Forecasting of New Energy Vehicle Sales and Evaluation of Regional Development Based on BP Neural Network and EWM-TOPSIS

Xieyang Wang [1, *, #], Zirui Xu [2, #], Yuewen Li [3]

[1] School of Physical and Mathematical Sciences, Nanjing Tech University, Nanjing, China, 211816

[2] School of Geomatics Science and Technology, Nanjing Tech University, Nanjing, China, 211816

[3] Reading Academy, Nanjing University of Information Sciences and Technology, Nanjing, China, 211544

* Corresponding Author Email: wxy_nancheng@njtech.edu.cn

#These authors contributed equally.

Abstract. This paper focuses on the new energy vehicle market, utilizing big data technology and artificial intelligence algorithms to perform statistics, analysis, and forecasting in both temporal and spatial dimensions. In the time dimension, the sales volume is forecasted by piecewise cubic Hermite interpolation, polynomial fitting, ARIMA model and BP neural network forecasting model, and the forecasting results between different models are compared and analyzed. Meanwhile, the factors affecting this market are analyzed using the entropy weight method. In the spatial dimension, the development level of each province is assessed using the TOPSIS comprehensive evaluation method, and the development stage in which different provinces are located is classified using K-means cluster analysis. The results show that this new energy vehicle market is developing rapidly, but there is still the problem of uneven development in some regions. At the same time, the study also found that BP neural network has higher credibility in sales prediction, the method of EWM-TOPSIS can effectively assess the market development level of each province and city, and K-means cluster analysis can intuitively show the differences in the development stage. The research of this paper can provide technical support and theoretical support for the industrial development of China's new energy vehicle market in the era of big data.

Keywords: New Energy Vehicle, BP Neural Network, EWM-TOPSIS, K-means Cluster Analysis, Sales Forecasting and Evaluation.

1. Introduction

With the progress of science and technology and changes in consumer demand,the new energy vehicle is gradually emerging and gradually changing the pattern of the traditional automobile market. This new type of automobile is mostly powered by electricity and is gradually becoming the trend of future automobile development.

However, what is the current development status of this new automotive market, what are the future development trends, and are there any differences in development between regions. These issues are of great significance in predicting and evaluating the future development of new energy vehicles in China.

Existing studies have given some of the factors affecting the development of new energy vehicles in China. Liu et al. (2022) [1] concluded that regional economic development, energy environment, infrastructure development and penetration rate of new energy vehicles would affect the development of new energy vehicles among different regions in China. Tang et al. (2019) [2] also suggested that the price of oil and the amount of oil stored, which are closely related to traditional fuel vehicles, would affect the sales and development of new energy vehicles. At the same time, Liu et al. (2021) [3] believe that the penetration of new energy vehicles in the automotive market also plays a key role in their development.

In terms of sales forecasting of new energy vehicles, Zeng et al. (2023) [4] applied a variable structure gray model, Yuan (2023) [5] applied multiple linear regression and time series models, and Yang et al. (2024) [6] applied an uncertain Bass model.

In recent years, in the studies about the development of new energy vehicles in China, there are few studies about the sales volume prediction using BP neural network prediction model, and few studies about the regional development assessment using EWM-TOPSIS for the development of each province. In this paper, we will focus on analyzing the sales forecast of new energy vehicles under the BP neural network prediction model and the regional development assessment under EWM-TOPSIS.

2. Materials and methods

2.1. Data acquisition and pre-processing

In order to make our statistical analysis more specific, we chose the relevant data of China's new energy vehicle market and pre-processed the data to make it more suitable for the models we chose.

2.1.1. Data sources

The data in this paper comes from the official website of the National Bureau of Statistics of China, the official website of the China Association of Automobile Manufacturers, the official website of the National Big Data Alliance of New Energy Vehicles of China, the Traffic Management Bureau of the Ministry of Public Security, the WeChat official account "Xingyuan Data", the WeChat official account "Passenger Vehicle Sales Query", and the RESSET database. The data in this paper includes the data information of most provinces in China, but the data of Hong Kong, Macao and Taiwan is not included. In addition, in the data classified by province, part of the data of Tibet is missing and is not included.

The data indicators selected in this paper mainly include the production of the new energy vehicle, the sales of the new energy vehicle, the total sales of automobiles, the penetration rate of the new energy vehicle, the number of public charging piles, and the sales of power batteries.

Among them, this paper introduces the penetration rate of the new energy vehicle as a key indicator to compare its market share in China's automobile market. It is calculated as the ratio of the sales volume A_0 of the new energy vehicle to the total sales volume A of automobiles in a certain period of time. That is, the penetration rate K is calculated as:

$$K = \frac{A_0}{A} \times 100\% \tag{1}$$

2.1.2. Pre-processing of data

The pre-processing of data in this paper mainly uses interpolation methods to fill the missing values. The interpolation method mainly utilizes the relationship between the variables to estimate the missing values by numerical methods with the help of the rest of the complete data in the vicinity of the missing values. In this paper, the new energy vehicle production data collected in the RESSET database has missing data in the time interval from 2016 to 2017, so the segmented cubic Hermite interpolation method is used here. Using the segmented cubic Hermite interpolation method, we can approximate the missing values of the production of tthe new energy vehicle in this period of time and apply them in the subsequent data analysis.

2.1.3. Preliminary analysis of data

In this section, we will use the data on the penetration rate of the new energy vehicles to conduct a preliminary analysis of the data in both the temporal and spatial dimensions.

1) Temporal dimension

Considering the concept of penetration rate mentioned in the previous section of this paper, we can get the penetration rate of the new energy vehicles in China's automobile market from 2015.1 to 2023.12 as shown in Figure 1 below.

Its penetration rate forms an inflection point around 2021 and starts to grow significantly. Its penetration rate has grown from less than 5% in 2015 to more than 35% today, and according to the trend of the image, the new energy vehicle still has more space for development and growth trend in the future in China. It can be predicted that, according to the trend of penetration growth at this stage, the new energy vehicles will be the leading product in the China's future automobile market.

2) Spatial dimension

In this paper, the penetration rate of the new energy vehicles of 31 provinces in China (excluding Hong Kong, Macao and Taiwan) in 2023 are selected as the data for the preliminary analysis of the spatial dimension.

The top five provinces in China in terms of penetration rate of the new energy vehicles in 2023 are shown in Table 1 below. In terms of the penetration rate of the new energy vehicles (as shown in Figure 2 below), its penetration rate in the eastern coastal provinces is higher, generally more than 30%, and the highest Zhejiang Province is even as high as 54.89%, but the penetration rate in the western region is generally less than 20%. This shows that, from the penetration rate to analyze the development of China's new energy vehicles, there is still an imbalance in the development of the eastern and western regions, which may also be inextricably linked with the local economy, policies and other factors. Often, the eastern coastal region is economically developed, consumers have strong purchasing power,while the western region is economically weaker, consumer purchasing power is also relatively weak. At the same time, with the economic development, environmental pollution has also intensified, the urgent need for environmental protection products such as the new energy vehicles to improve the environment, so the government of the eastern region to give preferential policies and supporting facilities for the construction of more, which also directly led to the penetration of new cars in the eastern region is much higher than in the western region.

Table 1. Top five provinces and penetration rate of new energy vehicles in China in 2023

Ranking	1	2	3	4	5
Province	Zhejiang	Hainan	Shanghai	Guangxi	Tianjing
Penetration Rate	54.89%	52.77%	48.70%	45.82%	39.20%

Figure 1. Penetration rate of the new energy vehicles in China from 2015.1 to 2023.12

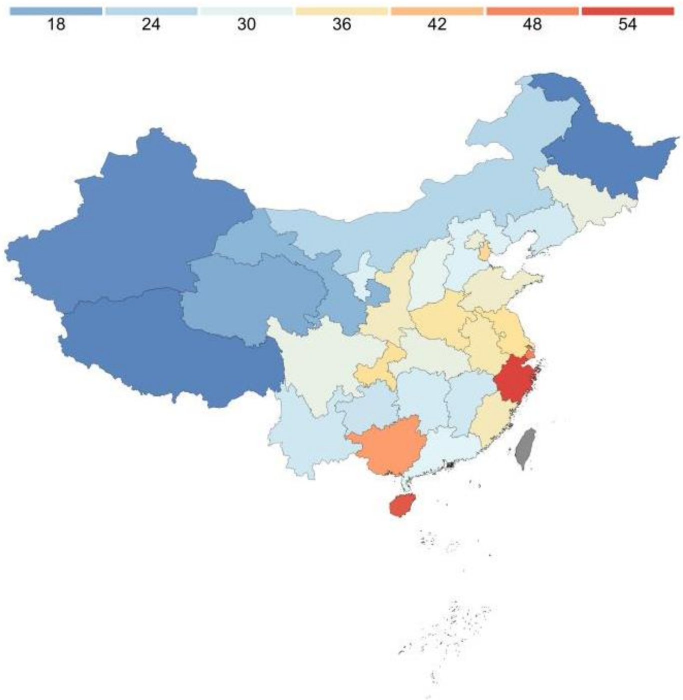


Figure 2. China's new energy vehicles penetration rate by region, 2023

2.2. Introduction to methods

2.2.1. ARIMA prediction model

ARIMA prediction model [7] is a commonly used time series model, also known as autoregressive integrated moving average model. It consists of three parts: the autoregressive (AR) part, which reflects the regression relationship between the lagged term of the series itself and the current term; the difference (I) part, which is used to transform the non-smooth series into a smooth series; and the moving average (MA) part, which takes into account the lagged effect of the error term. ARIMA model is widely used in the fields of economy and sales by identifying the data patterns and regularities, and predicting the future trend by using the past data. Its advantage is that it can better deal with linear correlation data. However, it requires high data smoothness and is difficult to capture non-linear relationships, which has limitations in complex data situations.

The ARIMA model is to treat the time series as a random sequence, draw samples reasonably and randomly for training, fit the model and use it to predict future values. ARIMA (p,d,q) can be written as:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t = (1+\sum_{i=1}^{q}\theta_i L^i)\xi_t \tag{2}$$

2.2.2. BP neural network prediction model

BP neural network prediction model [8] is an error back-propagation neural network model, which is a feed-forward network containing hidden layers. It mainly consists of input layer, hidden layer and output layer. The output is obtained by forward propagating the input signal, which is processed by the activation function of each neuron. Then the output is compared with the desired output, the error is calculated, and the error back propagation adjusts the connection weights between neurons.The schematic diagram of its working principle is shown in Figure 3 below:

The *f* activation function of the BP neural network prediction model selected in this paper is the sigmod activation function:

$$\begin{cases} f(x) = \dfrac{1}{1+e^{-x}} \\ f'(x) = f(x)(1-f(x)) \end{cases} \tag{3}$$

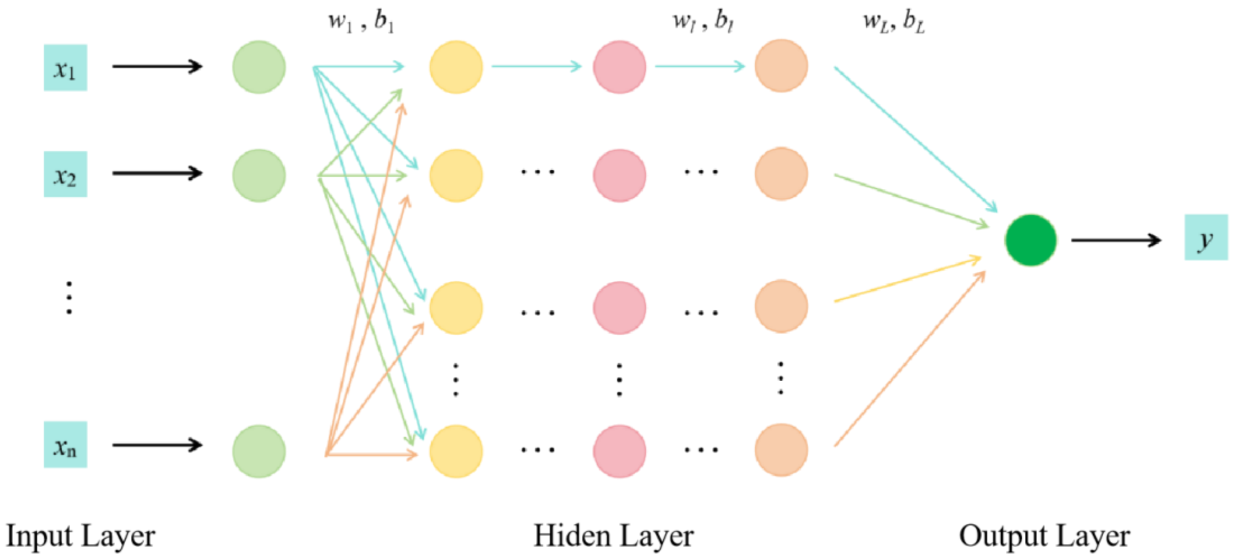


Figure 3. Schematic diagram of the principle of the BP neural network prediction model

2.2.3. EWM-TOPSIS

Entropy weighting method (EWM for short), i.e., exponentially weighted moving average, which can weight the data, is an objective assignment method commonly used in comprehensive evaluation. Entropy is a measure of the degree of disorder of the system, and the entropy value can be used to evaluate the degree of dispersion of an indicator in the scheme, and the smaller its entropy value is, the greater the degree of dispersion of the indicator, and the greater the degree of influence of the indicator on the comprehensive evaluation. The TOPSIS method [10], which is a commonly used method of comprehensive evaluation, is used to rank the evaluation objects by calculating the distance between them and the positive and negative ideal solutions, and can make full use of the information of the original data to accurately reflect the gap between the evaluation schemes, which accurately reflects the gap between the evaluation programs.

EWM-TOPSIS [11] combines the advantages of both, first using EWM to process the data, then using TOPSIS to evaluate, this method can be more reasonable for the comprehensive assessment of multiple objects.

2.2.4. K-means cluster analysis

K-means cluster analysis [12] is an unsupervised real-time clustering algorithm, which aims to divide the data into K different clusters, and has the advantages of simple algorithmic ideas, fast convergence, and easy to implement.

K-means cluster analysis works by first randomly determining K initial cluster centers, then calculating the distance from each data point to these centers, and dividing the data points into clusters corresponding to the centers with the closest distance. Then the centers of each cluster are recalculated and the process is repeated until the cluster centers no longer change or change little.

In this algorithm, the Euclidean distance between data objects in the space and the clustering center is calculated as:

$$d(X, C_i) = \sqrt{\sum_{j=1}^{m} (X_j - C_{ij})^2} \tag{4}$$

Where X is the data object, C_i is the ith clustering object, m is the dimension of the data object, and X_j and C_{ij} are the j-th attribute values of X and C_i.

3. Results and analysis

3.1. The new energy vehicles sales statistics and forecasting model based on ARIMA

In the ARIMA model, this paper select the sales of the new energy vehicles from 2015.1 to 2023.11 as the dataset and calculate the sales data from 2023.6 to 2024.9 according to the fitted model, and the prediction result is shown in Figure 4 below.

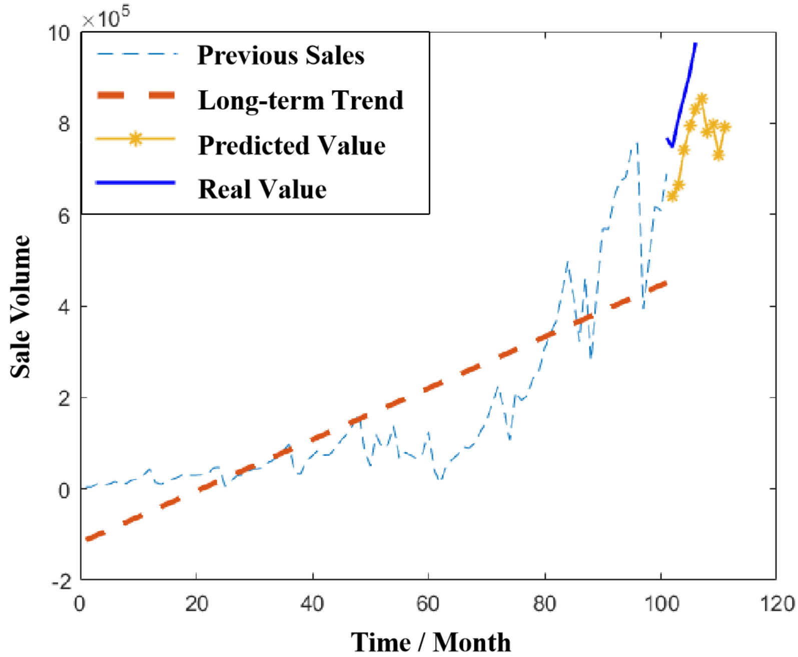


Figure 4. ARIMA prediction model results graph

As can be seen from the figure, for the future sales of the new energy vehicles, it is also rising in fluctuation. However, because of its comparison with the real values from 2023.6 to 2023.11, there is still some prediction error, so it is also necessary to correct this error adjustment model.

Meanwhile, in order to avoid the large prediction error using ARIMA model, this paper will use BP neural network prediction model in the next subsection to predict the sales of the new energy vehicles.

3.2. The new energy vehicles sales prediction model based on BP neural network

3.2.1. Polynomial data fitting

In order to predict sales of the new energy vehicles by BP neural network, we it is necessary to obtain the future (i.e., 2024.4 to 2024.12, which is not included in the collected data) of other factors affecting the sales of the new type of automobiles.

Here, this paper uses polynomial fitting [13] for the production of the new energy vehicles, the total sales of automobiles, the penetration rate of the new energy vehicles, and the number of public charging piles.

Polynomial fitting is done by finding an n ($n<m$) times polynomial for a set of data (x_0, y_0), (x_1, y_1), ..., (x_m, y_m) for (x, y):

$$y = f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i \tag{5}$$

Also, such that the sum of squares of the distances of the corresponding values y_0 to y_m to the curve, R^2, is minimized over the given range x_0 to x_m. Where R^2 is calculated by the following formula:

$$R^2 = \sum_{i=1}^{m} [y_i - f(x_i)]^2 \tag{6}$$

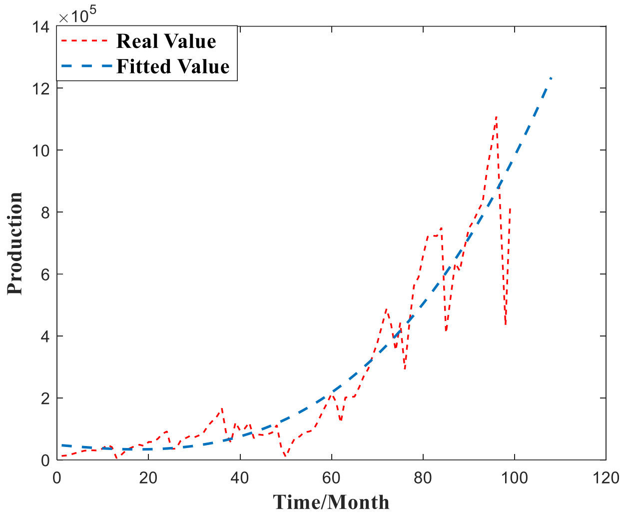
Considering that the higher the polynomial fitting order, the more complex the polynomial model is and the more flexible it is to reflect the data changes. However, the higher the order, the more prone to overfitting phenomenon, so this paper adopts the third-order fitting, and the curve of the fitting function of each index is shown in Figure 5 below.

According to the images, the third-order polynomial fitting is better for the above 4 types of indicator data, and the fitting function for each indicator data is shown in Table 2 below.

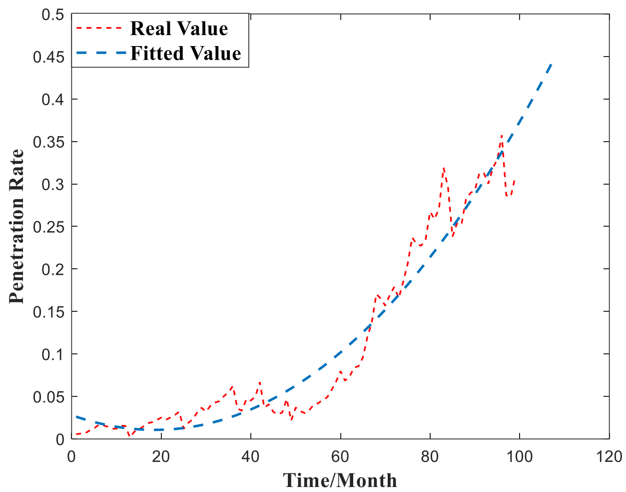
Table 2. Polynomial fit function for each indicator data

Indicator	Polynomial Fit Function
Production of the new energy vehicles	$y_1 = 0.93x^3 + 13.78x^2 - 1354.73x + 48511.26$
Total sales of automobiles	$y_2 = 0.000446x^3 - 0.05x^2 + 1.51x + 222.48$
Penetration rate of the new energy vehicles	$y_3 = 4.1322 \times 10^{-8} x^3 + 4.8941 \times 10^{-5} x^2 - 0.0019x + 0.028$
Number of public charging piles	$y_4 = 4.67x^3 - 299.44x^2 + 13340.01x + 13491.15$

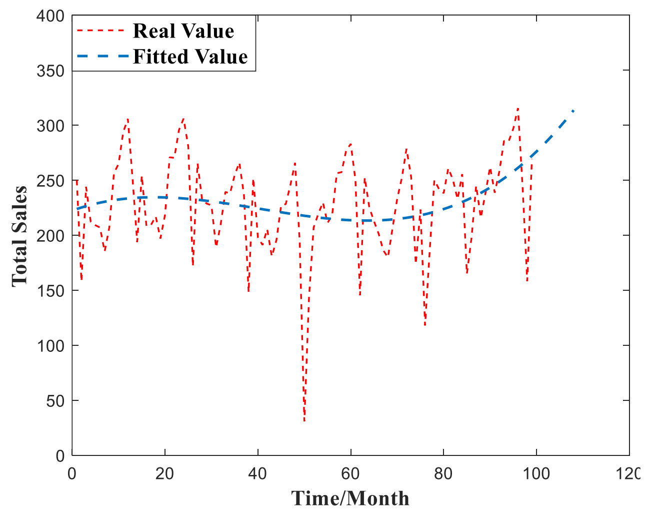
According to the functions, the fitted data for each indicator data from 2024.4 to 2024.12 can be calculated. Based on the collected data from 2016.1 to 2024.3 and the fitted data from 2024.4 to 2024.12, this paper performs a BP neural network prediction for the new energy vehicles sales.



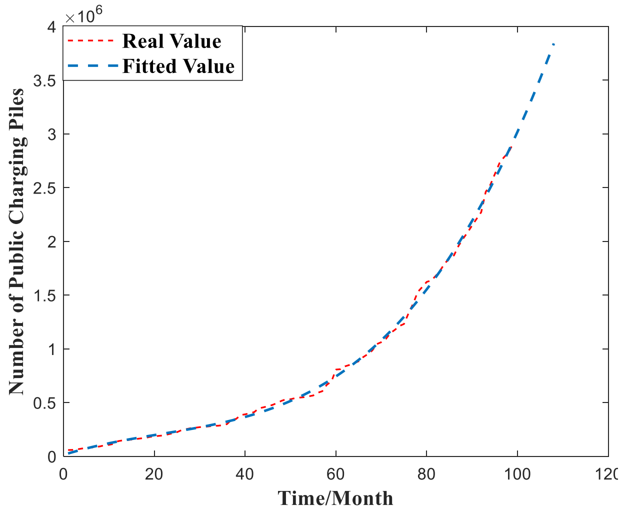
(a) Production



(c) Penetration rate



(b) Total sales



(d) Public charging piles

Figure 5. Polynomial fit for the data of each indicator

3.2.2. BP neural network prediction model

In the BP neural network prediction model, this paper selects the data of the production of the new energy vehicles, the total sales of automobiles, the penetration rate of the new type of automobiles, and the number of public charging piles from 2016.1 to 2024.12 as a means of training the model and predicting the data from 2023.1 to 2024.12, and compares the predicted values from 2023.1 to 2024.3 with the real values to analyze the BP neural network model for the prediction effect.

In training the BP neural network, this paper use the data from 2016.1 to 2022.12 as the training set to get the iterative process and the training state as shown in Figure 6 and Figure 7 below.

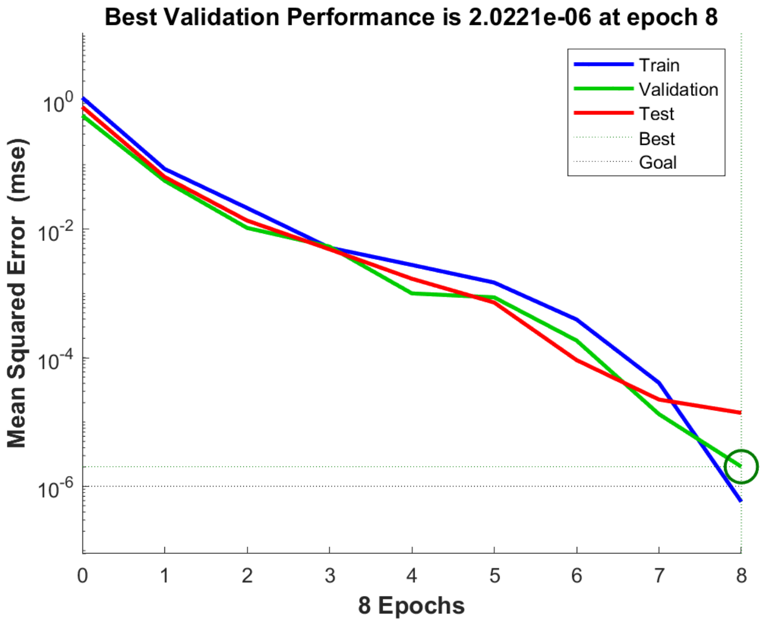


Figure 6. BP neural network iteration process diagram

According to Figure 6, it can reflect the error situation of the training set, validation set and test set during the training process in the process of continuous iteration, in which the validation set has the smallest error in the 8th iteration, that is, the best effect. The BP neural network training status graph shown in Figure 7 also shows that the present BP neural network prediction model has good training and prediction effects.

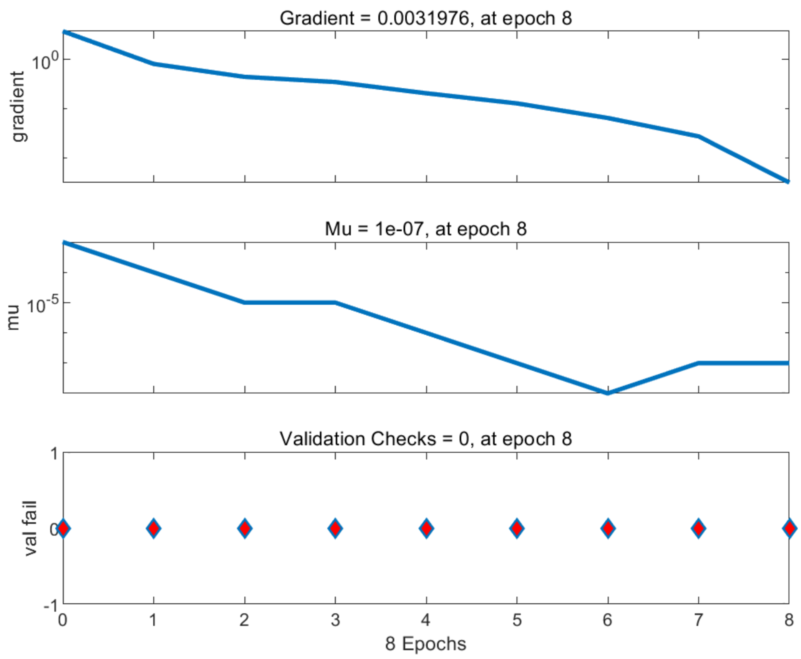


Figure 7. BP neural network training situation diagram

And the training effect for the sample data is shown in Figure 8 below, which shows the effect of regression coefficients for the training samples, validation samples, test samples and all samples, respectively. The closer the R-value is to 1, i.e., the closer the two lines are in the respective graphs, the better the training effect is. According to the R-value obtained in the figure, it can be considered that the BP neural network prediction model of new energy vehicles sales trained in this paper has good results.

Meanwhile, through the MATLAB program, the new energy vehicles sales prediction image can be obtained as shown in Figure 9 below, and its real and predicted values and errors from 2023.1 to 2024.3 are shown in Table 3 below.

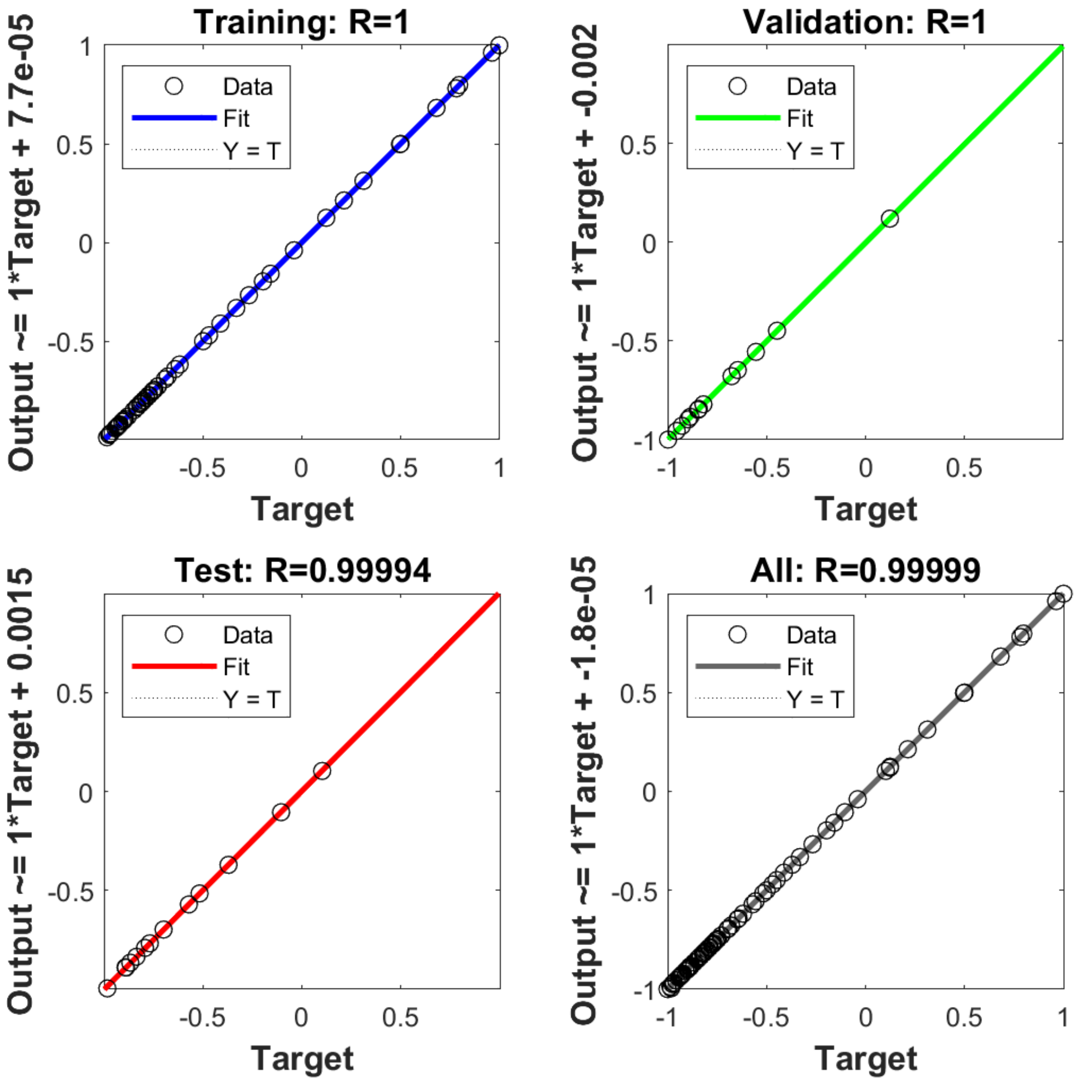


Figure 8. BP neural network regression chart

According to Figure 9 and Table 3, it can be learned that in the BP neural network prediction model, the prediction effect of the sales of the new energy vehicles is better overall, and the average absolute error of the prediction results is 1.36%, which is a better prediction, and it has certain reference value. At the same time, compared with the prediction effect of the ARIMA model in the previous section, the BP neural network prediction model is obviously better, so when assessing the development of new energy vehicle sales in China, this paper recommends the use of the BP neural network prediction model more.

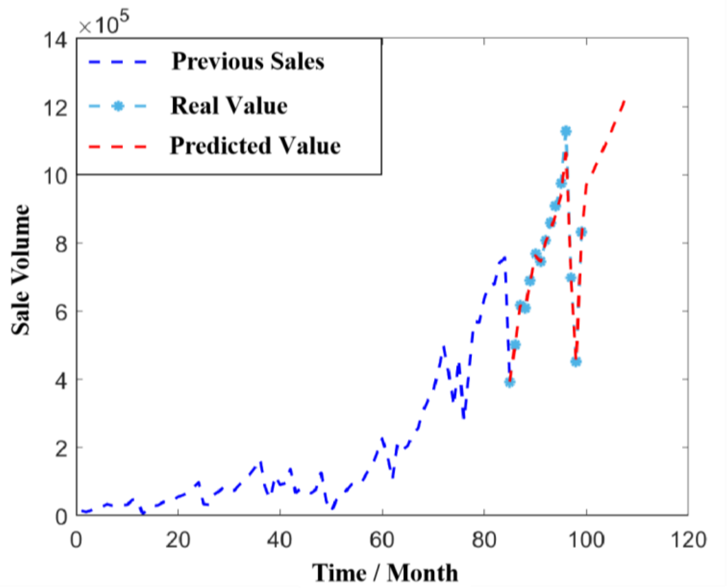


Figure 9. New energy vehicles sales prediction in BP neural network prediction model

Table 3. Forecasting errors of New energy vehicles sales in BP neural network forecasting model

Time	Real Value	Predicted Value	Error (%)
2023.1	392613	392270	− 0.09
2023.2	501118	502320	0.24
2023.3	617875	620370	0.40
2023.4	608758	613570	0.79
2023.5	689291	695450	0.89
2023.6	767338	772530	0.68
2023.7	746497	758490	1.61
2023.8	807555	817280	1.20
2023.9	858305	865460	0.83
2023.10	909808	918380	0.94
2023.11	975701	978130	0.25
2023.12	1127669	1101200	− 2.35
2024.1	698901	722710	3.41
2024.2	451344	471840	4.54
2024.3	831670	850350	2.25

3.3. Comprehensive evaluation of new energy vehicles development based on EWM-TOPSIS

3.3.1. Factors affecting the development of the new energy vehicles based on EWM

The six indicators selected for this paper are shown in Table 4 below. The data of each indicator from 2019.1 to 2023.12 are organized quarterly to form the data source of EWM. Here, production and sales are taken as the sum of 3 months, and penetration rate and charging pile ownership are taken as the average of 3 months.

Table 4. Indicators selected by EWM

Notation	Indicator
X_1	Power Battery Sales
X_2	Sales of the new energy vehicles
X_3	Production of the new energy vehicles
X_4	Penetration rate of the new energy vehicles
X_5	Number of public charging piles
X_6	GDP

Before weight calculation, the heat map of correlation between indicators was first obtained as shown in Figure 10 below. Except for the low correlation between GDP and other indicators, other indicators have high correlation with each other. But it is not difficult to see that the first five indicators have a key positive effect on the development of the new energy vehicles.

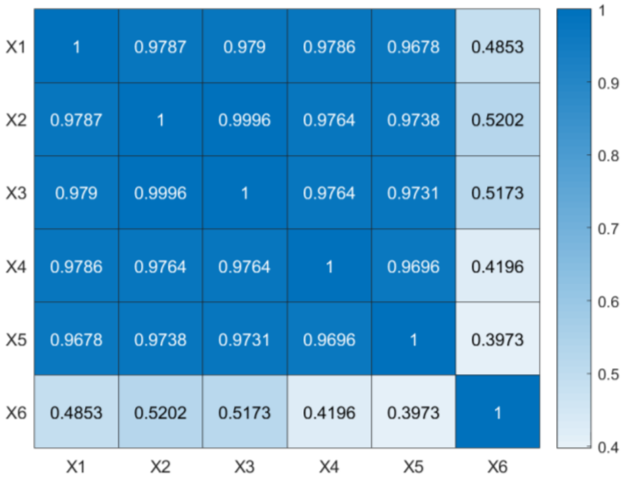


Figure 10. Heat map of correlation between indicators

First, the six indicators are standardized [14]. Considering that the larger each indicator is, the more conducive to the development of the new energy vehicles, the standardization processing formula for positive indicators is used here for all of them:

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)} \tag{7}$$

The weight of the j-th indicator in the i-th set of data P_{ij} is:

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}}, (i = 1,2,\cdots,n,\ j = 1,2,\cdots,6) \tag{8}$$

The information entropy E_j of each indicator is:

$$E_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} P_{ij} \ln P_{ij} \tag{9}$$

The weight W_j for each indicator is:

$$W_j = \frac{1 - E_j}{k - \sum E_j} (k = 6,\ j = 1,2,\cdots 6) \tag{10}$$

Through the program calculation, we can get the weights of the above six indicators as shown in Table 5 below. From the table, it can be analyzed that among the above indicators, the most critical factor affecting the development of the new energy vehicles is its penetration rate in the automobile market (weight up to 0.193), followed by the sales of power batteries (0.189) involved in its power.

Table 5. Weights of indicators from EWM calculations

Indicator	X_1	X_2	X_3	X_4	X_5	X_6
Weight	0.189	0.167	0.164	0.193	0.160	0.127

3.3.2. Comprehensive evaluation of regional development of the new energy vehicles based on TOPSIS

Considering that the sales volume of the new energy vehicles and its penetration rate are important factors reflecting the development of the new energy vehicles among regions, and the sales volume of this type of automobiles will also be affected by the population of each province, resulting in interregional disparities. In this paper, the sales volume and penetration rate of the new energy vehicles of 31 provinces in China (Hong Kong, Macao, and Taiwan are not included due to missing data) in 2023 are selected as data sources, which are comprehensively evaluated using TOPSIS to analyze the development of the new energy vehicles in each province.

Here, the data for the two selected indicators are both benefit-based indicators, so there is no need for additional data normalization.

The largest and smallest numbers of each indicator are taken to form the optimal solution vector z^+ and the worst solution vector z^- respectively, then for the i-th solution z_i, its distance from the optimal solution and the worst solution are:

$$d_i^+ = \sqrt{\sum_{j=1}^{2}(z_j^+ - z_{ij})^2} \tag{11}$$

$$d_i^- = \sqrt{\sum_{j=1}^{2}(z_j^- - z_{ij})^2} \tag{12}$$

Then the score (i.e., the fit of the optimal solution) for each solution is:

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{13}$$

Through the program calculation, this paper obtains the top five provinces in terms of TOPSIS composite score as shown in Table 6 below. The provinces with better development of new energy vehicles are mainly concentrated in the eastern coastal area. These provinces have high economic levels, large populations, and strong purchasing power of consumers. At the same time, there are numerous government policies given to the development and sales of local new energy vehicles [15], and the public infrastructure (e.g., power exchange stations, charging piles, etc.) is more numerous and better. These directly lead to a higher level of local development of the new energy vehicles.

At the same time, this paper also obtains the bottom five provinces in the TOPSIS composite score as shown in Table 7 below. Most of these provinces are located in the remote western and northeastern regions, where the economic development level is weak, the population is sparse, and the public infrastructure is not perfect, which directly leads to the low level of development of the new energy vehicles. In the future, if we need to improve the overall development level of the new energy vehicles in China, the government should first improve these provinces and regions, improve their infrastructure, help local economic construction, improve the living standard of the residents, so as to drive the development of the new energy vehicles in the local area.

Table 6. Top five provinces and their scores on the TOPSIS composite score for the development of new energy vehicles in China in 2023

Ranking	1	2	3	4	5
Province	Zhejiang	Jiangsu	Guangdong	Shandong	Henan
TOPSIS score	0.1115	0.0786	0.0743	0.0624	0.0592

Table 7. Bottom five provinces and their scores on the TOPSIS composite score for the development of new energy vehicles in China in 2023

Ranking	31	30	29	28	27
Province	Tibet	Heilongjiang	Xinjiang	Qinghai	Gansu
TOPSIS score	0.0004	0.0026	0.0048	0.0055	0.0071

3.4. Assessment of the development level of the new energy vehicles based on K-means cluster analysis

In order to more intuitively study the development level of the new energy vehicles among provinces and as a whole in China, this paper selects the sales and penetration rates of the new energy vehicles in China's 31 provinces (excluding Hong Kong, Macao, and Taiwan) in 2023 as the dataset under the K-means cluster analysis algorithm [17]. Through K-means cluster analysis, the clustering results can be obtained as shown in Figure 11 below, and its Silhouette Coefficient is 0.69896.

The clustering result divides the development level of each province into 3 categories by these two indicators, yellow represents that the development level of the new energy vehicles in the province is at the beginning stage, orange represents that the development level is at the middle stage, and blue represents that the development level is at the leading stage. Analyzing the development status of each province through this chart, the provinces marked with yellow and red are more, close to the origin and gathered, which indicates that most provinces in China are in the beginning and medium stage, and the difference in the development level between the provinces in these two stages is relatively small; the provinces marked with blue are fewer and dispersed, which indicates that there are fewer provinces in the new energy vehicles in the leading stage of development in China and there is a big difference in the development level of this type of provinces as well. There are large differences in the level of development of this category of provinces. On the whole, the development of the new energy vehicles in China is transitioning from the beginning to the middle stage.

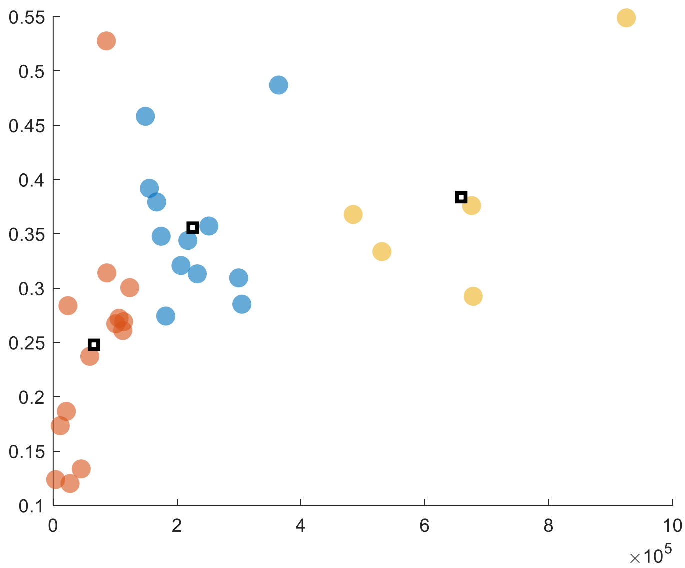


Figure 11. K-means cluster analysis result

4. Conclusions

With the introduction of relevant policies by the Chinese government and the advancement of related technologies, new energy vehicles have been developing rapidly in China in recent years.

In this paper, we make a comprehensive assessment of the development status of China's new energy vehicles by statistically analyzing the relevant data of China's new energy vehicles using EWM-TOPSIS, and carry out a comparative analysis of new energy vehicle sales forecasting through the ARIMA forecasting model and the BP neural network forecasting model. The research results show that: (1) China's new energy vehicles into the rapid development stage is around 2021, but at this stage of China's new energy vehicles still exists between the uneven development of the region, the provinces in the eastern coastal areas of new energy vehicle penetration rate can generally reach more than 30%, but the remote areas of the west is often less than 15%. (2) Compared with the ARIMA prediction model, the BP neural network prediction model has a better prediction effect in predicting the sales of new energy vehicles in China. (3) The data on the penetration rate of new energy vehicles plays a key role in predicting the future sales of new energy vehicles, followed by the sales of power batteries related to their power. (4) At this stage, the development of new energy vehicles in China is still in its infancy, and in most provinces, there is still a large gap between new energy vehicles and traditional fuel vehicles.

The innovation of this paper is that by comparing the ARIMA prediction model and the BP neural network prediction model, so as to get a more suitable method for predicting the sales of new energy vehicles in China. At the same time, EWM and TOPSIS are comprehensively applied to make a comprehensive assessment of the development of new energy vehicles in China.However, the factors selected in the EWM that affect the development of new energy vehicles have a certain correlation, which may have a certain impact on the assessment results. The relevant potential influencing factors and their relationships should be carefully analyzed in the future.

References

[1] Bingchun L ,Chengyuan S ,Qingshan W , et al.Research on regional differences of China's new energy vehicles promotion policies: A perspective of sales volume forecasting[J].Energy,2022,248

[2] Tang L ,Sun J .Predict the sales of New-energy Vehicle using linear regression analysis [J].E3S Web of Conferences,2019,11802076.

[3] Kexin L ,Hong S ,Bin L , et al.Research on New Energy Vehicle Market Penetration Rate Based on Nested Multinominal Logit Model [J].World Electric Vehicle Journal,2021,12 (4):249-249.

[4] Bo Z ,Hui L ,Cuiwei M , et al.Modeling, prediction and analysis of new energy vehicle sales in China using a variable-structure grey model [J].Expert Systems With Applications,2023,213 (PA):

[5] Yuan Y .Application of Multiple Linear Regression and Time-Series Models for Forecasting Sales of New Energy Vehicles [C]//Xiamen University,2023:

[6] Yang X ,Chen R ,Yang B , et al.Uncertain Bass model with application to new energy vehicle sales forecasting [J].Information Sciences,2025,689121450-121450.

[7] Dhankhar S ,Dhankhar N ,Sandhu V , et al.Forecasting Electric Vehicle Sales with ARIMA and Exponential Smoothing Method: The Case of India[J].Transportation in Developing Economies,2024,10 (2):32-32.

[8] Yan Z ,Jun Z .Research on automobile sales prediction model based on BP neural network[C]//Wuhan Business Univ. (China);China Automotive Technology and Research Ctr. Co.,Ltd. (China),2022:

[9] Xiang Z .Prediction of Purchase Volume of Cross-Border e-Commerce Platform Based on BP Neural Network [J].Computational Intelligence and Neuroscience,2022,20223821642-3821642.

[10] Kaczyńska A ,Sulikowski P ,Wątróbski J , et al.Enhancing Sustainable Assessment of Electric Vehicles: A Comparative Study of the TOPSIS Technique with Interval Numbers for Uncertainty Management [J].Energies,2023,16 (18):

[11] Hui Z ,Zeng J ,Zhang Y .Analyzing the Influence Factors of New Energy Vehicles on Traditional Energy Vehicle Industry in China Based on Entropy Weighted TOPSIS and Pearson Analysis [J].Industrial Engineering and Innovation Management,2024,7 (3):

[12] Choi J S ,Jiao J .Uncovering electric vehicle ownership disparities using K-means clustering analysis: A case study of Austin, Texas [J].Journal of Computational Social Science,2024,7 (3):2403-2456.

[13] Kang Y ,Wu C ,Yu B .Time series clustering based on polynomial fitting and multi-order trend features [J].Information Sciences,2024,678120939-120939.

[14] Shengyuan W .Exploring the Sustainability of China’s New Energy Vehicle Development: Fresh Evidence from Population Symbiosis [J].Sustainability,2022,14 (17):10796-10796.

[15] Ning T ,Lu B ,Ouyang X , et al.Prospect and sustainability prediction of China's new energy vehicles sales considering temporal and spatial dimensions [J].Journal of Cleaner Production,2024,468142926-142926.

[16] Li X ,Wang Z ,Jiang S , et al.Financial subsidy, government audit and new transportation technology: Evidence from the new energy vehicle pilot city program in China [J].Research in Transportation Economics,2024,106101447-.

[17] Khan F ,Khan O ,Parvez M , et al.K-means clustering optimization of various quantum dots and nanoparticles-added biofuels for engine performance, emission, vibration, and noise characteristics [J].Thermal Science and Engineering Progress,2024,54102815-102815.