html?index=1.

Table 1. Experiment parameters.

Parameter	Value
Transmitter Antenna Wavelength Spacing	5 m
Receiver Antenna Wavelength Spacing	5 m
Bandwidth	46.08 kHz
Number of Subcarriers	384
Number of Picked Carriers	5
Number of Nodes	30

the distribution of node locations obtained from signal power and orientation, representing the nodes comprising this communication network. The procedure involves acquiring the features of spectrum data using the VGG16 model from the dataset. Subsequently, we apply hierarchical clustering to different nodes based on the principle of maximum similarity. Nodes exhibiting high similarity indicate sustained communication behaviors and a generic relationship between them. All nodes sharing a generic relationship are mined to construct the wireless network topology. By analyzing the clustering tree, we identify edges and network nodes, setting the number of categories for clustering and obtaining the clustering effect graph. In the same category, a through relationship between two nodes is established, and based on this relationship, the corresponding nodes are connected to form the network topology.

In this study, feature vector extraction is performed using the VGG16 network and DenseNet121 network [14]. Hierarchical clustering and other clustering methods (DBSCAN, OPTICS, K-means) are compared using evaluation metrics such as Mutual Information (MI), Normalized Mutual Information (NMI), and Adjusted Mutual Information (AMI), as shown in Fig. 5 and Fig. 6. MI was first proposed in [19], NMI is a metric that normalizes MI, and AMI is a metric that adjusts mutual information to account for the effects of random clustering results [20]. These metrics are employed to assess the degree of similarity between clustering outcomes and ground truth labels, where elevated values denote superior performance. The results indicate that hierarchical clustering outperforms other methods under the NMI and MI indexes.

Specifically, under the NMI metric, the image feature vector extraction using the VGG16 network with hierarchical clustering yields 0.81, while K-means achieves 0.68. The remaining two methods result in NMI scores of 0.45 and 0.47, respectively. For the DenseNet121 network, the NMI for hierarchical clustering is 0.84, whereas K-means and OPTICS produce NMI scores of 0.37 and 0.31, respectively. Thus, hierarchical clustering demonstrates superior overall performance in this context.

Figures 5 and 6 illustrate that the hierarchical clustering method exhibits better performance. However, the evaluation metrics MI, NMI, and AMI under the two convolutional networks do not vary significantly. The DenseNet121 network requires more time to extract image feature vectors, and the dimension of the extracted feature vectors is much larger compared to the VGG16 network, as depicted in Fig. 7. Taking into account the time cost and memory considerations, the overall performance of the VGG16

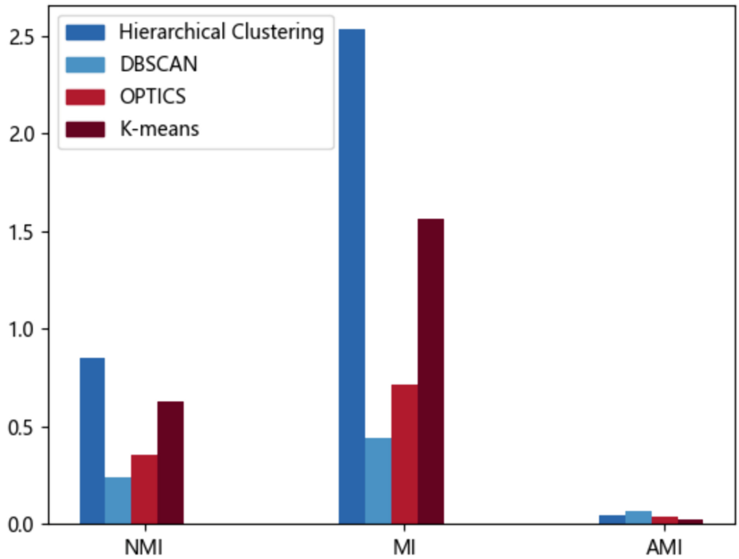

Fig. 5. The VGG16 method comparison diagram.

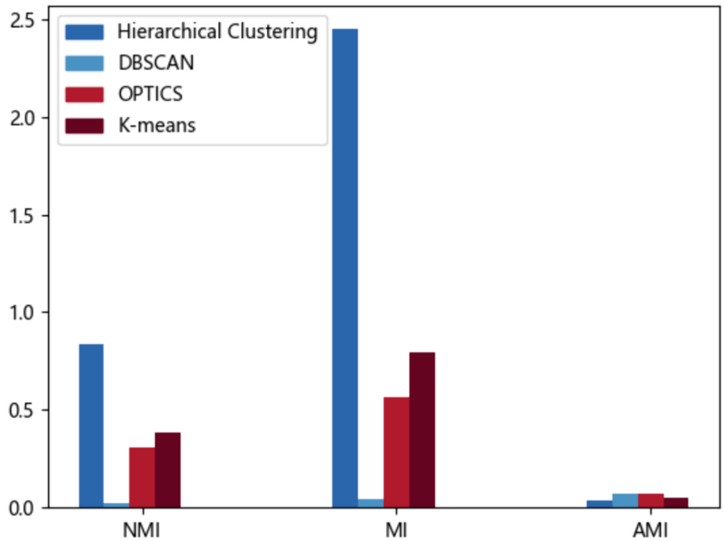

Fig. 6. The DenseNet121 method comparison diagram.

network surpasses that of the DenseNet121 network. Hence, it is preferable to extract the image feature vectors using the VGG16 network and then determine the number of clusters based on the clustering tree. After clustering, the categories can be filtered to identify the through-connections. By connecting the network nodes, a communication network structure can be formed. This approach strikes a balance between performance and resource efficiency.

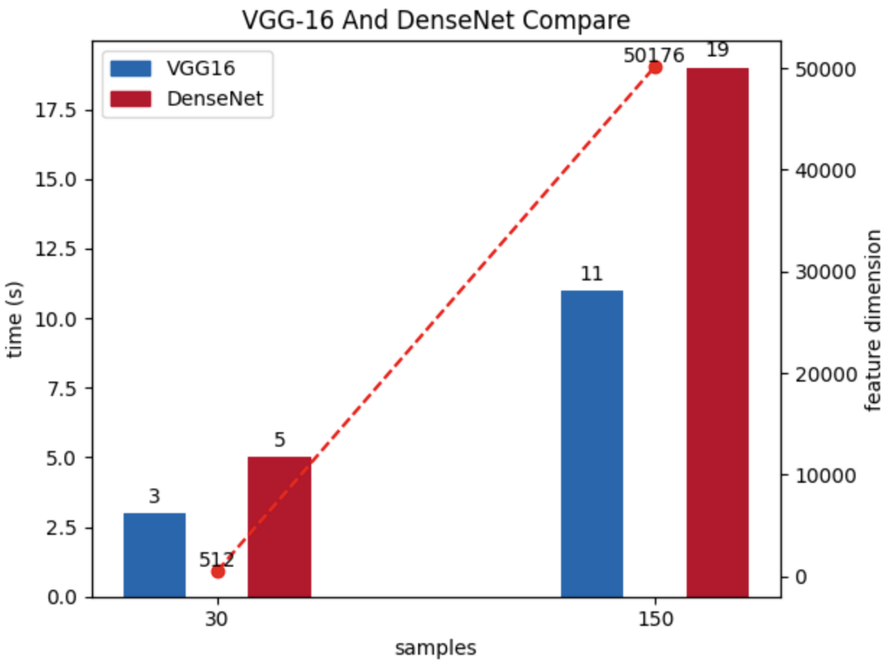

Fig. 7. VGG16 and DenseNet121 comparison diagram.

6 Conclusion

This paper aims to reduce the time and cost of deciphering communication contents by analyzing the characteristics of spectrum data to mine the communication behavior and network topology of nodes. Firstly, the features of spectrum data are selected, and preprocessing techniques are applied to calculate and uniquely determine the network nodes. Then, a CNN model VGG16 is utilized to extract the features of the spectrum data. With the maximum similarity method in hierarchical clustering, the samples with higher similarity are grouped into one class. This process helps to obtain the communication behavior of each node according to the clustering results. A network topology discovery algorithm is designed to mine of communication behavior and obtain wireless network topology from spectrum data. Experimental results demonstrate that the method is able to discover hidden communication behaviors that contribute to reveal the existing wireless network topology.

Acknowledgement. This research was supported by the National Natural Science Foundation of China (No. 62171449).

References

1. Yong, L., Zhou, Z., Xiong, L.: Research on electromagnetic spectrum management and control system of launch site based on big data. In: 21st International Symposium on Communications and Information Technologies 2022, Xi'an, China, pp. 27–30 (2022)
2. Hu, S., Pei, Y., Liang, Y.: Sensing-mining-access tradeoff in blockchain-enabled dynamic spectrum access. IEEE Wirel. Commun. Lett. **10**(4), 820–824 (2021)
3. Wang, C.: A spectrum feature-based security situation awareness algorithm for network coordinate system. In: 11th International Conference on Communications, Circuits and System 2022, pp.13–15 (2022)
4. Mario, B.: A deep learning-based signal classification approach for spectrum sensing using long short-term memory (LSTM) networks. In: 6th International Conference on Information Technology, Information Systems and Electrical Engineering 2022, Yogyakarta, Indonesia, pp. 13–16 (2022)
5. Liu, C., Wu, X., Zhu, L.: The communication relationship discovery based on the spectrum monitoring data by improved DBSCAN. IEEE Access **121**(10), 793–804 (2019)
6. Zeng, Y., Zhang, R.: Wireless information surveillance via proactive eavesdropping with spoofing relay. IEEE J. Sel. Topics Signal Process **10**(8), 1449–1461 (2016)
7. Han, Y., Duan, L., Zhang, R.: Jamming-assisted eavesdropping over parallel fading channels. IEEE Trans. Inf. Forensics Security **14**(9), 2486–2499 (2019)
8. Xu, J., Duan, L., Zhang, R.: Proactive eavesdropping via cognitive jamming in fading channels. IEEE Trans. Wireless Communication. **16**(5), 2790–2806 (2017)
9. Pan, T., Wu, X., Yao, C.: Communication behavior structure mining based on electromagnetic spectrum analysis. In: Proceedings of IEEE TAIC 2019, Chongqing, China, pp. 1611–1616 (2019)
10. Zhang, H., Yao, Y., Lei, L.: Study of end-to-end radio pass-through relationship based on spectrum data. Commun. Technol. **53**(11), 2745–2748 (2020)
11. Liu, C., Wu, X., Zhu, L.: Discover and research of communication relation based on communication rules of ultrashort wave radio station. In: Proceedings of IEEE ICBDA 2019, Suzhou, China, pp. 112–117 (2019)
12. Liu, C., Wu, L., Zhu, L.: Research on communication network structure mining based on spectrum monitoring data. IEEE Access **10**(1109), 3945–3959 (2019)
13. Franceschet, M.: PageRank: standing on the shoulders of giants. Commun. ACM **54**(6), 1–9 (2010)
14. Long, Z., Hong, C., Gke, L.: Recognizing automatic link establishment behaviors of a short-wave radio station by an improved unidimensional DenseNet. IEEE Access **10**(1109), 96055–96064 (2020)
15. Cheng, K., Zhu, L., Yao, C.: DCGAN based spectrum sensing data enhancement for behavior recognition in self-organized communication network. In: China Communications 2021, China, pp. 182–196 (2021)
16. Cheng, K., Zhu, L., Yao, C.: Squeeze-excitation based communication behavior recognition approach for spectrum sensing data. In: International Conference on Computer and Communications. 2021, Chengdu, China, pp. 682–686 (2021)
17. Zhang, H., Zhu, L., Yao, C.: Recognition of communication relationship based on the spectrum monitoring data by improved VGGNET. In: International Conference on Computer and Communications. 2020, Kunming, China, pp. 329–337 (2020)
18. Zhang, H., Zhu, L., Yao, C.: Recognition of connection relationship based on the spectrum monitoring data by deep convolutional neural networks. In: International Conference on Communication Technology, China, pp. 1285–1290 (2020)

19. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
20. Vinh, N., Epps, J., Bailey, J.: Information theoretic measures for clustering comparison: variants, properties, normalization and correction for chance. J. Mach. Learn. Res. **11**(10), 2837–2854 (2010)