International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0049

ABNORMAL ENERGY CONSUMPTION MODE DETECTION OF OFFICE BUILDING AIR CONDITIONING SYSTEM BASED ON INFORMATION ENTROPY

Xuan ZHOU[1,2], Xuehui ZI[2], Liequan LIANG[3*], Junwei YAN[2], Dongmei PAN[1,2]

1 Shenzhen Green Building Easy Home Co., Ltd. Shenzhen, 518100

2 South China University of Technology, Guangzhou, 510641, China

3 Guangdong University of Finance & Economics, Guangzhou, 510320, China (Lianglq@gdufe.edu.cn)

ABSTRACT

The Central Air-Conditioning System (CACS) in subtropical region is responsible for more than 50% of total energy consumption in public buildings. Improper operating modes of CACS often lead to abnormalities in DHECM (Diurnal Hourly Energy Consumption Mode), the detection of which is of great significance for energy conservation. However, It is difficult to detect the abnormal modes effectively by conventional feature extraction and single threshold anomaly detection methods due to its complicated operational condition. Two-year hourly energy consumption data of CACS in an office building collected by CACS monitoring and control platform are divided into types of typical working conditions by decision tree and the information entropy value is used as the characteristic parameter of uncertainty for diurnal hourly energy consumption time series to reduce their dimension. Furthermore, a clustering unsupervised algorithm was used to classify normal and abnormal DHECM which solve the problem that the threshold of the abnormal mode is difficult to determine. The abnormal detection results showed the effectiveness of this method in the field of abnormal DHECM detection of public buildings.

Keywords: air conditioning system, diurnal hourly energy consumption mode, abnormal detection, information entropy, clustering analysis

1. INTRODUCTION

The central air-conditioning system in subtropical accounts for more than 50% of total energy consumption in public buildings[1]. Improper air conditioning operation mode and equipment failure may cause more than 50% energy waste of CACS [2]. How to use advance data mining method to discover the abnormal energy use mode from the energy consumption data of CACS? How to use the domain knowledge to measure, interpret, analyze and extract valuable information from the data? These problem aroused interest of scholars in recent years.

In the field of anomaly detection of CACS, most of the research are focused on the fault diagnosis of equipment of CACS by modeling on data driven method. The advance data mining technology in recent years has provided a new direction for the abnormal diagnosis of air conditioning operation. At present, three main types of abnormal energy detection methods are used commonly: (1) the threshold method for detecting abnormal data which is preset by statistical methods. Daniel applied a data-driven CCAD-SW framework to detect building collective anomalous energy consumption data, this method take the quartiles in statistics as thresholds[3]; (2) the abnormal data is detected by clustering algorithm. Jiang used the SVM algorithm to identify the hourly energy consumption pattern and took the average energy consumption in per-hour as the characteristic parameters, then used LOF clustering algorithm to determine the hourly abnormal energy consumption[4]; (3) the energy consumption baseline is established based on predictive data or simulated data. Lina proposed a DET-Toa method to detect daily abnormal energy consumption data of buildings. This method detects abnormal energy consumption based on the deviation of simulated and actual data [5]. Chou et al. utilized the ARIMA algorithm

Selection and peer-review under responsibility of the scientific committee of the 11th Int. Conf. on Applied Energy (ICAE2019).
Copyright © 2019 ICAE

to predict the daily energy consumption data, and calculate the deviation between the actual data and the predicted data to identify the abnormal energy consumption according to the principle[6]. Meanwhile, the researches of energy consumption mode were mostly focused on the power industry. Peng et al. adopted clustering algorithm to intelligently identify customers' electricity usage patterns[7].

Because the operational conditions of CACS are complicated due to lots of influencing factors, such as meteorological parameters, operational parameters of CACS and so on, the DHECM data shows large fluctuation and high-dimensional and it is difficult to identify the abnormal modes by conventional abnormal detection methods. Therefore, the present work deals with the method of dividing the complex air conditioning working condition. Then a cluster-based thresholdless anomaly detection method is proposed which explored to take information entropy as the characteristic of DHECM data to reduce the dimension of time series and use clustering algorithm to classify the normal and abnormal modes.

2. ABNORMAL DETECTION METHOD OF DHECM

The abnormal detection method of DHECM proposed in this paper is a kind of semi-supervised learning method which mainly includes two processes.

（1）Normal data set establishment of DHECM

1)Historical data pre-processing, like data cleaning and Data normalization;

2)Complex operation condition division according to outdoor dry bulb temperature, data attribute (weekends or workday);

3)Characteristic parameter calculation (information entropy of DHECM);

4)Clustering DHECM data into normal and abnormal data set under each typical working condition, then establishing the normal data set of DHECM ;

（2）Abnormal detection of DHECM online

This process is similar to the first one. The difference is that the data is collected online, the working condition of this moment is also identified online, and as normal data set of DHECM was established in last process and in this process, the normal data set is updated constantly.

3. CASE STUDY

3.1 Data Source

In this paper, the energy consumption data of an office building central air-conditioning system in the subtropical region is taken to develop the abnormal detection technology of abnormal DHECM. The construction area of this building is about 87,000m^2. And the office area is cooled by central air-conditioning system. The working hours are more than 8 hours in summer and the cooling time is from May to November. So the annual electricity consumption of the air conditioning reached up to 1.7 million kWh. To reduce the operating costs, the CACS was retrofitted in 2013 and the monitoring and energy saving control platform was built by which the energy consumption data of the chiller plant could be collected online. The energy consumption data of 488 days from April to November of 2016 and 2017 (cooling season) were chosen to verify the effectiveness of the abnormal detection method of DHECM in office building.

3.2 Data Preprocessing

3.2.1 Data normalization

The historical data set of CACS includes energy consumption data of chiller plant and outdoor dry-bulb temperature data. Because the research object of this paper is DHECM, it is necessary to normalize the diurnal energy data into to a row vector format. The format is shown as follows:

$$Q_i = q_{i,1}, q_{i,2}, \cdots, q_{i,24} \tag{1}$$

$$T_i = t_{i,1}, t_{i,2}, \cdots, t_{i,24} \tag{2}$$

$i = 1,2,\cdots,N$, $j = 1,2,\cdots,24$. $q_{i,j}$ is the energy consumption for the jth hour of the ith day， $t_{i,j}$ is the outdoor dry-bulb temperature for the jth hour of the ith day， N indicates that there are N daily energy records in the data set.

3.2.2 Data Cleaning

There were always some obvious mutation, missing and abnormal zero-value abnormal data due to the sensor malfunction, signal transmission errors or some other faults. Data cleaning is necessary. The mutation data which were identified by the 3σ principle were deleted and handled as missing data.Then, the missing and abnormal zero-value data is interpolated according to Lagrange interpolation method. If there were more than 5 consecutive missing data in one day, the data record for this day would be deleted.

3.3 Establishment of Historical Data Set Of Normal DHECM

Copyright © 2019 ICAE

3.3.1 Division of the complicated working condition

DHECM depended on the working conditions and unreasonable division of working conditions could cause false detection or missed detection. Therefore, according to the analysis of the influence factor of energy consumption of CACS, the outdoor bulb temperature, daily operating hours and data attribute are taken as characteristic parameters to establish the decision tree to divide the complicated working conditions by trial and error method.

3.3.2 Establishment of normal DHECM data set from historical data set

The "mode" of energy consumption of CACS in this paper was defined for hourly energy consumption in one day with CACS on which represented the operating mode of CACS and control strategy. And, in order to deviate the abnormal modes from those normal modes under the same working conditions, it’s necessary to determine the characteristic parameter of the mode.

The concept of entropy stems from thermodyna mics which is a measurement of randomness and uncertainty. In 1948, Shannon first proposed the co ncept of "information entropy". For random variabl e, its information entropy value is defined as equat ion (3):

$$E(X) = -\sum_{x \in X} P(x) log P(x) \tag{3}$$

3.3.3 Calculation of characteristic parameter of DHECM under typical operational conditions

Distance-based similarity measurement is a kind of commonly-used method.The Euclidean distance is easy to calculate and suitable for high-dimensional data. In this paper, the euclidean distance is used to measure the similarity between different DHECMs, and build the energy consumption distance matrix D under different working conditions. The Euclidean distance calculation formula is as shown in equation (4):

$$d_{i,j} = (\sum_{k=1}^{n} \left| q_{i,k} - q_{j,k} \right|^2)^{1/2} \tag{4}$$

$q_{i,k}, q_{j,k}$ are the factors of distance matrix D , d_{ij} is the Euclidean distance between the mode in i th day and j th day.

The distance reflects the probability of the occurrence of the daily energy consumption pattern. The greater the distance is, the farther the mode from normal modes is, and the lower the probability of occurrence. The distance matrix D is shown as equation (5).

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,N} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ d_{N,1} & d_{N,2} & \cdots & d_{N,N} \end{bmatrix} \tag{5}$$

In the formula, the elements of the diagonal of the matrix are the distance of each modes with themselves whose value is 0. According to the distance matrix D , the distance contribution matrix P is established. The calculation of distance contribution is as following:

$$p_{i,j} = \frac{d_{i,j}}{\sum_{i=1}^{N} d_{i,j}} \tag{6}$$

$p_{i,j}$ is the distance contribution of the j th dimension data of i th day. The distance contribution can reflect the probability of the elements in different dimensions. The larger the value, the greater the contribution. Base on the distance contribution matrix calculate the information entropy value of each DHECM according to Equation 3.

3.3.4 Anomaly modes mining

The problem of abnormal DHECM detection was a two-category classification problem. The clustering method was chosen to cluster similar data into same cluster to distinguish between normal and abnormal DHECM. Commonly-used clustering algorithms include hierarchical clustering algorithm, K-Means algorithm and LOF algorithm. Each clustering algorithm has its advantages and applicability. The hierarchical clustering algorithm was chosen in this paper as it has the characteristics of simple implementation and high stability.

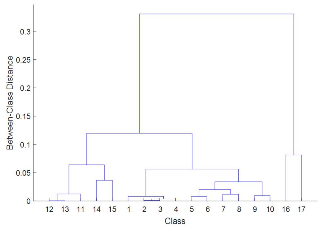
3.3.5 Establishment of the normal DHECM data set

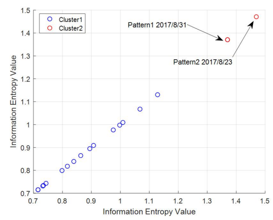
In order to establish normal DHECM data set, It is necessary to eliminate the abnormal DHECM from the historical data set under different working conditions. For example, when the CACS running time is 15 hours and the range of outdoor bulb temperature is [31℃ , 33℃), there are 17 records under this working condition, and by information entropy method the number of abnormal records of DHECM is two while by conventional parameter method it is 1 and by K-means method it’s 7.

It can be seen from Fig 1(a) that when the class spacing is 0.13, there are two DHECM records which are grouped into a single cluster, and in Fig 1(b), there are

Copyright © 2019 ICAE

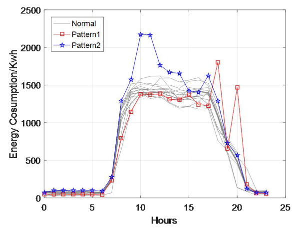
two obvious outliers shown in the scatter plot of the information entropy value.



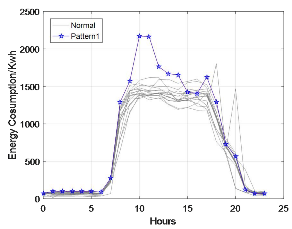


(a)clustering spectrum (b) scatter plot

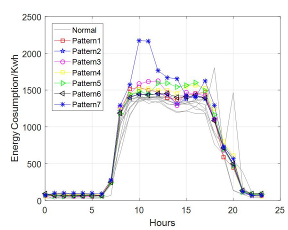
Fig 1 Information entropy method



(a)



(b)



(c)

Fig 2 The mining result by different methods by different methods

Fig 2(a) showed the hourly energy consumption curve for each DHECM record under the above condition by information entropy method, mode 1 and mode 2 were detected as abnormal DHECMs. After field investigation and operational parameters of CACS verification, on the day of mode 1, all the chillers were turned on when the CACS started up on the morning because of wrong parameters setting. On the day of mode 2, there are two mutations of this DHECM after 17:00 which might be caused by turning one more chillers because of incorrect operation. The detection results by Conventional parameter method and K-means method were shown as (b) and (C) of Fig 2. It showed that the information entropy method is the best one which avoided the influence of “Curse of Dimensionality” on the clustering algorithm. On the other hand, the information entropy value can accurately describe the data distribution of the DHECM, and retain the mode characters of hourly energy consumption records in one day.

4. CONCLUSION

A new thresholdless abnormal detection method was proposed in this paper which was applied for the abnormal DHECM detection of CACS in an office building. Firstly, this method can excavate the hidden DHECM of building air conditioning; Secondly, it could improve the detection effects, reduce false detection and missed detection rate due to inaccurate threshold. Thirdly, the information entropy was used as characteristic parameter of DHECM whose feature was represented well and by which the mode characteristics was retains to a large extent. This method provided a new idea for abnormal operation detection of CACS indirectly though energy consumption and it could also be used in other fields, like electricity power industry, or the abnormal detection of periodic time series.

ACKNOWLEDGEMENT

Supported by the Science and Technology Planning Project of Guangdong Province(2016B090918105) and the Natural Science Foundation of Guangdong Province (2017A030310162, 2018A030313352).

REFERENCE

[1] Kang L, Liu P. Tian Z. Research on the Factors Affecting Air Conditioning Energy Consumption of an Office Building in Shenzhen. Building Thermal Ventilation and Air Conditioning. 2017; 36(11): 44-46.

[2] Narayanaswamy B, Balaji B, Gupta R. Data driven investigation of faults in HVAC systems with model. In: Srivastava M, `editors. Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, Menphis: ACM; 2014, 50-59.

[3] Daniel B, Katarina G. An ensemble learning framework for anomaly detection in building energy consumption. Energy and Buildings. 2017;144: 191–206.

[4] Jiang H, Lu T. A dynamic real-time anomaly detection method for building energy consumption in colleges and universities.Computer engineering.2017;43(4):15-20.

[5] Lina G, David E, Claridge. A temperature-based approach to detect abnormal building energy consumption.Energy and Buildings. 2015; 93(04): 110–118.

[6] Chou JS, Abdi ST. Real-time detection of anomalous power consumption. Renewable and Sustainable Energy Reviews. 2014; 33(05): 400-411.

[7] Peng X, Lai J. Intelligent identification method of customer electricity mode based on cluster analysis. Power system protection and control. 2014; 42(19): 68-73.

Copyright © 2019 ICAE