Online Machine Learning System of Power Quality Information Technology Based on Big Data

Wei Wang*

The University of Queensland, Brisbane 4072, Australia

*Corresponding author: weiwang@uq.edu.au

Abstract. Power quality monitoring has the characteristics of large amount of collected data and high collection frequency. Long-term monitoring will form power quality big data, while conventional data storage and processing methods are difficult to effectively use big data and explore the value of big data. This paper proposes an online machine learning recognition and classification method for power quality disturbances based on inherent modal function singular value decomposition and least square support vector machines. Aiming at the characteristics of power quality data in distribution networks, this thesis proposes a power quality big data storage scheme based on Mongo DB + Redis, designs a power quality big data processing process, and builds an online machine learning system based on the Apache Spark power quality big data calculation framework.

Keywords: Big data, power quality, online machine learning, least squares support vector machine.

1. Introduction

The field of power quality research involves the sensitive and rapid detection of power quality disturbances, the classification and identification of complex power quality disturbances, the diagnosis and location of the cause of voltage sags, the optimal installation and arrangement of monitoring devices, the interaction mechanism between power quality and sensitive loads, and power quality Research on the relationship between disturbances and equipment status, control strategies for power quality governance, and other aspects, artificial intelligence algorithms are based on data-driven, showing strong application prospects in mining the internal relationship between multiple data, and solve power quality related problems. A large number of trial applications have been carried out on the Internet, and fruitful results have been achieved [1]. With the continuous deepening of informatization, cloud computing, and big data, the new generation of artificial intelligence technology will provide important support for solving all aspects of power quality problems.

In the past ten years, artificial intelligence represented by machine learning has continued to develop, such as deep learning, reinforcement learning, etc. The application of new methods has continued to expand in solving traditional power quality disturbance recognition, and with the increase of power quality monitoring data, it is based on artificial intelligence [2]. Algorithm mining power quality disturbance monitoring data information has been explored more, and good results have been achieved. Based on this research background, this paper uses the LS-SVM classifier based on the
radial basis kernel function to classify and recognize the extracted power quality feature sets. The classification results show that the method can effectively classify power quality signals such as voltage swells, sags, short-term interruptions, transient oscillations and pulses, with high recognition accuracy and small errors.

2. Based on big data power quality recognition online machine learning system architecture

2.1. Overall architecture
Referring to the characteristics of the conventional cloud storage platform architecture and power quality monitoring system, the architecture of the power quality big data cloud storage platform consists of the following parts: data access and protocol analysis module, MongoDB + Redis data storage layer, analysis and calculation layer and the user interface layer, its architecture is shown in Figure 1.

![Power quality big data cloud storage platform architecture](image)

The data access module is responsible for protocol analysis and compressed data decompression, and according to the user's request for real-time data or historical data, it can be displayed to the user or stored in the Hbase database. In order to reduce the amount of network data transmission, the acquisition device generally compresses the data before transmitting it to the main station [3]. At present, the mainstream compression algorithm uses zlib technology, and many high-level languages provide a complete zlib decompression interface, which is convenient for application calls.

2.2. Data structure design
First estimate the amount of data: each monitoring point has real-time data (collected once in 3s) and historical data (collected once in 1 minute). Each piece of data occupies about 80 Bytes, and 2550 indicators are collected each time [4]. Taking one collection as an example, there are about 10427 monitoring points, and the data volume is about 2G per minute, 120G per hour, 2.8T per day, and the data volume per year is 1026T (PB level). The core data of power quality analysis is the measurement data of the monitoring point. Any instance can be expressed as a four-tuple: monitoring point, measurement index, measurement occurrence time, and measurement value. Among them, the monitoring point coding standard code is a composite structure, which is composed of the provincial company code, the city code, and the difference code, and the length is 10 characters. The provincial
company code is composed of 2 characters, the city code is composed of 2 characters, and the distinction code is composed of 6 characters. The measurement index coding is more complicated, as shown in Figure 2.

![Figure 2. Quadruple](image)

Each item corresponds to a four-digit code, and the first four uniquely determine a measurement index.

### 2.3. Power quality big data processing flow

Distribution network voltage quality monitoring data has important characteristics of big data such as large volume, fast speed, diverse forms, and great value. The sources of power quality monitoring data are diversified, and the structure of power quality data is complex. Therefore, the analysis of power quality data with multiple modalities and complex structures must face this complex data environment.

Data pre-processing in the era of power quality monitoring big data, the first thing to do is to perform operations on the data obtained from various power quality data sources, including extraction and integration, obtain relations and entities from the results obtained, and use association and aggregation after analysis. The data is stored by using a well-recognized and well-defined structure [5]. Figure 3 is a technical roadmap for the construction and processing of big data on power quality in the distribution network.

![Figure 3. Power quality big data processing flow](image)
3. Power quality machine learning algorithm design

3.1. Feature extraction

Assuming that a certain power quality disturbance signal is \( x(t) \), when the disturbance signal \( x(t) \) is decomposed by the EEMD method, \( n \) intrinsic modal functions \( c_1, c_2, ..., c_n \) are obtained, which respectively contain different frequency components of the signal, and the \( n \) intrinsic modal functions are decomposed into two groups the initial eigenvector matrices \( A \) and \( B \) are respectively expressed as:

\[
A = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_J \\ \end{bmatrix}, \quad B = \begin{bmatrix} c_{J+1} \\ c_{J+2} \\ \vdots \\ \vdots \\ c_n \end{bmatrix}
\]  

(1)

In formula (1), when \( n \) is an even number and \( \frac{n}{2} \) is an odd number, \( \frac{n}{2} \leq J \leq n/2 \). The matrices \( A \) and \( B \) characterize the essential characteristics of the disturbance signal \( x(t) \). In matrix theory, the singular value of the matrix is the inherent characteristic of the matrix. A real matrix \( C \) with \( N \) rows and \( M \) columns is decomposed as follows, which is called singular value decomposition, namely

\[
C = U \Lambda V^T
\]

(2)

\[
U = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix} \in R^{N \times N}, U^T U = I
\]

(3)

\[
V = \begin{bmatrix} v_1 & \cdots & v_M \end{bmatrix} \in R^{M \times M}, V^T V = I
\]

(4)

\( \Lambda \in R^{N \times M} \) is the matrix \( [\text{diag}\{\sigma_1, \sigma_2, ..., \sigma_p\} : 0] \) or its transposed form, which depends on the form of \( N<M \) or \( N\geq M \). \( p = \min(N, M) \), \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0 \), \( \sigma_j, \sigma_j, ..., \sigma_p \) become the singular value of matrix \( C \). The singular value of the matrix has good stability, that is, when the matrix element changes slightly, the singular value of the matrix changes very little. Same as equation (2), we can divide the real matrix \( C \) composed of intrinsic modal functions into two groups to form the initial eigenvector matrices \( A \) and \( B \) to perform singular value decomposition, respectively, to find the singular values \( \sigma_{A,i} \) and \( \sigma_{B,i} \):

\[
\sigma_{A,i} = [\sigma_{A,i}^1, \sigma_{A,i}^2, ..., \sigma_{A,i}^J]
\]

(5)

\[
\sigma_{B,i} = [\sigma_{B,i}^1, \sigma_{B,i}^2, ..., \sigma_{B,i}^J]
\]

(6)

Among them, \( \sigma_{A,i}^1 \geq \sigma_{A,i}^2 \geq \cdots \geq \sigma_{A,i}^J \), \( \sigma_{B,i}^1 \geq \sigma_{B,i}^2 \geq \cdots \geq \sigma_{B,i}^J \).

The matrix singular values \( \sigma_{A,i} \) and \( \sigma_{B,i} \) conform to the stability, rotation and scale invariance required for the extraction of eigenvectors in pattern recognition, so they can be used as eigenvectors of power quality disturbance signals.
3.2. Least Squares Support Vector Machine
Among them, the SVM classifier based on the radial basis kernel function is more suitable for classifying high-dimensional data. When the data is linearly inseparable, the radial basis kernel function has a better classification effect. This paper uses LS-SVM based on the radial basis kernel function to classify and identify the power quality feature vectors.

\[ K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \]  

(7)

This paper uses the above four schemes to classify and recognize the power quality feature vector, compare the accuracy of classification and recognition, and the amount of calculation, and finally think that MOC is the optimal coding scheme.

4. System simulation design

4.1. Disturbance data collection
The deep learning algorithm model proposed in this paper requires a large amount of power quality disturbance tag data for training and testing. Therefore, it is planned to use a combination of simulation and reality. Firstly, a variety of disturbance signal waveform sampling data are obtained by MATLAB/Simulink software simulation [6]. The parameter setting of the composite disturbance signal is superimposed and compounded on the basis of the foregoing, in order to observe the influence caused by the simultaneous occurrence of multiple disturbance types. At the same time, build a simulation model based on the typical disturbance introduced in section 3.1 to simulate and record the disturbance signal. Figure 4 shows the setting window for the occurrence of sag disturbance in the aforementioned line short-circuit model, which is used to generate voltage sag disturbance waveform data caused by line short-circuit within a certain range.

![Figure 4. Line short circuit simulation model disturbance occurrence setting window](image)

At the same time, the order is shuffled before use to ensure the randomness of the training data. The total number of simulation samples generated for various types of power quality disturbances is shown in Table 1.
Table 1. Main types and parameter settings of power quality

| Disturbance type | Mathematical model                                                                 | Parameter settings                                                                 |
|------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1 Swell          | \( u(t) = (1 + \alpha (u(t_1) - u(t_2))) \sin(w_1t) \)                             | \( \alpha = 0.1 \sim 0.9; \ T < t_2 - t_1 < 9T \)                                  |
| 2 Sag            | \( u(t) = (1 - \alpha (u(t_1) - u(t_2))) \sin(w_1t) \)                             | \( \alpha = 0.1 \sim 0.9; \ T < t_2 - t_1 < 9T \)                                  |
| 3 Interrupt      | \( u(t) = (1 - \alpha (u(t_1) - u(t_2))) \sin(w_1t) \)                             | \( \alpha = 0.9 \sim 1; \ T < t_2 - t_1 < 9T \)                                  |
| 4 Transientimpulse | \( u(t) = \sin(w_1t) + \alpha e^{-\beta t} \sin(2\pi f_0 t - \phi) \) \( u(t_1) - u(t_2) \) | \( \alpha = 0.1 \sim 0.8; \ c = 2.5 \sim 5; \ \beta = 5 \sim 10 \)                |
| 5 Transientoscillation | \( u(t) = U_m \sin(\omega t + \phi) + \gamma \left[ e(t_2) - e(t_1) \right] \) | \( \gamma \in (1, 3); \ t_2 - t_1 \in (1ms, 3ms) \)                              |

4.2. Simulation results

Based on the inherent modal function singular value decomposition and LS-SVM, a total of 200 sets of signals of 5 types of power quality disturbances generated in Section 4.1 are extracted and classified using the method described above. The classification results are shown in Table 2.

Table 2. Comparison of classification results

| Feature amount         | Singular value of intrinsic mode function | Feature amount         | Energy value |
|------------------------|------------------------------------------|------------------------|--------------|
| Number of classification categories | 5                                        | Number of classification categories | 5            |
| Training samples       | 160                                      | Training samples       | 160          |
| Test sample            | 40                                       | Test sample            | 40           |
| Coding scheme          | MOC ECOC OneVsAll OneVsOne               | Output code digits     | 3 4(≤45) 5 10 |
| Training time/s        | 0.6864 0.312 0.7644 0.936               | Test time/s            | 0.0312 0.0312 0.0312 0.0624 |
| Classification accuracy rate/% | 100 97.5 92.5 25                        | Classification accuracy rate/% | 97.5 92.5 87.5 22.5 |

5. Conclusion

Based on the inherent modal singular value decomposition to extract the characteristics of power quality disturbance signals, the least squares support vector machine based on radial basis kernel function is used to classify and recognize disturbance signals and other common deep learning algorithm models, which are built on Mongo DB + Redis. The sequential model comparison experiment on the framework shows that the accuracy of the fusion model of inherent modal singular value decomposition extraction + least square support vector machine on the simulation data set is 92.5%, and the accuracy on the real data set is 97.5%. Taken together, this online machine learning system model can well identify the concentrated multiple types of single power quality data disturbance signals and composite disturbance signals. Although the accuracy rate is reduced in the actual data test, it still has a certain anti-noise performance. It has practical application value and will be continuously optimized and improved in the future.
References

[1] Ma, J., Zhang, J., Xiao, L., Chen, K., & Wu, J. Classification of power quality disturbances via deep learning. IETE Technical Review, 34(4) (2017) 408-415.

[2] Rodrigues, W. L., Borges, F. A., de Carvalho Filho, A. O., & de AL Rabelo, R. A Deep Learning Approach for the Detection and Classification of Power Quality Disturbances with Windowed Signals. SN Computer Science, 2(2) (2021) 1-14.

[3] Zhao, C., Li, K., Li, Y., Wang, L., Luo, Y., Xu, X., & Meng, Q. Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition. IEEE Transactions on Industrial Informatics, 15(5) (2018) 2915-2926.

[4] Bagheri, A., Gu, I. Y., Bollen, M. H., & Balouji, E. A robust transform-domain deep convolutional network for voltage dip classification. IEEE Transactions on Power Delivery, 33(6) (2018) 2794-2802.

[5] Liu, M., Song, T., Hu, J., Yang, J., & Gui, G. Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks. IEEE Transactions on Vehicular Technology, 68(1) (2018) 641-653.

[6] Zhang, D., Han, X., & Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE Journal of Power and Energy Systems, 4(3) (2018) 362-370.