

Empirical Study on Abnormal CNG Refilling Behaviors Based on Single Attribute

Yang Li^{1,3,*}, Yulian Zhao¹, Dengchao Jin¹, Jinhui Sun², Jian Shao³

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

² Tianjin Key Lab of Aqua-Ecology and Aquaculture, College of Fisheries, Tianjin Agricultural University, Tianjin 300384, China

³ Tianjin Huabei Gas & Heating Engineering Design Co. LTD, Tianjin 300384, China

Email address: liyang005@163.com(Y. Li)

Abstract. This study used mainly boxplot method to detect abnormal CNG refueling behaviors. The ownership of CNG vehicles in China has ranked the first in the world for many years. However, due to difficulties such as approval of land, CNG filling stations are difficult to meet the increasing demand for refueling. Therefore, the boxplot method was used to study more than 0.25 million CNG refilling records from a secondary CNG filling station in Tianjin to improve the refilling efficiency in this article. The study found that there were about 2.5% abnormal refilling behaviors based on volume, initial pressure and final pressure separately. To verify those results, another unsupervised learning method of kernel density estimation was used to estimate the probability density of each attribute. Finally, it was found that repeated refilling actions were the main reasons of abnormal behaviors detected based on volume attribute. The results suggested that the way to reduce abnormal CNG refilling behaviors was to avoid repeated refilling actions.

1. Introduction

Since the 1930s, Italy has developed CNG vehicles because of its oil depletion but rich in natural gas^[1]. In 2017, the number of compressed natural gas vehicles(CNGVs) in China will be more than 5 million units, with gas consumption exceeding 20 billion cubic meters, 10 million CNGVs by 2020^[2].

As an alternative fuel to transportation materials such as gasoline and diesel, CNG has been rapidly developed mainly due to its economy, environmental friendliness and the need for decentralized transport energy due to national energy security. CNG filling station occupies a significant position in the promotion of CNGV. However, CNG filling stations and other infrastructure develop slow since the refueling station as an infrastructure faced many difficulties, such as land approval, the relatively high land price and others^[3]. This leads CNGV drivers to wait for a longer time to refill their CNGVs. Therefore, under the existing CNG filling station conditions, it is a more urgent issue to optimize the equipment and human resource allocation to achieve the improvement of the refilling efficiency. One reason caused the poor efficiency is the abnormal refilling behaviors, which were empirically studied by data mining technologies and the corresponding measures were analyzed in this paper.

The rest of this paper is organized as follows. Section 2 introduces the dataset and methods used in this research. Section 3 describes the data mining results. And the main reason of these abnormal behaviors and discussion are given in section 4. Conclusions are provided in the final section.



2. Dataset and methods

2.1. Dataset

This research used dataset containing a total of more than 0.25 million refilling records, from one CNG filling station in Tianjin, which were generated from from January 1 to December 31, 2014.

Each record contains 24 attribute variables, such as the transaction time, refilling gun number, refilling action’s serial number, storage number, volume, amount of money, price, card number, card area belongs, user number, initial pressure, final pressure, filling time, etc..Table 1 is a sample with the first 15 records of the dataset.

The data, after cleaning and processing, was stored into the MongoDB database for subsequent analysis and mining.

Table 1 The first 15 records of the dataset

Index	Time	Volume	Money	Price	Initial Pressure	Final Pressure
1	1-1-2014 0:00	17.53	73.63	4.2	4.2	17.7
2	1-1-2014 0:01	13.53	56.83	4.2	7.2	18.6
3	1-1-2014 0:03	8.58	36.04	4.2	10.2	18.4
4	1-1-2014 0:05	14.47	60.77	4.2	6.4	18.3
5	1-1-2014 0:08	14.23	59.77	4.2	6.4	18.1
6	1-1-2014 0:11	14.75	61.95	4.2	6.3	18.5
7	1-1-2014 0:12	13.65	57.33	4.2	7.2	18.3
8	1-1-2014 0:14	8.63	36.25	4.2	8.9	17.6
9	1-1-2014 0:15	21.02	88.28	4.2	2.1	18.2
10	1-1-2014 0:17	16.05	67.41	4.2	5.1	17.6
11	1-1-2014 0:18	14.3	60.06	4.2	7	18.7
12	1-1-2014 0:19	6.53	27.43	4.2	10.8	18.4
13	1-1-2014 0:21	16.84	70.73	4.2	2.4	18.4
14	1-1-2014 0:22	9.14	38.39	4.2	8.7	18.3
15	1-1-2014 0:24	24.17	101.51	4.2	0	18.5

2.2. Methods

Volume(v), initial pressure(π_i) and final pressure(π_f) are the three main quantities describing each CNG refilling behavior. It is common sense: a normal and effective CNG refilling behavior is a refilling action with larger amount of refilling volume, lower initial pressure and higher final pressure. A major purpose of this paper is to quantify those critical values of abnormal refilling behaviors.

There are many methods to realize this quantifying purpose^[4]. Due to the lack of markers for abnormal behavior, unsupervised learning methods is the only choice. This research used mainly boxplot method based on a single attribute^[5] and kernel density estimation(KDE)^[6].

The core idea of the boxplot method^[5,7] is the five statistics for the calculation of the data, i.e. the median, the first and third quartiles q_1 , q_3 , and the upper critical value(F_U) and lower critical value(F_L) of outliers. The last two quantities are introduced by the interquartile range (IQR) as follows

$$IQR = q_3 - q_1$$

And F_U and F_L are defined as

$$F_U = q_3 + k_U \times IQR$$

$$F_L = q_1 - k_L \times IQR$$

The same values can be taken for k_U and k_L , generally 1.5^[5] and also different values^[8,9]

J.W.Tukey^[5] mentioned that the outlier values founded by the boxplot only given the possibility that these values become outliers, and further analysis was needed. Therefore, for data with non-Gaussian distribution, boxplot standard method can be used as a tentative identification firstly. Because the main objective is to find possible ways to improve the efficiency of CNG stations by data mining, the abnormal gas-filling behavior has been clearly defined in this research, and we can tentatively identify these abnormal behaviors by using the boxplot standard method with k_U and k_L taken 1.5.

In order to further verify the results of the boxplot, the unsupervised learning method of KDE is used to directly estimate the probability density function of each single attribute of CNG refilling behaviors. The core idea of KDE is to estimate the probability density function value at x using observations around the variable x (the number of observations used is determined by either the kernel density bandwidth h or the nearest neighbor number k)^[6]. The kernel density function is a function that takes into account that these observations have different effects on the estimation. This paper uses the Gaussian distribution as a kernel density function. Again, the pressure variables involved range from 0 to 20 MPa, while the volume variable is greater than zero. Therefore, it is required to cut off the boundary in KDE and to normalize it in the estimation range^[10], i.e. the `pyqt_fit.kde_methods.renormalization` method^[11] is adopted in this research.

3. Results

3.1. Abnormal refilling behavior detection based on volume

As can be seen from Figure 1, based on volume, the boxplot method detects the abnormal refilling behaviors with critical values F_U and F_L . The quantified critical values use superscripts i and 1 respectively to represent the result of iterative and single boxplot detection methods. Figure 1 shows the upper critical and lower critical values identified by the two methods, and draws the detection results in horizontal lines. The data points located outside these horizontal lines are abnormal behaviors.

In the case of small amounts of data, most of the literature using the boxplot only once. However, for this research involving 254841 data, the results of using the standard method once in F_U^1 and F_L^1 in Figure 1 are not satisfactory. Following the idea of iterative detection of outliers as provided by C.C. Agarwall^[7], this article iteratively uses the boxplot standard method until there is no more anomalous values left.

As an alternative to the boxplot method, the right panel of Figure 1 shows the results of the KDE method. The bin width in the bar chart of Figure 1 directly exploits the bandwidth in the kernel density estimation. In the smaller volume area, the probability density function has a local maximum, so it is reasonable to take refilling behaviors whose volume belongs to this area as abnormal actions with low effectiveness.

There are 6533 data points falling behind the critical volume value 1.020 m^3 , which means these behaviors are detected by iterative boxplot method as abnormal actions. The abnormal rate is 2.56%. In the right panel of Figure 1, values less than F_L^1 have a percentage of 2.25 by calculating the area under the pdf estimated by KDE.

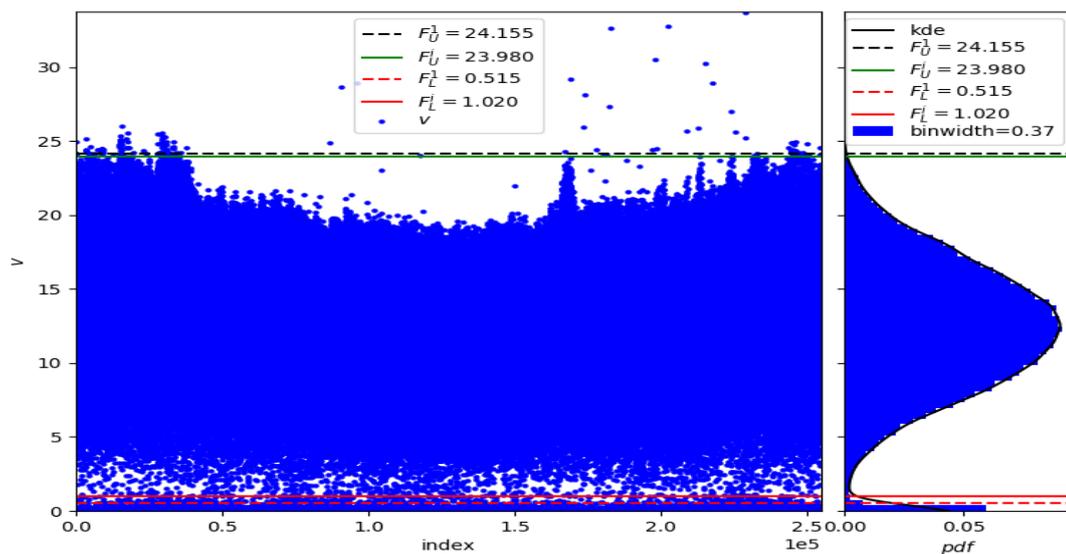


Figure 1. Abnormal detection results based on volume.

3.2. Abnormal refilling behavior detection based on pressure

Pressure of the CNG refilling behavior is another important attribute. Initial pressure and final pressure are used to describe the pressure state of CNGV storage before and after refilling. The detection results are shown in Figure 2 and 3 for initial and final pressure respectively.

In Figure 2, there is no abnormal behavior detected based on initial pressure by using boxplot method. But from the right panel of Figure 2, a bin between 0 to 0.29 MPa has the largest pdf value, which means abnormal behaviors. However, because of lower initial pressure meaning a CNGV needing to be refilled, this abnormal behaviors have higher CNG refilling efficiency as the common sense tells us. So this actions are not concerned in this paper.

In Figure 3, the iterative boxplot method and non iterative method have the same results. Especially, the abnormal behaviors with larger final pressure detected by boxplot method are not the purpose of this article, because a higher final pressure usually means a more volume would be filled. Just as the abnormal behaviors with larger volume detected by boxplot method in Figure 1, these actions have appreciate CNG refilling efficiency, and not need to be avoided.

The abnormal behaviors in Figure 2 contain 2.47% of total data points, and the same percentage in the pdf estimated by KDE. There are 5983 behaviors with final pressure less than 16.90 MPa in Figure 3. And the area under the pdf with value less than 16.90 MPa is 2.46%.

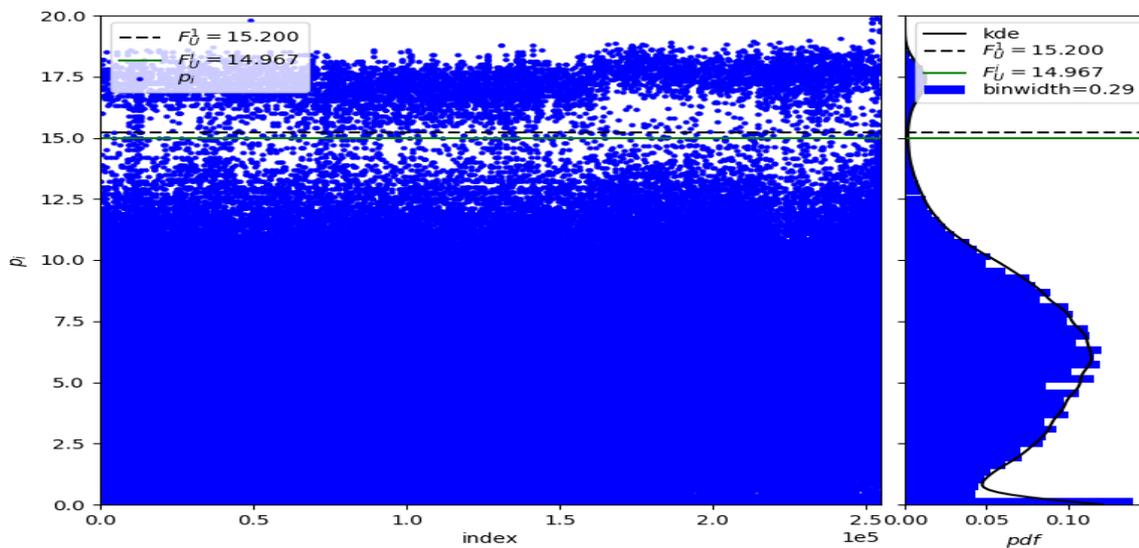


Figure 2. Abnormal detection results based on initial pressure.

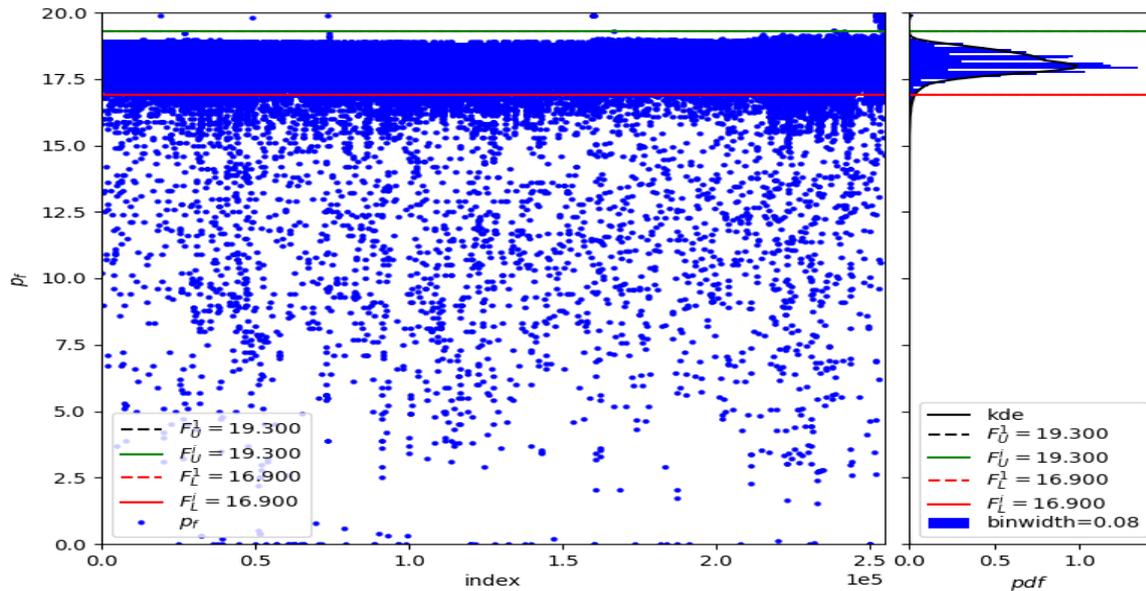


Figure 3. Abnormal detection results based on final pressure.

4. Discussion

By using boxplot and KDE abnormal detection methods, abnormal CNG refilling behaviors are detected based upon volume, initial pressure and final pressure separately. These data mining results are shown by Figure 1 to 3. For volume and final pressure attributes, the abnormal refilling behaviors detected based on each of them fall on the both outsides of F_U and F_L . However, from the point of improving refilling efficiency, only the F_L critical value are useful for volume and final pressure. A refilling action with volume larger than F_U or with final pressure higher than F_U deserves appreciation.

The volume is directly related to the money spent at each refilling action, and is the most important performance parameter in the national standard (GB/T 19273 2003) and verification regulation (JJG 996-2012) of CNG dispensers. Thus, volume is the main abnormal detection attribute in this research.

To explain why the abnormal behaviors detected based on volume happen, a python program is written and run for analysis. It is found that the abnormal actions detected by volume are caused mostly by repeated refilling actions. The statistics of abnormal behaviors and repeated refilling actions are summarized in Table 2. Values in the parenthesis are results of non iterative boxplot method. Value in each cell represent the number of CNG refilling actions under both its column and row attributes.

Table 2 Statistics of abnormal behaviors and repeat refilling actions.

Attributes	Volume ($v <= F_L^i$)	Initial pressure ($p_i >= F_U^j$)	Final pressure ($p_f <= F_L^i$)	Repeated
Volume ($v <= F_L^i$)	6533(5999)	5548	706	6293
Initial pressure ($p_i >= F_U^j$)		6291(6185)	184	5672
Final pressure ($p_f <= F_L^i$)			5983(5983)	1046
Repeated				12503

From Table 2, repeated refilling actions are the main reason of abnormal behaviors detected by volume. The best way to reduce abnormal behaviors is to avoid repeating, because these actions fill few volume of natural gas. The abnormal behaviors detected by initial pressure can be explained by volume,

and finally by repeated refilling actions. But the abnormal behaviors detected by final pressure are needed to be explained by other reason in future research.

5. Conclusions

With single attribute, the abnormal behaviors of CNG secondary filling station are studied in this paper by boxplot and kernel density estimation. Based on each refilling attributes of volume, initial pressure and final pressure, the data mining results show that there are about 2.5 percent of total refilling behaviors being considered as abnormal refilling actions. These abnormal behaviors have lower volume, higher initial pressure, or relative lower final pressure, which are considered as inefficiency. A further analysis shows that the abnormal behaviors detected based on different single attribute are not all the same. Especially, outliers based on volume and initial pressure have more than 85 percent actions in the same ones. But outliers based on final pressure have less than 10 percent same results with volume or initial pressure. At the same time, abnormal behaviors with small volume or relative higher initial pressure can be mainly explained by repeated refilling actions. Therefore, a way to improve the refilling efficiency of CNG secondary filling station is to avoid repeated refilling behaviors. However, the repeated actions account little for these abnormal refilling behaviors with relative lower final pressure, which means further research is needed.

Acknowledgments

This work was financially supported by Science and Technology Program of Tianjin (14ZXCXGX00392).

References

- [1] Guo JX, CNG application and safety[M], Beijing: China Petrochemical Press, 2015.
- [2] Oil and Gas Division of National Bureau of Energy of China, 2016 China Natural Gas Development Report [R], Beijing: Petroleum Industry Press, 2016.
- [3] Hao H, Liu FZ, Li WQ, Natural gas as vehicle fuel in China: A review [J], Renewable and sustainable energy reviews, 2016, 62: p521-533.
- [4] Han JW, Kamber M, Pei J, Data mining concepts and techniques(3rd)[M],Beijing: China Machine Press,2015:p573-575.
- [5] Tukey TW,Exploratory data analysis [M], Philippines: Addison-Wesley publishing company, 1977. p27-56.
- [6] Hastie T, Tibshirani R, Friedman J, The elements of statistical learning: data mining, inference, and prediction(2nd)[M],Springer, 2009, pp.191-211.
- [7] Agarwall CC, Outlier analysis (2nd) [M], Switzerland: Springer, 2016.
- [8] Carling K, Resistant outlier rules and the non-Gaussian case [J]. Comput. Statist. Data Anal. 2000, 33: 3. pp. 249-258.
- [9] Schwertman NC, de Silva R, Identifying outliers with sequential fences [J]. Computational Statistics and Data Analysis, 2007, 51, pp. 3800-3810.
- [10] Jones MC, Simple boundary correction for kernel density estimation [J], Statistics and computing, 1993, 3: p135-145.
- [11] https://pythonhosted.org/PyQt-Fit/mod_kde_methods.html#id1