



Segmented Time-Series Plot: A New Design Technique for Visualization of Industrial Data

Tian Lei, Nan Ni^(✉), Ken Chen, and Xin He

School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China
nanni@hust.edu.cn

Abstract. Time-series plots have been widely used in the fields of data analysis and data mining because of its good visual characteristics. However, when researching and analyzing the massive data formed in industrial, some shortcomings of the traditional time-series plot make the visualization of big data ineffective, which is not conducive to data analysis and mining.

In this paper, the traditional time-series graph is improved and a segmented time-series plot that can be used for massive industrial data analysis is proposed. In addition, this paper describes in detail the steps of making the segmented time-series plot. The method can reduce the information overload and the interface issues by limiting the amount of information presented.

Keywords: Data visualization · Design technique · Industrial data Segmentation method

1 Background

In recent years, data visualization has been widely used and put into function due to its convenience, comprehensibility and accuracy in industry. For example, through the integration of images, 3D animation and computer-controlled technology with the solid model, the visualization technology of operation simulation visualizes the state of the equipment so that the manager has a specific concept of the equipment as well as its position, shape and all other parameters [1]. However, how to display the industrial data better to the user has been a new topic in industry, with the further application of information visualization in industry, especially in the face of massive, widely distributed, complex, fast processed and uneven evaluated industry data [2].

2 Status of Industry Data Visualization

Most industrial data are time series which are collected at different points of time and reflect the change status of things or phenomena. Therefore, the time-series plot is usually used for visualization. Time-series plot is also called transition diagram that describes the variable in relation with time. Scatter plot, bar chart and broken line chart can all be used to visualize time-series data. For example, the graph below shows China's GDP and GDP growth rate from 1985 to 2015 in relation with time [3]. This

figure shows the data clearly to readers. However, different challenges appear when the data visualization in industry is concerned compared with Internet data visualization (Fig. 1).

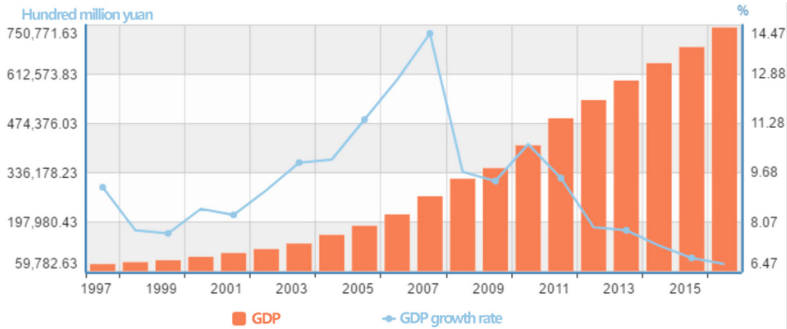


Fig. 1. China's GDP and GDP growth rate from 1985 to 2015

The first problem that needs to be solved is that the huge amount of data leads to a failure of effective and immediate display in the monitoring points. As industrial data comes primarily from various sensors and a common device may have been equipped with hundreds of sensors that update at a very high frequency, the amount of data in industry is much larger compared with traditional Internet data. How can so many categories of data be arranged and selected on a narrow screen?

The second problem is how to balance 'big' with 'small'. The problem is divided into two aspects, one of which is how the whole and the part can be effectively combined. The staff should pay attention to not only the overall trend of the data but also the abnormal values. In the face of such nested data, the staff needs to grasp the overall situation while capturing some local changes quickly. The other aspect is how to balance local and details. In a display of local data, we also want to get some specific details including related data, historical data, abnormal data, data trends, data forecasting, etc. And this requires designers to grasp the relationship between local and detail display.

The last problem lies in the translation efficiency from data into effective information for the searching and delivery of users. It has been common in Internet companies, such as Taobao and Jingdong from China, to use big data to analyze their users' habits and hobbies to find their favorite topics and push the relevant products [4]. However, this practice with big data is rarely utilized in industry.

These issues can lead to poor readability and low efficiency of the graphics, thus leading to poor user experience (Fig. 2).

Based on these problems, different scholars have made their own attempts to visualize time-series data. The main goal of visualizing time-series data is to reduce the dimensions and reduce the noise interference [5]. Among the various solutions, visualization method based on segmentation is frequently used for its advantages in data compression and noise filtering. Li proposed that the data should be sorted and

classified according to certain requirements and arranged according to a certain order so as to facilitate people to analyze and study the problems [6].

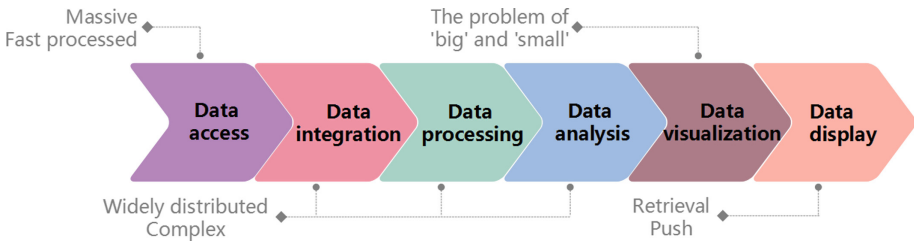


Fig. 2. Difficulties in data visualization

According to the differences in segmentation methods, segmentation-based visualization method can be divided into PAA [7] and PLR [8]. PAA approximates the entire sequence with the average of each segment by equally dividing the time series. PLR approximates the original time series with a number of straight line segments that are adjacent to each other whose interval are not necessary equal. The PAA method is rarely applied because it uses only equal division without considering the actual situation of the time-series plot, thus failing in retain the change trends of the original data.

Tian et al. described the superiority of segmentation-based visualization and improved the segmentation of time-series data at important points based on PLR theory. This method well retains the trend of the original data and clearly shows the abnormal and important points [9]. Similarly, Yu et al. also proposed a new feature-point-based segmentation method that can accurately represent time-series data, which preserves the main state features of the time series [10]. However, the PLR method does not guarantee that there will be only one basic trend in each segment, thus focusing too much on local details while ignoring the overall features.

In order to make up for the shortcomings of PLR, this study proposes a more effective method – segmented time-series plot – to achieve the visualization of time-series plot from the perspective of massive data display and balance between “big” and “small”.

3 Industrial Data Visualization Based on Segmented Time Series

This method proposes a segment processing to the traditional time-series plot based on the existing situation of presenting industrial data with a time-series plot. The specific steps are as follows.

- (1) The data is segmented with PLR as follows (R is the preset distance threshold);
 - Take the start and end point as the initial segment point;
 - Find the point that keep the largest distance from the segment line which should be greater than R ;

- Take the points that meet the previous conditions as new segment points;
 - If there is any point in the segments that has a distance greater than R , go to step 2. If not, the segmentation is then finished [11].
- (2) Process each segment into a separate time series. That is, to extract the data from each segment to remake a time-series plot, the height of which is constant.
 - (3) Distinguish each time-series plot according to the design elements such as graphics, color, text, etc. The way we recommend here is to differentiate each time-series plot with different color backgrounds. But pay attention that the background should not be so fancy that user's attention is transferred;
 - (4) Connect the remade time-series plot in chronological order. Starting from the second segment, the time-series plot is shifted longitudinally according to the positions of each segment point until the initial point of the second segment coincides with the end point of the first segment. And repeat this step until the initial point of all the segments coincides with the end of the previous segments. Gestalt psychology suggests that people tend to percept a part as the whole. According to the continuity principle in perceptual organization, these sections are relatively easy to be perceived as a whole because they are connected at segment points [12].
 - (5) Conduct interaction design for the connected time-series plot. Human-computer interaction technology in information visualization can be summarized into five main categories: Dynamic Filtering Technology, Overview + Detail Technology, Pan + Zoom Technology, Focus + Context Technology, Multi-view Correlation Technology [13].

4 Case Analysis

4.1 Data Source

The data presented comes from experiments on a milling machine under various operating conditions. There are 16 sets of experimental data with differences in the three independent variables. Three different types of sensors (sonic sensors, vibration sensors and current sensors) are utilized for data acquisition. The lateral wear is not measured constantly but at intervals.

The data is organized in a Matlab structure shown in Table 1.

Table 1. Data structure

Field name	Description
case	Case number (1–16)
run	Counter for experimental runs in each case
time	Duration of experiment (restarts for each case)
DOC	Depth of cut (does not vary for each case)
feed	Feed (does not vary for each case)
material	Material (does not vary for each case)
smcDC	DC spindle motor current

Take case 1 (depth of cut is 1.5 mm, feed is 0.5 mm, material A) as an example. The current curve of the DC spindle motor in relation with time is shown in Fig. 3.

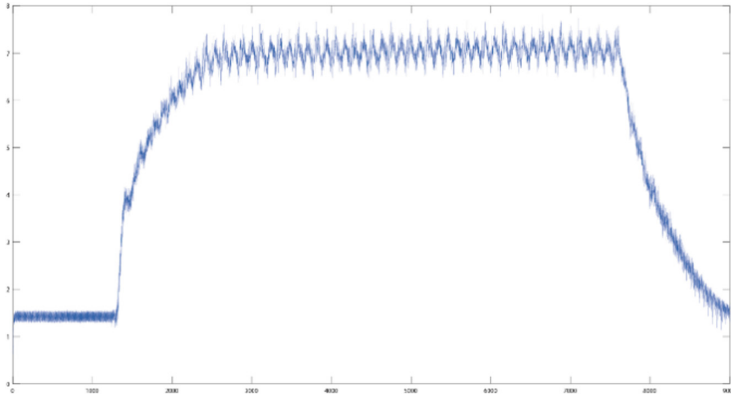


Fig. 3. The DC spindle motor of case 1

4.2 Visualization Process

Time-Series Plot Segmentation

Firstly, start point A and end point B are connected as shown in Fig. 4.

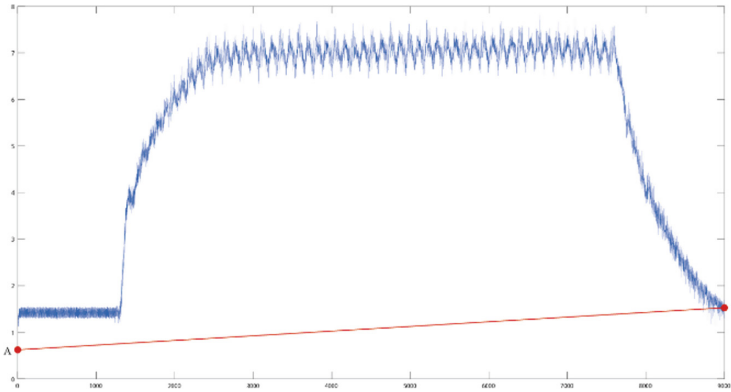


Fig. 4. Connect A, B

With R preset to 1.4, find the farthest point C from the line AB and tell whether the distance from point C to line AB is greater than R . If yes, point C will be chosen as a new segment point and connected with A and B respectively as shown in Fig. 5. If the distance is less than R , R should be properly adjusted. The value of R is determined by

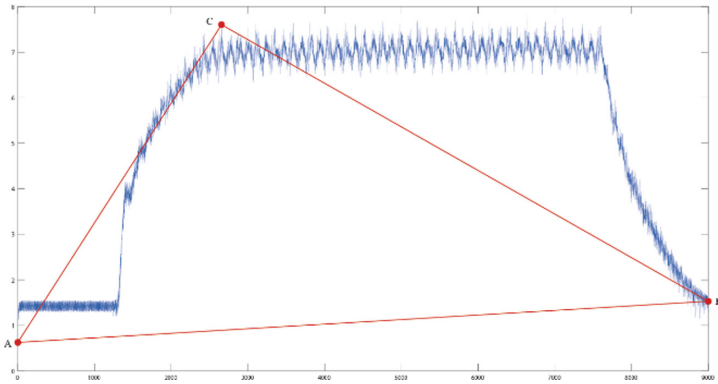


Fig. 5. Connect A, C and B, C

the degree of subdivision required and the ordinate of the time-series plot, which is proportional to the degree of subdivision.

Repeat this step until the farthest distance from all points to the segment line in each segment is less than R . Then the segmentation is finished. The time-series plot is finally divided into four sections, namely a, b, c and d, as shown in Fig. 6.

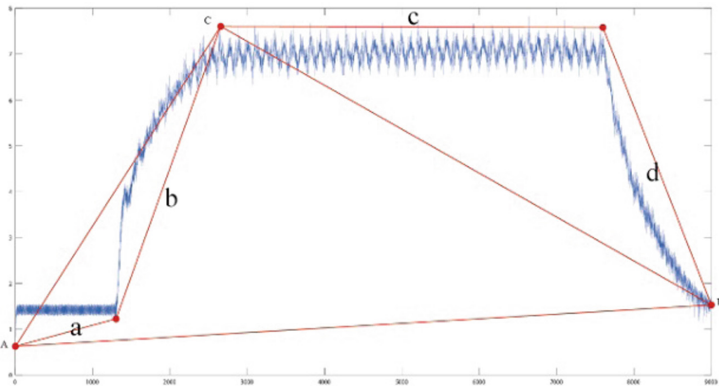


Fig. 6. Segmentation is completed

Independent Remake

First, extract each segment's data and make it into a time-series plot while ensuring that the height of each segment is approximately the same. It should be noted in this step that the same method is used to visualize each segment of the data. We still use the line chart here as shown in Fig. 7.

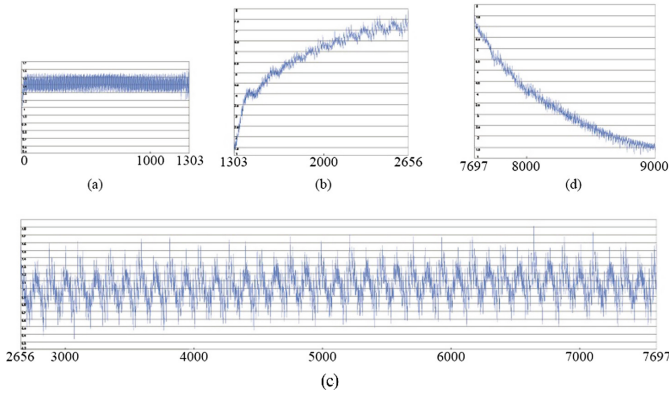


Fig. 7. Independent remake

Visual Distinction

At this step, the background of the line chart for each segment is reset. According to the range of ordinate values, the background color of the line chart with smaller ordinate value is replaced by the light blue while the darker blue is applied in line charts with larger ordinate value, as shown in Fig. 8. It should be noted that the background color and the polyline color should have a greater distinction.

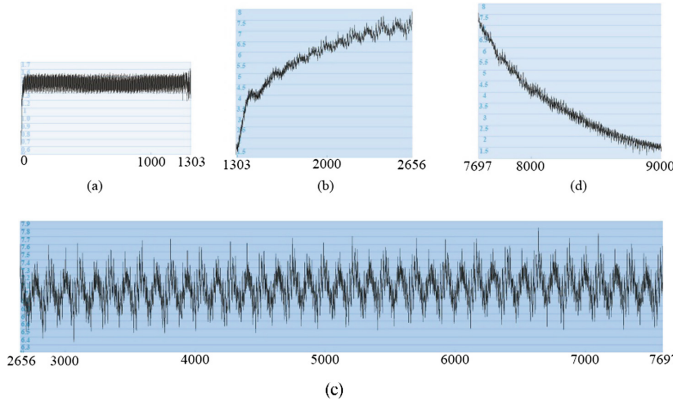


Fig. 8. Visual distinction. (Color figure online)

Connection

Connect the remake line charts together in the original chronological order. The initial point of the second segment is coincident with the end point of the first segment. And this step is repeated until the initial point of all the segments coincide with the end point of previous segments. The result is shown in Fig. 9.

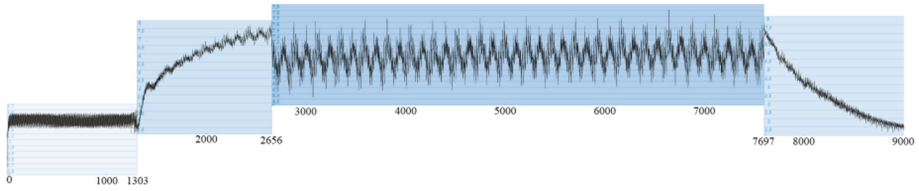


Fig. 9. Connection

Add Interactive Action

Then the time-series plot is basically completed and will be displayed as the main part in the interface of Data Viewer (DV) software. DV is a data visualization tool for data analysts and their superiors. The design of DV should meet three requirements of users: First, it should support switching among multiple variables and timelines and be able to explore and navigate data interactively. Second, it should be scalable to handle large amounts of time-series data. Third, it should be better readable (Fig. 10).

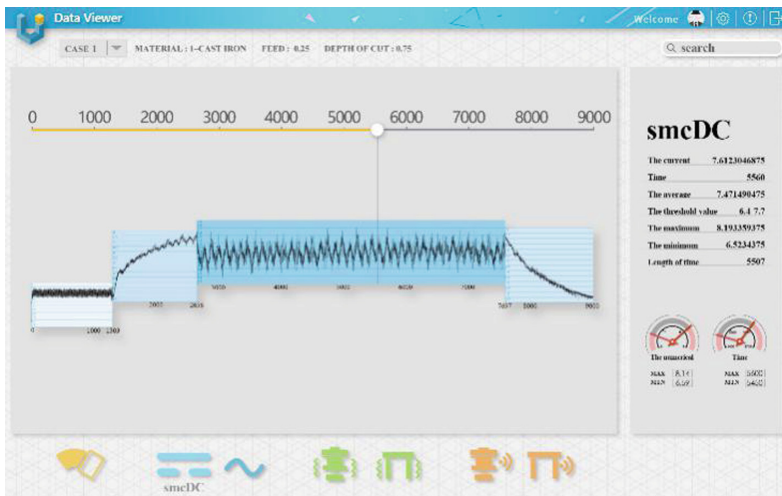


Fig. 10. Main interface of DV

The segmented time-series plot is displayed in the central part of the interface.

Limiting the amount of information presented can alleviate information overload and interface issues both overall and in part. Interaction is added here.

First, when clicking on a segment of the segmented time-series plot, the user can read the threshold lines and special values carefully by enlarging the selected area while compressing other areas. This method not only shows all the segments but also ensures a complete temporal context.

Second, the DV also provides a standard slider assembly and a timeline. The time slider divides the data in the time-series plot into two types: regular data and selected

data. The data pointed to by the slider is displayed and the rest of data except the outlier is hidden. The specifics of this area are shown on the right side of the screen.

Finally, the page shows only one case in the DV interface. But the user can click the upper left button to switch. Different types of data in the same case are symbolized and displayed at the bottom of the page. The data type can be switched by clicking buttons.

At the same time, DV also provides search capabilities. But the search interface provides the ability to find data instead of discovering information. It supports keyword searches on time-series data, such as time and values.

5 Summary

In this paper, we introduce the segmented time-series plot that visualize the massive time-series data from industry. The segmented time-series plot improves the traditional time-series plot, which shows the details of the data in a multi-level way while clearly showing the trends in the data. Furthermore, this method also provides interactive techniques to support the industrial data analysis in a visual way.

In addition, we are still considering optimizing the visual interface. This can provide a better way to compare time-series data. In the future design, we will extend the analytical capabilities to support more demanding tasks, for example, visualizing data from various data sources or using more mathematical analysis.

References

1. Akpan, I.J., Shanker, M.: The confirmed realities and myths about the benefits and costs of 3D visualization and virtual reality in discrete event modeling and simulation: a descriptive meta-analysis of evidence from research and practice. *Comput. Ind. Eng.* **112**, 197–211 (2017)
2. Jiang, X.F., Li, J.Y.: Research on the digital industrial design system integrated with product data management system. *Adv. Mater. Res.* **102**, 106–110 (2010)
3. 王孜弘: 新常态下中美综合经济实力对比——基于国内生产总值的分析. *美国研究* **30**(05), 31–49+6 2016. (in Chinese)
4. Gil, D., Song, I.-Y.: Modeling and management of big data: challenges and opportunities. *Future Gener. Comput. Syst.* **63**, 96–99 (2016)
5. Lu, Z., Zhangm, Q.: Clustering by data competition. *Sci. China (Inf. Sci.)* **56**(01), 65–77 (2013)
6. 李晓明: 采用Excel制作数据分段统计表的研究. *电脑知识与技术* **10**(23), 5572–5579 (2014). (in Chinese)
7. Lin, W.-M., Gow, H.-J., Tsay, M.-T.: A partition approach algorithm for nonconvex economic dispatch. *Int. J. Electr. Power Energy Syst.* **29**(5), 432–438 (2006)
8. Stoica, P., Soderstrom, T., Ahlen, A., Solbrand, G.: On the convergence of pseudo-linear regression algorithms. *Int. J. Control* **41**(6), 1429–1444 (1985)
9. 田野, 张忠能: 改进的基于重要点的时间序列数据分段方法. *微型电脑应用* **28**(2), 48–51 (2012). (in Chinese)
10. 喻高瞻, 彭宏, 胡劲松, 等: 时间序列数据分段线性表示. *计算机应用与软件* **24**(12), 17–18 (2007). (in Chinese)

11. 孙焕良,邱邦华,魏溯华. 一种优化的自底向上时间序列分段算法. 沈阳建筑大学学报(自然科学版) (06), 1049–1052 (2007). (in Chinese). Assessed 12 Oct 2017
12. Evergreen, S., Metzner, C.: Design principles for data visualization in evaluation. *New Dir. Eval.* **2013**(140), 5–20 (2013)
13. Cockburn, A., Karlson, A., Bederson, B.B.: A review of overview + detail, zooming, and focus + context interfaces. *ACM Comput. Surv. (CSUR)* **41**(1), 2 (2009)