Research Article

Electromagnetic Environment Portrait Based on Big Data Mining

Lantu Guo ^{1,2}, Meiyu Wang ³, and Yun Lin ³

¹Beijing Institute of Technology, Beijing 100000, China

²China Research Institute of Radiowave Propagation, Qingdao 266000, China

³Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Yun Lin; linyun@hrbeu.edu.cn

Received 26 February 2021; Revised 23 March 2021; Accepted 2 April 2021; Published 3 May 2021

Academic Editor: Xin Liu

Copyright © 2021 Lantu Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of IoT in smart cities, the electromagnetic environment (EME) in cities is becoming more and more complex. A full understanding of the characteristics of past spectrum resource utilization is the key to improving the efficiency of spectrum management. In order to explore the characteristics of spectrum utilization more comprehensively, this paper designs an EME portrait model. By checking the statistical information of the spectrum data, including changes in the noise floor and channel utilization in each individual wireless service, the correlation between the spectrum and time or space of different channels and the information is merged into a high-dimensional model through consistency transformation to form the EME portrait. The portrait model is not only convenient for storage and retrieval but also beneficial for transfer and expansion, which will become an important foundation for intelligent electromagnetic spectrum management.

1. Introduction

With the coming of the 6G era, many advanced technologies have been extensively researched, such as artificial intelligence technology [1], image processing technology [2], intelligent integration technology [3], and wireless communication technology [4], which have also promoted the development of the IoT field. The application of IoT technology is becoming more and more important. The scarcity of spectrum resources is the most important issue facing IoT technology [5], and the research of electromagnetic spectrum space is related to the advantage of IoT technology in future smart cities [6, 7]. And many technologies related to electromagnetic space have been developed, including wireless device identification [8], radio frequency fingerprint identification [9, 10], and electromagnetic signal identification [11].

Of course, electromagnetic environment (EME) monitoring activities are the basis of the above research, including marine monitoring, land monitoring, mobile monitoring, or fixed station monitoring. Various monitoring activities continuously monitor the environment and accumulate a large amount of electromagnetic data. Equally important, the era of big data belonging to the EME has arrived. Electromag-

netic big data has 4 basic characteristics, including volume, variety, velocity, and value. Volume refers to the continuous growth of electromagnetic big data in the dimensions of time, frequency, and space, making the volume of data larger and larger. Variety refers to a wide range of electromagnetic data sources, and different equipment, locations, or tasks in monitoring activities bring data diversity. Velocity refers to the rapid generation of electromagnetic data and requires timely processing. Value refers to the experience of electromagnetic management and activities hidden in historical data. These valuable information in electromagnetic data need to be fully excavated. Therefore, it is very important to use big data mining technology to analyze the electromagnetic space.

Due to the spectrum resources are limited and have fixed propagation characteristics [12], they are also easily interfered by man-made noise and natural noise, resulting that there are some challenges in the comprehensive management of the EME [13, 14].

1.1. The Hidden Information Mining and Comprehensive Utilization of Electromagnetic Data Is Difficult. On the one hand, there are many types of electromagnetic data [15], showing multisource heterogeneous characteristics [16]. As

the types of electromagnetic equipment and stations continue to increase, spectrum usage methods continue to evolve [17], resulting in an increasing number of the electromagnetic data [18]. At the same time, many electromagnetic signal data automatic modulation classification and recognition methods have developed [19]. On the other hand, electromagnetic data management methods are backward and lack of unified electromagnetic data collection services, and processing methods are single, and valuable information urgently needs to be deeply explored [20]. At present, the commonly used database management technology only processes a single type of data, and cannot realize the fusion analysis and correlation analysis of multiple types of data [21]. Therefore, it is difficult to perform secondary analysis on the original monitoring data to obtain useful value information.

1.2. The Multidimensional and Omnidirectional Visual Description of EME Is Very Difficult. Electromagnetic environment data includes multidimensional domains such as “time domain,” “frequency domain,” “spatial domain,” and “energy domain” [22]. Multidimensional visualization of EME is a comprehensive description of multidimensional domain data information [23]. However, the existing visual description methods, whether it is point drawing or surface drawing, will encounter one-sided problems [24], and it is difficult to break through to high-dimensional visual description.

1.3. The Application Expansion of EME Visualization Is Difficult. On the one hand, electromagnetic big data has a large time scale, a large span of space, and dynamic changes in environmental information [25]. On the other hand, software development is task-oriented and goal-oriented, leading to scenarios and functions that are easy to solidify [26]. EME visualization software has constraints on the mobility and application scalability of different electromagnetic scenarios.

Therefore, a model that can mine electromagnetic big data from multiple angles and deep levels [27], describe the environment in multiple dimensions, realize correlation analysis, and extend application services is urgently needed. This paper proposes an EME portrait model, which combines portrait technology with big data mining. The main research content of this paper is shown in Figure 1.

First, the EME portrait is defined. There are three parts to construct a portrait: the first step is elements mining of EME portrait, the second step is interval number conversion of EME portrait elements, and the third step is integration of multimodal EME portrait. Finally, the future application of the EME portrait is prospected.

The main contributions of this paper are as follows:

- (1) The definition of EME portrait is proposed. The EME is invisible. The EME portrait proposed in this paper is aimed at visualizing the EME and laying a theoretical and practical foundation for the combination of portrait technology and big data mining technology
- (2) Designed a multilevel mining process for EME big data. Data mining levels include direct label extrac-

tion, data preprocessing, shallow information mining, secondary mining, and deep association mining. Through this mining process, the elements of constructing the portrait have been fully discovered

- (3) A consistent representation and fusion model is established for multidomain elements. In order to obtain more representative mining elements, the multidomain elements are converted into interval numbers, and then, the high-dimensional tensor model is used to fuse the multidomain elements to form the representation of portrait

1.4. Definition of EME Portrait. The portrait system emerged from big data in the field of e-commerce. Similar to the field of communications, accurate signal identification is equivalent to label signal [28]. In the era of big data, user information is flooded in the Internet. E-commerce abstracts each user's specific information into tags, and uses these tags to embody the user's image, thereby providing users with targeted services.

This paper combines the concept of user portrait into the description of EME. Based on the characteristics of multidimensional dynamic changes of EME, a model of EME portrait is designed, which is defined as follows.

1.4.1. Definition. The EME portrait is a virtual portrait formed by the combination of the characteristic and the relationship information obtained from the EME big data mining according to the environment adaptation.

The EME portrait model considers the five aspects of multiscale, multilevel, multidimensional, multiangle, and multigranularity, and uses multilevel mining algorithms to obtain rich environmental characteristics to depict the accurate and changeable EME, as shown in Figure 2.

The time scale of the portrait is based on the minimum time slot of the monitoring data and extends to minutes, hours, days, weeks, and months. The spatial scale of the portrait is based on the monitored coverage area. The bandwidth scale of the portrait is based on the service business division. The granularity of portraits is related to user needs, and can range from coarse-grained macrocontrol to fine-grained observation. Multiangle and multidimensionality cater to the multisource heterogeneous characteristics of electromagnetic big data. The multilevel portrait depends on the multilevel design of the mining algorithm, and the deep-level mining information comes from the secondary mining of the shallow-level mining information.

The EME portrait information obtained based on big data mining includes regularly changing vectors, semantic information, and some irregularly changing variables, so it also has complex and diverse characteristics. These information must be combined through high-dimensional fusion modeling to form a portrait model that is convenient for storage and application expansion. This paper combines interval number conversion with high-dimensional tensor models.

The EME portrait model has the following advantages.

FIGURE 1: The main research content of this paper.

FIGURE 2: Construction of EME portrait.

1.4.2. Rich Multidimensional Features. With the accumulation of massive amounts of data, the EME portrait model will fully cover the subtle features of multiple domains, including time, frequency, space, and energy, to help users fully understand the EME.

1.4.3. Real-Time Data Connection. The EME portrait model supports real-time connection, which can not only realize data backtracking but also meet the needs of real-time screening of current scenes.

1.4.4. Rapid Generation of Portrait. The EME portrait model supports label customization, which can output labels in real time or complete different environment portrait labels, with strong modeling capabilities and high generation capabilities.

1.4.5. Customized Joint Model. The EME portrait model can be customized for joint modeling, supporting existing data combined with demand-side data for joint modeling, and generating personalized feature labels to meet shared applications.

1.5. Elements Mining of EME Portraits. Data mining methods include association analysis, classification, clustering, outlier detection, evolution analysis, fuzzy function model and other methods, mining data frequent patterns, classification patterns, clustering patterns, abnormal patterns, evolution laws, and other knowledge. In this paper, big data mining in electromagnetic environment is divided into six stages, as shown in Figure 3.

FIGURE 3: Big data mining process of EME.

TABLE 1: Collect location information.

Short name		Short description
Location 1	IN	Office building, Aachen, Germany
Location 2	NE	Residential area, Maastricht, Netherlands
Location 3	AB	Third floor balcony, Aachen, Germany

The six stages include data source, direct label extraction, data preprocessing, shallow mining, secondary mining, and deep mining. Among them, the data source used in this paper is from Aachen University in Germany. The direct label extraction part extracts the direct label of data. The data preprocessing part deals with label data, noise data, redundant data, and abnormal data items. In the shallow-layer mining stage, after the mining target is defined, the mining algorithm technology is used to carry out the shallow layer information mining calculation and analysis of the data from the three aspects of bottom noise extraction, occupancy analysis, and signal analysis. On the basis of shallow mining, deep mining refers to further extracting and mining knowledge. Among them, data preprocessing, noise analysis, spectrum utilization analysis, and correlation analysis are the key points of data mining, which will be described in turn in the following paragraphs.

The data used in this paper are from data sets collected by Aachen University in Germany during 2006-2007. The data set was collected from three different locations with a collection frequency range of approximately 20 MHz to 6 GHz. And it is divided into four subbands within the frequency

TABLE 2: The technical parameters of the data set.

Center frequency	Band 1: $f_c = 770$ MHz
	Band 2: $f_c = 2250$ MHz
	Band 3: $f_c = 3750$ MHz
	Band 4: $f_c = 5250$ MHz
Frequency span	1500 MHz
Resolution bandwidth	200 kHz
Number of measurement points	8192
Sweep time	1 s
Measurement duration	About 7 days per subband

range. The collection location information is given in Table 1. And some important technical parameters are given in Table 2.

In this paper, some frequency bands are used as examples to experiment, such as TV band (614-698 MHz), ISM band (2400-2485 MHz), LSA band (2300-2400 MHz), GSM1800u band (1710-1785 MHz), and GSM1800d band (1820-1875 MHz). The energy changes of the above frequency bands in different acquisition locations are shown in Figure 4. It can be seen that the spectrum utilization is different in different service frequency bands or different urban areas.

1.6. Data Preprocessing. Data preprocessing is the first important link in data mining. There are three reasons why original

data cannot be directly used for mining: First, the original data has the problem of multisource heterogeneity, which requires data fusion. Second, there are some missing and abnormal data in the original data. Missing data can be filled through fitting and other methods, while abnormal data need to be detected and searched. Third, the amount of original data is often large, so it will greatly reduce the mining efficiency if all the data are used for mining. It is necessary to filter and simplify the data first.

Therefore, the following three methods are adopted to solve the above three problems: consistency processing, anomaly detection, and resolution selection in the data preprocessing stage. The complete process of data preprocessing is shown in Figure 5.

1.6.1. Consistency Processing. Because of the multisource heterogeneous characteristics and large amount of electromagnetic data, consistent processing is an essential part of data preprocessing. According to the storage form and characteristics of data, consistency processing may include digitization, discretization, normalization, and quantification. The storage of the Aachen data set is relatively standardized, so this work involves less digitization, discretization, and normalization, and mainly involves quantification of the power spectral density. The specific methods and applications of quantification will be discussed in the later section.

1.6.2. Abnormal Detection. In the part of anomaly detection, this paper mainly uses the interquartile range detection method. The first quartile (Q1), also known as the “smaller quartile,” is the 25% of the number in the sample from small

lest to largest. The second quartile (Q2) is the 50% number in the sample from smallest to largest. The third quartile (Q3), also known as the “higher quartile,” is equal to the 75% of all values in the sample in descending order. The difference between the third quartile and the first quartile is also called the quartile distance (IQR). The quartile distance is used to represent the data fluctuation distance, and the formula is as follows:

$$IQR = Q3 - Q1. \quad (1)$$

The maximum and minimum observed values that can be trusted are as follows:

$$\begin{aligned} \max &= Q3 + k * IQR, \\ \min &= Q1 - k * IQR, \end{aligned} \quad (2)$$

wherein k is the measurement factor and k is set as 1.5 in this paper.

Data is abnormal when it is out of range of trust. This method is an anomaly detection method based on the statistical characteristics of data. It is suitable for the long time, and the whole time abnormal and frequency points are disturbed. It has low computational complexity and good engineering practicability. Some of the results are shown in Figure 6. Figure 6(a) is to determine whether there is an abnormal event from the perspective of frequency points, and the box plot and its nearby points represent the data within the normal range, and the floating beard points above are the abnormal data. Figure 6(b) is to determine the

FIGURE 5: Process of data preprocessing.

location of the abnormal event from the perspective of the time slot; the point 36 is obviously abnormal.

1.6.3. Select Resolution. The data resolution needs to be determined according to the original data and the content of the mining. For EME data, the resolution can be selected from three perspectives: time domain, frequency domain, and spatial domain. The original data resolution is time interval 1.8 seconds and frequency interval 200 kHz. For most of the mining content, the resolution is only set from the time domain perspective. Generally, the time interval is 75 seconds, and the other resolutions are the same as the original data. For some special mining content (such as noise mining), this paper will choose a smarter resolution. The previous experience and the experimental results of this paper have shown that this resolution setting method can compress the amount of data while retaining the information more completely.

1.7. Mining the Elements of Noise Characteristic. For the description of the EME, we generally ignore the specific signal and pay more attention to the overall description. The analysis of noise is an important part of electromagnetic environment description. This part contains 2 contents: constructing noise data set through preprocessing and feature information mining.

As mentioned earlier, this article chooses a rougher resolution in the noise analysis part: time interval 37.5 minutes; frequency interval 6 MHz. The reason for this choice is that the noise changes relatively flat in the time domain and frequency domain, and at the same time, this resolution makes it easier for us to extract noise. The time-frequency distribution of noise is shown in Figure 7.

The time-frequency distribution of the noise spectrum signal can be further processed to mine characteristic information. First, pay attention to the distribution of noise in the time domain and frequency domain, and visualize them separately. And use intelligent sorting algorithms to search for feature points, such as bubble sorting, selection sorting, or insertion sorting. The result is shown in Figure 8.

The periodicity of the change of the noise signal is more obvious. Specifically, the noise floor amplitude during the day is relatively large, while the noise floor amplitude at night is relatively small, which has different effects on useful signals. Due to this, there are more frequency-using devices during the day, such as the occurrence of spectrum leakage, which will increase the noise floor level; the opposite is true

at night, making the noise floor signal level lower than that during the day. The noise signal is related to the use of frequency equipment, and there is no obvious change rule in the frequency domain.

In order to further dig out the noise information, the high-order spectrum method is used to obtain the high-order features of the noise signal. The third-order and fourth-order spectrum formulas of vector x are as follows:

$$B_x(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3x}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)},$$

$$T_x(\omega_1, \omega_2, \omega_3) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} C_{4x}(\tau_1, \tau_2, \tau_3) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2 + \omega_3 \tau_3)},$$
(3)

where the k -th order cumulant of vector x is as follows:

$$C_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \text{cum}\{x(n), x(n + \tau_1), \dots, x(n + \tau_{k-1})\}.$$
(4)

From this, the statistical characteristics, variance, skewness, and kurtosis of the noise signal can be extracted as follows:

$$\sigma_x^2 = \frac{1}{N} \sum_{k=1}^{N_x} [x(k) - \bar{x}]^2,$$

$$\gamma_x = \frac{1}{\sigma_x^3 N_x} \sum_{k=1}^{N_x} [x(k) - \bar{x}]^3,$$

$$K_x = \frac{1}{\sigma_x^4 N_x} \sum_{k=1}^{N_x} [x(k) - \bar{x}]^4.$$
(5)

Based on the above formula, we can obtain various statistical characteristics of the noise signal to model the structure of the noise element.

1.8. Mining the Elements of Spectrum Utilization. In this article, we measure spectrum utilization by calculating occupancy. First, quantify the power spectral density data. It is found through research that the power spectral density in the data set is relatively evenly distributed between -80 and -120 dBm/200 kHz. Therefore, uniform quantization is used

FIGURE 6: Interquartile range algorithm anomaly detection box plot.

FIGURE 7: The time-frequency distribution of noise.

to quantify; it selects the maximum and minimum values of the global data as the quantization boundary and divides the middle data into multiple segments equally.

Then, we judge whether the frequency point is occupied or idle according to the quantized result. With reference to some previous experiences, choose the boundary of two quantized segments as the threshold value; the specific value is about -114 dBm/200 kHz. The frequency points above the threshold are considered to be occupied, and frequency points below the threshold are considered to be idle. The formula is as follows:

$$CS = \begin{cases} 1, & P_c \geq P_0, \\ 0, & P_c < P_0, \end{cases} \quad (6)$$

where CS is the channel state, P_c is the power spectral density, and P_0 is the threshold.

FIGURE 8: The distribution of noise in the time domain and frequency domain.

1.8.1. Considering the Calculation of Occupancy. Two types of occupancy are involved in this article: channel occupancy and frequency band occupancy (time domain occupancy and frequency domain occupancy). The channel occupancy refers to the ratio of the duration of the signal exceeding the threshold level of the receiver to the total test time during a scan of the monitoring frequency band. The calculation formula is as follows:

$$\text{FCO} = \frac{T_0}{T}, \quad (7)$$

where FCO is the channel occupancy, T_0 is the duration of the signal exceeding the threshold level of the receiver, that is, the time the channel is occupied, and T is the total monitoring time.

The channel occupancy of two different frequency bands at the NE location is shown in Figure 9. The spectrum utilization of different service bands is different. The utilization of other frequency bands has obvious tidal effect, with a great difference between day and night. However, the utilization rate of other frequency bands is relatively stable.

Frequency band occupancy refers to the ratio of the number of occupied channels to the total number of channels in the frequency band during a scan of the monitoring frequency band. The calculation formula is as follows:

$$\text{FBO} = \frac{N_0}{N}, \quad (8)$$

where FBO is the frequency band occupancy, N_0 is the number of occupied channels, and N is the total number of channels.

The frequency occupation of three locations in the GSM1800d band (1820-1875 MHz) is shown in Figure 10. It can be seen that the frequency utilization between different cities is different. For example, there are still some idle channels in Aachen, while Maastricht's channel is already very tight, which confirms that the frequency utilizations are different between cities.

On the basis of the above research, some parameters of the city electromagnetic environment are obtained, and the shallow excavation of the electromagnetic environment is completed.

1.9. Correlation Analysis of Elements. Spectrum correlation is divided into positive correlation and negative correlation. Positive correlation means that when one channel is busy (idle), the other channel is also in the busy (idle) state; negative correlation is opposite, when one channel is busy (idle), and the other channel is in the idle (busy) state. We use channel correlation factor (CCF) to express the correlation degree between two channels. Its calculation method is shown in Equation (9), where M is the total number of channels, C_i^m is the occupied state of channel i in time slot m , $I\{A\}$ is the index function, and $I = 1$ when A is true.

$$\text{CCF} = \frac{\sum_{m=1}^M I\{C_i^m = C_j^m\}}{\sum_{m=1}^M I\{C_i^m = C_j^m\} + \sum_{m=1}^M I\{C_i^m \neq C_j^m\}}. \quad (9)$$

First, study the results of channel correlation factor in time domain. The channel correlation factor of the two frequency bands at the NE location is shown in Figure 11. Obviously, the overall similarity of ISM band is high, while the similarity between day and night of TV band is low. This result is the same as that in Figure 9. At the same time, the evolution of electromagnetic spectrum is closely related to human activities.

Next, study the correlation of evolution situation between frequency points. In addition to the above channel correlation factor, we also use a complex network method to show the correlation between frequency points more intuitively by building a relational network model.

Complex network is a network with very complex structure. A typical network is composed of nodes and adjacent edges between nodes. Nodes usually represent individuals with practical significance, and adjacent edges represent the connectivity between nodes. When the theory of network is

FIGURE 9: Comparison of channel occupancy in different frequency bands.

FIGURE 10: Comparison of frequency occupancy in different locations.

used to study problems, we usually choose the network with different characteristics according to different actual situations. The degree of a node refers to the number of edges connected to the node itself. The complex network is concerned about the structure of individual interaction in the system, and it is a way to understand the nature and function of complex system, which is more suitable for real system modeling. Therefore, a complex network is established to study the measurement of frequency similarity. The frequency points

are regarded as the nodes in the network, and the relative relationship between the frequency points in the situation evolution is regarded as the connection or not relationship between the nodes in the network. This paper mainly studies the characteristics of the spectrum situation evolution in the frequency domain. The state sequence of frequency point evolution in a period of time reflects the attributes of frequency point. The processing of frequency point sequence is no longer only focused on the sequence itself but on the

FIGURE 11: Comparison of time domain channel correlation factor between different frequency bands.

correlation between sequences. Because the physical meaning of similarity measure index is distance, the larger the index value is, the weaker the similarity is. When the similarity index does not exceed the decision threshold, the elements in the corresponding adjacency matrix are set to 1; otherwise, they are set to 0. For undirected graph, diagonal element of adjacency matrix must be 0; other elements in matrix only have two values of 0 and 1.

$$\begin{cases} \text{if } \text{corr_ED}s_{ij} \leq \delta, w_{ij} = 1 \\ \text{if } \text{corr_ED}s_{ij} > \delta, w_{ij} = 0, \end{cases} \quad (i \neq j), \quad (10)$$

where $\text{corr_ED}s_{ij}$ is the similarity value of frequency i and frequency j calculated based on Euclidean distance method, which is between 0 and 1; δ is the adjustable decision threshold; w_{ij} is the adjacency matrix element of constructing relational network.

According to the theory of complex networks, choose the appropriate decision threshold δ to make the degree of nodes in the network obey the power law distribution. At this time, the network presents scale-free characteristics.

As shown in Figure 12, for the data of ISM frequency band at the NE location, channel correlation factors and complex networks are used to describe the correlation of evolution among frequency points. Gephi software is used to draw the network topology. The nodes of the network are represented by a labeled circle. The label is the serial number of the frequency point within the frequency band involved, and the connection between the nodes represents the similarity between the two frequency points. The figure shows that the similarity of frequency point evolution is universal. Most frequency points are similar to their adjacent frequency points. This conclusion accords with the general cognition of people. However, the similarity between different frequency bands or different cities (which can be represented by the mean of channel correlation factors or the clustering degree in the complex network) is different.

1.10. Interval Number Conversion of EME Portrait Elements. The elements of the EME portrait have different characteristics for different frequency bands or different areas, and these

features are the key to describing the EME. With the deepening of data mining, the types and numbers of EME elements will continue to increase, and these elements must be processed consistently before a fusion representation model can be established. As we all know, although consistent processing can damage the original data, the value brought by the fusion characterization is far greater than the value of the original element alone. Therefore, in this section, we try to extract these features by means of interval number conversion, which lays a foundation for the construction of the final EME portrait.

Constructing fuzzy function is an important way to realize interval number transformation [17]. First, give the final domain U , any mapping μ_A from U to the closed interval of $[0, 1]$:

$$\begin{aligned} \mu_A : U &\rightarrow [0, 1], \\ u &\rightarrow \mu_A(u). \end{aligned} \quad (11)$$

A fuzzy subset of U can be determined, called u . The membership degree of fuzzy subset A can also be denoted as $A(u)$. Correctly determining the membership function is the key to using fuzzy set theory to solve practical problems. The membership function is a quantitative description of fuzzy concepts. The fuzzy concepts we have encountered are numerous, but the membership function of the fuzzy set that accurately reflects the fuzzy concepts cannot find a unified model. In this article, we use trapezoid to construct the membership function.

Next, we consider how to use the existing electromagnetic environment elements to construct membership functions. First, we select a series of factors that vary with time, such as field strength, noise energy, time domain occupancy, frequency domain occupancy, time domain CCF, and frequency domain CCF, and try to use these elements to construct membership functions. However, through research, we find that these elements themselves usually have a wide range of values, and the time variation is very strong. However, we find that their distribution rules are relatively stable, so we finally use the statistical characteristics of these elements (mean, peak, and valley) to construct fuzzy functions, and get good results. This process is shown in Figure 13.

FIGURE 12: The correlation of evolution state between frequency points.

FIGURE 13: Membership function model of the mean value of noise power.

The result is that we transform the EME elements of multiple domains into the vector form of interval numbers. In the next section, we will discuss how to fuse these vectors to construct an EME portrait.

1.11. Integration of Multimodal EME Portrait Elements. The fusion method adopted in this paper is to use the high-dimensional tensor model to fuse the elements of the multidomain EME portrait. Since the three-dimensional tensor is easy to display, take the three-dimensional tensor as an example; the three dimensions X , Y , and Z are represented as noise characteristic, spectrum utilization, and elements correlation, respectively. The fusion function is as follows:

$$h_m = \begin{bmatrix} h_x \\ 1 \end{bmatrix} \otimes \begin{bmatrix} h_y \\ 1 \end{bmatrix} \otimes \begin{bmatrix} h_z \\ 1 \end{bmatrix}. \quad (12)$$

The fusion result h obtained by Equation (12) is shown in Figure 14. Here, the submatrix $h_x \otimes h_y$ represents the fusion feature of noise characteristic and spectrum utilization; the submatrix $h_x \otimes h_z$ represents the fusion feature of noise characteristic and elements correlation; the submatrix $h_z \otimes h_y$ represents the fusion feature of elements correlation and spectrum utilization.

The tensor-based fusion model can fuse more than 3 domain elements. The fusion of elements in different domains will not only break the traditional spectrum analysis of only time and frequency analysis but also truly realize the vision of multidomain fusion EME portraits.

1.12. Future Application View of EME Portrait. The EME portrait is aimed at establishing a model that integrates data services, business applications, and visualization. In the future, facing more and more IoT devices in smart cities, the application of EME portrait in spectrum management

FIGURE 14: Multidomain elements fusion model based on tensor.

will become more and more important. This section will introduce two expected future applications: rapid response based on EME portrait library and long-distance transfer learning based on EME portrait model.

1.13. Rapid Response Based on EME Portrait Library. Since EME portraits are composed of some simplified elements extracted, EME portraits are more convenient to store and retrieve than big data. In the future, with the accumulation of different environment portraits, it is very necessary to build an EME portrait library. When electromagnetic equipment is facing a new environment or unexpected unknown event, it can quickly match the historical event or environment in the portrait library through search technology and strive for the best time to resolve the emergency.

1.14. Long-Distance Transfer Learning Based on EME Portrait Model. In general, each electromagnetic spectrum monitoring station is independent of each other. Therefore, it is difficult to perform correlation analysis on the collected data and the environment of multiple stations on a large time scale and large spatial span. The EME portrait model obtains a unified representation of each EME by determining the mining algorithm flow and the interval number conversion of different domain elements, which provides a great help for long-distance correlation analysis and transfer learning.

1.15. Guide New Devices to Better Adapt Based on EME Portraits in IoT Scenarios. In the future, the data of wireless devices will increase rapidly, and the IoT scenario will be very complicated [29]. When a new device is connected to a new scenario, it is very critical to be able to quickly adapt to the environment and find a suitable channel and communication method. Through the description of the environment in historical time, the EME portrait allows the new device to quickly understand its current environment and predict the future changes of the environment. In this way, the portrait helps the new device quickly adapt to the environment.

2. Conclusions

In this paper, we use spectrum big data to construct an EME portrait. Although the EME is invisible and the spectrum resources are very abstract, the EME portrait describes the environment from multiangles, multilevels, multigranularity, multidimension, and multiscale. The portrait model uses big data mining algorithms, interval number conversion, and fusion methods to achieve a unified multidomain representation of EME characteristics. The subsequent development of the EME portrait can not only provide help for the efficient use of spectrum resources but also can quickly correlate abnormal events, maintain spectrum order, and protect the electromagnetic ecological environment.

Data Availability

In this paper, the spectrum data used in the experiment comes from the public spectrum data set of Aachen University in Germany. The data acquisition URL is <https://download.mobnets.rwth-aachen.de/index.php?id=registration>.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the Key Laboratory of Advanced Marine Communication and Information Technology, Ministry of Industry and Information Technology, Harbin Engineering University, Harbin, China.

References

- [1] Y. Lin, H. Zhao, X. Ma, Y. Tu, and M. Wang, "Adversarial attacks in modulation recognition with convolutional neural networks," *IEEE Transactions on Reliability*, vol. 70, no. 1, pp. 389–401, 2020.
- [2] Y. Zhao, H. Jiang, Q. Chen et al., "Preserving minority structures in graph sampling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1698–1708, 2021.
- [3] G. Gui, M. Liu, F. Tang, N. Kato, and F. Adachi, "6G: opening new horizons for integration of comfort, security, and intelligence," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 126–132, 2020.
- [4] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5703–5706, 2020.
- [5] X. Liu and X. Zhang, "Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 5971–5980, 2019.
- [6] X. Liu and X. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5379–5388, 2020.
- [7] W. Lu, Y. Gong, X. Liu, J. Wu, and H. Peng, "Collaborative energy and information transfer in green wireless sensor

- networks for smart cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1585–1593, 2018.
- [8] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *The Journal of Supercomputing*, vol. 75, no. 6, pp. 3010–3027, 2019.
 - [9] J. Yu, A. Hu, G. Li, and L. Peng, "A multi-sampling convolutional neural network-based RF fingerprinting approach for low-power devices," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 1–6, Paris, France, 2019.
 - [10] Z. Zhang, J. Chang, M. Chai, and N. Tang, "Specific emitter identification based on power amplifier," *International Journal of Performability Engineering*, vol. 15, no. 3, pp. 1005–1013, 2019.
 - [11] Y. Tu, Y. Lin, C. Hou, and S. Mao, "Complex-valued networks for automatic modulation classification," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10085–10089, 2020.
 - [12] Z. Ji and K. J. Liu, "Cognitive radios for dynamic spectrum access - dynamic spectrum sharing: a game theoretical overview," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 88–94, 2007.
 - [13] G. Ding, F. Wu, Q. Wu et al., "Robust online spectrum prediction with incomplete and corrupted historical observations," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8022–8036, 2017.
 - [14] S. Liu and G. Sun, "Analysis of the concept and effects of complex electromagnetic environment," *Journal of the Academy of Equipment Command & Technology*, vol. 19, no. 1, pp. 1–5, 2008.
 - [15] R. Streich, "3D finite-difference frequency-domain modeling of controlled-source electromagnetic data: direct solution and optimization for high accuracy," *Geophysics*, vol. 74, no. 5, pp. F95–F105, 2009.
 - [16] C. Hou, X. Zhang, and X. Chen, "Electromagnetic signal feature fusion and recognition based on multi-modal deep learning," *International Journal of Performability Engineering*, vol. 16, no. 6, pp. 941–949, 2020.
 - [17] K. Przystupa, I. Vasyukivskyi, V. Ishchenko, V. Pohrebennyk, and O. Kochan, "Electromagnetic pollution: case study of energy transmission lines and radio transmission equipment," *Ecosystems*, vol. 1, no. 3, pp. 52–55, 2020.
 - [18] E. Haber and C. Schwarzbach, "Parallel inversion of large-scale airborne time-domain electromagnetic data with multiple OcTree meshes," *Inverse Problems*, vol. 30, no. 5, article 055011, 2014.
 - [19] Y. Tu, Y. Lin, J. Wang, and J. U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *CMC-Computers Materials & Continua*, vol. 55, no. 2, pp. 243–254, 2018.
 - [20] Y. Zhao, X. Luo, X. Lin et al., "Visual analytics for electromagnetic situation awareness in radio monitoring and management," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 590–600, 2020.
 - [21] N. Correa, T. Adali, Y.-O. Li, and V. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.
 - [22] G. Ding, Y. Jiao, J. Wang et al., "Spectrum inference in cognitive radio networks: algorithms and applications," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 150–182, 2017.
 - [23] Y. Wu, L. Zhang, L. Zhang, and W. C. Zhang, "Survey on electromagnetic environment simulation and visualization," *Journal of System Simulation*, vol. 21, no. 20, pp. 6332–6338, 2009.
 - [24] L. Zhang, Y. W. Shen, Y. N. Wu, L. Mu, H. B. Wang, and Y. D. Lv, "Three-dimensional simulation of complex electromagnetic environment and visualization system," *Journal of China Academy of Electronics and Information Technology*, vol. 8, no. 2, pp. 111–118, 2013.
 - [25] A. Ryabov and P. Baum, "Electron microscopy of electromagnetic waveforms," *Science*, vol. 353, no. 6297, pp. 374–377, 2016.
 - [26] D. Bolchini and J. Mylopoulos, "From task-oriented to goal-oriented web requirements analysis," in *Proceedings of the Fourth International Conference on Web Information Systems Engineering (WISE 2003)*, pp. 166–175, Rome, Italy, 2003.
 - [27] X. Liu, R. Shi, B. Hee, and M. Chen, "Detection on abnormal usage of spectrum by electromagnetic data mining," in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, pp. 182–187, Suzhou, China, 2019.
 - [28] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 34–46, 2021.
 - [29] X. Liu, X. B. Zhai, W. Lu, and C. Wu, "QoS-guarantee resource allocation for multibeam satellite industrial Internet of Things with NOMA," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2052–2061, 2021.