Research on Discovery of Radio Communication Relationship Based on Correlation Analysis

Changkun Liu¹, Xinrong Wu¹, Changhua Yao¹, Jibin Guo¹, Haoren Fan¹, Ting pan¹, Kefeng Wei²
¹The Army Engineering University of PLA, Nanjing, China
²Unit 61623 of PLA, Beijing, China
¹liuchangkun@126.com, bNJWXr@163.com

Abstract. As the medium of wireless communication, the physical characteristics of electromagnetic spectrum signals and the connection between signals also potentially reflect the intelligence information related to the communication intention of communication individuals. For military communications, spectrum management, communication security and other fields, it is of great significance to mine and analyze the communication behavior of communication individuals from spectrum data. This paper proposes a method for mining the communication relationship between stations in massive spectrum signals based on association analysis. Firstly, we improved the peak density clustering algorithm and used it to study the distribution rules and statistical information of spectrum signals. Then, we carried out correlation analysis on the time range of clustering data to mine of the communication relationship. The experimental results show that this method can effectively discover the communication relationship from chaotic and missing spectrum signals, which lays a foundation for further discovering and studying the communication network structure between radio stations and provides ideas for the analysis of spectrum data.

1. Introduction

The electromagnetic spectrum has become an indispensable national strategic resource in the information age and plays an important role in all aspects of society. The rapid development of wireless communication technology makes the spectrum environment more complicated and also makes the spectrum management more difficult. It is of great significance to strengthen the monitoring and analysis of spectrum signals in the fields of spectrum management and communication security, especially in the fields of military affairs and anti-terrorism, the dominance of spectrum resources is the basis for winning the future electromagnetic spectrum war [1].

From the open literature, the mining and analysis of spectral data are mainly focused on the study of the signal itself, such as signal parameter estimation, signal classification, and abnormal signal detection. For massive spectrum signals, there are relatively few studies to obtain the connection between the communication individuals that generate the signals and the communication behavior of the communication individuals by analyzing the relationship between the signals, the physical characteristics of the signals, and the statistical laws. On the other hand, the study of wireless individual communication behavior is more through intercepting communication content. Such methods usually
require a huge price for signal cracking. However, by mining and analyzing the physical characteristics of the spectrum signal, the distribution law based on certain features, the statistical law, and the relationship between the spectrum signals, it is also possible to obtain hidden information such as communication relationships, location deployment, and behavioral relationships among communication individuals in the massive spectrum data.

Based on the massive spectrum signals generated by radio communication, this paper proposes a method to discover the communication relation between radio stations in the spectrum data based on correlation analysis. This method firstly extracts the characteristics of the signal such as center frequency, signal power, signal direction and signal monitoring time from the spectrum data, and then studies the distribution rule of the spectrum data through the improved peak density clustering algorithm. Finally, the radio communication relationship in the spectrum monitoring data is mined based on correlation analysis. The method studies the relationship between the communication individuals that generate the spectrum signals according to the physical characteristics of the signals, the connections between the signals, and the method of data mining. This method avoids the difficulty of intercepting the analysis signal content to obtain communication behavior, and provides a new perspective for the analysis of spectrum data.

2.  Related Work

According to the existing literatures, the analysis and mining of spectral data is mainly aimed at signal anomaly detection [4] [5], signal classification [6] [12] and spectrum sensing [7] [8] research. These studies are more limited to the spectrum signal itself, but there are few studies on further mining communication individuals [1] through spectrum data. However, through the in-depth analysis of the spectrum signal, it is of great significance to acquire the communication behavior and communication relationship of communication individuals within the monitoring range, especially for the future spectrum warfare and other fields [1].

Liu et al [1] proposed a method for mining communication relationships between communication individuals based on ideal spectrum monitoring data. The method classifies only the ideal spectrum monitoring data through communication rules, and further analyzes the relationship between the data to mine the communication relationship between the stations in the spectrum monitoring data. However, the actual spectral data is missing, and the classification method cannot handle the confusing spectral data. Pan et al [2] monitors’ data through multiple monitoring stations, estimates the relative geographical position of the communication individual according to the attenuation changes of the signals monitored by different monitoring stations, and then constructs the communication network according to the communication relationship between the signals. However, there are errors in the fusion processing of data from multiple monitoring stations, and this method has higher requirements for monitoring data, and it is difficult to mine missing spectrum data. Liu et al [3] proposed a method based on the improved DBSCAN algorithm to study the communication relationship in spectral data. However, this method still has high requirements for data integrity. As the communication environment deteriorates, the accuracy of the method will decrease. This method has preprocessed the data multiple times, thus losing some of the original information of the data.

Based on the above research background, we propose a method for obtaining the communication relationship between stations using the original spectral data. This method still has a good mining effect on the missing spectrum monitoring data.
3. Data Feature Analysis

3.1. Spectrum Monitoring Data Analysis

![Figure 1](image1.png)

**Fig. 1** Spectrum signal monitored during a scan period, from which the center frequency, signal bandwidth, and signal power can be extracted.

In order to mine the communication relationship between communication individuals from the spectrum monitoring data, we need to preprocess the spectrum monitoring data to extract the features that can characterize the signal and apply to the mining connection relationship. As shown in Figure 1, the original spectrum monitoring data includes frequency, signal bandwidth, signal power, and monitoring time. On the other hand, by monitoring the calculation of the device, we can also obtain the direction information of the signal.

The actual monitored signal is missing due to fading during signal propagation and the scanning period of the monitoring device. As shown in Fig. 2, the yellow area indicates the scanning period monitoring, and the black line indicates the signals of different lengths. During the scanning period, the monitoring device can only capture part of the signal, so the actually monitored signal has a large number of missing.

![Figure 2](image2.png)

**Fig. 2** Spectrum signals monitored by the spectrum device

3.2. Feature Selection and Analysis

Features such as carrier frequency, signal power, signal monitoring time, and signal direction represent the physical characteristics of the spectral signal. Although the fading of signal propagation is different in different environments, the signal power can also reflect the relative distance of the source relative to the monitoring device. The direction of the signal relative to the source can be obtained by the calculation of the monitoring device, and the direction of the signal reflects the direction information between the sources. The signal monitoring time directly reflects the time and time range of communication and the time relationship between signals. For point-to-point communication stations, complex and poor communication environments will cause large error rates and fading. In order to ensure reliable data
transmission, the station usually stops using the ARQ protocol [11] for error control at the data link layer. Therefore, after the sender sends the data once, the receiver needs to reply to the ACK for confirmation. This transmission-recognition mechanism makes the specific time correlation and alternation between the data. Therefore, the two sets of stations that maintain communication have a similar time range for the spectrum signal set generated during communication.

Due to the error caused by the monitoring equipment and transmission process, the actual monitored signal shows a normal distribution in the dimension of signal power, signal direction and signal monitoring time. Therefore, the clustering method can be used to study the distribution of massive spectrum signals. We selected the carrier frequency, the average signal power, the signal direction and the monitoring time of the signal to identify and mine the communication relationship.

4. Communication Relationship Discovery Method

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar. Suppose spectrum monitoring data sets \( X = \{x_1, x_2, \ldots, x_n\} \), where \( x_i = \{F_i, P_i, \theta_i, B_i, t_i\} \). \( F \) is the central frequency point of the signal, \( B \) is the bandwidth, \( P \) is the signal power, \( t \) is the signal monitoring time, and the direction of the signal is \( \theta \).

4.1. Improved Density Peak Algorithm

Because the spectral data show clustering properties in the space composed of signal power, signal direction and signal monitoring time, we introduce column coordinates and adopt density peak algorithm [9] to study the clustering and distribution of the data. In order to better adapt to the characteristics of the data, we improved the peak density clustering algorithm.

\[
\text{dist}(y_i, y_j) = \sqrt{(\theta_i - \theta_j)^2 + (P_i - P_j)^2 + \delta(t_i - t_j)}
\]

(1)

\[
\delta(t) = \begin{cases} 0, & |t| \leq h \\ \infty, & |t| > h \end{cases}
\]

(2)

Where \( y_i = \{P_i, \theta_i, t_i\} \), and \( h \) is time threshold.

4.2. Correlation Analysis

The spectral data is clustered by the improved density peak clustering algorithm, and each cluster set represents a spectrum signal set generated by different communication individuals in different communication processes. Since the spectrum signals generated by the stations with the communication relationship are similar in time, we match the cluster sets by correlation analysis to discover the connection relationship between different cluster sets.

4.3. Communication relation discovery algorithm

**Input:** \( Y = \{y_1, \ldots, y_n\} \), where \( y_i = \{P_i, \theta_i, t_i\} \)

**Time threshold** \( \epsilon \), **cut-off distance** \( d_c \)

**Output:** Number of communication relationships;

Spectrum data set corresponding to the communication relationship;
1. Improve the distance of the density peak algorithm according to (1)(2);  
2. The clustered spectral data set is clustered by the improved density peak algorithm to obtain the cluster set \( U = \{U_1, U_2, \cdots, U_m\} \).

3. For \( i = 1: m \)

4. Record the start time \( t_{\text{begin}}_i \) and end time \( t_{\text{end}}_i \) of the data set \( U_i \).

5. \( \text{end} \)

6. \( k = 1 \)

7. While \( U \) is an empty set

8. if \( \left| t_{\text{begin}}_i - t_{\text{begin}}_j \right| < \varepsilon \)

9. if \( \left| t_{\text{end}}_i - t_{\text{end}}_j \right| < \varepsilon \)

10. \( U_i \) and \( U_j \) have a communication relationship, denoted as \( r_k \);

11. Remove \( U_i \) and \( U_j \) from \( U \);

12. The spectrum set corresponding to the communication relationship \( r_k \) is recorded as \( V_k = \{U_i, U_j\} \);

13. \( \text{end} \)

14. \( \text{end} \)

15. \( \text{end} \)

Time complexity analysis of the algorithm: the time complexity of the density peak algorithm is \( O(n^2) \), the matching of the cluster set is linear, and the time complexity is \( O(n) \), so the total time complexity is \( O(n^2) \).

5. Verification experiment

5.1. Scene Settings

We randomly arranged nine ultrashort wave stations within the monitoring range, and the stations communicated with each other. The communication sequence between the stations and the parameter settings are shown in Table 1

Based on the above scenario and parameter settings, the spectrum data of the communication between the stations is acquired by the monitoring device. By preprocessing the data, we obtain the center frequency, signal power, bandwidth, signal monitoring time, and signal direction of the spectrum signal, which are recorded as data set \( X \), as shown in Table 1.

Table. 1 Ultrashort wave radio communication group parameter setting

| \( x \) | Frequency (KHz) | Bandwidth (KHz) | Power (dBm) | Time (ms) | Signal direction |
|-------|-----------------|-----------------|-------------|-----------|-----------------|
| \( x_1 \) | 63349 | 9.7 | 18.234 | 23 | 45.6 |
| \( x_2 \) | 82100 | 9.8 | 27.654 | 5675 | 120.74 |

Note: The monitoring time starts from 0.
5.2. Data display

Figure 3 shows the distribution of spectrum monitoring data in the time domain, where two black lines represent the spectral signals produced by fixed frequency communication and the discrete points represent the spectral signals of the frequency hopping communication. Since the carrier frequency of the frequency hopping communication is constantly changing, the fixed frequency communication relationship and the frequency hopping communication relationship can be distinguished according to whether the carrier frequency changes. Figure 4 shows the distribution of the spectral signal in a cylindrical coordinate system consisting of signal power, signal direction, and signal monitoring time. The data has clustering properties.

Table 2 Example of data characteristics

| Communication group number | Communication type | Carrier frequency/range | Communication duration | Color annotation |
|----------------------------|-------------------|-------------------------|------------------------|-----------------|
| 1                          | Fixed frequency   | 60000KHz                | 2s                     | Light blue      |
| 2                          | Frequency hopping | 80000KHz                | 4s                     | Purple          |
| 3                          | Frequency         | 30---88MHz              | 1.8                    | Red             |
| 4                          | hopping           | 30---88MHz              | 3s                     | Blue            |
| 5                          | Frequency         | 30---88MHz              | 2s                     | Green           |
| 6                          | hopping           | 30---88MHz              | 3s                     | Yellow          |

Fig. 3 Time-frequency distribution of spectrum monitoring data

Fig. 4 Time-frequency distribution of spectrum monitoring data
5.3. Analysis of Results

Figure 5 shows the spectral data clustering results based on the improved density peak algorithm, with different color data representing different cluster sets. According to the time distribution of the cluster set data, the clustering set is correlated and matched to obtain the communication relationship between the stations. The cluster set of the same color in Fig. 6 represents the spectrum set produced by the station having the communication relationship. Projecting the data of Figure 6 to the polar coordinate system results in Figure 7, which shows the communication relationship between the stations. The communication relationship recognition results are shown in Table 3.

Fig. 5 Clustering results of the improved density peak algorithm

Fig. 6 Communication relationship matching results

Fig. 7 Results of the communication relationship shown in the polar coordinate system
Table 3 Communication relationship recognition results

| Communication group number | Communication type   | Carrier frequency/range | Communication duration | Color annotation |
|----------------------------|----------------------|-------------------------|------------------------|------------------|
| 1                          | Fixed frequency      | 60000KHz               | 2s                     | Light blue       |
| 2                          | Fixed frequency      | 80000KHz               | 4s                     | Purple           |
| 3                          | Frequency hopping    | 30----88MHz             | 1.8                    | Red              |
| 4                          | Frequency hopping    | 30----88MHz             | 3s                     | Blue             |
| 5                          | Frequency hopping    | 30----88MHz             | 2s                     | Green            |
| 6                          | Frequency hopping    | 30----88MHz             | 3s                     | Yellow           |

Fig. 8 Recognition accuracy comparison

Finally, in order to verify the validity of the proposed method, we compare the method with the method proposed in [1] [2] [3]. The recognition accuracy is shown in Figure 8. Obviously, the method proposed in this paper has higher accuracy. The communication relationship discovery method based on association analysis proposed in this paper can mine the communication relationship between stations from the missing chaotic spectrum monitoring data, and has higher recognition accuracy.

6. Conclusion

The mass spectrum signal hides the intelligence information related to the communication behavior of the communication individual. Especially in the military and communication security fields, it is of great significance to dig deep into the hidden information of the massive spectrum signal. In this paper, based on the physical characteristics of the spectrum signal, the distribution law and the relationship between the signals, the communication relationship between the stations is analyzed. The discovery of the communication relationship laid the foundation for the next step of analyzing the communication network and communication behavior of communication individuals.

Based on the massive spectrum signals generated by radio communication, this paper proposes a method to discover the communication relation between radio stations in the spectrum data based on correlation analysis. This method firstly extracts the characteristics of the signal such as center frequency, signal power, signal direction and signal monitoring time from the spectrum data, and then studies the distribution rule of the spectrum data through the improved peak density clustering algorithm. Finally, the radio communication relationship in the spectrum monitoring data is mined based on correlation analysis. The experimental results show that this method has better adaptability than the existing mining methods.

The research of this paper realizes the mining and analysis of the missing and disordered spectrum monitoring data, and can accurately mine the communication relation between radio stations from the
spectrum monitoring data, and lays a foundation for further mining the communication behavior characteristics of radio stations.

Acknowledgments
The work is supported by the National Natural Science Foundation of China (No. 61971439), the Natural Science Foundation of Jiangsu Province of China (No. BK20191329), Postdoctoral Science Fund of China (No. 2018M633684).

References
[1] C.-K Liu et al., "The Communication Relationship Discovery Based on the Spectrum Monitoring Data by Improved DBSCAN," in IEEE Access, vol. 7, pp.121793-121804, 2019.doi: 10.1109/ACCESS.2019.2938296
[2] C.-K Liu et al., "Discovery and Research of Communication Relation Based on Communication Rules of Ultrashort Wave Radio Station," in Proc. IEEE ICBDA, Suzhou, China, 2019, pp. 112-117.
[3] T. Pan et al., "Communication Behavior Structure Mining Based on Electromagnetic Spectrum Analysis," in Proc. IEEE TAIC, Chongqing, China, 2019, pp. 1611-1616.doi: 10.1109/ITAIC.2019.8785571
[4] R. Sreeraj et al., "SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features." in Proc. IEEE DySPAN, Seoul, 2018, pp. 1-9.
[5] M. Ahmed. "Abnormal behavior detection of jamming signal in the spectrum using a combination of compressive sampling and intelligent bivariate k-means clustering technique in wideband cognitive radio systems." in Proc. IEEE ICEE, Boumerdes, 2015, pp. 1-4.
[6] M. Tuma, V. Rørbech, M. K. Prior and C. Igel, "Integrated Optimization of Long-Range Underwater Signal Detection, Feature Extraction, and Classification for Nuclear Treaty Monitoring," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 6, pp. 3649-3659, June 2016.
[7] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, “Wideband spectrum sensing for cognitive radio networks: a survey,” IEEE Wireless Communications, vol. 20, no. 2, pp. 74-81, April 2013.
[8] E. Axell, G. Leus, E. Larsson, and H. Poor, “Spectrum Sensing for Cognitive Radio: State-of-the-Art and Recent Advances,” IEEE Signal Processing Magazine, vol. 29, no. 3, pp. 101–116, May 2012.
[9] Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. Science, 344(6191), 1492-1496.
[10] V. S. Avramchiks and Tran Viet Chau, "Time-frequency correlation analysis," International Forum on Strategic Technology 2010, Ulsan, 2010, pp. 182-183.
[11] R. Fantacci, "Performance evaluation of some efficient stop-and-wait techniques," in IEEE Transactions on Communications, vol. 40, no. 11, pp. 1665-1669, Nov. 1992.doi: 10.1109/26.179927
[12] He Weigang et al, "Study of improved multiple signal classification algorithm based on coherent signal sources," 2016 World Automation Congress (WAC), Rio Grande, 2016, pp. 1-4.