# Analysis and Implementation of Multidimensional Data Visualization Methods in Large-Scale Power Internet of Things

Zhoubin Liu[1], Zixiang Wang[1], Boyang Wei[2], and Xiaolu Yuan[3](✉)

[1] State Grid Zhejiang Electric Power Research Institute, Hangzhou 310014, Zhejiang, China
{liuzhoubin, wangzixiang}@zj.sgcc.com.cn
[2] Georgetown University, Washington DC 20007, USA
bw558@georgetown.edu
[3] RUN Corporation, Wuxi 214000, Jiangsu, China
rubin0513@gmail.com

**Abstract.** In the large-scale power Internet of things, a large amount of data is generated due to its diversity. Data visualization technology is very important for people to capture the mathematical characteristics, rules and knowledge of data. People tend to get limited and less valuable information directly form large data when rely only on human-being's cognition. Therefore, people need new means and technologies to help display these data more intuitively and effectively. Data visualization mainly aims at conveying and communicating information clearly and effectively in term of graphical display, which can make data more human-readable and intuitive. Multidimensional data visualization refers to the methods to project multidimensional data to two-dimensional plane. It has important applications in exploratory data analysis, and verification of clustering or classification problems. This paper mainly studies the data visualization algorithm and technology in large-scale power Internet of things. Specifically, the traditional Radviz algorithm is selected and improved. The improved radviz-t algorithm is designed and implemented, and the unknown information of data transmission is obtained by analyzing its visualization effect. Finally, the methods used to study fault detection ability of radviz-t algorithm are discussed in detail.

**Keywords:** Data visualization · Radviz algorithm · Power Internet of Things

## 1 Introduction

Visualization is used to facilitate people's understanding of data in term of graphics display. With the advent of the computer age, the concept of visualization is more enriched, and the process of data visualization is more complex [1, 2]. Visualization technology is the most effective way to interpret data information by manipulating huge data processing in computer into a form that can be intuitively understood. The development of visualization has mainly gone through three stages: scientific


Q. Li et al. (Eds.): BROADNETS 2019, LNICST 303, pp. 135–143, 2019.
https://doi.org/10.1007/978-3-030-36442-7_9

visualization, information visualization and data visualization. Visualization technology is becoming more and more mature, as more tools emerged. In programming field, visualization software can directly translate algorithms into dynamic graphics. Visualization covers a broad range, not limited to data visuals, visuals from daily electrical power production such as electrical substation monitoring videos, furnace flame monitoring videos, and security monitoring videos can also be means of visualization.

The foundation of visualization builds upon data. Exploring the meaning of data, acquiring the information transmitted by data, and expressing data with appropriate visualization methods have also become the core of visualization inquiry [4]. Data are categorized and data processing methods are categorized. Exact same data could focus on dramatically different perspectives. Understanding the data itself, determining the center of gravity of the data, choosing the appropriate method of data transmission, and viewing the effect of data transmission are particularly important in each step of data visualization. Only by accomplishing the steps above could guarantee “deep meaning” contained in the data be displayed, and the data can be better managed, transmitted and saved, and more “secrets” about unknown data be explored.

This paper focuses on the research of visual clustering analysis (Radviz algorithm). Through the recognition of the principle and design implementation, Radviz algorithm still has room for improvement. Two classical large data visualization tools, D3 and R, are used to implement the algorithm, and some indexes are used to analyze the algorithm and explore the significance of improvement. Meanwhile, through the application of multi-dimensional data in the Internet of Things, the strength of Radviz and improved algorithm for fault detection are explored in depth.

## 2 Visualized Background and Radviz Algorithms

Data visualization can be divided into four stages: the era of poor computing, the era of first computing, the era of later computing and the era of big data at present. The era of poor computing refers to the era before the invention of calculators. The era of computing first refers to the era from calculator to modern computer. Post-computing era refers to the era of rapid development of the Internet. The era of big data is the era in which massive data are acquired and processed. In the earliest era of poor computing, people often used numbers and simple graphics to label data; first, in the computing age, formal formats gradually formed; then, in the computing age, Internet computers developed rapidly and charts became very popular; then, in the era of big data, the means of visualization are more varied [6].

With the popularity of computers and the development of the Internet, increasing attentions are paid to data visualization as the system has improved and technology is more cutting-edge. Visualization has become a complete subject and field of study. In different fields, it has related applications [7]. The annual IEEE VIS conference held in the UK is a summary and discussion conference on the field of visualization research. At present, the core technology of the conference is virtual reality technology, which has the characteristics of three-dimensional interaction and provides a good medium environment for the study of more complex structures.

Large and complex dimensions of data limit the amount of information people can directly acquire, thus application or technology to display multi-dimensional data visuals is needed [8]. Because the static state cannot meet the requirements of multi-dimensional data, we use dynamic visuals that can change or display processing through graphic changes. The core of such application is to reduce data dimensionality to dynamically display the information.

Radviz, a radial coordinate visualization method, is a multi-dimensional data visualization technology. It reduces the dimensionality of multi-dimensional data and projects it in low-dimensional space. Through certain characteristics, its clusters and analyses the similar data points. It has a circle arranged in the order of dimensions. Different rankings have a great impact on the final performance. The method also applies the idea of projecting the dimensions into several new dimensions, expanding the sorting space, to obtain better visual clustering effect.

The important concept is called dimension anchor, which fixes dimension to a circle. There visually serves as a spring-like pulling force between the anchors of each dimension. There is a special case when the resultant force becomes zero. At this time, the pulling force of non-neutral dimension is proportional to the value of dimension. When the values of dimensions between data points are similar, their positions in the circle are similar. By describing the clustering structure of data points, the visual clustering effect is displayed. The number of anchors on the ring is high scalable, so that the number of data points in the garden can be increased. The radial projection mechanism of data points using spring force can also maintain the original characteristics of multi-dimensional data in low-dimensional space. At the same time, it can keep the process simple and beautiful.

The prototype of Radviz algorithm was first proposed by Huffman et al. Radviz technology was proposed by Ankerst in 1996. Since then, various methods of multi-dimensional data analysis related to Radviz have been elvolved. Radviz is widely used in biology and medicine because of its outstanding ability to express genetic data. Radviz algorithm also has certain strength for fault detection, making it wide-applicable, such as sensor networks [10].

However, there are many constraints on the Radviz algorithm itself. Firstly, as the core content of the algorithm, dimension anchor cannot be operated at will, whether to add, delete, move or rearrange. This feature facilitates the exploration and application of Radviz in the field of interaction, but at the same time, the order of finding dimension anchors also increases the workload of users. It is a NP problem to find the best ranking method of dimension anchors. Some researchers point out that the quality of dimension anchors ranking can be guaranteed by heuristic strategy. They have creatively designed a vectorized Radviz called VRV. VRV is the number of dimensions of the extended algorithm, which makes the space of dimension anchor sorting become larger, and the way to obtain it become more flexible, from which we can find a more optimized result [11].

## 3 Radviz's Principle and Performance Evaluation

The dimensionality reduction of multidimensional data can be classified into linear and non-linear categories. Radviz is a non-linear visualization technology based on geometric technology. The dimension of multi-dimensional data is m. The results of these data are mapped to a circle through $R_m \sim R_2$. The number of springs obtained is n and the distances between springs are qual. Each dimension of data has its own characteristic variables, which are the observed values after mapping each point. Each spring is fixed at both ends, with one end at the observation point of the circumference and the other end $\{S_1 \sim S_m\}$ on the variables mapped by the data points. As shown in the figure (m = 6) (Fig. 1):

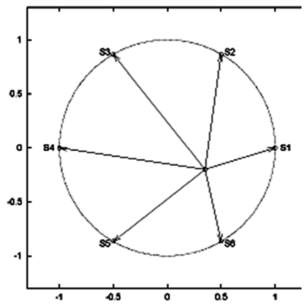


**Fig. 1.** Traditional Radviz model

The m-dimensional data mapping has a total of M data features, which are mapped to he circumference of two-dimensional space. The circumference is evenly divided into m parts. The data points are mapped into variables $\{S_1 \sim S_m\}$ on the circumference, that is, the equivalence points of the circumference. After the m-dimension mapping, every equal point will produce a spring. There are m springs in total, and the elastic coefficients of each spring are different. Assuming that the coordinates of equilibrium point are $(X_i, Y_i)$ in the circle, the point is the projection point of the m-dimensional data space point $\{S_1, S_2, S_3\ldots, S_m\}$ mapped to the two-dimensional space, and the equilibrium position of the point is $Ai = \{X_i, Y_i\}\ T_j$. The tension relationship between the observation point of the equilibrium position and the point on the circumference of the other end of the spring is expressed as $u_{ij} = S_j - A_i$. Among them, $S_j = (d_{1j}, d_{2j})$ represents the $j$*th* variable mapped to the circumference of two-dimensional space, and $A_i = (X_{1i}, Y_{2i})$.

According to the nature of spring, the spring force is determined by the stretching length and coefficient of spring. $K_{ij}$ represent stiffness constant of a scalar spring. When the spring is in a static state at the equilibrium point, the resultant force of the spring is 0.

$$\sum\nolimits_{j=1}^{m} f_{ij} = 0,\ that\ is\ \sum\nolimits_{j=1}^{m} \left(S_j - A_i\right)k_{ij} = 0$$

According to the above formula, the position of equilibrium point can be obtained.

$$A_i = \frac{\sum_{j=1,2,\ldots,m} k_{ij} S_j}{\sum_{j=1,2,\ldots,m} k_{ij}}$$

Suppose there is a number w:

$$w_{ij} = \left(\sum_{j=1}^{m} k_{ij}\right)^{-1} k_{ij}$$

Then the formula of the value of the observation point is obtained.

$$A_i = \sum_{j=1}^{m} w_{ij} S_j$$

In order to avoid the possible negative value of the coordinate axis, the initial data should be processed. According to Radviz visual measurement model, the evaluation indexes of the algorithm are mainly divided into three aspects: data scale, visual effect and feature preservation.

## 4 Improved Radviz-t Algorithm and System Implementation

Due to the impact of the era of big data, the application of traditional Radviz algorithm has become extremely limited, which is unable to deal with big data [12]. For this reason, Radviz algorithm is improved.

To avoid the disadvantage of the original Radviz algorithm, which maps all data points directly to two-dimensional space, occlusion coincidence occurs between the points. Although the data dimension points correspond to the circumference, the arrangement order on the circumference also has a great impact on the final rendering. The criteria to select the RVM model indicators to improve the visualization effect mainly involve three indicators of the visualization effect: density, chaos and coverage. According to the previous model, in order to optimize the visualization effect mapped by Radviz, the chaos must be increased, the density must be reduced, and the coverage must be low. The new algorithm is called Radviz-t algorithm.

Indicator formulas for visual effects: Assuming the data set is $\{B_1, B_2, B_3, \ldots, B_m\}$ and the projection point is $A_i = (X_i, Y_i)$, the total coverage of $B_i$ for $B_j$ is as follows:

$$OCC(B) = \sum_{i=1}^{m} \left[ \sum_{j=1, i \neq j}^{m} OCC(B_i, B_j) \right] i \neq j = 1, 2, \ldots, m$$

$B_i$ density: _defaults to 2_:

$$CLO(B_i) = \begin{cases} 1, \exists A_j \in \{A_1, A_2, \ldots, A_m\} \text{ and } \sqrt{(A_{i1} - A_{j1})^2 + (A_{i2} - A_{j2})^2} < \eta, i \neq j \\ 0, \text{ other} \end{cases}$$

Total density:

$$CLO(B) = \sum_{i=1}^{m} CLO(B_i)$$

Visual effects are expressed by ß, data coverage, density and confusion are expressed by c, g and l, respectively. According to the research, the overall relationship between them is as follows:

$$\beta = 1 - (0.228c + 0.403g + 0.369l)$$

From the formula above that there are three main indicators for visualization effect, namely, density, coverage and confusion. There is a certain relationship among these three indicators. When the density, coverage and chaos are given a certain set of values, the relationship is the most balanced and the visualization effect is the best. This balance is called "view effect value" and is expressed in Q.

$$Q_{c,g,l} = Q\{c, g, l\}$$

Given the multi-dimensional data is visualized, in order to present the best results, *m* times of random experiments are carried out and Monte Carlo algorithm is used to find the optimal case. Suppose that in the first experiment, there is a mode $a_i$, which keeps the optimal projection performance within a certain number of times.

$$\exists \mathrm{a} = \sum_{i=1}^{m} x_{max}$$

The final visual effect of the algorithm is determined by remove the visual effect value in this mode, in which the optimization is realized. Assuming that in AI mode, the probability of optimal projection effect is x, the visual effect value in this mode is $Q_{max}$, and the data coverage, data density and data confusion of the three index values are expressed by c\ g\ l, respectively, the formula is derived:

$$Q_{max} = \sum_{i=1}^{m} a_i\{x_{max}\}$$

From this, we can see that the visualization effect of this algorithm reaches the best state, which satisfies the modification of the visualization effect of Radviz algorithm. Because the improved algorithm needs to be designed for the optimal selection, the probability is selected to a certain extent, and the maximum probability is selected, from the probability distribution histogram. Through the visualization design of probability distribution histogram, we can directly get the function of the algorithm for data exploration and selection. According to the principle, probability distribution histogram is the necessary content for the improvement of Radviz-t algorithm. We use probability distribution histogram to calculate the index of visual view effect, and then select the index value that can present the optimal visualization effect for implementation.

The improved algorithm adds random experiment module and view effect evaluation module to the original algorithm. Random experiment module randomizes the order of array; first it randomizes the order of dimension ranking to ensure the randomness of the experiment, then randomly carries out multiple groups of experiments, and chooses the best visual effect map according to the view effect value. After starting the random experiment, in order to disrupt the dimension ranking, the original Radviz Implementation dimension ranking is chosen to numbering the variables of each dimension from 1 to m to form an array, and the random algorithm is used to disrupt the processing, so that the ranking of arrays is different in every second, and the visual effect of the output is also different to ensure the random experiment to follow. The visual effect changes when the order of dimension changes every second. To describe this effect, the view effect value is used. In the current second dimension order, the coordinates of the observation points in the current view are established and printed. Then, according to the formula principle, four modules are set up to calculate the data density, coverage, confusion, and the final calculation of the view effect value.

The data are processed by random module and then entered the optimal selection module. The module uses Monte Carlo algorithm to call the calculated visual effect value according to the characteristics of smaller view effect value and better visual effect. The idea of bubble sorting is used to ensure the minimum value of view effect, and the dimension corresponding to the visual effect map and the current visual effect map is sorted. In each second of random experiment, if a smaller view effect value appears, the original view effect value and its visual effect map and dimension ranking will be saved. Until the end of the experiment, if there is no smaller view effect value, it will output the view effect value saved in the previous second and the corresponding dimension ranking and visualization effect of the value, which is called optimal visualization, the whole process served as optimal selection.

For the start and end of the random experiment, the number and time of the experiment can be controlled manually by the button. In order to better explore data, when improving, the default choice of dimensions is changed to all choices, and the exploration of dimensions is abandoned. The advantage is faster exploration speed. The overall design page diagram is shown in the following Fig. 2:

Characteristic analysis:

(1) There is no need to select dimensions;
(2) Random ranking of dimensions is changed;
(3) The beginning and end of the experiment can be controlled at any time;
(4) Retaining color selectivity;
(5) Record the current optimal view effect value;
(6) Output the number of random experiments and the optimal visualization effect.

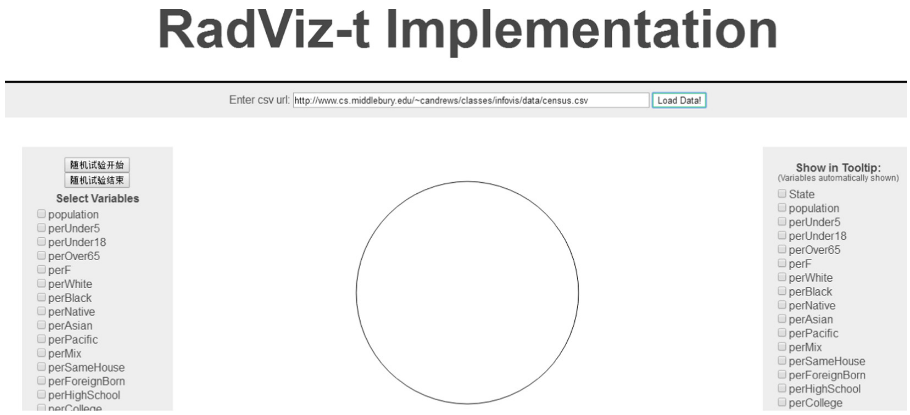


**Fig. 2.** The improved page is shown as Radviz-t Implementation

Radviz Implementation proposed that the smaller the view effect value, the better the view effect is. When number of random trials increases, the view effect value gradually decreases. Therefore, increasing the number of random experiments can improve the visualization effect. This means it can also increase the probability of finding the optimal visualization effect map of the whole data. Radviz algorithm itself has the ability of fault diagnosis. When identifying data points, it chooses to mark outliers with purple color. The ID of all fault points is 0. However, when dealing with high-dimensional data, the fault state of outliers cannot be observed intuitively, so Radviz-t is generated based on the original algorithm. Radviz-t Implementation based on Radviz-t increases the probability of choosing the optimal visualization effect map with the increase of random experiment numbers, by which the display of fault state becomes more and more obvious. The fault point begins to separate from the convergent normal area and produces an independent area, that is, the coverage between the fault point and the observation point in the normal area is less than 0. Finally, user can directly observe the purple point to get the information of the point in the current failure state (Fig. 3).

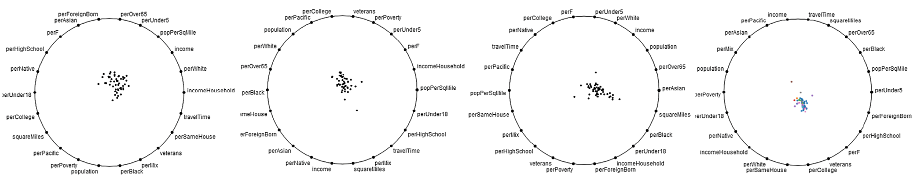

**Fig. 3.** Radviz-t dimension sorting change per second (uncolored and colored)

## 5 Conclusion

This paper first introduces the importance of data visualization in the power Internet of Things. Then, the principle and design of original Radviz algorithm and improved Radviz-t algorithm are discussed. An algorithm module is added to the original model to improve the original algorithm. When presenting the overall design, in addition to improving the original algorithm, the function of the original algorithm has been adjusted accordingly, and the Radviz-t implementation corresponding to the improved algorithm has been realized. This paper also gives a brief introduction to the new functions implemented by Radviz-t Implementation and analyses its properties. Finally, through the application of actual power data, the implementation of the algorithm and the improved algorithm is tested, and the correctness of the algorithm is verified.

**Acknowledgements.** This work was supported by State Grid Zhejiang Electric Power Corporation Technology Project (Grant No. 5211DS16001R).

## References

1. Sharko, J., Grinstein, G., Marx, K.A.: Vectorized Radviz and its application to multiple cluster datasets. IEEE Trans. Vis. Comput. Graph. **14**(6) (2008)
2. Novikova, E., Kotenko, I.: Visual analytics for detecting anomalous activity in mobile money transfer services. In: Teufel, S., Min, T.A., You, I., Weippl, E. (eds.) CD-ARES 2014. LNCS, vol. 8708, pp. 63–78. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10975-6_5
3. Lehmann, D.J., Theisel, H.: General projective maps for multidimensional data projection. In: Computer Graphics Forum, vol. 35, no. 2 (2016)
4. Daniels, K., Grinstein, G., Russell, A., Glidden, M.: Properties of normalized radial visualizations. Inform. Vis. **11**(4), 273–300 (2012)
5. Ravichandran, S., Chandrasekar, R.K., Uluagac, A.S., Beyah, R.: A simple visualization and programming framework for wireless sensor networks: PROVIZ. Ad Hoc Netw. **53**, 1–16 (2016)
6. Orsi, R.: Use of multiple cluster analysis methods to explore the validity of a community outcomes concept map. Eval. Program Plann. **60**, 277–283 (2016)
7. Keim, D.A., Mansmann, F., Schneidewind, J., Ziegler, H.: Challenges in visual data analysis. In: Tenth International Conference on Information Visualization, IV 2006, pp. 9–16. IEEE (2006)
8. Rao, T. R., Mitra, P., Bhatt, R., Goswami, A.: The big data system, components, tools, and technologies: a survey. Knowl. Inform. Syst. 1–81 (2018)
9. Li, P., Chen, Z., Yang, L.T., Zhang, Q., Deen, M.J.: Deep convolutional computation model for feature learning on big data in internet of things. IEEE Trans. Ind. Inform. **14**(2), 790–798 (2018)
10. Senaratne, H., et al.: Urban mobility analysis with mobile network data: a visual analytics approach. IEEE Trans. Intell. Transp. Syst. **19**(5), 1537–1546 (2018)
11. Zhou, F., Huang, W., Li, J., Huang, Y., Shi, Y., Zhao, Y.: Extending dimensions in radviz based on mean shift. In: 2015 IEEE Pacific Visualization Symposium (PacificVis), pp. 111–115. IEEE (2015)
12. Shi, L., Liao, Q., He, Y., Li, R., Striegel, A., Su, Z.: SAVE: Sensor anomaly visualization engine. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 201–210. IEEE (2011)