A Novel Systematic Algorithm Paradigm for the Electric Vehicle Data Anomaly Detection based on Association Data Mining

Yan Wang, Mengnan Wu

School of Economics and Management, Xi'an University of Technology, Xi'an 710054 PR China

Abstract

The electric automobile electric regenerative braking control strategy request achieves two goals, as namely enhances the entire vehicle the energy returns-ratio as well as optimizes the pilot to the feel. When the former request brake the electric regenerative braking first, in can satisfy the brake request under the premise first to then use the electric regenerative braking recycling brake energy and before the latter request rational distribution the trailing wheel braking force uses in tracking pilot's deceleration intention. Under this basis, this paper proposes the novel systematic algorithm paradigm for the electric car data anomaly detection based on association data mining. With the continuous improvement of the intelligent vehicle and the continuous improvement of the function of the system, the system can be used as a hybrid vehicle detection and vehicle maintenance standard equipment, has broad application prospects. The effectiveness of the proposed system is verified through the experiment.

Keywords: Electric Car, Systematic Algorithm (SA), Association Data Mining, Anomaly Detection.

1. Introduction

With the increasingly serious problems of energy and pollution, electric vehicles have become the focus of research in various countries, and hybrid vehicles are one of the core technologies of pure electric vehicles[1]. As a new type of electric vehicle, the electric car has advantages such as clean, pollution-free and energy-saving, but it has a short distance[2]. Therefore, improving the distance traveled by electric vehicles is an urgent problem to be solved. Regenerative braking is a unique technology of electric vehicles, its function is to ensure the stability of electric vehicles under the premise of the electric vehicle braking part of the mechanical energy through the basic regenerative braking system into electrical energy storage unit. Therefore, regenerative braking plays an important role.

Electric vehicle regenerative braking control strategy requires improvement of vehicle energy recovery rate and driver feel optimization When the former request brake the electric regenerative braking first, in can satisfy the brake request under the premise first to then use the electric regenerative braking recycling brake energy and before the latter request rational distribution the trailing wheel braking force uses in tracking pilot's deceleration intention[3-4]. When applies the brake increases the electric regenerative braking the proportion possibly to cause in front of excessively the automobile the trailing wheel braking force assignment not to be balanced, affects the brake stability. In a sense, the increase electric regenerative braking and enhances the brake stability is mutually contradictory and the author in satisfies in the brake stable foundation to take the maximum limit recycles the brake energy as the goal as proposed one kind of new electric regenerative braking control strategy and this strategy can adapt the different road surface to adhere to stick cohere the condition, as according to the pilot expected the braking force the request assigns the automobile the braking force, and aimed at some front-wheel drive electric automobile to carry on the simulation analysis[5].

In the analysis of the regenerative braking system for pure electric vehicles, the driving and driven wheels of the regenerative braking system and the regenerative braking force and their distribution are studied. The braking strength and the state of charge state of regenerative braking system, the fuzzy control model of the regenerative braking system of pure electric vehicle is designed and the braking force distribution is designed as finally that on the platform of electric vehicle simulation software and the control strategy. In the following sections, we will introduce the proposed system in detail and to begin, in the figure 1, we show the general organizations and the architectures of the electric car[6-7].
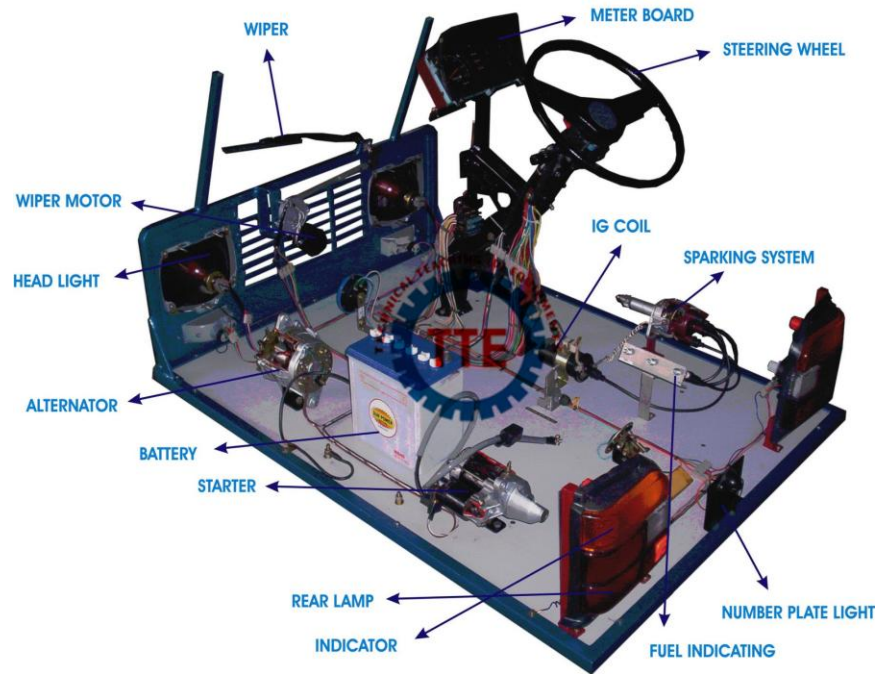

Figure 1.The Electric Car Inner Organization and Architecture

2. Association Data Mining Prior

At present, while also day by day becomes the curing about the data mining algorithm research. Because data mining itself is many kinds of technical synthesis is realizing together, therefore, the data mining algorithm also presents many kinds of the core characteristic. The commonly used data mining algorithm includes: Decision tree algorithm, neural network algorithm, connection rule algorithm. And for this condition, the commonly used algorithms can be generally summarized as the follows[8].

- Apriori algorithm. In 1993, Agrawal et al. Proposed the AIS and SETM algorithms to mine the association relationships among item sets in the core customer transaction database. Many researchers have done a lot of research on association rule mining, and improved the original algorithm (including Apriori algorithm), and improved the efficiency of mining rules, and extended the application of association rules.

- AprioriTid algorithm. AprioriTid algorithm is an improved version of Apriori. Since the Apriori algorithm has to check the database for each loop, the number of support for the candidate set is calculated. With the increasing number of the item sets, the

number of candidate item sets decreases gradually, and the transaction of candidate itemsets is less and less[9]. Therefore, if we can reduce the number of transactions to be scanned in the cycle, the efficiency will be greatly improved. AprioriTid algorithm is proposed to solve this problem. AprioriTid algorithm is able to use the advantages of gradually reduce the candidate options set table instead of the original database, but at the beginning of the cycle, especially a few times before, the number of the candidate sets tend to be greater than the number of data items in the database, lead to construct candidate options set table is bigger than the original transaction database.
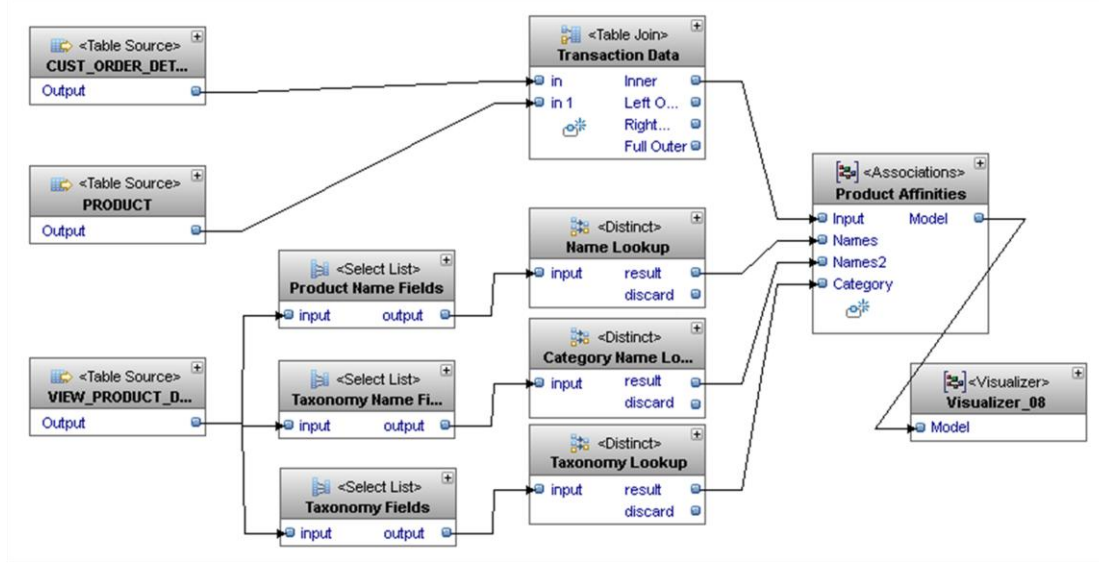
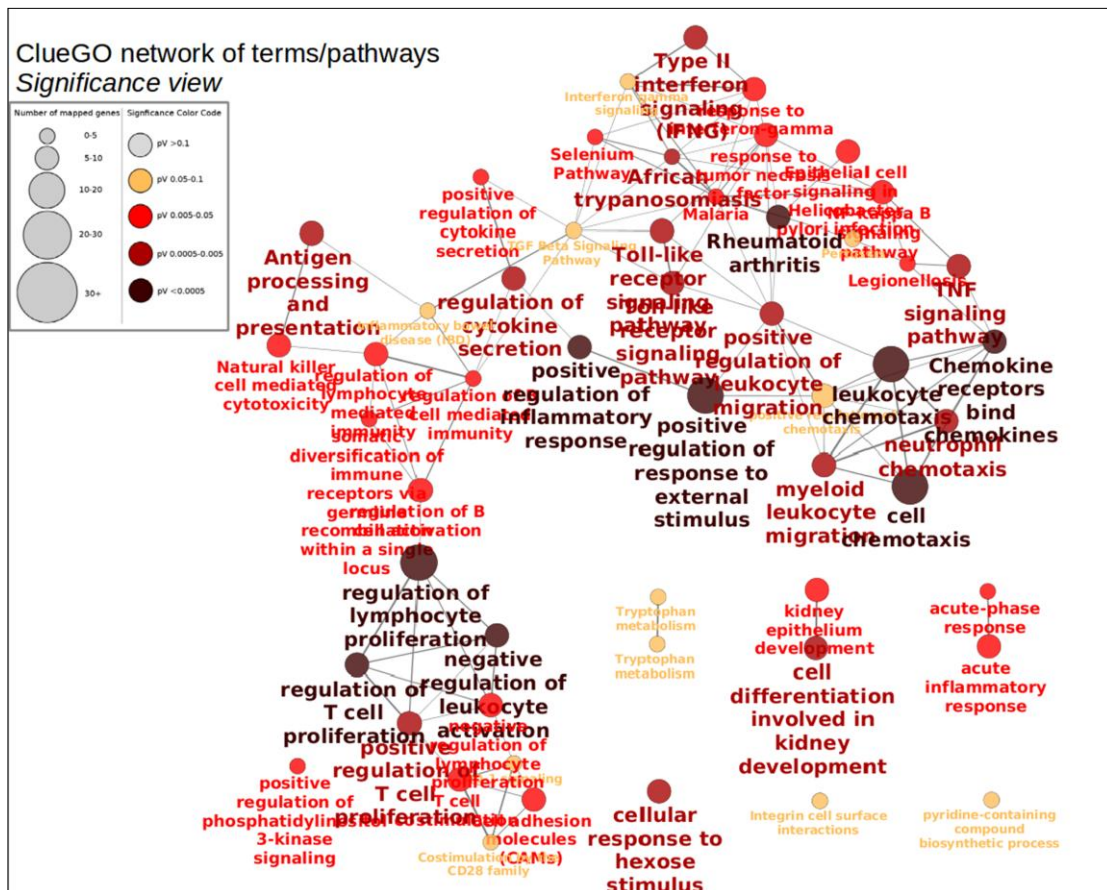**Figure 2.The Association Data Mining Systematic Architecture Flowchart**

FP-growth algorithm is an effective algorithm for mining association rules. The advantage of FP-growth algorithm is that it does not need to generate a large number of some frequent candidate sets or multiple scan datasets. The disadvantage is that a frequent pattern tree based on frequent item sets must be built in memory. FP algorithm adopts the following divide-and-conquer strategy: compresses the datasets that provide frequent item sets into the frequent pattern tree, but retains the item set information, and then divides the compressed data sets into a set of conditional data sets and the data set is associated with a frequent item, and each data set is mined separately which can be modelled as the following items.

$$p = \sum_{i=l}^{k} \frac{\binom{k}{i}\binom{n-k}{m-i}}{\binom{n}{m}} \tag{1}$$

$$p = 1 - \sum_{i=0}^{l-1} \frac{\binom{k}{i}\binom{n-k}{m-i}}{\binom{n}{m}} \tag{2}$$

$$\frac{\binom{k}{l}\binom{n-k}{m-l}}{\binom{n}{m}} \tag{3}$$

From the effective forecast rate angle, the damage data is at the forecast stage the damage data situation to have to miss obviously in is at the training stage the damage data situation. Because not only after this mainly a damage existence causes the restructuring phase space to select and waits a pre-frequency measurement previous time to then select between the basic interrelatedness to change, while moreover to treats the pre-frequency measurement take the damage data point as previous time under datum value (damage data point) also is wrong, and thus creates the forecast effect to become worse.


Figure 3.The Association Data Mining Demonstration of Samples

3. The Proposed Framework

3.1. Independent Principal Components Analysis

Principal component analysis (PCA) is one of the research methods of the multivariate statistical process monitoring[10-11]. It is widely used in process monitoring and fault diagnosis based on historical data. Kernel principal component analysis (KPCA) proposed by Scholkopf

et al. It is an effective nonlinear fault detection method and has been widely used. KPCA transforms the complex nonlinear problem in the original input space into a linear problem in the high-dimensional feature space, and achieves good results in the practical nonlinear fault detection. However, KPCA is the nonlinear transformation method. The use of non-linear mapping may mask the original space in the linear relationship and the combination.
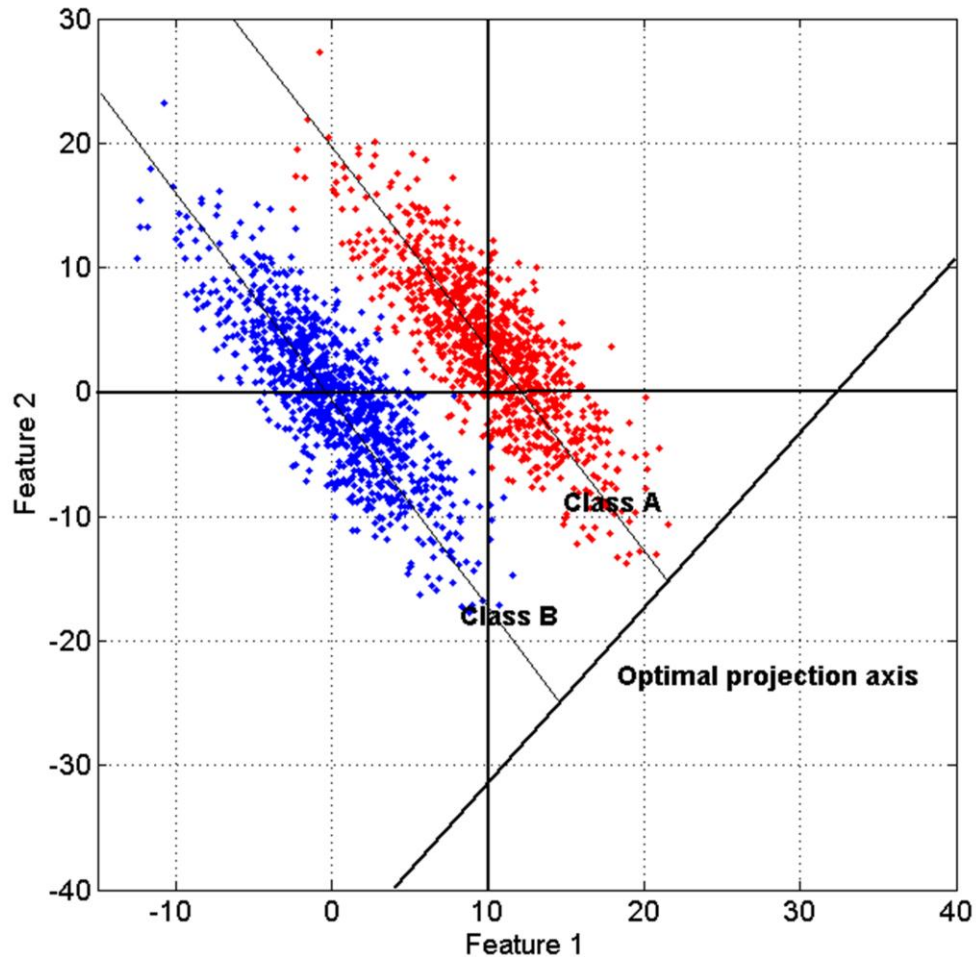
Figure 4.The Independent Principal Components Model

The results of principal component analysis and factor analysis are directly proportional to the degree of correlation between the evaluation indexes. The higher the general degree of correlation, the better the evaluation. As it is possible to choose more principal components, at this time the principal component analysis of the dimension reduction is not obvious, which is the traditional principal component analysis and basic factor analysis of a shortcomings; another problem is that they can only deal with linear problems just a linear dimensionality reduction technology. In real life, the relationship between indicators of non-linear

relationship, if at this time to use a linear relationship to reflect will get the incorrect conclusion and the modelling procedures can be modelled as the following aspects.

$$\varepsilon(W) = \sum_{i=1}^{n} \left\| \mathbf{X_i} - \sum_{j=1}^{K} W_{ij} \mathbf{X_{ij}} \right\|^2 \tag{4}$$

$$\Phi(Y) = \sum_{i=1}^{n} \left\| \mathbf{Y_i} - \sum_{j=1}^{K} W_{ij} \mathbf{Y_{ij}} \right\|^2 \tag{5}$$

Where the $\mathbf{X_i} - \sum_{j=1}^{K} W_{ij} \mathbf{X_{ij}}$ is the common judgement term, and the $\mathbf{Y_i} - \sum_{j=1}^{K} W_{ij} \mathbf{Y_{ij}}$ is the revised term, set up the comprehensive evaluation index system and evaluation objects through indexes between the two is the relative importance of the each and then through the characteristic value method and least square method and logarithmic least-squares method, objective operation such as the upper triangular element method to determine each evaluation index weight.

$$s(\mathbf{P},a) = s(\mathbf{P},0) + a \frac{\partial s(\mathbf{P},a)}{\partial a} + O(a^2) \tag{6}$$
$$\approx \mathbf{P} + a\mathbf{T}$$

When the correlation between primitive variables is not strong, or the number of variables is large, the variance contribution rate of the first principal component is often unsatisfactory, so in this case, evaluation, the variance contribution rate is not ideal. To this end, we consider the construction of the original variables on the nonlinear comprehensive evaluation function. And we propose the revised term as the follows.

$$L(u,v) = \begin{cases} d_u & \text{if } u = v \\ -1 & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The $L(u,v)$ is represented as the judgement term, smallest two rides support vector machines is the standard support vector machines one kind of basic expansion. Its loss function defines directly as the erroneous sum of squares will optimize in the question the inequality constraint transformation equality constraint from this will solve two plan questions to transform as the solution system of linear equations question, reduced the computation complexity, sped up the solution speed, enhanced the counting yield. For this article main research is LSSVM uses in the function regression estimate, therefore under only introduced uses in the function regression estimate LSSVM.

$$\varepsilon(W) = \sum_{i=1}^{n} \left\| (\mathbf{X_i} + a'\mathbf{T'}) - \sum_{j=1}^{K_i} W_{ij} \mathbf{X_{ij}} \right\|^2 \tag{8}$$

$$\varepsilon_{\text{new}} = \left\| (\mathbf{X}_{\text{new}} + a'\mathbf{T'}) - \sum_{j=1}^{K_{\text{new}}} W_{\text{new},j} \mathbf{X_{\text{new},j}} \right\|^2 \tag{9}$$

PCA is a high-dimensional space is mapped to the low-dimensional space to eliminate the correlation between variables, but mainly for the linear systems, the nonlinear process feature extraction is not sufficient. The KPCA maps the original data to the high-dimensional feature space through the kernel transformation method. The PCA processing in the high-dimensional feature space is essentially a nonlinear mapping.

3.2. Manifold Learning Model

Neighbor propagation algorithm is appeared in the recent years, a widely attention of the clustering algorithm compared with other traditional clustering algorithm, the AP algorithm to each data point as a class representative of the candidate point, to avoid the clustering results are limited by the choice of initial class represents points. From the general perspective, the starting methodology of the model can be modelled as the follows. (1) Along with the differential geometry development, the geometry and the analysis situs research technique is even more widespread in the correlation discipline application[12]. Mr. Chen Shengshen pointed out that future mathematics research object as will have to be a manifold. The traditional real number or the plural number space only the partial situation (although under many situations it can be the most important situation). (2) For large scale data sets, the data structure is nonlinear and the data distribution is obviously curved. When the curvature is large, the sample cannot be considered as a subspace. In this case, all the methods of finding the linear subspace cannot describe the structure of the data correctly. (3) Regarding actual in very many questions, the majority of high dimension observation data variable may use the few several influence factor to indicate that, this explained is containing the massive redundant informations, between various ingredients usually also has the strong relevance, in this kind of phenomenon geometry displays for the data distributes on low Manifold, or is nearby low manifold. But must promulgate its latent structure effectively, needs to study and to discover inserting in the high dimension space low content as the follows.

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \gamma_i k(\mathbf{x}, \mathbf{z}_i) + b \tag{10}$$

$$f(\mathbf{x}) = \mathbf{x} \cdot \left(\sum_{i=1}^{n} \alpha_i \gamma_i \mathbf{z}_i \right) + b \tag{11}$$

This article uses the description information to measure the service the similarity, namely expresses the service similarity with the description concept collection similarity. After the gain space information service description information, carries on the participle and the word frequency statistics to the description text, as the word length and the word frequency data extract characteristic word which the use produces, chooses the corresponding domain main body to carry on main body concept labelling to the characteristic word, after the labelling characteristic phrase becomes the service the description concept collection. Because space information service description information are few, the gain description concept centralism contains the concept is extremely sparse, in order to further carries on the description to the service which can be modelled as the follows.

$$d_E^2(u,v) = (y_u - y_v)^T (y_u - y_v) = \sum_{i=1}^{|V|} \exp[-\lambda_i t] (\varphi_i(u) - \varphi_i(v))^2 \tag{12}$$

$$k^2(u,v) = \frac{1}{r_s^2(u,v)} = d_G(u,v) - \frac{24(d_G(u,v) - d_E(u,v))}{d_G(u,v)^3} \tag{13}$$

$$h(A,B) = \frac{1}{|V_1|} \sum_{i \in V_1} \left(\frac{1}{|V_1|} \sum_{j \in V_1} \min_{I \in V_2} \min_{J \in V_2} \left\| k_2(I,J) - k_1(i,j) \right\| \right) \tag{14}$$

In order to solve the above problem, a Neighborhood method is used to create the neighborhood graphs of the distance matrix, and the connectivity of the graph is then determined. The Floyd shortest path algorithm is used to get the shortest path distance matrix D. Since the spatial information service concept matrix is not a numerical matrix, the matrix H in the MDS dimensionality cannot be obtained

by the service concept matrix X, but can only be calculated using the distance matrix. The Euclidean distance matrix in the MDS is replaced by the matrix D, and the D is transformed by the corresponding formula. According to the eigenvalues of the transformed matrix and the corresponding eigenvectors, the service concept space is transformed in some of two dimensions. The Euclidean distance between the dimensional objects approaches the shortest path between services, that is, the service objects in the original spatial information service concept space are two-dimensionally digitized, and the similarity relation among the spatial information services is maintained.

3.3. The System Review of the Electric Car

Electric vehicles should have good performance, the use of power motor should have a wide speed range and high speed, large enough to start the torque, small size, light weight, high efficiency, dynamic braking ability, and can feedback energy. From the direct current machine characteristic curve analysis, the electrical machinery in the process which the power rise unceasingly is the permanent torque movement, achieved maximum work rate later will enter the permanent power area. This characteristic suits the automobile slow-speed of basic revolution high torque, as the high speed low torque request very much, moreover the direct current machine structure is simple, smoothly modulates velocity easy, the control technology is mature, the early electric car research mainly uses the direct current machine to actuate[13-14]. The direct current machine shortcoming is the efficiency is low when movement needs the commutator installment, not easy to have the high speed, maintains the trouble and also the commutation structure easy to produce the electric spark in the movement, to automobile on other electronic device easy influential. The power big direct current machine volume and the weight are big, at present along with electric power electronic technology development, direct current machine in electric automobile already basically by some induction motor, permanent magnetism electric motor as well as switch magnetic resistance electric motor substitution.
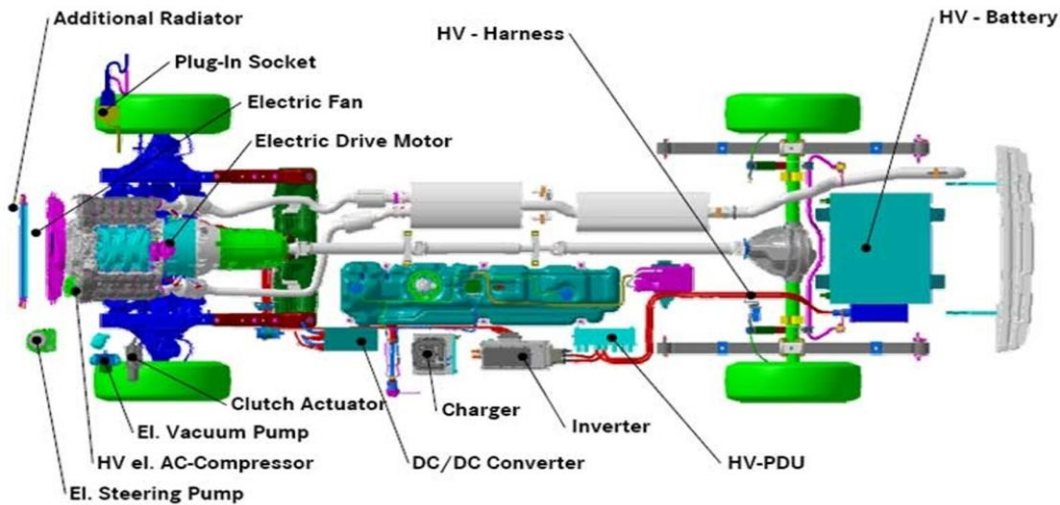

Figure 5.The Systematic Architecture of the Electric Car

The mechanism aspect, the primary coverage that has the controller shell mechanical strength, the water-cooling system bears the hydraulic pressure and the axis vibration intensity as well as the system vibration-proof mobility can and so on. The electricity safety performance mainly includes the electrical machinery stator winding the cold condition DC resistance the electrical machinery stator winding to the cabinet dielectric

resistance, the standing voltage, the electrical machinery controller protection function, the safe earth inspection as well as the contact current. The environmental testing including the temperature, the humidity and the thermal state dielectric resistance survey, as well as bear the salt-fog test. The electrical machinery performance mainly inspects the electrical machinery under the rated load and the peak value load rotational speed torque characteristics and the efficiency, electrical machinery regeneration energy back coupling ability, the highest work rotational speed and the over-speed ability, the operating mode movement temperature rise as well as the noise size[15]. The electromagnetic compatibility including the electromagnetic radiation and the general electromagnetism interference rejection two aspects, the former is inspects the electrical machinery and the controller work when to other part influence and the latter is inspects product itself the anti-jamming ability as the follows.
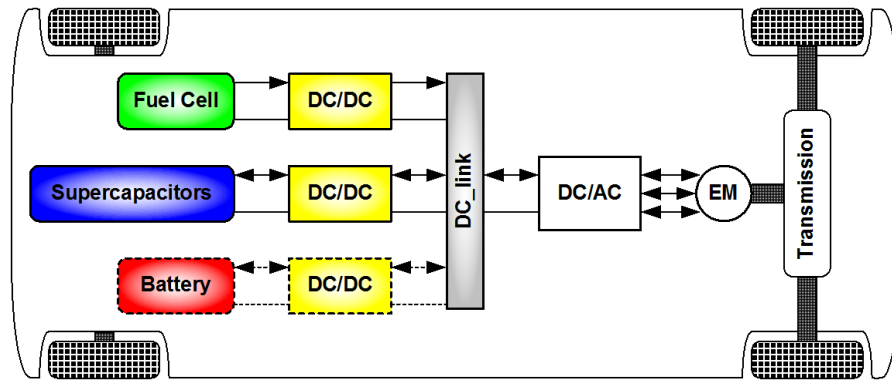

Figure 6.The Energy Supply System

3.4. Electric Car Data Analyzing Paradigm

An embedded system is one of the hottest research fields today and its momentum of the development has caused extensive concern of the society from all walks of life. It is used as the center, on the basis of computer technology, software and hardware can be cut, is suitable for the application system. The hardware architecture of embedded system hardware platform is divided into two parts: one is the system board, the smallest system based on ARM, including ARM CPU SLASH, SDRAM, serial port, keyboard and other basic parts; Board, provides for the completion of some various hardware modules. In the software architecture embedded software, including hardware-related underlying software, operating systems, graphical interface, communication protocols, database systems, standardized browsers and application software. Among them, the embedded operating system is used to support the embedded system software should usually include hardware-related underlying drivers, with system kernel, device driver interface, communication protocols, graphical user interface, etc.

The correctness guarantees that the model can simulate the actual working process of the vehicle correctly, and the accuracy reflects the error between the running characteristic of the model and the actual working characteristic of the vehicle. The model running on ADVISOR platform is not a physical model but a mathematical model based on the various equilibrium equations and operating characteristic equations of the vehicle and its components. In the ADVISOR platform to achieve the kind of simulation analysis only model is not enough, but also have to run with the data file. The main contact surface is in the pure electric automobile intelligence measuring appliance is most important also the frequency of use highest contact

surface[16-17]. The main contact surface includes the speedometer, the basic electrical machinery ratemeter and the accumulator cell SOC, about changes information and so on lamp, far low-beam lights, accelerator pedal opening and brake pedal opening.

Through touches " dynamic to monitor ", the "energy recycles " and the " breakdown self-checks buttons and so on ", may from advocate contact surface skipping to the corresponding contact surface, provides the more exhaustive driving information to the driver. In addition, but also may in input in the frame to input the test run serial number, the driving information preservation in the intelligent measuring appliance internal hard disk, clicks on " the preserved data " button, stops the data record and preserves the core current data[18]. When data quantity oversized, but also may meet through outside the measuring appliance behind USB connection moves the hard disk, data preservation to migration storage device in has realized the driving data memory and the reappearance as the follows.
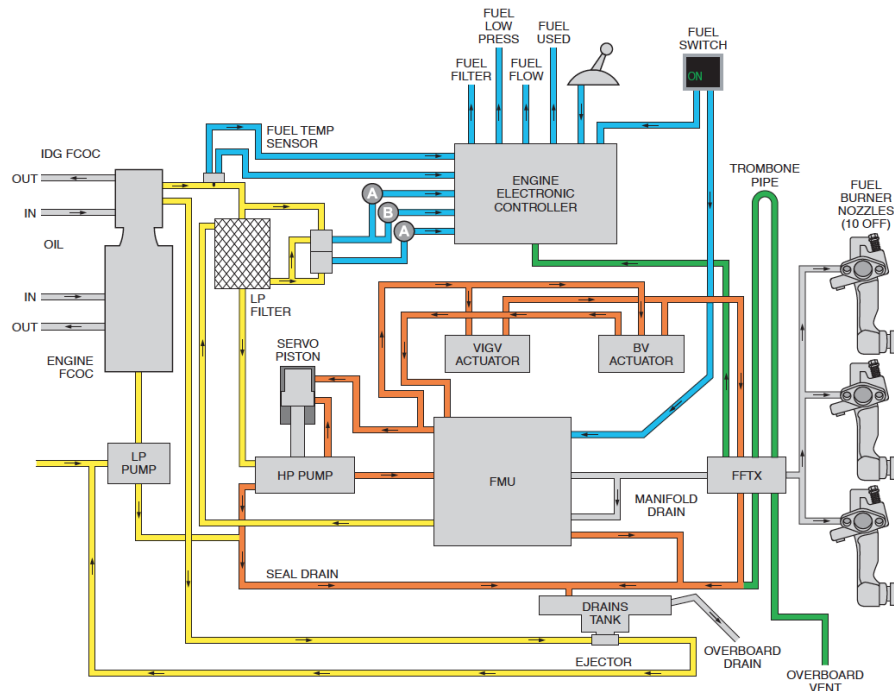

Figure 7.The Electric Car Data Analyzing System

4. Experiment

In this section, we conduct a series of simulations. The CAN bus is an important method to solve the problem of large amount of data exchange between the electronic control modules of the car passing through data communication protocol in modern life, and it is also an important channel for information sharing between smart meters and pure electric vehicles, in order to use the mini2440 as the main control module, it is first necessary to determine that it is not part of the external interface of the CAN bus interface. Therefore, the AIOOO chip acts as a CAN controller and the CAN transceiver PCA82C250 chip of the smart meter system can be used as an expansion module of the CAN interface[19-20]. Because of the TPC7062K ARM9 kernel smart device, the address bus and data bus are always separated. In order to fully consider the needs of drivers and researchers, the data on the module is updated via the serial port on the DSP. The update period is set to 50ms to better display the running status of

the entire vehicle and the electronic control unit in real time. When the DSP processor passes the fault information, the acousto-optic alarm receives the parameter information and fault information during transmission of the hybrid vehicle and issues a corresponding fault warning. The figure below fully demonstrates the results of this simulation.
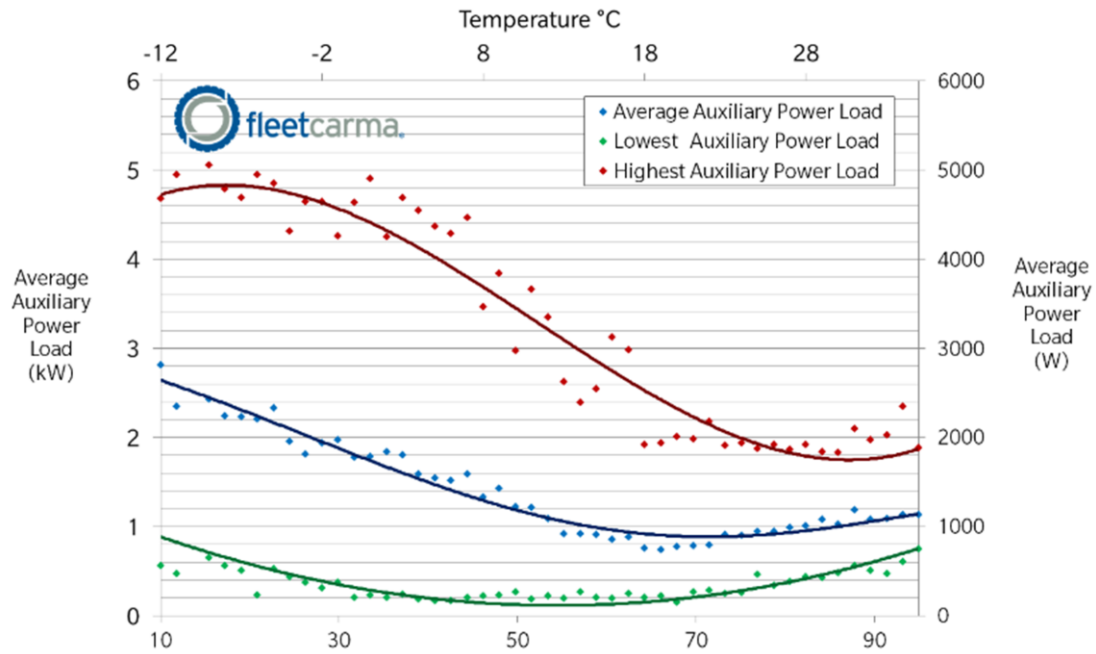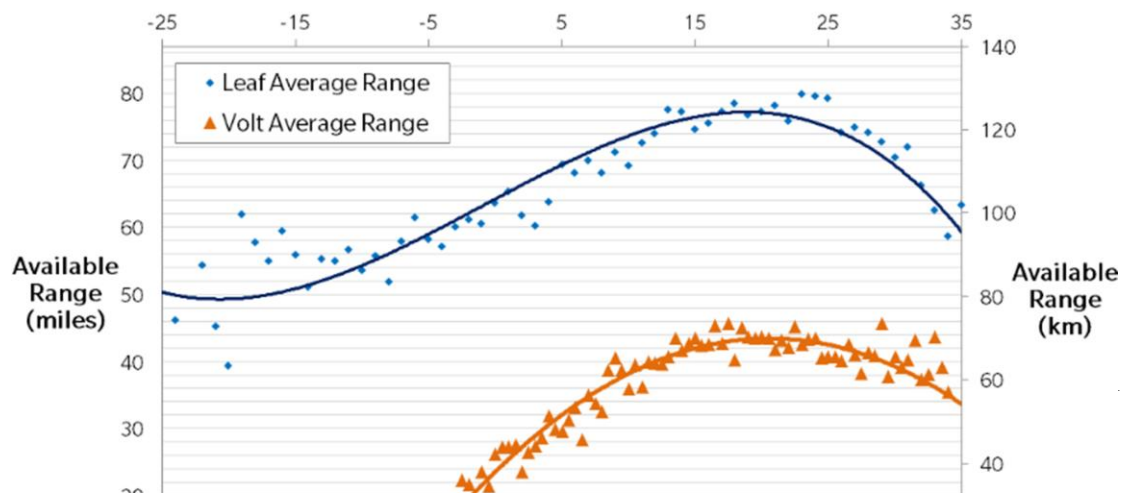

Figure 8.The Data Simulation Set One
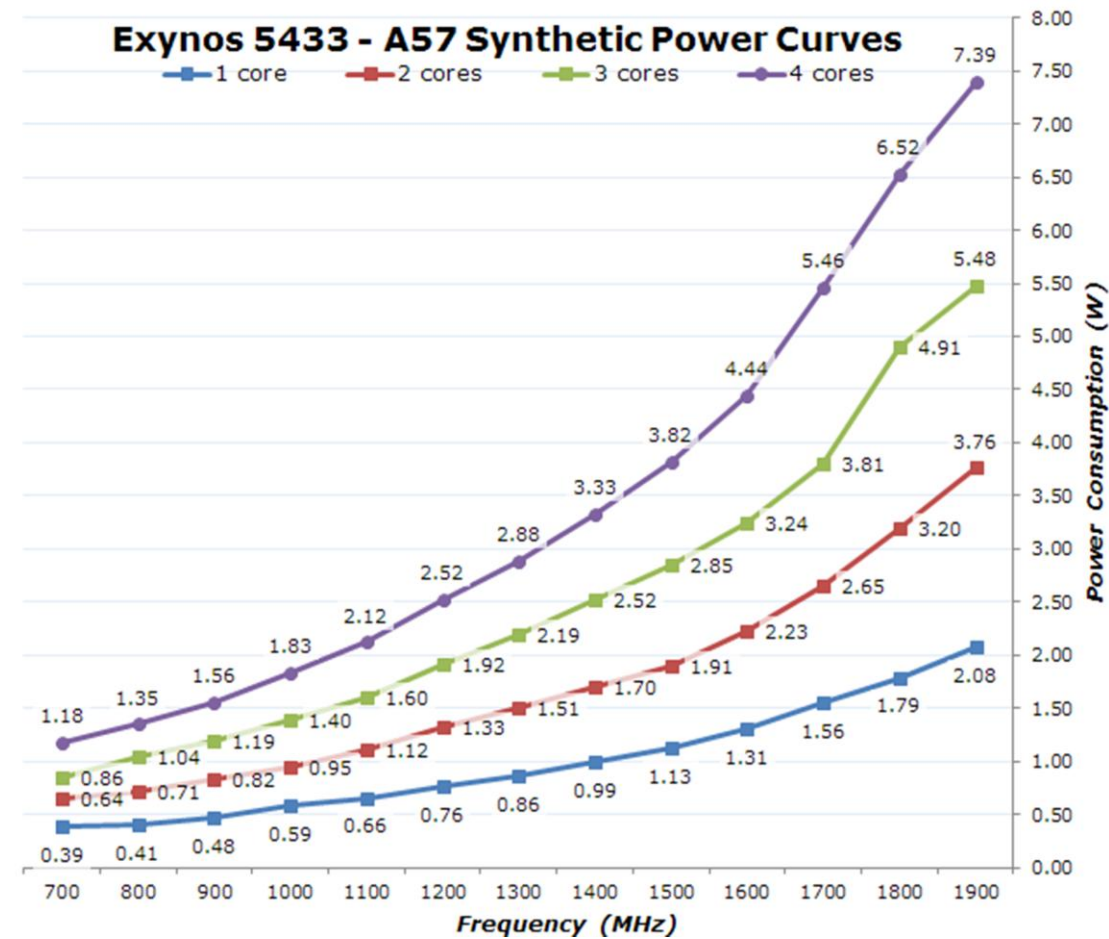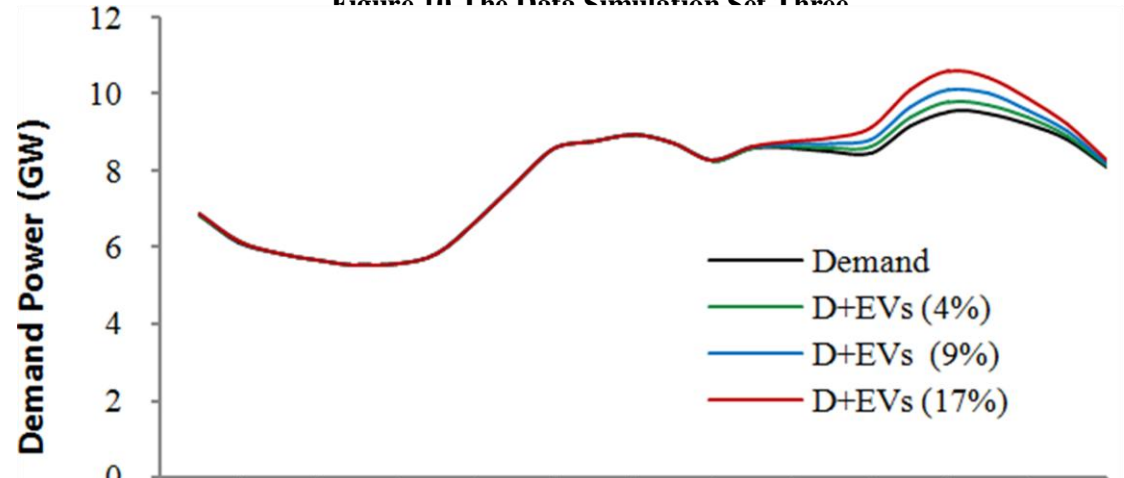
Figure 9.The Data Simulation Set Two

Figure 10 The Data Simulation Set Three

Figure 11.The Data Simulation Set Four

5. Conclusion

Intelligent instrument and equipment can carry out a basic visual curve display of vehicle-related parameter information, and store the displayed parameter information accordingly, which perfectly realizes the vehicle during the driving process of pure electric vehicles with electricity as the whole energy supply energy. Real-time monitoring and visual display of data-related information. At the same time, the history data curve function of the smart meter can be used to store and extract the data and reproduce the historical data, which enhances the readability of the monitoring data of the pure electric vehicle and facilitates further analysis and prediction of the vehicle's recent operating conditions. With the convenience and safety of data acquisition and storage, data assurance can be provided for fault diagnosis and vehicle development. With the continuous improvement of intelligent vehicles and the continuous improvement of system functions, the system can be used as a standard equipment for hybrid vehicle detection and vehicle maintenance. In the future, we will discuss more application scenarious for the technology.

6. ACKNOWLEDGEMENTS

7. References

[1] Patil, Nishad, Diganta Das, and Michael Pecht. "Anomaly detection for IGBTs using Mahalanobis distance." Microelectronics Reliability 55, no. 7 (2015): 1054-1059.
[2] Attia, Mohamed, Hichem Sedjelmaci, Sidi Mohammed Senouci, and El-Hassane Aglzim. "Game model to optimally combine electric vehicles with green and non-green sources into an end-to-end smart grid architecture." Journal of Network and Computer Applications 72 (2016): 1-13.
[3] Taylor, Adrian, Nathalie Japkowicz, and Sylvain Leblanc. "Frequency-based anomaly detection for the automotive CAN bus." In Industrial Control Systems Security (WCICSS), 2015 World Congress on, pp. 45-49. IEEE, 2015.

[4] Joss é, Gregor, Matthias Schubert, and Ludwig Zellner. "EasyEV: Monitoring and Querying System for Electric Vehicle Fleets Using Smart Car Data." In International Symposium on Spatial and Temporal Databases, pp. 497-502. Springer International Publishing, 2015.

[5] Ma, Mathew X., Henry YT Ngan, and Wei Liu. "Density-based Outlier Detection by Local Outlier Factor on Largescale Traffic Data." Electronic Imaging 2016.14 (2016): 1-4.

[6] Freeman, Paul, et al. "Model-based and data-driven fault detection performance for a small UAV." IEEE/ASME Transactions on mechatronics 18.4 (2013): 1300-1309.

[7] Li, Yue, et al. "Real-time estimation of lead-acid battery parameters: A dynamic data-driven approach." Journal of Power Sources 268 (2014): 758-764.

[8] Purarjomandlangrudi, Afrooz, Amir Hossein Ghapanchi, and Mohammad Esmalifalak. "A data mining approach for fault diagnosis: An application of anomaly detection algorithm." Measurement 55 (2014): 343-352.

[9] Tian, Zhiyong, et al. "Real-Time Charging Station Recommendation System for Electric-Vehicle Taxis." IEEE Transactions on Intelligent Transportation Systems 17.11 (2016): 3098-3109.

[10] Sedjelmaci, Hichem, and Sidi Mohammed Senouci. "An accurate and efficient collaborative intrusion detection framework to secure vehicular networks." Computers & Electrical Engineering 43 (2015): 33-47.

[11] Wu, Chao, David Birch, Dilshan Silva, Chun-Hsiang Lee, Orestis Tsinalis, and Yike Guo. "Concinnity: A Generic Platform for Big Sensor Data Applications." IEEE Cloud Computing 1, no. 2 (2014): 42-50.

[12] Zeitlin, Nancy P., Gregory R. Clements, Stanley J. Schaefer, Michael K. Fawcett, and Barbara L. Brown. "Ground and launch systems processing technologies to reduce overall mission life cycle cost." In Aerospace Conference, 2013 IEEE, pp. 1-20. IEEE, 2013.

[13] Zhang Chengjia, Liu Junyong, Xiangyue, et al. Configuration and two-phase charging optimization dispatching of electric vehicles based on data mining[J]. Proceedings of the CSEE, 2018, 38(4).

[14] Zhang Wenyan. Data-driven analysis of charging behavior of urban distribution electric vehicles and model establishment [D]. Beijing Jiaotong University, 2017.

[15] Wang Xin, Zhou Buxiang, Tang Hao. An Ordered Charge and Discharge Control Strategy for Electric Vehicles Considering User Factors[J]. Power System Protection and Control, 2018, 46(4).

[16] Liu Guang, Zeng Chengbi, Miao Hong. Comprehensive Evaluation of Electric Vehicle Charging Station Planning Based on Rough Set Theory[J]. Modern Electric Power, 2018(1).

[17] Li Tao, Tang Sixiao. Status Quo and Future Development Trend of Electric Vehicle Industry[J]. Agricultural Machinery Usage and Maintenance, 2018(1):11-12.

[18] Research on DC Distribution Network Planning for Electric Vehicle Charging and Discharging Stations.

[19] Rongguang Z. Current status and development trend of electric vehicle battery. Times Auto.2018;3.

[20] Ming D, Shengliang S, Pan H, et al.Planning of AC/DC hybrid microgrid with integration of electric vehicles charging load. Autom Electr Power Syst.2018;42(1):32-38

Author's brief introduction:

Yan Wang, Associate professor, school of economics and management, Xi'an University of Technology, got PHD from Northwest University. Focus on the study of regional economic growth.

Mengnan Wu, Graduate student, school of economics and management, Xi'an University of Technology. Focus on the study of regional economic growth.
.