Interactive Visualization of DGA Data Based on Multiple Views

Yujie Geng¹, Ying Lin¹, Yan Ma¹, Zhihong Guo¹, Chao Gu¹ and Mingtao Wang²

¹State Grid Shandong Electric Power Research Institute, 2000 Wangyue Road, Jinan, Shandong, China
²Shandong Zhongshi Yitong Co. Ltd., Xinluo Road, Jinan, Shandong, China

E-mail: gengyujie1225@126.com

Abstract. The commission and operation of dissolved gas analysis (DGA) online monitoring makes up for the weakness of traditional DGA method. However, volume and high-dimensional DGA data brings a huge challenge for monitoring and analysis. In this paper, we present a novel interactive visualization model of DGA data based on multiple views. This model imitates multi-angle analysis by combining parallel coordinates, scatter plot matrix and data table. By offering brush, collaborative filter and focus + context technology, this model provides a convenient and flexible interactive way to analyze and understand the DGA data.

1. Introduction

Traditional dissolved gas analysis (DGA) has a long analysis period and couldn’t continuously monitor equipment. Application of on-line monitoring apparatus for dissolved gases in transformer oil makes up for those weaknesses [1]. However, due to the limitation of sensor technology on-line monitoring device is not satisfactory on its reliability and sensitivity [2]. So it is difficult to make a correct judgment for condition of equipment by norms threshold or gas ratio boundaries. With the large range commission and operation of online monitoring device, online monitoring DGA data accumulates continuously. The data analysis capacity of people has lagged far behind the ability to obtain data. Volume and high-dimensional has brought a huge challenge to the Monitoring and analysis of data.

With the development of computer technology, information communication, polymer materials and chromatographic analysis, online dissolved gas analysis monitoring is becoming more and more advanced. But due to the limitation of sensor technology, on-line monitoring device is not satisfactory on its reliability and sensitivity. Besides, there are seven kinds of gas which can contribute to judge the internal fault of the transformer, so more than one gas need to be comprehensively considered in order to judge the status of the transformer correctly.

Visual analysis is an important method used in data analysis. A picture wins a thousand words. For human beings more than 80% outside information are obtained by the visual system. In this paper, we present a novel interactive visualization model of DGA data based on multiple views. This model imitates multi-angles analysis model of users by combining parallel coordinates, scatter plot matrix and data table. We organize massive data into multidimensional data by using multiple views collaborative visual analysis. Users can manage and analysis data directly from different dimension, so that information hidden in the data can be truly understand by users and reflect the status of the transformer correctly.
2. Related Work
The combination of oil and insulation paper is widely used in power transformer insulation. Transformer oil is a kind of mineral oil which is distilled and refined. Insulation oil and paper will be aging and decomposing gradually when affected by the thermal, electrical and mechanical force during transformer operating [3]. Decomposition speed and gas production rate will increase when transformer fault occurs. The decomposed gas will be dissolved in the oil and gas relay after diffusion and convection.

For the diagnosis analysis of dissolved gas, scholars puts forward many visual analysis methods. Triangle method is one of them. In 1971, Davies proposed a diagnosis method based on three elements (H, O, C) according to experiment and experience [4]. In 1974, Duval improved Davies’s method and diagnose the fault of transformer by measuring the relative components of CH4, C2H4, C2H2 [5]. As shown in Figure.1, by checking the position of the measured data in the map, people can diagnose the type of fault according to the fault-location look-up table.

Duval triangle method can realize the quick judgment of equipment fault. However, it depends on the accuracy of the data heavily. Since on-line monitoring device is not satisfactory on its reliability and sensitivity, it is difficult to make an effective judgment by using Duval triangle method. At the same time, the characteristics of high dimension and multi element bring a big challenge to the traditional low dimensional visualization technology.

In addition to the challenges brought by the expansion data scale, the problem caused by high dimension is also a focus of research under the background of big data times. Multidimensional data refers to the data variables with multiple dimension attributes. The goal of multidimensional data analysis is to explore the distribution patterns of multidimensional data items, and to reveal the hidden relationships among different dimension attributes. Keim [6] summarized the basic methods of multidimensional visualization including geometric projection techniques, pixel-oriented techniques, icon-based techniques and graph-based techniques. Among them, the multidimensional visualization method based on geometric projection techniques is the main research direction in recent years.

3. Visualization
There are seven kinds of gas which can contribute to judge the internal fault of the transformer. Single visual analysis technology just analyzes and displays data from a certain aspect and couldn’t judge the status of the transformer correctly. So in this paper, we make use of the advantages of various visualization techniques and analyze complex multidimensional data by multiple views. According to the features of online dissolved gas data, we use parallel coordinates, scatter plot matrix and data table to display and analyze information from different aspects.
Our Multiple views visualization model contains 10 dimensions. They are device ID, monitoring time, hydrogen (H2) content, methane (CH4) content, ethane (C2H6) content, ethylene (C2H4) content, acetylene (C2H2) content, carbon monoxide (CO) content, carbon dioxide (CO2) content and total hydrocarbon content respectively. We use color to code the device ID. H2 content, CH4 content, C2H6 content, C2H4 content, C2H2 content, CO content, CO2 content and total hydrocarbon content are mapped as the eight properties of parallel coordinates. In scatter plot matrix, we constitute a 8 * 8 scatter matrix using the above 8 kinds of gases. In data table, we list all the monitor data in accordance with the order from near to far. Figure 2 shows the overall of our visualization model.

3.1 Parallel Coordinates

Parallel coordinates is an important visualization technique used to display multidimensional data and analyze its relationship. The main idea of parallel coordinates is to map the N-dimensional data attribute to two-dimensional plane by N equidistant parallel axis. Each axis represents a dimension attribute. Its range is well distributed according to the min and max value of the corresponding attribute. A set of points on a line in n-space is transformed to a polyline in parallel coordinates intersecting at all the n−1 points. Those points represent the data points of the N dimensional data and the polyline can be expressed by N-1 linear independent equations as follows:

\[
\frac{x_1-a_1}{k_1} = \frac{x_2-a_2}{k_2} = \frac{x_3-a_3}{k_3} = \cdots = \frac{x_n-a_n}{n} \tag{1}
\]

From formula (1) we can know:

\[x_{i+1} = m_i x_i + b_i , i = 1, 2, ..., n - 1 \tag{2}\]

Where \( m_i = k_{i+1}/k_i \) is the slope and \( b_i = a_{i+1} - m_i a_i \) is the intercept of flat \( x_i/x_{i+1} \) on axis \( x_{i+1} \).

The sequence of dimension could affect people's analysis and judgment for the data in parallel coordinates visualization. Good dimension order allows people to find relevant features and abnormal data efficiently. The usual method is to calculate the similarity between each dimension attribute and rearrange axis. Arrange the biggest similarity axes together can effectively reduce disorder of the data in parallel coordinates and allow users to view the trends and characteristics clearly. However, the time complexity to calculate similarity is O(N!). With the increase of the data dimension and amount, the time and space cost to calculate similarity will increase exponentially. Besides, the relationship between two dimensions will be weaker when the two axes are far apart.

In order to solve that problem, we arrange the two feature gases which content ratio are useful as the adjacent axis according to the transformer oil dissolved gas analysis and judgment guidelines. By doing that, we can not only ensure a relationship of adjacent data dimension but also reduce the disorder of data between the axes. Our final order for the parallel coordinates is C2H6, C2H4, C2H2, H2, CH4 , TN, CO and CO2.

3.2 Scatter Plot Matrix

Scatter plot matrix is one of the most popular multidimensional visualization methods. Scatter plot maps only two of all the dimensions into two axes and draw the scatter among them. It can effectively reveal the relationship or connection between two variables. Scatter Plot Matrix makes use of the above idea and builds a matrix of all two combinations of the multidimensional data. Variables of multidimensional data are plot in the corresponding panel of the matrix. People can get the implicit message from any two variables in each panel. Scatter plot matrix can explain the relationship between each two dimensions intuitively without being affected by size of data set and the number of dimension. It plays a an irreplaceable role in showing pairwise relations of multidimensional data. We use color to code different devices and use a scatter plot shows the cluster distribution of different devices. By doing this, we can make comparison and judgment for the state of device.

3.3 Data Table
All the information of DGA data is shown by a two dimension data table in our paper. The table contains 10 columns which are device ID, H$_2$ content, CH$_4$ content, C$_2$H$_6$ content, C$_2$H$_4$ content, C$_2$H$_2$ content, CO content, CO$_2$ content, total hydrocarbon content and monitoring time respectively. The displayed data items are arranged sequentially from top to bottom according to the monitoring data from near to far.

4. Interaction
In order to analyze and display data from multiple views effectively and conveniently, we apply interaction technology including brush, collaborative filter and focus + context to the multiple views visualization model. Those interactions can associate professionals to analyze data consistently and interactively.

4.1. Brush Technology
Brush technology is a useful visual interactive technology which can reflect the multidimensional data subset intuitively. The basic function is to protrude those data users concerned by highlights and hide those unconcerned data. As shown in Figure 3, click the left button of mouse on the C$_2$H$_2$ axis and drag to a certain point to release, then the parallel coordinates applies an overall filter to DGA monitoring data and only shows the monitoring data whose acetylene values in the chosen ranges.

Users can simply double-click the left mouse on the acetylene axis to cancel the filter. The visualization model also supports multiple brush filters which can apply brush functions to more than one axis for multiple condition analysis. The multiple brush technology can effectively reduce the interference of irrelevant data. It is easy for users to analyze the distribution as well as the relation of interest data.

![Figure 3. Brush technology](image)

4.2. Collaborative Filter Technology
Collaborative filter technology is the core technology to realize cooperative multi-view analysis. In a single view, filter technology can help to analyze the selected core data from the large amounts of data and reduce the impact of unconcerned information. Collaborative filter technology applies the filter of a view to other related views. For example, when placing the mouse on a row of the data table (the gray item of the data table), visualization model will highlight that row. At the same time, the corresponding data item will also be highlighted in both parallel coordinates and scatter plot matrix (as shown in Figure 4). Since the items in data table are arranged sequentially from top to bottom according to the sampling date monitoring data from near to far, when the mouse move over rows of data table from top to bottom successively, we can view the trends of feature gases from the evolution of data in the visualization model. When using brush technique to select core data from parallel coordinates, the scatter plot matrix and data item table will also filter out non-core data so as to analyze the core data by multiple views (As shown in Figure 5).

4.3. Focus + Context Technology
The focus + context technology (F + C) is originated from fisheye view. This technology shows the
concerned data (focus) and the overall context (context) in the same views simultaneously. By the focusing function, the technology apply a selective distortion to the object in the views. The focus object will stand out and the surrounding environment in the context will be gradually reduced. The basic cognitive psychology technology is that people often need to maintain the visibility of whole information space when exploring the local information. If the information space is divided into two display areas, people need to constantly switch attention and memory when exploring information and lead to inefficiencies in cognitive behavior. We apply focus + context technology to parallel coordinates and scatter plot matrix. When placing the mouse on any row of the data table, the corresponding data item will become bigger and highlight in parallel coordinates and scatter plot matrix and the surrounding environment data will became light. (As shown in Figure 4)

5. Conclusion

In this paper, we present a novel interactive visualization model of DGA data based on multiple views including parallel coordinates, scatter plot matrix and data table. This model manages and analyzes data directly from different dimension imitating human.to reveal multidimensional data. And by offering brush, collaborative filter and focus + context technology, this model provides a convenient and flexible interactive way to analyze and understand the data. Our model can improve analysis ability of observers and has a great significance for the development of electrical equipment of online monitoring system in substation.

References

[1] Cai-xin, S. 2005 Present Situation and Development of Condition On-line Monitoring and Diagnosis Technology for Power Transmission and Transformation Equipment. Electric Power, 2005, 38(2): 1-7.

[2] Jian-zhuang, Z., Chun-lei, N. and Jin-hua. 2007 H. The Status of Monitoring On-line Technology of Transformer in Henan Province. Henan Electric Power, 2007, 4:31-33.

[3] Yan-cai, X., Heng-jun, Z., Xiao-yuan, Z., et al. 2006. On-line Monitoring and Diagnosis of Power Transformer Based on Dissolved Gas Analysis. Electric Power Automation Equipment, 2006, 26(6): 93-96.

[4] Leibniz, Uniz. 2008 A Software Implementation of the Duval Triangle Method [C]. Electrical Insulation, 2008. ISEI 2008. Conference Record of the 2008 IEEE International Symposium on, 2008:124-127

[5] Diego, M., Satish, D. 2010 Gas Analysis of a Thermally Overloaded Oil Immersed Current Transformer[C]. IEEE/PES Transmission and Distribution Conference and Exposition. 2010: 645-653

[6] Keim, DA., Kriegel, HP. 1996 Visualization techniques for mining large databases: A comparison. IEEE Trans. on Knowledge and Data Engineering, 1996,8(6):923–938.