Research of Electrical Equipment State Identification Based on K-means Clustering Algorithm

Liting Lei

College of Electronic Information

Nantong University

Nantong, China

1531790519@qq.com

Hui Xu*, Rongrong Fan

College of computer science and technology

Nantong University

Nantong, China

xu.h@ntu.edu.cn

Abstract—**For the current heavy manual monitoring work and frequent failure of electrical equipment, an online monitoring scheme for equipment state based on current sensor, WIFI communication module and embedded development board is designed. K-means clustering algorithm is used to analyze the current data set collected from different working status of the equipment, and the corresponding characteristic value is obtained. And then the automatic identification of the equipment working status is realized. The experimental results show that this method can quickly recognize the equipment operation status with high accuracy, and working stability.**

Keywords—electrical equipment; on-line monitoring; current signal; clustering algorithm; state recognition

I. INTRODUCTION

Nearly 30 years, with the rapid development of computer technology, microelectronics technology, sensor technology, intelligent electromechanical systems and their large number of applications to a variety of electrical equipment, electrical equipment plays a huge role in the daily production work, greatly improves the production efficiency, and its complexity and failure rate is getting higher and higher. So the reliability and safety of the equipment has become a key factor to insure the economic and social benefit, which has been highly valued in the engineering field. Monitoring the equipment working state and recognition analysis of working state can be more intuitive, accurate and real-time understanding of the real state of the electrical equipment, to anticipate equipment failure or safety hazard and provide effective help for production and operation, fault diagnosis, etc.

There are many domestic and foreign scholars have done a lot of research in recognition technology of equipment status, and a number of recognition methods are proposed. For example, with infrared temperature image and temperature gradient distribution images as the judgment of equipment state [1], using the frequency resolution performance evaluation model to construct the Hilbert time-frequency spectrum of the diesel engine vibration signal, based on this, the status of the diesel engine is identified [2], and using the classification method of support vector machine to judge the working state of equipment [3], adopting two-stage architecture of Savizky-Golay filter and neural network for real-time classification online of machine operation state, and using the finite state machine model to simulate the equipment working process in manufacturing industry, in order to achieve equipment management and reduce the energy consumption [4]. These identification technology are very useful for judging the state of the equipment, but there are still many shortcomings and limitations, such as many parameters, complex realization process, etc. Based on previous, this paper proposes a clustering analysis of the collected current data of electrical equipment by the K-means clustering algorithm to realize the automatic identification of the equipment working status. The method has fast calculation speed, high recognition rate and good stability. It can provide reliable basis for judging equipment failure and equipment daily maintenance.

* Corresponding Author.

II. K-MEANS CLUSTERING THEORY

K-means clustering is a Cluster Analysis, Cluster Analysis divides a data set into different categories or clusters in accordance with a specific standard (such as distance criteria),and makes the data objects in the same cluster have high similarity, and the similarity between data objects in different clusters as far as possible low [5]. Cluster Analysis as an effective data analysis tools has been widely studied and successfully applied in many fields, such as pattern recognition, image processing, information retrieval, data mining, customer segmentation, Web document classification [6]. K-means clustering is one of the classic clustering algorithms. Because of the high efficiency of the algorithm, it is widely used in the clustering of large scale data.

K-means algorithm with k as the parameter, the n objects is divided into k clusters, so that the cluster has a high degree of similarity, and the similarity between clusters is low. According to the actual need to select one of the Euclidean distance, Manhattan distance, Hamming distance as the measure of similarity, the most common use is Euclidean distance. The processing procedure of K-means algorithm is as follows: First, randomly selecting k objects, each object initial represents the average or center of the cluster; for each remaining object, according to the distance to each cluster center, it is assigned to the nearest cluster; and then recalculating the average value of each cluster. This process is repeated until the criterion function is convergent [7]. In general, the squared error criterion is used as the criterion function, which makes the generated clusters as compact as possible.

III. THE SYSTEM DESIGN

The current is an important characteristic to reflect the working status of the circuit and equipment, although there are many parameters for the state monitoring of the electrical equipment, the current signal has unique advantages such as convenient to be obtained, stable, not easy to be disturbed and so on. In general, the equipment current data under the same working state is similar, the characteristic quantities under different working states can be extracted from collected current data by using the K-means clustering

analysis. For subsequent current data, only need to compare with the previous state characteristics based on a certain rule, can infer the equipment running status in real time.

A. The Hardware Structure of The System

The hardware of equipment remote monitoring system is shown in Fig.1. The hardware system is mainly composed of remote monitoring platform, communication network, embedded development board and acquisition terminal. The tasks of this system mainly include the design of the data acquisition module and the WIFI data communication module, the monitoring software development of the upper computer and so on. The model of embedded development board is Cortex-M0+ ARM, which has reliable performance, low power consumption and strong expansibility. The current sensor as data acquisition module, can capture the equipment working status. The WIFI module is an embedded module in line with the WIFI wireless network standard based on UART interface, built-in wireless network protocol IEEE802.11 protocol stack and TCP/IP protocol stack. MCU transmits the collected data to the "Serial to WIFI Module" through serial and eventually transfers to the client through wireless. The client is developed by Java, which can be used by any PC and notebook computer and has strong portability. The client receives the state parameters through the WIFI receiver, and displays the working current in the screen in real time. According to the real-time current data, the client gives the status identification, and the data can be transmitted to the server in a wired way and stored for a long time.
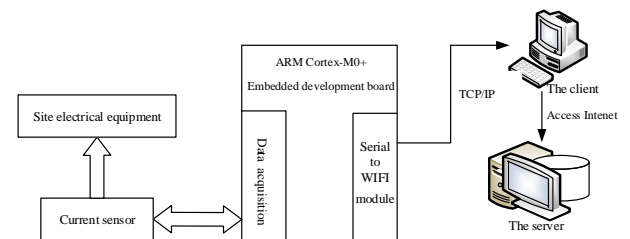

Fig.1.   Hardware framework of remote monitoring system

B. Recognition Method of Equipment State Based on K-means Clustering

1) Equipment state division

In this paper, the motor is used as the experimental object, using current sensor to capture the effective value of the

current changes in the motor running. As is known to all, the number of motors has increased substantially in recent years they are almost used in every corner of our daily life, for example, the manufacturing industry system, air transportation, ground transportation, household energy conversion system and mobile phone vibration system and so on. It has become the main requirements of the motor in many industrial applications that achieving the reliability of motor operation in harsh industrial environment. An unexpected shutdown could lead to a disruption of critical services, such as in medical, transportation, or military operations [8].

According to the distribution of working current value, the working state of the electromotor can be simply divided into normal operation state and abnormal state, the normal state refers to the operation process that the motor no-load operation to stress different mechanical load. The no-load operation is the stator of the motor has a symmetrical three-phase voltage attached to it, and the rotor without mechanical load. The load current is the actual detecting current value of the stator at this time, this value is proportional to the size of the load change, generally no more than rated current. The abnormal state includes overload operation (the operating current exceeding the rated current) and fault state (the operating current increasing singularly). The long-term motor overload is the current exceeds the rated value, will lead to the motor index such as winding temperature, shaft temperature roses more than normal allowable value, serious motor heating, which will reduce the service life of the motor, might cause motor burn in more severe situation. The working current will increase sharply when there are failures of motor like lack of phase, bearing damage, axis bend or winding burned and other faults. Therefore, the motor working state will be divided into three states for normal load, overload and failure, the number of clusters k adopts 3, using C++ Visual to achieve the algorithm.

2) Clustering learning
Use K-means clustering algorithm to cluster the dataset of motor state, specific process shown in Fig. 2.
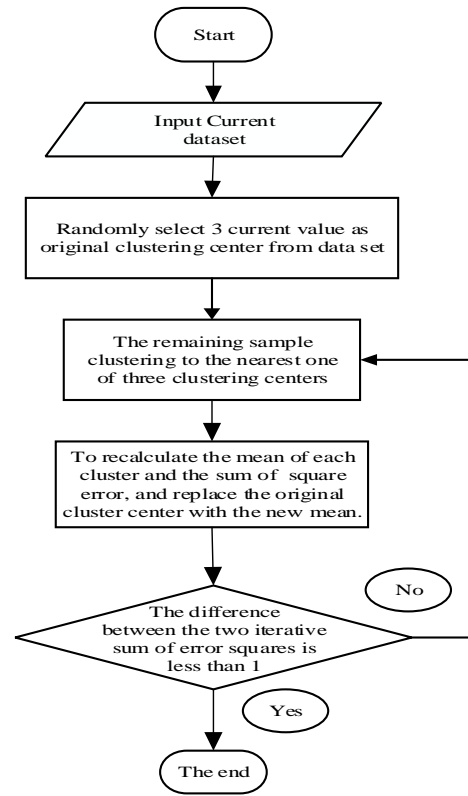

Fig.2.   K-means clustering algorithm flowchart

a) First of all, determine the cluster centroid, we randomly selected 3 current values in the input current data as the initial cluster centers of the normal load, overload and fault state, and j is the number of iterations of the algorithm.

b) According to each cluster center, calculate the Euclidean distance between each object x of the current data set and these three centers , and the corresponding objects are re-clustered according to the minimum distance, and the error sum of squares $E(j)$ is calculated. The calculating formula of Euclidean distance and error sum of squares are shown in (1) and (2).

$$d(x, m_i) = \sqrt{\sum (x - m_i)^2} \qquad (1)$$

$$E = \sum_{i=1}^{k} \sum_{x_{ij} \in C_i} |x_{ij} - m_i|^2 \qquad (2)$$

c) Calculate the new cluster centroid after distribution, the centroid is calculated shown in (3).

$$m_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \quad (i = 1, 2, 3) \qquad (3)$$

Above these formulas, x_{ij} is the object of the cluster C_i in current dataset, n is the number of objects of the cluster C_i[9].

d) Cycle steps *b)~c)* until $\|E(j+1)-E(j)\|$ is less than the minimum deviation threshold value which is taken 1, algorithm stops when $\|E(j+1)-E(j)\|$ is less than 1 that means the criterion function value does not change significantly. Now, the three cluster centers are the cluster centers of the three working state of motors correspond to normal, overload and fault.

3) Automatic recognition of working state

Through clustering learning the motor current data set, three cluster centers of different working states are trained. According to the training characteristics, the client conducts pattern recognition of the received remote current data, using the distance criterion to judge each current data belonging to which one of three kinds of operation state and giving the real-time status recognition. For example, normal state, overload, fault, the cluster centers of these three kinds of motor running state are I_1, I_2, I_3. For the input current value I, by the distance formula (1), the distance between I and I_1, I_2, I_3 can be calculated respectively for d_1, d_2, d_3. Comparing the size of d_1, d_2, d_3, the minimum distance means the maximum similarity between the present current and that cluster. For example, when d_1 is less than d_2, d_3, the current motor working state is normal state; if d_2 less than d_1, d_3, the working state belongs to overload, the client should remind of the staff to motor power check timely to avoid motor fault.

IV. THE EXPERIMENT

The current sample data acquired from system is divided into two categories, one is the training sample to train the cluster center in the K-means clustering algorithm, the sample size is 50. The other is the test sample, which is used for identification test of equipment working state utilizing the K-means clustering algorithm. The sample capacity is 40, and the sample sizes of load status, overload status and fault are 20, 10 and 10 respectively. On the beginning of clustering training samples by K-means algorithm, randomly selecting 3 cluster centers, the clustering results obtained are shown in Fig.3.
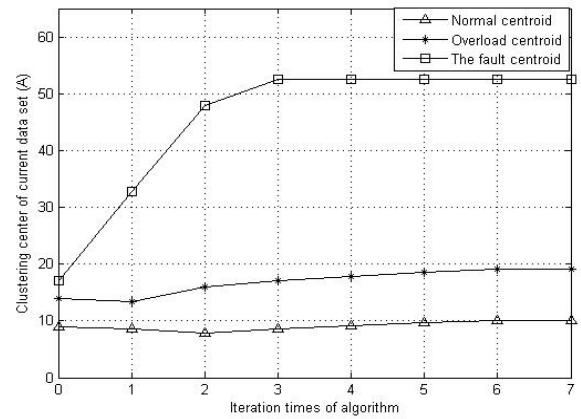

Fig.3. The clustering results of training sample by using the K-means clustering algorithm

Through repeated training on sample data, the maximum number of iterations of the algorithm is seven times, the resulting cluster centers are 10.0435(A),19.1765(A) and 52.5(A) respectively, representing three operating states clustering means of normal, overload and fault. Finally, the recognition results of equipment working status can be achieved according to the cluster centers, and the recognition results are shown in Tab. I. The experimental results presented in Tab. I show that there is a fault sample identified by mistake is classified into the overload cluster, and the correct rate is 97.5%. The reason for error is about the limit of training and testing sample data as well as the choice of initial clustering center. If cluster centers generated randomly inappropriate, it is possible to extend the running time, even it is impossible to obtain meaningful clustering results [10]. The method is very suitable for the automatic identification of the working status of other electrical equipment, as long as the collected data set contains appropriate characteristics.

TABLE I.   RECOGNITION RESULTS OF EQUIPMENT WORKING STATUS

Actual Sample	Recognition Sample		
	Normal	*Overload*	*The fault*
normal(20)	20	0	0
overload(10)	0	10	0
The fault(10)	0	1	9

Note: the misclassification error is 2.5%.

V. CONCLUSION

In this paper, current sensor, embedded development board and wireless communication network are used to monitor the working status of the equipment in real time, therefore, this status information can be quickly collected and transmitted. Based on this, using K-means clustering algorithm for equipment condition recognition analysis can well distinguish three kinds of typical equipment working state under different working current value, the correct rate is 97.5%. This method is also very suitable for working state recognition of other electrical equipment, as long as the data set containing the appropriate features can be collected. In order to further improve the accuracy of the recognition and the speed of the algorithm, need to make further research on the determination of the initial cluster center of K-means clustering algorithm in the future.

REFERENCES

[1] Zhang Lei, "Research of equipment status recognition technology based on infrared thermal imaging," Beijing: Beijing University of Chemical Technology, 2011.

[2] Li Hongkun, Ding Jian, Lian Xiaoting. "Diesel status identification based on the frequency resolution evaluation model," The proceedings of the tenth national conference on vibration theory and application in 2011 Vol.1.

[3] Tian Y, Yu L, "An Image Recognition Method of the Electric Equipment Operation States," Power and Energy Engineering Conference, 2009, APPEEC 2009,Asia-Pacific, IEEE, 2009,pp.1-4.

[4] Pang C K, Le C V, Gan O P, "Intelligent Energy Audit and Machine Management for Energy-efficient Manufacturing," 2011 IEEE 5th International Conference on Cybernetics and Intelligent Systems,2011, pp.142-147.

[5] Ma Li, "The comparative study of clustering analysis and traditional clustering based on data mining,"Journal of Mathematical Medicine,2008, pp.530-531.

[6] Sun Jie, "The analysis and application of clustering algorithm in gene expression data,"Jiangsu: Jiangsu University of Science and Technology,2013.

[7] Li Haoran, Wu Chaozhong, Chu Duanfeng, "Car collision accident analysis in highway based on k-means clustering," Chinese journal of safety science, 2013,pp.91-96.

[8] Hamid a. Toliyat, Subhasis Nandi, Seungdeog Choi, Motor Modeling, Condition Monitoring and Fault Diagnosis, Zhang Weiping, Yu Fei, Zhang Chao, Beijing: Mechanical Industry Press, 2014.

[9] Chen Zuoping, Ye Zhenglin, Zheng Hongchan, "A fast fractal coding method based on K-means clustering," Chinese Journal of image and graphics, 2007, PP.586-591.

[10] Wang Xiaofei, Nie Shengdong, Wang Yuanjun, "Improved K- means clustering algorithm and its application in brain tissue segmentation," Chinese Journal of Medical Physics,2014,PP.4760-4764.