

## Analysis and Application of Sensor Alarm Threshold Adaptive Technology Based on Big Data Analysis

Tong Zhu<sup>1,2, a</sup>, Zhanbin Wang<sup>2, b</sup>, Ke Shao<sup>2, c</sup>

<sup>1</sup>College of Electronic and Information Engineering, Tongji University, 4800 Cao'an Highway, Shanghai, China, 201804

<sup>2</sup>Anti-counterfeiting Technology Division, The Third Research Institute of The Ministry of Public Security of The P. R. C., Bisheng Road 339, Shanghai, China, 201204

<sup>a</sup>fw339zt@126.com, <sup>b</sup>wzb\_wzb\_wzb@126.com, <sup>c</sup>fw339sk@126.com

**Keywords:** Gas Sensor; Fast Big Data Analysis; Data Hierarchy and Knowledge Map Model; Adaptive Technology

**Abstract:** In China, a multi-sensor intelligent alarm system based on the sensor identification technology has been widely used across the country and produces a wealth of information data. In the process of actual application, however, the alarm threshold of a traditional gas sensor is set at a fixed value, which often leads to false alarm. In some big cities, such false alarm has reduced the work efficiency of the danger-removal personnel, and affected the further promotion and application of the multi-sensor intelligent alarm system. The new data acquired by tens of millions of sensors are added every day and stored for a long time, so it is difficult to perform the analysis and judgment on the data with manual methods, and intelligent analysis and processing are urgently needed on this aspect. With the "Sensor Alarm Threshold Adaptive Technology Based on Big Data Analysis" presented in this paper, the prediction and judgment of the gas sensor threshold can be implemented through the "fast big data analysis", thereby realizing the self-adaption of the sensor alarm threshold. The actual application effects prove that it can effectively improve the efficiency of the multi-sensor intelligent alarm system.

### Working Principles and Deficiencies of a Single Gas Sensor Alarm System

In the field of intelligent alarm systems, the sensor-based automatic identification technology is essential. China has been trying to explore and practice for many years in this respect. Currently, the widely-applied and mature technology is the automatic recognition technology system based on the image video sensor. In terms of the gas sensor, however, most of the automatic recognition technologies are still stuck in the stage of setting of a fixed threshold as an alarm basis, which results in certain deficiencies in the actual application.

Inflammable, explosive, and toxic gas leaks may occur during processing, production, transportation, and storage for industrialized large-scale production, resulting in frequent occurrence of fire, explosion, or poisoning. The combustible/toxic gas alarm has been widely used as an effective safety protection device <sup>[1]</sup>. In some big cities, however, false alarm have reduced the work efficiency of the danger-removal personnel, and affected the further promotion and application of the multi-sensor intelligent alarm system.

Gas sensors can be roughly divided into three categories by the detection principle, namely: gas sensors making use of physical and chemical properties; gas sensors making use of physical properties; and gas sensors making use of electrochemical properties <sup>[2]</sup>. A single gas sensor alarm system is generally employed according to the working principle of directly monitoring whether the gas concentration exceeds the preset gas concentration threshold. The traditional method of

detecting combustible gas or toxic gas concentration is neither conducive to the centralized management and control, nor helpful to the prevention of fire and explosion accidents <sup>[3]</sup>.

The reasons for false alarm of the gas sensor intelligent alarm system in the actual application include: it is susceptible to environmental conditions such as temperature, humidity, thunderstorm, and the performance reduction of the gas sensor due to its service time and operating environment. In addition, the main factors influencing the application of the gas sensor intelligent alarm system are that normal human operations cause the gas concentration to change normally, but it is difficult to identify and distinguish this circumstance by the setting of the traditional fixed threshold, thus resulting in a higher rate of false alarm. In China, for example, the cooking wine is used when people are cooking in the kitchen. The cooking wine contains only 10-15% alcohol, but it can also cause false alarm <sup>[4]</sup>. In addition, false alarm may also occur due to the presence of propane or isobutane in insecticides used in kitchens <sup>[4]</sup>.

### Architecture and Working principle of the Multi-sensor-based Alarm System

The multi-sensor-based alarm system (hereinafter referred to as the System) consists of a front-end multi-sensor device (in a large number) at the local installation site, a data analysis platform in the data computer room, a remote operation management terminal, and a remote alarm terminal. The system architecture is shown in the figure below:

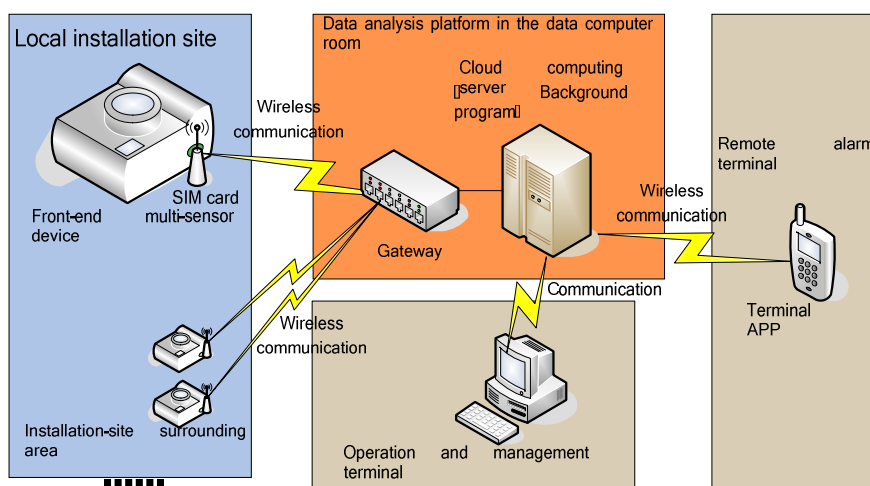


Fig.1 Architecture of the Multi-sensor-based Alarm System

In addition to the traditional single gas sensor, the front-end multi-sensor device at the local installation site also integrates a temperature sensor, a humidity sensor and an infrared sensor, which can separately acquire temperature information, humidity information, and nearby personnel status information. The front-end multi-sensor device and its installation effect are as shown in the figure below:



Fig. 2 Front-end Multi-sensor Device and Installation Effect

Working principles of the multi-sensor-based alarm system:

1. The front-end multi-sensor device at the local installation site detects local air information, temperature information, humidity information, and nearby personnel status information and transmits the information to the data analysis platform in the data computer room through wireless communication.
2. The data analysis platform in the data computer room obtains the information of multiple front-end multi-sensor device and combines it with the "current air quality information", "current climate condition information" and "history record information" to calculate the dynamic alarm threshold through the cloud computing server. And then, it will perform analysis and judgment by comparing the local air information and the dynamic alarm threshold for a period and send the alarm information to the remote alarm terminal.
3. After the remote alarm terminal obtains the alarm information, it can perform subsequent alarm disposal.
4. The "calculation of the dynamic alarm threshold through the cloud computing server" in this system is the key technology. It will be described in detail in Sections III. and IV. .

### Building of the dynamic alarm threshold model based on the data hierarchy and knowledge map rules

The cloud computing server can not only acquire the "current sensor record information" from the front-end multi-sensor device to form the "history record information", but also obtain the "current air quality information", the "current climate condition information" and so on through other information systems.

In the cloud computing server analysis system, the various types of information mentioned above are composed of original data/operational data and derived data/DSS data. The natural extension process of data separation caused by the difference between the original data and the derived data is as follows:

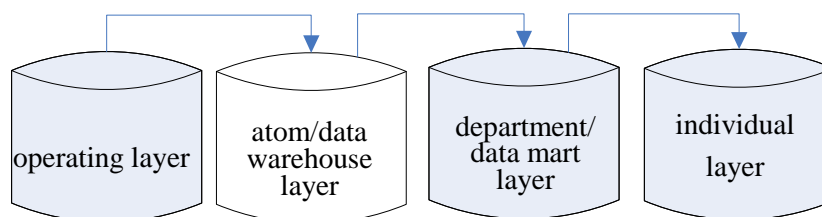


Fig. 4 Architecture Hierarchy

Through the operating layer, atom/data warehouse layer, department/data mart layer and individual layer, a basic big data set is formed for subsequent processing. Among them, the data of the operating layer contains only the application-oriented original data, and mainly serves the high-performance transaction processing field. In the data warehouse layer, non-updatable integrated original historical data and some derived data can be stored. The department/data mart layer is built to meet the final alarm needs. Most heuristic analysis is done in the data individual layer [6].

In this system, we used the multi-sensor device and air conditions as the main entity composition of the strong coupling; used the external environment and history information such as temperature information, humidity information, nearby personnel status information, air quality information, and climate condition information as the composition of the loose coupling; and used the location and

time as the edge of the semantic relationship between the associated entities, the resulting knowledge map model is as shown below:

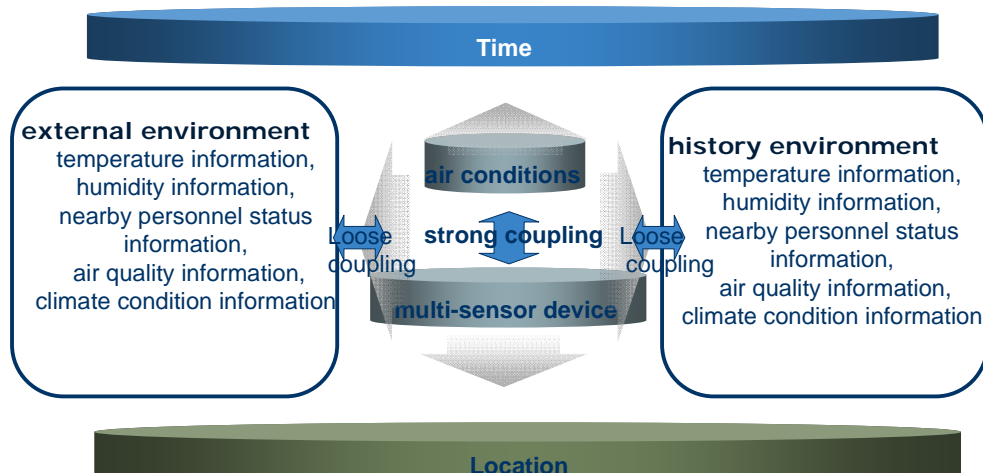


Fig.5 Knowledge Map Model with Multi-sensor Device and Air Conditions as Strong Coupling Conditions

On the basis of that the basic knowledge map model is completed and the knowledge is acquired based on the knowledge acquisition process, the semantic link model with the "multi-sensor device" and "air conditions" as the strong coupling conditions can be finally implemented through the triple storage and maintenance with extended attributes; the addition of the time series features of the extensive labeling and precise labeling methods, and the trustworthy computing of subsequent relationship.

When the conditions of concern change, we can also extract the elements of the loose-coupling entities to add or replace the elements of the original strong coupling to form a new semantic link model.

### Realization of the prediction of the dynamic alarm threshold with the fast big data analysis method.

In actual applications, the amount of data of the front-end multi-sensor device acquired by the cloud computing server is large. Large-scale industry/enterprise big data has far exceeded the processing capacity of existing traditional computing technologies and information systems. Therefore, it is an urgent need in the real world to search for effective big data processing technologies, methods, and means <sup>[7]</sup>.

In this system, based on the preprocessing with the knowledge map method, the big data can correlate all the factors in the "air condition alarm major event and composition" and "air external environmental conditions and attributes" and perform the fast big data analysis according to the input information of each sensor's automatic dynamics. Through the analysis, the "prediction of dynamic alarm thresholds" results can be obtained. Specifically, there are three stages in the implementation process.

In this system, the corresponding letters and numbers are used for the nodes of different types and locations. For easy description, all the nodes are subsequently unified to the same type of nodes in this paper. For the nodes at different locations, they are simplified and then seven of them are selected, of which, Node 0 is the center node, and Nodes 1, 2, 3, 4, 5, and 6 are the adjacent 6 nodes at different locations in the same geographical area as Node 0.

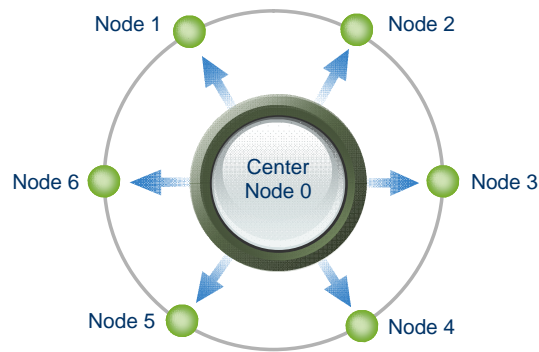


Fig. 6 Seven Nodes Centered on Node 0

**First Step:** we used the Hadoop computer cluster technology to process the big data acquired at each node, and used R language (hereinafter referred to as R) of perform the data statistics. In this system, the R program will perform the data analysis through original functions such as data loading, exploration, analysis, and visualization, while Hadoop will implement the parallel data storage and computes the distributed data [8]. In the example of this system, each sensor node is namely the node that supports Hadoop data storage and parallel computing, and performs the statistical analysis on the value information acquired by the multiple sensors through R, including gas concentration value, temperature, humidity, and infrared sensing values, gas concentration change value per unit time, temperature change value per unit time, humidity change value per unit time, and whether any person has been present in the proximity of the infrared sensor area. Through the existing weighted value of the “history record information” of the node, the standard adjustment can be done on the values to obtain comparable gas concentration change value per unit time, temperature change value per unit time and humidity change value per unit time.

The mathematical principles and formulas are as follows:

Time: current time  $t$ , previous 1-unit time  $t-1$ , previous 2-unit time  $t-2$ , and previous 3-unit time  $t-3$ ...

**Definition:**

gas concentration values  $Q$  acquired by nodes 1 to 6 at the current time can be expressed as:  $Q_0(t)$ ,  $Q_1(t)$ ,  $Q_2(t)$  ...  $Q_6(t)$

The temperature value  $T$  acquired by Nodes 0 to 6 at the current time may be expressed as:  $T_0(t)$ ,  $T_1(t)$ ,  $T_2(t)$  ...  $T_6(t)$

The humidity value  $HUM$  acquired by Nodes 0 to 6 at the current time may be expressed as:  $HUM_0(t)$ ,  $HUM_1(t)$ ,  $HUM_2(t)$  ...  $HUM_6(t)$

The infrared sensor value  $R$  acquired by Nodes 0 to 6 at the current time may be expressed as:  $R_0(t)$ ,  $R_1(t)$ ,  $R_2(t)$  ...  $R_6(t)$

**Calculation:**

(1) The gas concentration change value  $C_{Q0}(t)$  of Node 0 during the current two-adjacent-unit time:

$$C_{Q0}(t) = (Q_0(t) - Q_0(t-1))$$

In the same way, we can obtain the values of  $C_{T0}(t)$ ,  $C_{HUM0}(t)$  and  $C_{R0}(t)$  by calculation.

(2) The gas concentration average value  $V_{Q0}(t-10 \sim t)$  of Node 0 during the current ten-adjacent-unit time:

$$V_{Q0}(t-10 \sim t) = (Q_0(t) + Q_0(t-1) + Q_0(t-2) \dots + Q_0(t-10)) / 10$$

In the same way, we can obtain the values of  $V_{T0}(t-10 \sim t)$ ,  $V_{HUM0}(t-10 \sim t)$  and  $V_{R0}(t-10 \sim t)$  by calculation.

(3) Through the existing weighted value of "history record information" of Node 0, the standard adjustment can be made on all the values:

**3-1** According to the history record information, the non-person standard value of the infrared sensor of Node 0 during the 10-unit time:  $V_{R0}(\text{non-person standard value}_{10t})$



If  $V_{R0}(t-10 \sim t) \geq V_{R0}(\text{non-person standard value } 10t)$ , then the standard adjustment value  $|V_{R0}(t-10 \sim t)| = 0$  (indicating that the infrared sensor has detected that person exits)

If  $V_{R0}(t-10 \sim t) \leq V_{R0}(\text{non-person standard value } 10t)$ , then the standard adjustment value  $|V_{R0}(t-10 \sim t)| = 0$  (indicating that the infrared sensor has detected person doesn't exits)

**3-2** According to the history record information, 100-interval standard values of the humidity sensor of Node 0 during the 10-unit time can be divided in the following sequence:

$V_{HUM0}(\text{interval-1 standard value } 10t) < V_{HUM0}(\text{interval-2 standard value } 10t) \dots < V_{HUM0}(\text{interval-99 standard value } 10t) < V_{HUM0}(\text{interval-100 standard value } 10t)$

Compare the interval standard value of this node according to the actual value of  $V_{HUM0}(t-10 \sim t)$ , to determine the standard adjustment value  $|V_{HUM0}(t-10 \sim t)|$ .

For example:

If  $V_{HUM0}(\text{interval-43 standard value } 10t) < V_{HUM0}(t-10 \sim t) < V_{HUM0}(\text{interval-44 standard value } 10t)$ , then  $|V_{HUM0}(t-10 \sim t)| = 44$

If  $V_{HUM0}(\text{interval-100 standard value } 10t) < V_{HUM0}(t-10 \sim t)$ , then  $|V_{HUM0}(t-10 \sim t)| = 101$

If  $V_{HUM0}(t-10 \sim t) < V_{HUM0}(\text{interval-1 standard value } 10t)$ , then  $|V_{HUM0}(t-10 \sim t)| = 1$

**3-3** Similar to the calculation method of 3-2, according to the history record information, the standard values of 100 intervals of temperature sensors and gas sensors of Node 0 during the 10-unit time can be divided and the standard values can be obtained respectively by calculation:  $|V_{T0}(t-10 \sim t)|$  and  $|V_{Q0}(t-10 \sim t)|$ .

(4) With the "sorting through distribution<sup>[5]</sup>" method, by taking  $|V_{Q0}(t-10 \sim t)|$  as the first priority sorting basis, and  $C_{Q0}(t)$  as the second priority sorting basis, the current status is statistically determined at node 0. The situation in all "history record information" of Node 0 is determined. If the current state is located in the position of the extreme value of the situation distribution, the second step verification for starting the alarm discrimination will be triggered.

(5) The calculated values of  $|V_{R0}(t-10 \sim t)|$ ,  $|V_{T0}(t-10 \sim t)|$ ,  $V_{HUM0}(t-10 \sim t)$ ,  $C_{T0}(t)$  and  $C_{HUM0}(t)$  will be taken as the reference values for the judgment of the second and third steps instead of for the first step sorting.

**Step 2** When  $|V_{R0}(t-10 \sim t)| = 0$ , we can carry out the alarm value optimization processing under the unmanned condition. In this system application example, we used the accumulator to optimize the weighted gas concentrations at the node of the other six sensors located at the same time within the same location of the central node. The temperature information, humidity information, nearby person status information, air quality information, climate condition information, and the like in the loose coupling were broadcast as a "read only lookup table" to the center node based on the location coordinates, thereby correcting and eliminating the influence of the environmental factors in the loose coupling on the interference generated by the sensor. In the process of big data processing in this system, because of the high real-time requirements of the alarm data on the results, we adopted the Spark data engine for fast and universal cluster calculations to quickly implement big data analysis. We specifically used two types of shared variables: accumulator and broadcast variables to optimize the big data analysis process. Of which, the accumulator was used to aggregate information, and the broadcast variable was used to efficiently distribute larger objects<sup>[9]</sup>.

**Step 3** When  $|V_{R0}(t-10 \sim t)| = 1$ , we conducted alarm value optimization under the manned conditions. Based on the data preprocessing and exploratory data analysis of the loose-coupling variables of the "history record information", we built a prediction model for the gas concentration at this node in the subsequent period, that is, pre-judge the subsequent normal  $|V_{Q0}(t-10 \sim t)|$ ,  $C_{Q0}(t)$  value. In terms of the model building method, we referred to the classical prediction algorithm in "Predicting the amount of seaweeds"<sup>[10]</sup>, regarding different types of nodes as different types of seaweeds, and predicting the background environmental impacts of the nodes in the subsequent period to perform the analysis and judgment on the trend of the normal base value for the threshold value of gas sensors.

After the fast big data analysis and processing, the results of the normal base value of the gas sensor threshold of the node were obtained. The basic value of the gas sensor threshold and the actual measured value were analyzed for a period, and whether to make an alarm is determined based on the result of the final calculation.

## Conclusion

With the rapid development of technologies such as big data technology, information transmission technology, and sensor technology, the use of a multi-sensor-based alarm system architecture can make up for the shortcomings of the original single gas sensor alarm system. On this basis, the data fusion and decision making through artificial intelligence technology can provide more accurate and effective early warning of disasters.

## References

- [1] Liu Yang. Analysis of Problems in Practical Application of Gas Detection and Alarm Device, Architectural Engineering Technology and Design, 2016, vol.15. International Publication Number: 2095-6630. Domestic Issue: 43-9000/ TU
- [2] Feng Cuicui. Principle of Combustible Gas Sensor and Its Application in the Fire Alarm System of Crude Oil Storage Tank, Science and Technology 4 (2014): 107-107.
- [3] Lu Xuming. Combustible Gas Detection System Based on AVR Microcontroller and MCGS Software, Laboratory Research and Exploration 33.12(2014):126-130.
- [4] Zhao Dali et al. Analysis of the Causes of Sensors Affecting the Long-term Stability of Combustible Gas Alarms and False Positives, Sensor World 18.6 (2012): 20-22.
- [5] (U.S.) Donald E. Knuth. The Art of Computer Programming volume 3 Sorting and Searching Second Edition [M]. Compiled and Translated by Su Yunlin, Beijing: National Defense Industry Press, 2002.
- [6] (U.S.) written by Yin Hong (Inmon.W.H), and translated by Wang Zhihai et.al. Data Warehouse (4th edition of original book). - Beijing: Mechanical Industry Press, 2006.8 (2014.5 reprint).
- [7] Huang Yihua, Miao Kaixiang, et al. In-depth Understanding of Big Data: Big Data Processing and Programming Practices. - Beijing: Machinery Industry Press, 2015.6.
- [8] (Indian) written by Vignesh Prajapati, and translated by Li Ming, Wang Weiyang, Sun Sidong et al. R. Hadoop Big Data Analysis and Practical Application. Beijing: Mechanical Industry Press, 2015.10.
- [9] (U.S.) Holden Karau (U.S.) Andy Konwinski (U.S.) Patrick Wendell (Canada) Matei Zaharia with; Wang Daoyuan. Spark fast big data analysis (Learning Spark: Lightning-fast Data Analysis). - Beijing: People's Posts and Telecommunications Press, 2015.9.
- [10] (Portuguese) written by Luis Torgo and translated by Li Hongcheng, Chen Daolun and Wu Liming, Data Mining and R Language. Beijing: Mechanical Industry Press, 2015