Visualization for compressed natural gas(CNG) secondary station refilling behaviors

Yang Li^{1,2,*}, Yulian Zhao¹, and Jian Shao²

¹College of Engineering and Technology, Tianjin Agricultural University, 300384 Tianjin, China

²Tianjin Huabei Gas & Heating Engineering Design Co. LTD, 300384 Tianjin, China

Abstract. Data visualizations are recently used for providing an access to complex data and information. In this paper, the visualization method is adopted to find possible ways of improving CNG secondary filling station's refilling efficiency. Based on single attribute of volume, initial and final pressure, the visualization results show that one possible way to improve refilling efficiency is to by avoid abnormal inefficient refilling behaviors with small volume, higher initial pressure or lower final pressure. Based on time, the visualization results, considering the periodic property of refilling behaviors, present another possible way to improve refilling efficiency through reducing inefficient use of time. From the refilling log's visualization, it is found that there is room for improvement in refilling efficiency and possible improving ways are primary studied. But the reasons of these abnormal actions are needed to be studied in future research.

1 Introduction

The transaction log data collected from a compressed natural gas(CNG) secondary filling station contains plenty of valuable information. These refilling logs record comprehensive and concrete refilling behavioral information in a CNG secondary filling station, which thus are of great value in mining the refilling pattern, detecting the abnormal refiling behaviors, and improving the CNG secondary filling station's efficiency.

Data visualization is a useful technique to extract information from complex raw data such as log data[1]. Visualization provides some means to see what lies within, find relations between attributes, and perhaps apprehend things which could not be seen directly in other forms.

Within this paper we apply the concept of visualization to present CNG secondary filling station's refilling log data. This presentation is used to identify patterns of these refilling behaviors, and is used to show that there are room to improve the refilling efficiency.

Traditional liquid automobile fuels, such as petrol and diesel, suffer problems of environmental pollution and high cost of production. As an alternative, CNG usually is used as clean fuel for vehicles[2]. In additional, the relative low price of natural gas makes tend to adopt natural gas as

transportation fuel. There has been more than 5.3 million CNGVs in China by the end of 2016[3-5], which means that China has been on the first position for three consecutive years. However, CNG filling infrastructures, interference with urban land-use planning and long-time administrative approval are the major barriers of this industry to scale up. Especially, limited CNG filling stations result in long waiting time for CNGVs to refilling. Therefore, finding a way to improve CNG filling stations' refilling efficiency is very important under the current infrastructure condition[4]. In this article, visualization methods are used to find out the possible sections of CNG filling station in improving its refilling efficiency.

The rest of this article is structured as follows: Section 2 overviews the data set visualized. In Section 3, the visualization methods are introduced. And the visualization results are presented in Section 4. Finally, discussions and conclusions are given in the last section.

2 Dataset

The data set used in this article is from a CNG secondary filling station in Tianjin. This data set contains 125,021 transaction logs of this station over a period of 3 months from May to July in 2014. The data log record each transaction with 24 attributes,

* Corresponding author: author@e-mail.org

such as transaction time, No. of dispenser gun, refilling volume, initial pressure, and final pressure, etc. Table 1 is part of the first 10 records of this data set.

Table 1. Part of the first 10 records of the data set used in this article.

Transaction time	Volume	Initial pressure	Final pressure
2014-5-1 0:01	14.11	4.5	18.6
2014-5-1 0:01	18.95	0.6	18.9
2014-5-1 0:01	17.89	1.5	18.7
2014-5-1 0:01	14.12	2.4	18.8
2014-5-1 0:01	15	4.4	18.7
2014-5-1 0:04	12.66	6.4	19
2014-5-1 0:04	18.66	1.6	19.5
2014-5-1 0:04	13.12	1.5	19.2
2014-5-1 0:04	15.34	4.3	19.4
2014-5-1 0:05	11.59	7.6	19.9

In Table 1, the first column is the transaction time, which is described by t in this article. And the 2nd column is the volume of natural gas filled in the CNGV. The 3th and final columns record the pressure of the CNGV before and after the refilling behavior. Volume, initial pressure and final pressure of each refilling action are abbreviated to v, pi, pf respectively in this article.

3 Methodology

In this study, the CNG dataset is displayed in excel format (.xls). Before visualization, the data set is cleaned first. The data cleaning process is to ensure that the CNG refilling dataset were clean by removing inconsistent and missing data.

After data cleaning, the cleaned data set is loaded into a Mongoddb database. Then, the visualization methods from matplotlib is used to visualize these refilling behaviors.

Firstly, for each attribute of volume, initial pressure and final pressure, its scatter plot is given. For attribute of time, its periodicity of each day is fully

considered in visualization. CNG filling station's efficiency is closely related with the time interval between adjacent refilling behaviors. Therefore a derived variable, dt, is introduced by subtraction of the adjacent refilling action's time. Finally, then an scatter plot matrix of multiple attributes is presented.

Fig. 1. Research methodology

4 Visualization results

4.1. Visualization based on single attribute

To improve a CNG secondary filling station's refilling efficiency, a simple way is to see whether there is some refilling actions abnormal. And the attributes of volume, initial and final pressure are description of a refilling behavior. Therefore, visualizations based on these three attributes are plotted mainly with Matplotlib, a package of Python. Figure 2 shows the visualization results.

In Figure 2, visualization results of initial pressure, volume and final pressure are presented in each row. In the first column, scatter plot of each attribute is given. And the boxplot method[6-8] is used to detect outliers based on each single attribute in the second column. The abnormal refilling behaviors detected based each attribute are labeled by red color. Kernel density estimation methods[9,10] are adopted to estimate distribution of each attribute with and without the outliers detected by boxplot. Taken the bandwidth in the kernel density estimation as the bin width, the histograms are plotted in the 3rd column in Figure 2.

Fig. 2. Visualization results of v, pi and pf.

4.2 Visualization based on time

Time has a special role in improving efficiency because of each day has limited 24 hours. Therefore, time attribute in these dataset is visualized in this section. For its periodicity, the visualization is designed carefully to reveal any shortage that needed to be improved. The visualization result fully considered periodicity is shown in Figure 3.

Fig. 3. Visualization results of time.

In Figure 3, the small blue vertical line segments represent the refilling actions. The horizontal and vertical coordinates are hour and day of each refilling behavior. The white space between small blue line segments means that there is no refilling action, which provided the possible room to improve this station's filling efficiency. Meanwhile, the small line segments with red color are represented the abnormal refilling behaviors detected by volume. From the randomly temporal distribution of red lines, there is

no clear regularities of distribution of the abnormal refilling behaviors.

4.3 Visualization based multiple and derived attributes

From the above visualization results of the two subsections, it is found that possible ways to improve the CNG secondary filling station's refilling efficiency are from the abnormal behaviors detected based on the behavioral attributes of volume, initial and final pressure and ineffective use of time. Therefore, it is necessary to visualize their relationships.

However, transaction time is not a good variable describing time use. Three variables are introduced. The first one is the time interval between adjacent refilling behaviors, which is defined as

$$dt(i) = t(i) - t(i-1) \quad (1)$$

where $t(i)$ means the i th refilling action of this station. And the time interval is divided into two parts. One is the time span used by the $(i-1)$ th refilling action, which is described by dt_{eng} in this research, and the remaining time span, dt_{pre} , is used for preparation for the i th refilling action. The relation between the three time intervals is

$$dt = dt_{\text{eng}} + dt_{\text{pre}} \quad (2)$$

At the same time, volume is a important variable for both seller and buyer, which is the base for settling accounts. From the perspective of physics, the directive relation with volume is the change of pressure during the refiling process, which is defined as

$$dp = pf - pi \quad (3)$$

Finally, the relationships of volume, dp , and the three time intervals is visualized in Figure 4.

In Figure 4, the subfigure in each cell represents the scatter plot of its horizontal and vertical coordinates. The subfigure in the diagonal line is the density distribution of each attribute, which is estimated by kernel density estimated method.

Fig. 4. Visualization results of Scatterplot matrix of v , dp , and three time intervals.

5 Discussions and conclusion

5.1 Discussions

For its environmental friendship and economy, CNG is appreciated by both the government and consumers. But the infrastructure of CNG is not meeting the need. Therefore, the need in time is to improve the CNG secondary station's refilling efficiency. To find the possible ways to increase this efficiency, visualization is a primary exploration on this issue.

In this article, the attributes of volume, initial and final pressure, transaction time are adopted in visualization study. From the behavior variables of volume, initial and final pressure, the scatter plots show that there are possible abnormal refilling behaviors. Then boxplot method is used to detect these suspicious actions. To validate this detection results, the density distributions of these behavior attributes is estimated by kernel density estimated method. From this visualization results, it is found there are really abnormal refilling behaviors with small volume, higher initial pressure or lower final pressure. These behaviors are ineffective and needed to be avoided.

On the other hand, one possible way may be to improve the efficiency use of time. The visualization result of time shows that there are too long time intervals without refilling actions.

Therefore, abnormal refilling behaviors and unusually long time interval without refilling are need to be further studied. From the temporal distribution of abnormal refilling behaviors, there is no clear regularity. A scatter matrix plot shows that volume,

pressure change and different time intervals have abnormal relationships, which need to be studied in future research.

5.2 Conclusion

In conclusion, to improve refilling efficiency, based on single attribute, multiple attributes and derived variables, visualization is applied to dataset from a CNG secondary filling station in Tianjin.

Firstly, through visualization of volume, initial and final pressure, it is found that there are possible abnormal inefficient refilling behaviors. By boxplot method and kernel density estimated method, these abnormal behaviors with little volume, higher initial pressure, or lower final pressure, are detected. Then one way to improve the CNG secondary filling station's efficiency is by avoiding these abnormal inefficient refilling actions.

Secondly, by visualization of time, considering the periodicity, it is found that there are inefficient use of time, which is represented in the visualization result by white space. Therefore, another way to improve the CNG secondary filling station's efficiency is to optimize the operation of the station, and make the time use more efficient.

Meanwhile, in the visualization of time, the irregular temporary distribution of abnormal behavior with tiny volume is found. Further, the relationships of volume, derived variables of pressure change, and three time intervals are visualized by scatter matrix plot. From this visualization result, there are unusual relations between these variables. To find the reasons of abnormal behaviors with small volume, and to describe and explain these unusual relations, further research is needed after the primary study by visualization in this article.

Acknowledgments

This work was financially supported by Science and Technology Program of Tianjin (14ZXCXGX00 392).

References

1. J. W. Han, M. Kamber, J. Pei, *Data mining concepts and techniques(3rd)*(China Machine Press, Beijing, 2015)
2. J. X. Guo, *CNG application and safety* (China Petrochemical Press, Beijing, 2015)
3. Oil and Gas Division of National Bureau of Energy of China, *2016 China Natural Gas Development Report* (Petroleum Industry Press, Beijing, 2016)
4. H. Hao, F. Z. Liu, W. Q. Li, *Renewable and sustainable energy reviews* **62**, (2016)

5. <http://www.ngvchina.com/exhibition-review/activities/>.
6. T.W. Tukey, *Exploratory data analysis*(Addison-Wesley publishing company, Philippines, 1977)
7. T. Hastie, R. S. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*(2nd)(Springer, 2009)
8. C. C. Agarwall, *Outlier analysis (2nd)* (Springer, 2016)
9. M. C. Jones, Statistics and computing, **3**, (1993)
10. https://pythonhosted.org/PyQt-Fit/mod_kde_methods.html#id1