Research on a New Method of Signal Data Recognition and Acquisition

Kai Zhou [a], Xueling Zhang and Jiajia Chen

The State Radio Monitoring Center, Beijing 100000, China.

[a] zhoukai@srrc.org.cn

Abstract. With the development of 5G communication, radio monitoring data will have an explosive growth. Due to the large size, huge variety and strong specialty of radio data, data acquisition is becoming a concerned difficulty. The key links in radio service data recognition and acquisition are studied, and a new method of data acquisition is proposed. the improved BP neural network algorithm is used to improve the accuracy of data identification, enhance the generalization ability of the system, and realize the automatic identification and storage of data. Experimental results show that the new method can reduce the system complexity and has high efficiency.

Keywords: Radio, data acquisition, neural network, monitoring.

1. Introduction

Radio data acquisition is of great significance for radio management [1]. Because it can provide multi-angle signal parameters and trace signal variation, radio data acquisition is playing an increasingly important role in radio monitoring. The acquisition system of different data formats based on heterogeneous monitoring equipment is the basis and one of the difficulties. Data acquisition is difficult mainly due to data heterogeneity: the data formats of different monitoring equipment are not unified and the structures of different radio services are different. Heterogeneous data brings many difficulties for data collection and application, which is also not conducive to data mining and analysis in the later stage [2].

At present the data acquisition method of monitoring equipment is mainly through the API interface, which is opened by monitoring equipment [3]. The API inter face controls the monitoring equipment and store the data returned by the equipment after analyzing it [4]. With equipment exclusive controlled, this kind of method has poor generalization ability, high development cost and long operation cycle. Its implementation theory is interpreted as in figure 1.
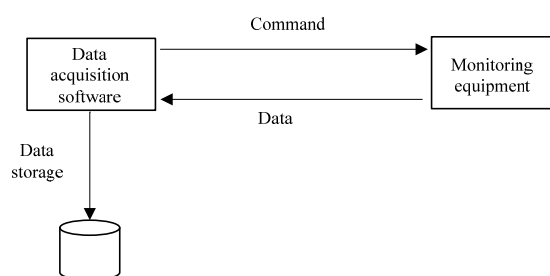

Fig. 1 Data acquisition implementation based on API or command set

2. Network Data Identification and Acquisition

In reference [5], neural network was introduced into network intrusion detection, and network anomalies and viruses were analyzed and detected for data characteristics. Enlightened by this idea, in addition to the detection of data features in network data, we believe that the data acquisition system can be realized by further analyzing and automatically storing data of same characteristic in a certain type or several types. It is also theoretically feasible to bring data recognition based on neural network into data acquisition work:

1. The data acquisition based on network data flow can be actually attributed to the recognition and decoding of data. The fuzzy mapping relationship between neural network and data is suitable for solving such problems.

2. Neural network has the ability of autonomous learning and automatic recognition, and has strong generalization ability for different equipment and data types.

3. The neural network is simple in construction, widely used as a classification or recognition model in various fields, which has a good effect.
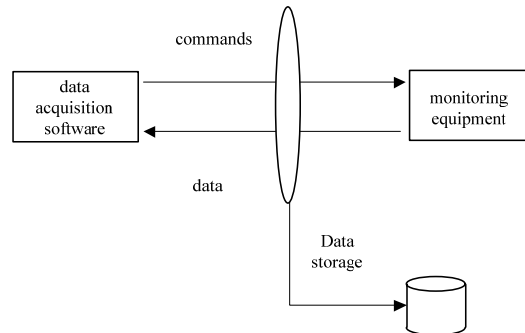

Fig. 2 Data acquisition implementation based on network analysis

Based on the view points above, we set data packets transmitted over the network as the unit, and focus on the recognition of radio service data. The specific steps are as follow: set the network card into a hybrid mode, capture the network data packets, classify the data packets by their features, take the classified data sets as the input of the neural network, recognize the data sets automatically, and collect and store involved radio services data automatically. In this method monitoring data on the network is collected automatically, which won't affect the normal operation of monitoring. The method has the advantage of large data volume and complete collection type. The implementation method is shown in figure 2:

3. Data Acquisition and Recognition

3.1 Data Acquisition Procedure

Data capture and data recognition are the key technology of data acquisition system. In our method, Winpcap technology is used to capture data packets from network adaptor. Data acquisition and recognition mainly includes the following steps:

Data normalization

Convert network data samples to ASCII format. The network transmission data is based on binary transmission. The monitoring data generated by some radio equipment is encrypted, so it would be easy for data classification and analysis when the data is converted into ASCII format.

Data classification and feature extraction

Data characteristics refers to data package size, transmission time, protocol type, etc. Data in the monitoring process can be divided into service metadata, service parameter data and service main data. Service metadata mainly refers to the description of data, including data packet source address, data destination address, protocol name, transmission time, data amount, network connection status and so on. Service parameter data includes monitoring frequency, bandwidth, parameter type, identification code, etc. Service main data generally has a large volume and a long transmission time, such as the original level data, audio data and IQ data.

Data recognition

The data recognizer applies BP neural network [6], which is a model based on error back propagation and is the most widely used neural network model at present. One or more hidden layers are added between the input layer and the output layer, and the gradient search algorithm is adopted to make the error mean square of the actual output and the expected output gradually close to the minimum. The basic model of BP neural network is shown in the figure 3.
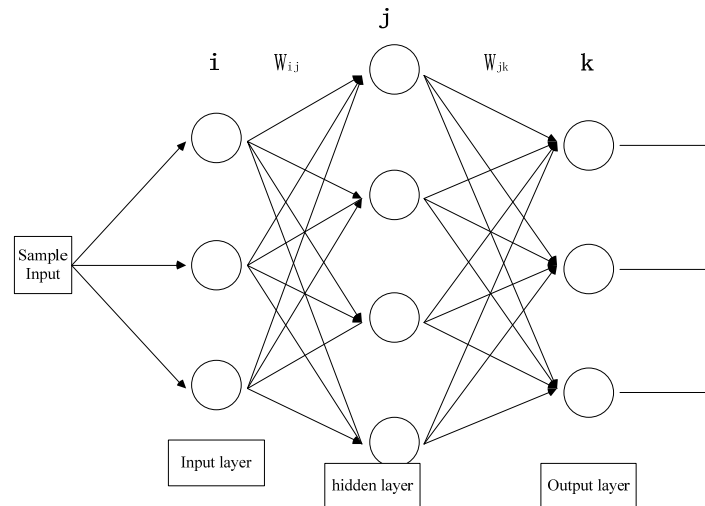
Fig. 3 BP neural network model

3.2 Design of Recognizer

In BP neural network, input layer M-dimension X = (X1, X2, ...) Xm), hidden layer S = (S1, S2Sn), the output layer Y = (Y1, Y2, Y3Yk). The connection weight of the input layer to the hidden layer is w_{ij}, the threshold is θ_j; the connection weight of the hidden layer to the output layer is w_{jk}, the threshold is θ_k.

Both the hidden layer and the output layer are the weighted sum of the neurons in their previous layer, that is:

$$S_j = \sum_i X_i * w_{ij} \tag{1}$$

$$Y_k = \sum_i S_j * w_{jk} \tag{2}$$

S_j Is the j-th neuron of the hidden layer, X_i Is the i-th neuron of the input layer, w_{ij} Is the connection weight of the i-th input layer neuron and the j-th hidden layer neuron .Similarly, Y_k is the k-th neuron in the output layer, S_j is the j-th neuron in the hidden layer, and w_{jk} is the connection weight of the j-th hidden layer neuron and the k-th output layer neuron.

The BP neural network design model is as follow:

1. Select the number of hidden layers. According to Kolmogoro theorem, three-layer network can complete data mapping of any dimension, which means only one layer of hidden layer is needed. As too many layers will increase the complexity of the algorithm and lead to "overfitting", the hidden layer is selected to have only one layer.

2. Determine the number of input and output neurons. According to the data set categories extracted from the above features, the number of units in the input layer is determined to be 12, while the number of units in the output layer is determined to be 9.

3. Determine the number of neurons in the hidden layer. According to the two methods proposed in the reference: 1) the number of hidden layer units is half of the total number of input and output units; 2) the number of hidden layer units is the square root of the number of input and output units. Considering the cross-test results, the number of hidden layer units is determined to be 11.

4. Initializethe weights. Small random numbers are generally selected to ensure the system convergence. Here we select random numbers in (-1, 1).

5. In the learning process, error value between calculating output and actual output is defined as:

$$E = (Y_k - D_k)^2 / 2 \tag{3}$$

3.3 Improvement of BP Neural Network

The classification and recognition of service data can be realized by inputting the feature set into the neural network recognizer.However, the basic BP algorithm is make the result reach the local minimum.Insufficient training sample will cause the phenomenon of "underfitting".In fact, it's

difficult for the training sample to cover all data formats and data forms. When the input characteristics change significantly, the BP model enters the flat area of the error curved surface, resulting in the recognition error and making the system difficult to converge.

In order to improve the accuracy and generalization ability of data extraction and recognition, we proposed an improved BP neural network algorithm. In the case of small sample size, the system model is self-corrected by using the weak learning algorithm.The radio monitoring data has a specific service meaning in itself and a relatively stable range, which objectively provides a yardstick for the correction of the system.On the one hand itcan reduce the machine learning time, on the other hand it can reduce the complexity of data recognizer.

The improved BP neural network algorithm in this paper defines the range before data identification, in other words, to define the service meaning and range of the output data. Service metadata and service parameter data like protocol name, connection status, monitoring frequency, bandwidth, and parameter types have certain output options, such as transmission time, data volume size and electric level value; they can also be determined in a certain range.
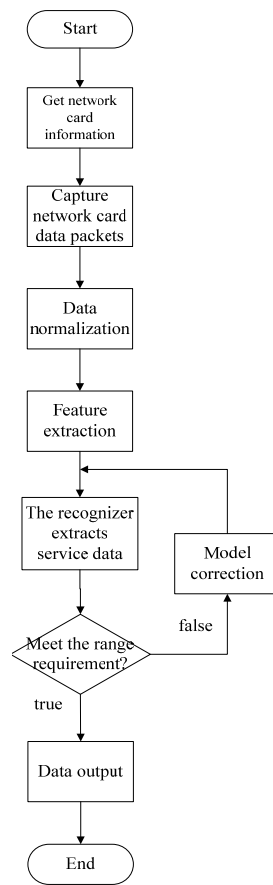
Fig. 4 Data acquisition and recognition process

Therefore, the output data is sentenced within the range to achieve the purpose of automatic system correction. If the output data does not meet the setting range, adjust the weight, and then re-identify the collection until the recognition result falls within the range. Data acquisition and recognition process is as Fig 4.

The weight determination and correction of BP model during training follow the steps below:

Calculate the error of the j-th output unit when the single input value changes, that is, the difference between the actual output and the expected output.

$$E_{S_j} = Y_j - d_j \qquad (4)$$

Calculate the error changing rate of the j-th output unit when the values of multiple input units change.

$$E_{X_j} = Y_j \, (\, 1 - E_{S_j} \,) \, \star \, E_{S_j} \tag{5}$$

Calculate the error changing rate of each output unit when the single weight changes.

$$E_{w_{ij}} = Y_i E_{X_i} \tag{6}$$

$$E_{w_{jk}} = Y_j E_{S_j} \tag{7}$$

In the actual identification phase of BP model, if thecalculated output value meets the given range, maintain the connection weight of w_{ij} and w_{jk} remains; if not, compare the error with the previous output error.

$$\Delta w_{jk}^s = \alpha * \Delta w_{jk}^{s-1} + \eta * \sum_{n=d}^{s-1} E_{w_{jk}} * (d_j^n - Y_j^n) \tag{8}$$

$$\Delta w_{ij}^s = \alpha * \Delta w_{ij}^{s-1} + \eta * \sum_{n=d}^{s-1} (d_j^n / \Delta w_{jk}^s - X_j * w_{ij}^n) * E_{w_{ij}} \tag{9}$$

In the formula above, α is the correction constant, which is in 0.7-0.9. The smaller α is, the bigger the correction will be.η is the adjustment constant, which is in 0.1-0.3. The biggerη is, the bigger the correction will be.The adjustment factor is the sum of the error change rates of the previous times that do not meet the threshold range. Current connection weight can be obtained by adding the previousconnection weight to the adjustment factor.In the adjustment process, the connection weights of the hidden layer and the output layer are calculated first according to the output results, and then the connection weights between the input layer and the hidden layer are adjusted according to the same principle.

The adjustment factor makes use of the memory of the change of serviceparameters and compares the change trend of this error with that of the previous error to determine the adjustment quantity of connection weight.Thus, it prevents the BP model entering the flat area of the error curved surface,enabling the system to self-correct.

4. Algorithm Implementation and Platform Construction

Based on our above-mentionedmethod, we developed a data acquisition system using C# programming language, to collect data from monitoring receiver.

In the process, the monitoring parameters are changed for data collection, data packets are classified and analyzed, and the standard formatted monitoring data is finally stored in the file.

In this experiment, improved BP neural network algorithm is compared with BP neural networkalgorithm. When the training number reaches 2000, an example of under sampling, the recognition error rate is significantly lower than ofBP neural network, and the system tends to converge faster.When the number of training is more than 6000,the error rate of the improvedBP neural network algorithm reaches the minimum.When the number of training is greater than 10000, "overfitting" phenomenon also occurs, but the error rate is also significantly lower than usingBP neural networkalgorithm.

Table 1. Performance comparison of BP neural network algorithm and improved BP neural network algorithm

	BP neural network algorithm		Improved BP neural network algorithm	
	Training number	Error number (10000times)	Training number	Error number(10000times)
1	1000	128	1000	94
2	2000	97	2000	62
3	5000	47	5000	31
4	6000	28	6000	19
5	10000	39	10000	25
6	20000	44	20000	24

We conduct a comparison and analysis of the performance of improved BP neural network algorithm, k-means clustering algorithm [7] and BP neural network algorithm. The experiment results show that training time is reduced in improved BP neural networkalgorithm, the data amount is slightly lower but the accuracy of collected data was obviously higher than using the other two algorithms. The performance comparison of these algorithms is shown in table 2.

Table 2. Performance comparison of three different algorithms

	Training number	Training time (S)	Collected data	Error times
k-means clustering algorithm	6000	9	5324.8MB	1676
BP neural networkalgorithm	6000	21	5011.8MB	869
Improved BP neural networkalgorithm	4000	16	4924.8MB	226

Several experiments proved that our data acquisition system is easy to develop, and has strong adaptive error-correcting ability and strong generalization ability for different data formats. It can accurately recognize and acquis the radio service data in the network.

5. Summary

In this paper, a new data acquisition method based on network data analysis is proposed to identify and extract the data transmitted in the monitoring network. An improved BP neural network algorithm is adopted in our data acquisition method to shorten the learning time, greatly reduce the error rate of data recognition and improve the generalization ability of the system.This method proposed a new idea to solve the identification and acquisition of heterogeneous data. The data acquisition system developed on the data acquisition method verified its feasibility and has a high practical value.

Acknowledgements

References

[1]. FukuoLiu, Wei Li, HuanjiaPeng. Research on big data prediction and storage strategy for radio management monitoring [J]. China radio.2016 (7): 62-64.

[2]. Fangli Ma. Radio monitoring in the era of big data [J]. China information industry network.2014 (4): 55-58.

[3]. Zhongzhong Xu. Taizhou_data storage innovation in the era of big data [J]. Shanghai informatization.2014 (8): 30-32.

[4]. Wei Wei. Design and implementation of the acquisition and analysis system for radio monitoring data[D]. Harbin Institute of Technology. Master's thesis.2011 (7).

[5]. Shouye Zhu. Intrusion detection based on BP neural network and Bagging algorithm. Computer engineering and applications [J].2009, 45(18) :123-125.

[6]. Jingxin Wang, Kui Dai, HuiSong, Zhiying Wang. Intrusion detection system based on neural network. Computer engineering and science[J].2003 (25): 28-30.

[7]. Qi Chen, Shiqiong Song, PengfeiDeng. Research on communication signal reception and data acquisition system based on software radio technology [J]. Ship science and technology. 2010,32 (5): 59-63.