International Journal of Applied Science and Mathematics
Volume 2, Issue 5, ISSN (Online): 2394-2894

Feature Extraction and Forecasting Model of Gas Vehicle Energy Consumption Based on Time Series

Zhang Yan
Email: whw2046@126.com

Jiang Yaming
Email: jiangym@ms.xjb.ac.cn

Xue Huajian
Email: xuehj@ms.xjb.ac.cn

Zeng Wenxiao
Email: zy@ms.xjb.ac.cn

Wang Haiwei
Email: zwx@ms.xjb.ac.cn

Zhang Ronghui
Email: zrh1981819@126.com

Xinjiang Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China;
Xinjiang IOT Perception and Control Laboratory, Urumqi 830011, China.

***Abstract* – The thesis carried out the research work of the scale level data analysis method, mainly based on mass perception data that Vehicle Cylinder IOT Electronic Supervision System accumulated. The large data analysis platform was built based on Hadoop. Then, the method of building data warehouse and data layout schem was designed. Using the platform with large-scale cylinder gas-adding data, statistical analysis was performed. With the analysis methods of energy consumption for time series, feature extraction and feature selection was worked on. With Hidden Markov Model method, vehicle gas prediction model was established. And the data layout scheme was put forward under the Hadoop environment that is applicable to the perceptual data features rapid extraction. At last, the technical reference was provided for the gas vehicle IOT intelligence oversight, which was the development and application of CMIS technology in gas vehicle energy consumption prediction research.**

***Keywords* – WSN, Data Analysis, Markov Chain, The Internet of Things, Vehicle Engineering, CMIS.**

I. Introduction

The source of Big Data is wide, its application requirements and data types are different, but the basic process consistent. The entire processing flow of the Big Data can be defined as: Under the appropriate auxiliary tool, a wide range of heterogeneous data source is extracted and integrated, the results is stored according to certain standard. The data stored is analyzed with the use of proper data analysis technology, from which the useful knowledge is extracted, and using the appropriate ways the end users are shown results. Specifically data analysis technology can be divided into data extraction and data integration, data analysis and data interpretation[1-3].

The energy forecasting and early warning in order to study the security of energy supply as the main content, is an emerging discipline, which has been developed in the field of energy system science in recent years[4]. Many scholars and institutions at home and abroad have conducted more extensive studies on the energy consumption problem, also put forward a number of effective energy consumption forecasting method, which mainly include input and output, MEDEE, LEAP, trend extrapolation method, elasticity coefficient method, exponential smoothing method, time series model, etc. In view of the complicated nonlinear and uncertainty in energy systems, in recent years many scholars has began to research the characteristic and uncertainty of non-linear system, to introduce nonlinear method to forecast the demand for energy. Commonly-used method such as chaos theory, genetic algorithm, the grey theory and neural network, etc., which can make up the shortfall for the linear method in the prediction of complex energy requirements[5-7].

Natural gas that is used by gas vehicle cylinders as a clean energy, is an important component of energy supply in china. Early warning of the consumption of natural gas by the way of scientific method, is also an important aspect of energy forecasting and warning. Time series of the consumption of natural gas for vehicle is Modeled in Turpan with HMM model, the traditional modeling method of the HMM model is improved using Big Data technology, the vehicle gas consumption was forecast, and which provide technical support for the government's energy policy[8].

II. Data Warehouse and Data Layout Scheme of Cylinder Filling Information Based and Built on Hadoop

The construction of the data warehouse of gas cylinder filling information, aimed at efficient storage and analysis of continuously generated massive cylinder filling data, and from which valuable regional energy consumption information is generated, so as to achieve the mastery and use of regional energy consumption comprehensively and effectively. Hive can be used to construct data warehouse in the Hadoop platform, that Facebook proposed, is a large data warehouse based on Hadoop, and its goal is to simplify the Hadoop data aggregation, AD hoc query and analysis of large data set and etc., in order to alleviate the burden of the programmer[10]. It uses the technology of Schema management, SQL interface and etc. of relational database, to map the structured data file to a database table, and provide HiveQL similar to SQL descriptive language for programmers to use, and automatically parse HiveQL statement into an optimization execution sequence of MapReduce task. In addition, it also supports user-defined MapReduce function[11-12].

In the data warehouse, how effective large-scale gas cylinder filling data is distributed to different locations of it, is an important research subject. Reasonable data layout can effectively reduce the data transmission at the Hadoop platform, improve the computing speed of the data[13]. At

present, the layout method of large-scale data in Hadoop environment mainly are studied at the three aspects on reducing data transfer across the data center, data dependency and global load balancing. In this paper, by the way of the data dependency and global load balancing, research and development on distribution scheme of large-scale gas cylinder filling data was carried on, and it studied the data layout method that can quickly extract gas cylinder filling data feature[14-16].

Fig.1. Experiment platform of automatic filling of gas vehicle

III. Feature Extraction and Feature Selection of Time Series of Energy Consumption

The feature is the basic unit of data analysis and processing. In dealing with of large-scale gas cylinder filling data, the feature of the filling data was obtained by using statistical methods, such as in the month forecast, then monthly the statistics of the following features was Carried on:

Gas vehicle number this month, total number of times of gas-adding this month, volume of gas at most a day of this month, volume of gas at least a day of this month, volume of gas at most a week of this month, volume of gas at least a week of this month, total gas volume this month.

As is noticed, the feature set has the characteristics of high dimension. Usually, the high dimensional feature contains many irrelevant, redundant features. The emergence of the redundant features, not only increased the dimension of feature space and reduce the efficiency of learning, but also increased the possibility of noise data, which interfered with the learning process of the learning and the mining algorithm, and ultimately affected the structure of the classification model. Therefore, in order to reduce the adverse factors, the dimension of feature space was decreased and irrelevant or redundant features was eliminated. According to certain evaluation criteria, an optimal or the most effective feature subset was selected instead of the original feature space, which achieved the purpose to cut the dimension of feature space.

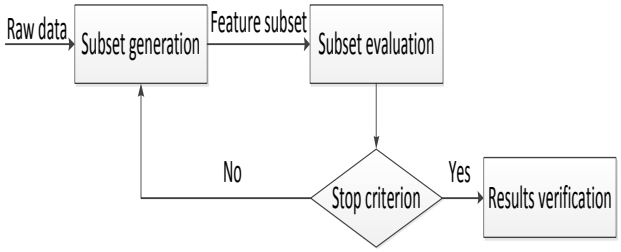

Fig.2. Flow chart of algorithm on feature dimension reduction

IV. The Prediction Model of Time Series of Gas Vehicle Energy Consumption Based on the HMM Model

Based on the time series data that was observed through the system, time series analysis is a kind of theory and method of the mathematical model which is established by curve fitting and parameter estimation. Time series analysis commonly used in the national economy macro control, integrated regional development planning, market latent capacity prediction, etc. Through the cloud platform, the time series of features was obtained by Statistical analysis on behalf of the features monthly. As in the case of Turpan, statistics can be shown in the table below:

Table 1: Feature sequence on energy consumption monthly

Time	Gas vehicle number monthly	Gas volume monthly	Volume of gas at most a week of a month	Volume of gas at least a week of a month
201301	15000	25000	10000	6000
201302	15100	29000	11000	6500
201303	15300	30000	13000	7500

As the modeling tool of a kind of time series, HMM has been widely applied in the field of speech recognition, ECG analysis, stock price time series prediction and etc. Hidden Markov Model (referred to as HMM) is used to describe a Markov process with unknown parameters implied. The difficulty was to determine the hidden parameters of the process among the observable parameters. Then these parameters were applied to further analysis. HMM is suitable for dealing with complex time series data. The state of the HMM could not be directly observed, but could be obtained through the observation of vector sequence, each observation vector showed various state through certain probability density distribution and it was generated by the state sequence of a corresponding probability density distribution.

Fig.3. Schematic diagram of motor vehicle gas-adding

V. The Experimental Results on the Platform

The filling data in the security system of gas cylinders includes gas vehicle filling volume and filling time, figure 4 shows the curve graph of gas vehicle installation monthly, figure 5 does the line chart of gas volume total, Figure 6 does the comparison chart of forecasting gas volume total in the same period. Through the statistical analysis of gas-adding data stored in the system, monthly vehicle gas volume of consumption at pilot areas was got, time-series data analysis technology was applied to these statistics, and based on the actual situation to make some improvements, which was used for the establishment of the vehicle gas consumption prediction model at the pilot areas, and provided an important basis for the energy scheduling of the government at the pilot areas. At the same time, these statistics also reflected the LPG stations operating situation, it helped to improve the operation of LPG stations, and could be taken as the basis for whether to expansion of LPG stations.

Fig.4. The curve graph of gas vehicle installation monthly

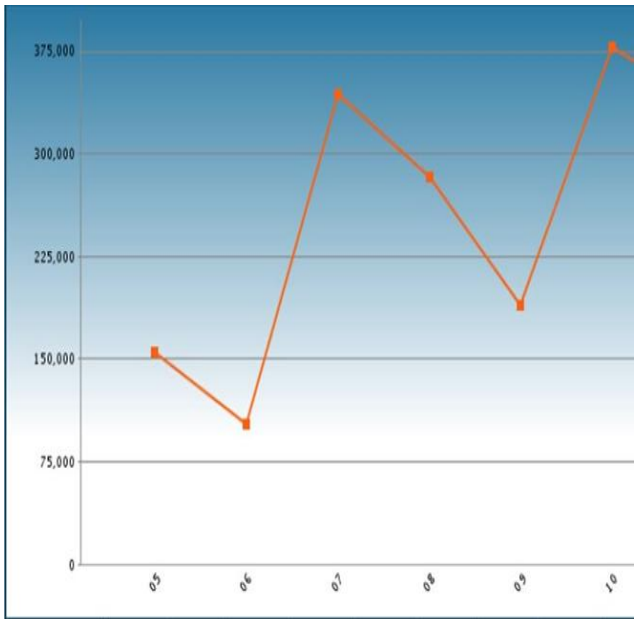

Fig.5. The line chart of gas volume total

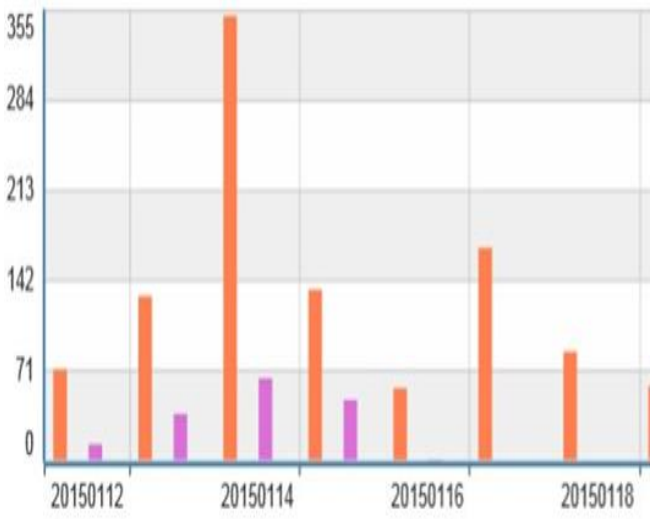

Fig.6. The comparison chart of forecasting gas volume total in the same period

VI. Discussion

In this paper, we proposed a manage method for gas vehicle. It's very useful for city safety, especially for Gong' an and Transpiration Government. For the future's research work, we'll pay more attention to big data analyses.

Acknowledgement

This work is partially supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (The Granted NO. 2014211B44). The first author would like to appreciate Dr.Ronghui Zhang and the anonymous reviewers for their valuable comments on the earlier versions of the paper.

REFERENCES

[1] SARMA S, BROCK D, ASHTON K. The networked physical world–proposals for engineering the next generation of computing, commerce & automatic identification[J]. Auto-ID Center, MIT, 2000.
[2] ATZORI L, IERA A, MORABITO G. The internet of things: A survey[J]. Computer Networks, 2010, 54(15): 1–19.
[3] SUNDMAEKER H, GUILLEMIN P. Vision and challenges for realising the Internet of Things [J]. Luxemborg: publications office of the European union, 2010, 54(15).
[4] XIAOFEI Z, LEI C, MIN W. Efficient multi-way theta-join processing using MapReduce[J]. Proceedings of the VLDB Endowment, 2012, 5(11): 1184–1195.
[5] THUSOO A, SARMA J Sen, JAIN N, et al. Hive-a petabyte scale data warehouse using hadoop[C]//2010 IEEE 26th International Conference on Data Engineering (ICDE). 2010: 996–1005.
[6] ABOUZIED A, BAJDA-PAWLIKOWSKI K, HUANG J, et al. HadoopDB in action: building real world applications [C]//Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. 2010: 1111–1114.
[7] CHEN S. Cheetah: a high performance, custom data warehouse on top of MapReduce[J]. Proceedings of the VLDB Endowment, VLDB Endowment, 2010, 3(1-2): 1459–1468.
[8] YOU Feng, ZHANG Rong-hui, WANG Hai-wei, et al. "Dynamic model and tracking control of tractor semi-trailer vehicle with underactuated system". Journal of Jilin University (Engineering and Technology Edition), 2014, 44(5): 1296-1302.
[9] Ronghui Zhang, Pingshu Ge, Xi Zhou, et al, "An Method for Vehicle-Flow Detection and Tracking in Real-Time Based on Gaussian Mixture Distribution," Advances in Mechanical Engineering, vol. 2013, Article ID 861321, 8 pages
[10] Kim, S., Oh, H., Suk, J., & Tsourdos, A. (2014). Coordinated trajectory planning for efficient communication relay using multiple UAVs. Control Engineering Practice(29), 42-49.
[11] Hou, Y., Edara, P., & Sun, C. (2015). Situation Assessment and Decision Making for Lane Change Assistance Using Ensemble Learning Methods. Expert Systems With Applications.
[12] Harman D: *In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (1992). p.1-10.
[13] Buckley C and Salton G Allan J: *In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (1994). p.292-300.
[14] X.S. Duan: *Artificial Intelligence* (Renmin University Press, China, 1993).
[15] Buckley, C: *Implementation of the SMART Information Retrieval System (*Cornell Univesity,1985).
[16] H. Nakashima, H. Fujii, A. Oida, et al.: *Discrete element method analysis of single wheel performance for a small lunar rover on sloped terrain*. Journal of Terramechanics, Vol.47 (2010), pp.307-321.

AUTHOR'S PROFILE

Pro. Zhang Yan
was born in 1962, xinjiang province China. He received the B.Sc. (Eng.) degree from the Department of Automation Science and Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1984.
He is currently an professor with the Department of Computer Science & Engineering, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China. His current research interests include intelligent control and information technology.

Mr. Jiang Yaming
was born in 1968, xinjiang province Chins. received the M.Sc.(Eng.) degree in Computer Application Technology from University of Chinese Academy of Sciences, Beijing, China in 2012.
He is currently an associate professor with the Department of Computer Science & Engineering, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China. His current research interests include intelligent control and information technology.

Dr. Xue Huajian
was born in 1978, henan province China. He received the Ph.D.(Eng.) degree in Computer Application Technology from University of Chinese Academy of Sciences, Beijing, China in 2012.
Now, he works as an assistant professor with the Department of Computer Science & Engineering, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China. His interests include: machine learning, natural language processing and internet of things.

Dr. Zeng Wenxiao
was in 1974, anhui province China. He received the Ph.D.(Eng.) degree in Computer Application Technology from University of Chinese Academy of Sciences, Beijing, China in 2013.
He is currently an associate professor with the Department of Computer Science & Engineering, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, China. His current research interests include IOT and information technology.