Research of Machine Learning Algorithm for Broadcasting Spectrum Signal Processing

Jinyu Sun¹, Qun Zhou², Jianping Shang³, Shuai Sun (Corresponding Author)³

¹College of Automation, Wuhan University of Technology, Wuhan, China
²Radio Management Office of Hubei, Wuhan, China
³Huawei Technology Co., Ltd., Wuhan Institute, Wuhan, China
1219863633@qq.com(Jinyu Sun), 1072086501@qq.com (Corresponding Author)

Abstract. FM broadcasting is mainly used to transmit sound and other signals in the form of wireless transmission. This thesis is based on a project for a radio department, and mainly focuses on the broadcasting frequency band, uses radio monitoring equipment to scan the spectrum signal of the broadcasting frequency band, and performs the data preprocessing, signal feature extraction and classification processing of the radio frequency spectrum signal on the extracted spectrum signal, so as to extract abnormal signals such as pseudo base stations, black radio, cheats in exams, etc. Firstly, relevant pre-processing is performed on the spectrum information of the broadcast frequency band. Through the improved K-Means algorithm, the original sample data including the glitch signal is eliminated, and the original signal is processed through the wavelet analysis of the spectrum signal. The original signal was decomposed by wavelet and wavelet reconstruction, so as to achieve the purpose of denoising the original signal. Secondly, based on the analysis of signal characteristics and the comparison of a large number of spectrum signals, a method for extracting individual features of spectrum signals is summarized. Finally, the grey relational degree cluster Analysis is used to extract the features of the spectrum signal, which provides a certain basis for the subsequent classification algorithm.

1. Introduction

At present, the radio management monitoring department collects a large amount of radio frequency spectrum data of the broadcasting band, and the system displays the data and presents it to the user, mostly in the form of spectrum diagram, waterfall diagram and the like [1]. However, people may need to analyze massive spectrum data based on historical records, geographic locations, and various environmental parameters. In this case, the theory of machine learning and data mining needs to be used to solve the problem. At the same time, with the development of the times, the accumulated spectrum data is also exponentially growing. The radio management department urgently needs more effective methods for the management and utilization of radio spectrum data in the broadcasting band. It is especially necessary to explore the machine learning algorithms for the signals in the broadcasting band.

On the basis of analyzing the radio spectrum monitoring and machine learning research, this paper applies the support vector machine and the over-limit learning machine algorithm to the monitoring and identification research of the radio frequency spectrum overall signal, and takes the spectrum information of the radio frequency band in the radio as the research object. Through the research on
the feature extraction and classification algorithms of data, an effective machine learning algorithm is proposed to process the radio spectrum data to prevent the resources from being illegally exploited for some exploration.

Due to the high dimensionality of the spectral feature data and the certain correlation of the feature data, the gray correlation degree clustering analysis is used to extract the effective feature data with less relevance as the input of the subsequent classifier. Then, the paper uses the over-limit learning machine ELM, support vector machine SVM and the combination algorithm of the two algorithms to classify the spectrum data, and test and analyze the actual samples, and obtain the model simulation effect, and the physical equipment monitoring. The history is consistent. Finally, the feature fusion of GBDT and LR is applied to the classification algorithm of spectrum signals, and corresponding simulation analysis is carried out.

2. Research on Data Preprocessing and Feature Extraction Algorithm

2.1. Research and Implementation of Spectrum Data Preprocessing Algorithm

As shown in Figure 1, the spectrum signal in one of the radio frequency bands is displayed. Near the 625th point of the original sample, a glitch signal with an amplitude of about -78 appears. This glitch signal is only in one or two times in a random time, this has a certain interference on signal feature extraction. In radio data monitoring, the inclusion of glitch signals is sporadic, although the probability of occurrence is low, once the glitch interference occurs, this will affect the analysis of data and the extraction of features. This section focuses on the original radio data of the monitoring, and proposes an algorithmic study of the spectral data of the spurs. Specifically, the improved K-mean algorithm is used for cluster analysis in different time dimensions, and the wavelet adaptive noise reduction algorithm is used to filter out Gaussian white noise in the spectrum signal.

The wavelet transform algorithm is used to decompose the original spectrum signal, then the wavelet after reconstruction is wavelet reconstructed, and finally the reconstructed spectrum signal is smoothed, and the spectrum after denoising as shown in Fig. 2 is obtained. The signal, as can be seen from the comparison results, has formed a significant contrast effect before and after denoising, indicating that the wavelet denoising algorithm has a certain effect on the spectrum preprocessing.

![Figure 1. Comparison of glitch signals with normal signal details.](image1)

![Figure 2. Comparison of the original signal and the spectral signal after wavelet transform.](image2)

2.2. Research and Implementation of Spectrum Feature Extraction Algorithm

Feature extraction mainly extracts the feature information describing the signal by analyzing the sample after the noise point is removed [2]. The spectrum characteristics of different frequency bands are analyzed, then the overall characteristics of the radio broadcast spectrum signals are studied, and 11 corresponding signal characteristics are proposed.

In the pre-processed radio broadcast spectrum data, some of the data are selected, and the statistical characteristics of the corresponding spectral signals are extracted, and some frequency domain statistical features such as variance, mean, peak, and deviation kurtosis are collected. Perform data
The original extracted feature information may be related to the spectrum signal, but some vectors have no strong correlation. Therefore, the principal feature vector of the spectrum signal is used to filter the original feature vector to select the feature vector with strong correlation. Purpose of data compression.

The basic flow of the principal component analysis algorithm is shown in Figure 3:

![Figure 3. Principal component analysis flow chart.](image)

3. Research on Single Classifier Algorithm for Spectrum Signal
The radio broadcast spectrum signal classifier is used to identify whether the broadcast signal is a normal broadcast signal or an abnormal signal attributed to the black broadcast. The quality of the classification is not only related to the data preprocessing and spectral feature extraction algorithms. And you need to choose the right classifier. Reasonable classifiers play a crucial role in the classification of spectral signals [3].

3.1. SVM classifier
The SVM support vector machine is used to classify the spectrum signals. The selection of parameters will have a great influence on the effect and quality of the classification. The corresponding parameters include the penalty factor C, the kernel function, etc. [4]. In the process of analyzing the spectral data, first select the appropriate kernel function for the data-related characteristics, and then select the penalty factor of the system model. The figure below is the parameters trained by the network cross-validation algorithm. It can be seen from Figure 4. When parameter c is selected as 0.43528 and g is selected as 1.3195, the classification accuracy can reach 92.1348%, which achieves the best results.

![Figure 4. SVC parameter selection result graph (contour map).](image)

3.2. ELM overrun learning machine
Applying the Overrun Learning Machine (ELM) algorithm to the classification of radio broadcast spectrum data requires two data processing processes: a sample training process and a sample testing process. In the sample training process, the spectrum information of the radio is first collected, and the original spectrum data is preprocessed, that is, the abnormal noise signal and the interference signal
are eliminated by the clustering algorithm, and the edge of the original data is preprocessed. Then, the pre-processed spectrum signal is subjected to corresponding feature extraction, so that the feature signal is used as the input data of the input layer of the over-limit learning machine.[5] The flow chart of the ELM algorithm is as follows:

![ELM algorithm flow](image)

**Figure 5. ELM algorithm flow.**

### 3.3. Comparison and summary of experimental results

In this experiment, for the broadcast frequency band data, the original spectrum data collected is subjected to relevant pre-processing operations, and the characteristics of normal and abnormal signals are extracted. Finally, 2000 normal spectrum signals are collected, and 1000 groups have abnormal broadcast signals. The raw data of the spectrum is divided into two groups, which are divided into training set and test set. The corresponding grouping situation is shown in Table 1:

| Grouping of training samples and test samples. |
|-----------------|-----------------|----------------|
|                   | Normal signal   | Abnormal signal |
| Training section  | 1600            | 800             |
| Testing section   | 400             | 200             |
| Total semaphore   | 2000            | 1000            |

This experiment is mainly to study the difference between the two classification algorithms in terms of the overall spectrum characteristics and PCA data dimension reduction in training time and test time and test accuracy, and further analysis of the test results. Table 2 shows the pros and cons of different feature extractions and different classifiers in terms of complexity and accuracy:

| Feature extraction | Classifier | Training Time (s) | Testing Time (s) | Classification result |
|--------------------|------------|-------------------|------------------|-----------------------|
| Spectrum overall characteristic signal | SVM       | 37.271            | 4.382            | 92.7                  |
| PCA feature extraction | ELM       | 6.549             | 1.678            | 93.3                  |
|                     | SVM       | 31.626            | 3.691            | 94.3                  |
|                     | ELM       | 4.396             | 1.098            | 95.1                  |

It can be seen from Table 2 that for the same classifier, different feature extractions will cause differences in training results. It can be seen from the results that the classification effect after PCA feature extraction is better than the characteristic signal of the spectrum overall, and training Both time and test time have been significantly reduced, and the classification effect has also been significantly improved. For the same feature extraction, the classification effect of the over-limit learning machine is better than that of the support vector machine, which has certain advantages. Finally, by selecting different samples to train the number, and performing a series of precision tests on the spectrum test samples, the sample trend shown in Figure 6 is compiled.
It can be seen from the figure that the SVM classification effect is better when the training samples are less, but with the increase of the number of training samples, the classification effect of the ELM classifier is better than that of the SVM classifier[6].

4. Research on Spectrum Signal Combinatorial Classifier Algorithm

Suppose the data set of the original training sample is \((x, y) = \{(x_1, y_1), ..., (x_n, y_n)\}\), where \(x_n\) is the original sample instance and \(y_n\) is the attribute tag data of each sample data \(x_n\). The combined classifier can be represented as \(D = \{D_1, D_2, ..., D_n\}\), where \(D_n\) represents the base classifier. Table 3 shows the classification results for the combined classifier for test sample \(x\).

| \(y_1\) | \(y_2\) | \(\cdots\) | \(y_n\) |
|---|---|---|---|
| \(D_1\) | \(p_{11}\) | \(p_{12}\) | \(\cdots\) | \(p_{1n}\) |
| \(D_2\) | \(p_{21}\) | \(p_{22}\) | \(\cdots\) | \(p_{2n}\) |
| \(\cdots\) | \(\cdots\) | \(\cdots\) | \(\cdots\) |
| \(D_m\) | \(p_{m1}\) | \(p_{m2}\) | \(\cdots\) | \(p_{mn}\) |

In the experiment, the combined classifiers designed in three combinations are denoted as \(C_1, C_2, C_3\), and two single classifiers, denoted as \(S_1\) and \(E_1\). The three base classifiers of the combination classifier \(C_1\) are all SVM classifiers, the three base classifiers of the combination classifier \(C_2\) are ELM classifiers, and the three base classifiers of the combination classifier \(C_3\) include an ELM classifier. And two SVM classifiers. \(S_1\) represents a single SVM classifier and \(E_1\) represents a single ELM classifier. Table 4 makes corresponding statistics on the results of the experiment according to different conditions. The specific results are as follows:

| Number of samples | \(C_1\) | \(C_2\) | \(C_3\) | \(S_1\) | \(E_1\) |
|---|---|---|---|---|---|
| 400 | 57.6% | 56.1% | 59.3% | 51.3% | 49.1% |
| 800 | 76.9% | 77.2% | 79.2% | 79.1% | 73.6% |
| 1000 | 82.1% | 85.0% | 86.1% | 81.3% | 84.7% |
| 1200 | 89.5% | 91.2% | 92.0% | 86.2% | 87.3% |
| 1600 | 93.3% | 94.1% | 95.9% | 88.6% | 91.9% |
| 2000 | 95.7% | 96.5% | 97.9% | 91.2% | 93.5% |

It can be found from Table 4 that under the premise of the lower training samples, the obtained training model has certain defects, and the accuracy of the model prediction is low. For the combined classifier \(C_1\), it can be known that when the number of sample training is 400 The accuracy of the classification is only 57.6%. However, with the increase of the number of training samples, the accuracy of the algorithm has been significantly improved. Therefore, in order to obtain the completed training model, it is necessary to train based on a large number of samples, thereby improving certain accuracy. Degree[7].

When selecting the number of different sample trainings, the corresponding result statistics are performed for different classification models. The accuracy rate is shown in Figure 7:
The experimental results show that the classification effect of the combined classifier based on SVM classifier is higher than that of the combined classifier based on ELM classifier, but the classification effect of hybrid combination classifier based on SVM classifier and ELM classifier is higher than that. The classification effect of the first two classifiers, the accuracy of the combined classifier based on the SVM classifier is about 94.1%, and the accuracy of the combined classifier based on the ELM classifier is about 93.1%, and the classification accuracy of the hybrid combination classifier is 94.7%. Thus, it is proved that the combined classifier has a high practical value in the classification processing of the broadcast spectrum data.

5. Conclusion

In this paper, the machine learning algorithm is applied to the monitoring and processing of broadcast spectrum signals in a radio management department, which improves the accuracy and timeliness of clustering identification of broadcast spectrum signals, reduces the cost of manual detection, and realizes the radio spectrum monitoring system. Artificial intelligence + has been a useful exploration. The SVM/ELM combined model for broadcast spectrum signal classification processing is proposed. The system is compared with the training models of different training samples and different combination classifiers. It is proved that the SVM/ELM combined classifier is efficient.

References

[1] K. Ellis, N. Serinken. Characteristics of Radio Transmitter Fingerprints[J]. Radio Science. 2001, 36. pp:585-597.
[2] Rongze Yang, Yi Liu. Multi-classifier selection integration method for abnormal data flow[J]. Computer Engineering and Applications, 2018, 54(02): 107-113.
[3] Yi Bao, Tao Wang, Guoyong Qiu. Research on Applicability of SVM Kernel Functions Used in Binary Classification[M]. Springer India:2014-06-15.0.
[4] Fei Wang, Xiaolin Qin, Liang Liu, et al. K-means Clustering Algorithm Based on Data Stream in Cloud Environment[J]. Computer Science, 2015,42(11):235-239+265.
[5] Jie Wang,Chenlong Guo. Wavelet kernel overrun learning machine classifier[J]. Microelectronics & Computer,2013,30(10):73-764+80.
[6] Zhaozhen Zhou. Optimization and application research of algorithm for over-limit learning machine [D]. Nanjing University of Information Science and Technology, 2016.
[7] Hee-Joong Kang. A framework for probabilistic combination of multiple classifiers at an abstract level[J]. Engineering Applications of Artificial Intelligence,1997,10(4).