ISSN 2706-655X Vol.6, Issue 1: 15-21, DOI: 10.25236/IJFET.2024.060103

The Influence Model of New Energy Vehicle Development Based on Correlation Analysis

Ye Zhu[1,a,#], Zhicheng Lian[2,b,#], Yikun Chen[3,c,#], Chengguan Wei[4,d,#], Runze Yu[5,e,#], Junchang Chen[6,f,*]

[1]*Textile College, Donghua University, Shanghai, China*
[2]*School of Big Data, Guangzhou City University of Technology, Guangzhou, China*
[3]*Department of Economic and Management, Zhanjiang Preschool Education College, Zhanjiang, China*
[4]*School of Economics and Management, Communication University of China, Beijing, China*
[5]*College of Chemical and Biological Engineering, Shandong University of Science and Technology, Qingdao, China*
[6]*School of Fire Protection Engineering, China People's Police University, Langfang, China*
[a]*dhu_zhuye@163.com,* [b]*15819491559@163.com,* [c]*chen954223@163.com,* [d]*13407772880@163.com,* [e]*19862273635@163.com,* [f]*15041461131@163.com*
[#]*These authors contributed equally to this work and should be considered co-first authors*
[*]*Corresponding author*

Abstract: *This article provides a comprehensive and in-depth look at the development of America's new energy electric vehicle industry, employing multiple mathematical modeling techniques to analyze in depth various key aspects of the industry's development. First, we focus on the market acceptance and distribution of NEV sales. In order to improve the generalization ability of the model and prevent overfitting, the data normal distribution check and possible data transformation are carried out. Then we explore the main factors affecting the development of new energy vehicles by calculating the Pearson correlation coefficient between the data. Next, we calculate the evaluation scores of new energy vehicles in different years by constructing TOPSIS model. Finally, we introduce how the LSTM model solves the gradient disappearance or explosion problem of traditional RNNS when dealing with long sequences. The analysis includes the decrease of MSE and RMSE values of the test set and the training set, which shows that the model performs well in the training process and the performance of the test set is also improved.*

Keywords: *New Energy Electric Vehicle, Pearson Correlation Coefficient, TOPSIS, LSTM*

1. Introduction

In dealing with environmental and energy issues brought about by the automobile industry, the international community has reached a consensus to develop new energy vehicles as a strategic choice for the country in the 21st century. Countries have regarded the development of new energy vehicles as an important measure to seize emerging markets, and have adopted preferential policies such as subsidies and tax cuts to upgrade their technological level and strive for leadership in the field. The American government attaches great importance to the development of the new energy vehicle industry, and has put forward a number of policy measures and incorporated them into the national development strategy. Although America's new energy vehicles have broad prospects, the market is still at an initial stage of development. Since 2012, sales of new energy vehicles in America have gradually increased, but there is still a gap compared with foreign markets. It is still difficult to meet the 2020 target set by the government. In the United States, 96,000 new energy vehicles were sold in 2013, and the sales of a single Tesla brand exceeded the total number of new energy vehicles in America. America still faces major challenges in achieving "curve overtaking" in the auto industry.

2. Related Works

The new energy vehicle industry has been a focus in America in recent years, and its development is influenced by various factors. Here are some important references on the American new energy vehicle industry, covering spatial distribution to policy impact, providing valuable insights into this industry. [1] Cao et al. (2022) studied the spatial distribution of the new energy vehicle industry, revealing development patterns and influencing factors in different regions of America. [2] Su et al. (2020) conducted field surveys in Shanghai and Nanjing, investigating factors influencing user satisfaction with new energy vehicles. Their findings contribute insights for marketing and product improvement. [3] Zhang and Wu (2021) focused on the spatial association network structure for innovation efficiency in America's new energy vehicle industry, offering a unique perspective on understanding innovation networks. [4] He et al. (2022) examined the impact of oil prices on innovation in American new energy vehicle enterprises, providing clues on how oil price fluctuations shape innovation strategies. [5] Liu et al. (2023) explored the multi-objective development path evolution of new energy vehicle policies through big data analysis, offering guidance for policymakers in promoting sustainable development. These studies collectively reveal various aspects of the American new energy vehicle industry, from regional characteristics to user satisfaction, innovation, and policy factors [6]. These research works provide a rich reference for the future development of the industry.

3. Theory and Method

3.1 TOPSIS

We analyze the main factors affecting the development of new energy electric vehicles in America, including policy support, technical level, market demand, and infrastructure and so on. These factors can be described by mathematical models, for example, we can use regression analysis to study the impact of policy support and technology level on the development of new energy electric vehicles.

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a decision support method used to identify the best scheme. Its steps are as follows:

Step 1: Using the standard 0-1 transformation, the data are non-scale and normalised to obtain the normalised decision matrix:

$$B = \left(b_{ij}\right)_{mxn} \tag{1}$$

For a benefit attribute j, let

$$b_{ij} = \frac{a_{ij} - a_{jmin}}{a_{jmax} - a_{jmin}} \tag{2}$$

For a costly attribute j, let

$$b_{ij} = \frac{a_{jmax} - a_{ij}}{a_{jmax} - a_{jmin}} \tag{3}$$

Step 2: Composition of weighted canonical array. Using the coefficient of variation method the weight vector for each attribute is obtained as.

$$\begin{gathered} W = [W_1, W_2, \dots, W_n]^T \\ C_{ij} = W_j \cdot b_{ij}, i = 1,2,\dots,m, j = 1,2,\dots,n. \end{gathered} \tag{4}$$

Step 3: Determine the positive ideal solution C^* and negative ideal solution $C°$

Let the first attribute value of the positive ideal solution be j attribute of the positive ideal solution is C_j^* , and the value of the first attribute of the negative ideal solution is j The value of the first attribute of the negative ideal solution is C_j^* , then the positive ideal solution.

$$C_j^* = \begin{cases} \overset{max}{i} C_{ij} \\ \overset{min}{i} C_{ij} \end{cases} j = 1,2,\dots,n. \tag{5}$$

Negative Ideal Solution.

$$C_j^{\circ} = \begin{cases} \overset{min}{max} c_{ij} \\ {}_i c_{ij} \end{cases} j = 1,2,\dots,n. \tag{6}$$

Step 4: Calculate the distance from each scenario to the positive and negative ideal solutions: scenarios d_i. The distance to the positive ideal solution is.

ISSN 2706-655X Vol.6, Issue 1: 15-21, DOI: 10.25236/IJFET.2024.060103

$$S_i^* = \sqrt{\sum_{j=1}^{n} \left(C_{ij} - C_j^*\right)^2}, i = 1,2, \ldots, m \tag{7}$$

The distance to the negative ideal solution is

$$S_i^\circ = \sqrt{\sum_{j=1}^{n} \left(C_{ij} - C_j\right)^2}, i = 1,2, \ldots, m. \tag{8}$$

Step 5: Calculate the importance indicator value (i.e., composite evaluation index) for the vegetable category.

$$CEI = \frac{S_i^\circ}{S_i^\circ + S_i^*}, i = 1,2, \ldots, m \tag{9}$$

3.2 Pearson Correlation Analysis

Pearson correlation coefficient is essentially a kind of linear correlation coefficient in statistical methods, its analysis is usually used to measure the linear relationship between fixed-distance variables, and this paper is used to measure the possible linear correlation between different categories of vegetable products or different single products. However, since the results of questionnaire survey are attribute data, which cannot be directly used for mathematical calculation, it is necessary to quantify them into numbers and then estimate the sample correlation coefficient. Given two samples $= (x_1, x_2, \cdots, x_n)$, Pearson's correlation coefficient is defined as $Y = (y_1, y_2, \cdots, y_n)$, the Pearson correlation coefficient is defined as

$$r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right) \tag{10}$$

Among others $\bar{x}$、$\bar{y}$ are x and y the respective sample means, and σ_x、σ_y are the sample mean of x and y. The larger the absolute value of the Pearson correlation coefficient, the higher the degree of correlation between the dependent variable and the independent variable. $|r| < 0.3$, denotes weak correlation, denotes low correlation, denotes high correlation; and $0.3 < |r| < 0.5$ indicates significant correlation, $0.5 < |r| < 0.8$ indicates high correlation, $0.8 < |r| < 1$ indicates high correlation.

3.3 LSTM

A short term memory network (LSTM) is a special form of recurrent neural network (RNN) that specializes in processing time series data with long-term dependencies. Compared with traditional RNN, LSTM introduces three key gating structures--forget gate, input gate and output gate, which help to optimize long-term dependence problems. And the LSTM neural network structure diagram is shown in Figure 1.

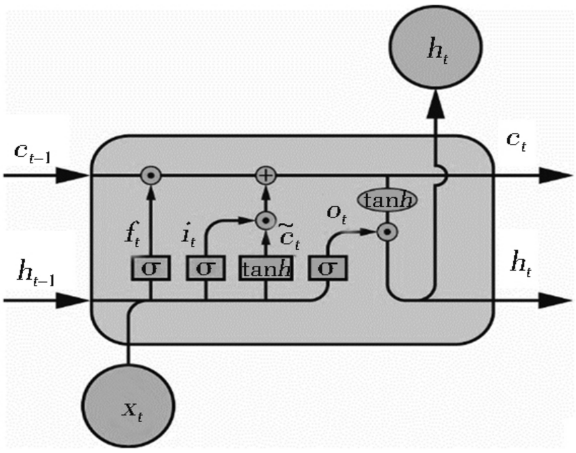

Figure 1: LSTM neural network structure diagram

The principle of the LSTM neural network is as follows:

1). Forget the gate

ISSN 2706-655X Vol.6, Issue 1: 15-21, DOI: 10.25236/IJFET.2024.060103

When a new input enters the LSTM network, a forget gate decides what information should be forgotten or retained. The forget gate uses a sigmoid function to determine which information should be forgotten. At each time step t , the forgetting gate is calculated as follows.

$$f_t = \sigma\left(W_f[x_t, h_{t-1}] + b_f\right) \tag{11}$$

2).Input Gate

When a new input enters the LSTM network, the input gate decides which information should be retained and updates the cell state. The input gate uses a sigmoid function to determine which information should be retained. At each time step t , the input gate is calculated as follows.

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \tag{12}$$

3).Cell State

The cell state can be regarded as the core of the entire LSTM network, which can store and transfer information, as well as control the flow and update of information. The cell state of the LSTM will be updated and transferred to the next time step. At each time step, the cell state is updated as follows t The cell state of LSTM is updated and passed to the next time step.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{13}$$

4).Output Gate

When information from the current time step needs to be passed to the next layer or output layer, an output gate is needed to control which information should be output. The output gate uses a sigmoid function to determine which information should be output. At each time step, the output gate is calculated as follows.

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o) \tag{14}$$

The prediction process of LSTM is as follows: take the data of 90% is the training set and 10% is the test set. Normalise the data set to remove the scale, set 200 hidden units, set the solver to 'adam' and perform 300 rounds of training. The advantages are: the step size of the update can be limited to an approximate range (initial learning rate) and the step annealing process can be implemented naturally (automatic adjustment of the learning rate).

4. Results and Discussions

4.1 Results of TOPSIS

The results are calculated as follows in the TOPSIS Score column of the table, as shown in Table 1.

Table 1: TOPSIS results

Year	2011	2012	2013	2014	2015	2016
TOPSIS_Score	0	0.000868	0.002383	0.010368	0.042266	0.064161
Year	2017	2018	2019	2020	2021	2022
TOPSIS_Score	0.109901	0.182526	0.218875	0.266008	0.548898	1

4.2 Results of Pearson Correlation Analysis

We visualized how each indicator correlates with all other indicators and development scores using a correlation heat map, which is a method of visualizing a matrix of correlation coefficients. It uses color changes to show the magnitude of the correlation coefficients be-tween different variables, making the correlation analysis more intuitive. In order to judge the relationship between the indicator and the development score, we only need to look at the rightmost column of this heat map. The correlation analysis result chart is shown in Figure 2.

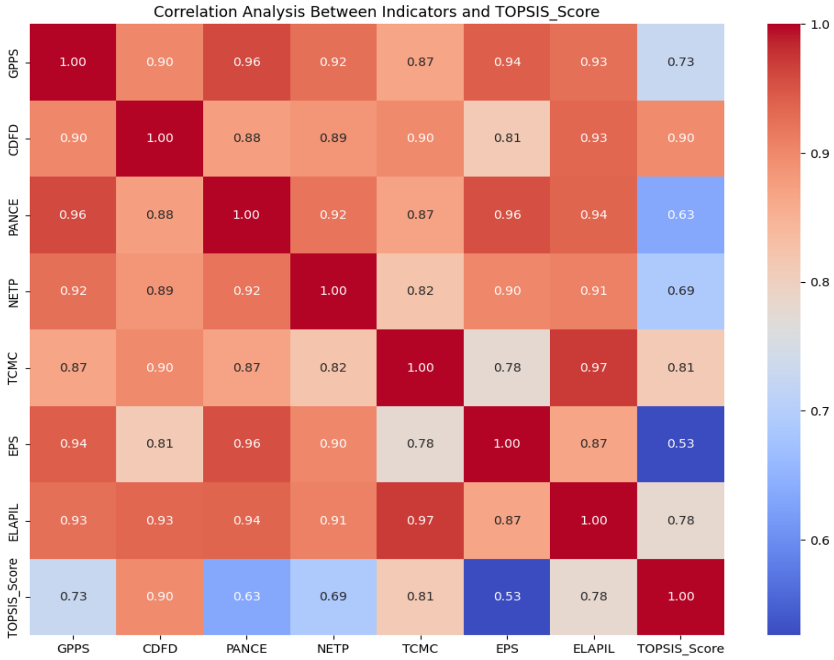

Figure 2: Correlation analysis result chart

The analysis results show that EPS (Environmental Protection Sense) has a low correlation with TOPSIS Score, but because its value is not extremely low, it is not sure whether it should not be taken as an independent variable of multiple regression. Therefore, we separately studied the fitting effect with or without EPS as the independent variable.

4.3 Results of LSTM

In this model, the data collected in TOPSIS method was used. Mean square error (MSE) and root mean square error (RMSE) are used as loss functions to evaluate and optimize the model performance during training. Through 23 rounds of training, an early stop strategy was used to prevent over fitting. The model training iterative loss results is shown in Figure 3.

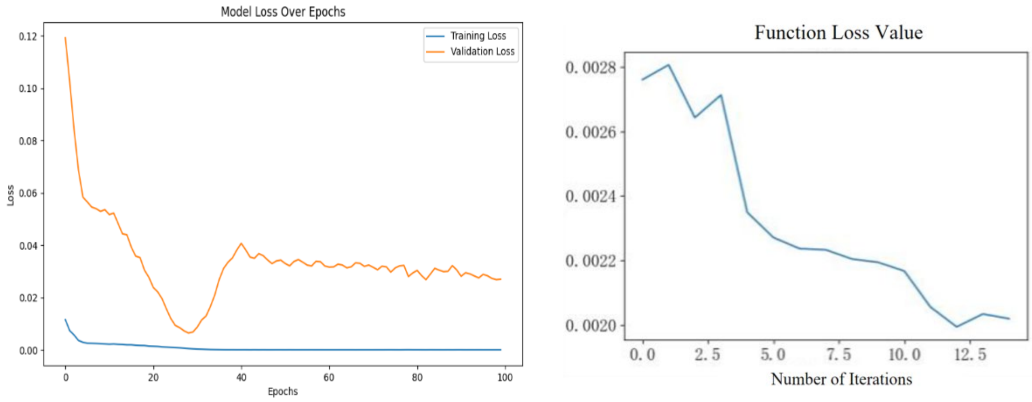

Figure 3: Model training iterative loss results

Changes in loss values show a downward trend in both training losses and validation losses over the course of training as the number of training sessions increases. This indicates that the performance of the model is gradual while learning the data set.

The forecast results of the LSTM model show that America's new energy electric vehicle sales will continue to grow in the next decade. By comparing historical sales and model forecast sales, we found that the model can effectively capture the dynamics of market development. This provides a strong forecast for the future development trend of new energy electric vehicles. The LSTM model prediction results are shown in Figure 4.

ISSN 2706-655X Vol.6, Issue 1: 15-21, DOI: 10.25236/IJFET.2024.060103

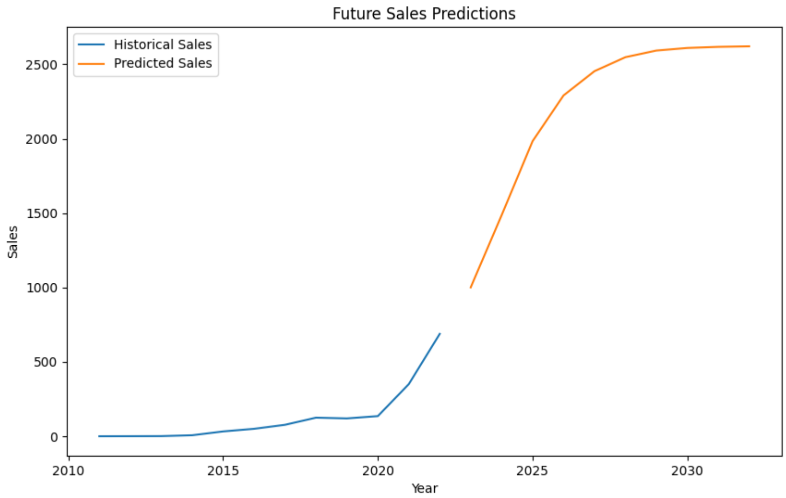

Figure 4: LSTM Model Prediction Results

Detailed forecast results for the next ten years are shown in the table 2 below:

Table 2: Forecast results of LSTM

Year	CNEVS	Stock	CNEVMS
2023	1048.9	2271.3	36.4
2024	1601.1	3355.4	54.6
2025	2172.5	4434.3	72.1
2026	2510.9	5063.6	81.5
2027	2681.4	5380.2	85.8
2028	2771.9	5548.3	88
2029	2812.5	5624.6	88.9
2030	2827.4	5653.4	89.3
2031	2833.3	5665.1	89.4
2032	2836	5670.4	89.5

America's new energy vehicle market is showing strong growth momentum in the next decade. According to the forecast of the LSTM model, it will climb rapidly from about 10.489 million units in 2023 to about 28.36 million units in 2032, almost achieving three-fold growth. The huge growth is driven not only by the maturity of the domestic market and technological innovation, but also by the global consensus on reducing carbon emissions and achieving a green energy transition. The forecast for ownership also shows strong growth, soaring from about 22.713 million vehicles in 2023 to 56.704 million vehicles in 2032.

5. Conclusion

The rapid growth of NEV market share is highlighted, highlighting the significant upward trend of NEV in the global automotive market. The growth of market share has a very high positive correlation with the year-on-year growth rate of NEV sales, indicating that the increase of NEV sales is the main driver of market share growth. However, indicators related to technological innovation, such as the amount of patent disclosure and the amount of patent grants, show a negative correlation with the growth of market share, indicating that the impact of technological innovation on market share may be more reflected in the medium and long term.

ISSN 2706-655X Vol.6, Issue 1: 15-21, DOI: 10.25236/IJFET.2024.060103

References

[1] Cao L, Deng F, Zhuo C, et al. Spatial distribution patterns and influencing factors of America's new energy vehicle industry. Journal of Cleaner Production, 2022, 379: 134641.
[2] Su D, Gu Y, Du Q, et al. Factors affecting user satisfaction with new energy vehicles: A field survey in Shanghai and Nanjing. Journal of Environmental Management, 2020, 270: 110857.
[3] Zhang Y, Wu Z. Research on the spatial association network structure for innovation efficiency of America's new energy vehicle industry and its influencing factors. Plos one, 2021, 16(8): e0255516.
[4] He J, Li J, Zhao D, et al. Does oil price affect corporate innovation? Evidence from new energy vehicle enterprises in America. Renewable and Sustainable Energy Reviews, 2022, 156: 111964.
[5] Liu Q, Wen X, Cao Q. Multi-objective development path evolution of new energy vehicle policy driven by big data: From the perspective of economic-ecological-social. Applied Energy, 2023, 341: 121065.
[6] Nasser Liser. New Energy Vehicle Development Strategy Integrated with Finite Volume Method. Kinetic Mechanical Engineering, 2020, 1(2): 45-53.