Optimizing the Performance of Factor Analysis Model by Using Clustering——Take Electric Vehicle as an Example

Ziyu Wang*

Department of Mathematics and Statistics, Dalhousie University, Halifax, Canada

*Corresponding author: zy855449@dal.ca

Abstract. As people pay more attention to environmental protection, electric vehicles occupy a larger share of the entire automobile sales market. However, due to the rapid development of electric vehicles and their structure being quite different from that of traditional fuel vehicles, many potential consumers still have some misunderstandings when buying electric vehicles, they still use the logic of traditional fuel vehicles to examine electric vehicles. This paper uses factor analysis and EM cluster analysis to analyze the data on electric vehicles (EVs) and try to maximize the information on EVs at a smaller cost. This paper first introduces the ideas of factor analysis and EM cluster analysis as well as the data related to EVs and then states the generated models. Then, analyzes the data of electric vehicles using two methods respectively and obtains the analysis results. Finally, this paper combines the two analysis methods to get the final model with higher universal applicability and pertinence.

Keywords: Factor analysis; EM cluster; dimensionality reduction.

1. Introduction

1.1 Background

Electric vehicles have the characteristics of high efficiency, energy saving, and zero emission. More than 17.4m new electric vehicles will be registered in the eu, US and China by 2030, accounting for nearly 27 percent of total vehicle sales, according to a recent report by PWC [1], the consultancy. After several years of development, The electric vehicle market has shifted from being policy-driven to market-driven.

However, most potential consumers of electric cars do not know as much about electric vehicles as they do about their gas-powered counterparts. The sense is that electric cars are cheaper to build because they do not have engines, transmissions, fuel pumps, etc [2]. In addition, the driving cost of electric vehicles is much lower than that of traditional fuel vehicles. Moreover, most electric vehicle manufacturers pursue a simple design style in appearance and interior decoration, so electric vehicles leave the impression of being cost-effective and lower cost to most people [3].

In fact, according to Oliver Wyman's statistical analysis, this is not the case. The total manufacturing cost of building a compact (C-class) fuel car is around 13,900 euros, while the total cost of building an electric car is up to 20,200 euros, which is 45% more than a fuel car [4].

1.2 Objection

Both fuel vehicles and electric vehicles in the market are subdivided into many different models, so it is completely insufficient for us to only use factor analysis to try to get a universal model. Therefore, this paper first uses cluster analysis to get different clusters and then applies factor analysis to these clusters to build models, which can improve the interpretability of the whole model, in order to achieve the purpose of using fewer variables to better feedback and complete information. This paper also uses the obtained model to calculate the estimated price of different models and then compares it with its real price to judge the relationship between the real price and the expected price of electric vehicles in the data set under this model system.

1.3 Data

The data set is a sample of electric vehicles on the market until 2022, there are 194 pieces of data and 11 variables in the data set, most of these variables are the performance parameters of electric vehicles, which can be roughly divided into the following aspects: motor, battery, and static parameters. At the same time, this paper also obtained the selling price of these electric vehicles from the Internet [5]. Through the above two methods, this paper aims to maximize the display of all information by using fewer factors and by clustering the data.

1.4 Models

1.4.1 Factor analysis

Factor analysis can help us identify underlying variables, or factors, that explain the correlation patterns in a set of observed variables. Factor analysis is often used for the dimensionality reduction of data. The purpose is to identify a few factors that can explain most of the variance observed in many dominant variables. The idea of factor analysis is to recombine the original variables without a trade-off and finally get multiple common factors affecting these variables to achieve the purpose of dimensionality reduction and data simplification. At the same time, the interpretability of the model can be improved through factor rotation.

1.4.2 K-means clustering method

The k-means clustering method is an iterative clustering analysis algorithm. It divides the data into k groups, randomly selects k objects as the first cluster center, calculates the distance of each object from each seed cluster center, and assigns each object to the nearest cluster center. After assigning each sample, the center is recalculated according to the cluster objects until the center of the cluster does not change anymore. Finally, the effects of classifying the data can be achieved and adequacy can be ensured for subsequent modeling.

2. Methodology

In this paper, the factor analysis method is first used to observe whether the data set can maximize feedback and complete information by using fewer variables through dimensionality reduction, and at the same time to judge whether the newly generated factors have certain interpretability. Then, this paper uses K-means clustering, aiming to cluster this unclassified data set by simple machine learning and make up for the lack of data classification by vehicle type in this data set. Finally, in order to further improve the interpretability of the whole model, by combining the above two methods, firstly clustering the data set and then factor analysis of each cluster, so as to obtain a more universal and targeted model.

2.1 Factor analysis

There were some missing sales prices of electric vehicles in the data set, and they could not be processed by the general method of replacing missing values for the time being since the added value brought by such factors as the brand effect cannot be estimated. Therefore, the data that did not contain the selling prices of electric vehicles were deleted and obtained the observed random vector $x = (x_1, x_2, \dots, x_p)'$ with p dimensions, for which the mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$ and covariance matrix $\Sigma = \sigma_{ij}$, and the general model of the data set is [9]

$$\begin{cases} x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 \\ x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + \varepsilon_p \end{cases} \tag{1}$$

Where $f = (f_1, f_2, \dots, f_m)'$ is a vector of $m < p$ underlying latent variables, $A = (a_1, a_2, \dots, a_m)'$ is a $m \times p$ matrix of factor loadings, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)'$ are uncorrelated zero-mean disturbances. The equation (1) can be denoted by matrix or vector as:

$$\tilde{x} = x - \mu = Af + \varepsilon \tag{2}$$

The covariance of x is as follows:

$$Cov(x) = ACov(f)A' + Cov(\varepsilon) = AA' + \psi \tag{3}$$

Assuming that
$E(f) = 0$,
$E(\varepsilon) = 0$,
$V(\varepsilon) = D = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$,
$Cov(f, \varepsilon) = E(f\varepsilon') = 0$.

The above equation (2) and the assumptions made constitute the orthogonal factor model. According to the above assumption, f is independent of each other, ε is independent of each other, and is independent of f.

Since ψ is a diagonal matrix, the non-diagonal elements of Σ can be determined by the elements of A, that is, the factor loading A completely determines the covariance between original variables. By standardizing each component of random vector x, then Σ can be the correlation matrix R, which is

$$R = AA' + \psi \tag{4}$$

In order to simplify the structure of the factor load array obtained by factor analysis, the factor load array is often transformed. Both orthogonal and oblique rotation is applicable [6]. Orthogonal rotations artificially force factors to be uncorrelated, whereas oblique rotations allow factors to be correlated.

The expression of the variance of x_i as

$$Var(x_i) = Var\left(\sum_{i=1}^{m} a_{ij} f_j + \varepsilon_i\right) = \sum_{i=1}^{m} a_{ij}^2 Var(f_j) + Var(\varepsilon_i) = h_i^2 + \sigma_i^2 \tag{5}$$

h_i^2 is the contribution of all the common factors f to the total variance of the variable x_i, which is called the common factor variance, as well, as the sum of squares of the elements in the factor load matrix A:

$$h_i^2 = \sum_{i=1}^{m} a_{ij}^2 \tag{6}$$

The factor analysis model obtained is $\tilde{x} = x - \mu = Af + \varepsilon$, assume $\Gamma = \gamma_{ij}$ as an orthogonal matrix, by doing the orthogonal transformation,

$$h_i^2 B = h_i^2(A),\ g_j^2 B = \sum_{k=1}^{p} \gamma_{ki}\, g_k^2(A) \tag{7}$$

Where $B = b_{ij}$.

This indicates that after orthogonal rotation, the common factor variance h_i^2 does not change, but the variance contribution of the common factor g_j^2 is no longer the same as the original one. So, a reasonable explanation for the factors can be obtained [7].

After the establishment of the factor analysis model, we can use this model to evaluate the status of each variable in the whole model, that is, to conduct a comprehensive evaluation.

Factor score function as the linear combination of common factor f represented by variable x is

$$f_j = a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jp}x_p, j = 1, 2, \ldots, m \tag{8}$$

thus, the common factor score of each component can be obtained according to equation (8).

2.2 K-means cluster

K-means is the most used clustering algorithm based on Euclidean distance, which believes that the closer the distance between two targets, the greater the similarity.

In the data set, by setting $D = \{x_1, x_2, \ldots, x_m\}$, and to minimize square error $W(C)$ of the division of clusters by clustering $C = \{C_1, C_2, \ldots, C_k\}$ as [10]

$$W(C) = \sum_{i=1}^{k} \sum_{x \in c_i} ||x_i - \mu_i||^2 \tag{9}$$

For each cluster C_i, recalculate its cluster center $\mu_i = \frac{1}{|c_i|}\sum_{x \in c_i} x$. Then repeat the above steps until a certain termination condition (using the minimum change in error in the model) is reached.

3. Data analysis

In this part, the above-mentioned factor analysis and K-mean analysis are applied to analyze the data. At the beginning the correlations between each variable should be tested as follow.

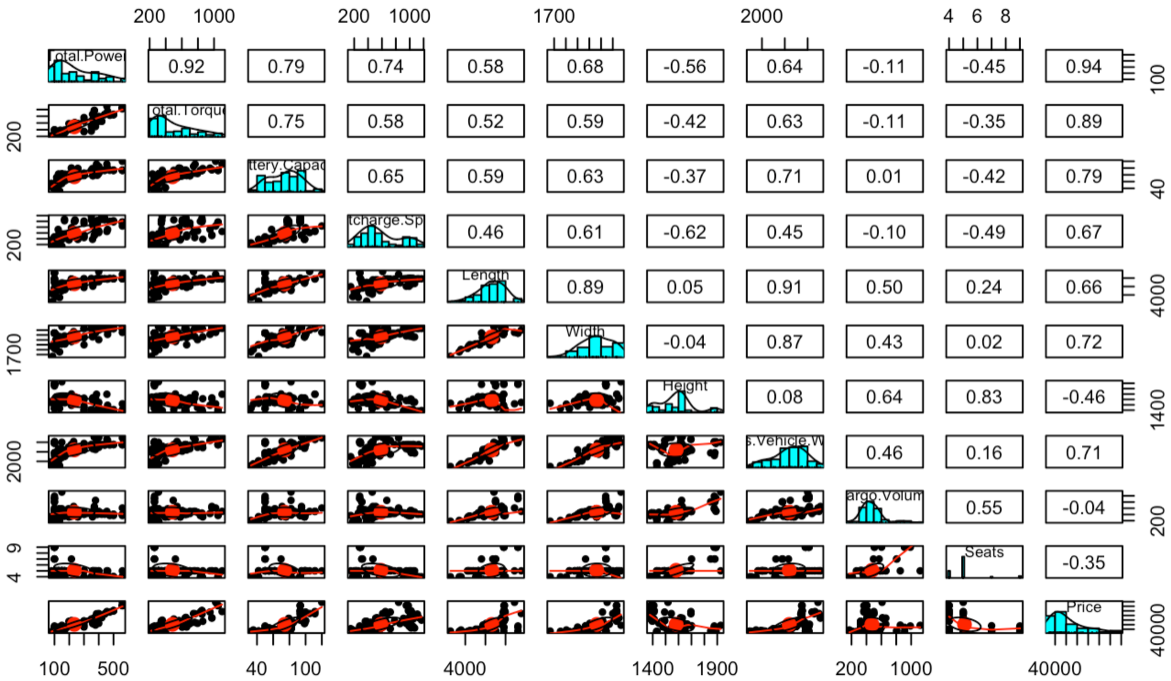

Figure 1. Correlation coefficient test (Photo credit: Original)

The correlation coefficient is between $(-1, 1)$, and the closer the value is to 0, the worse the correlation is. In contrast, the closer every two variables get to ± 1, the stronger the correlation. According to Fig. 1, we can obtain most variables get strong correlations, which is a prerequisite for us to conduct factor analysis and cluster analysis.

In the above discussion of factor analysis, we know that the first step in factor analysis is to get the number of factors, the scree plot below indicates the optimal number of factors in this model.

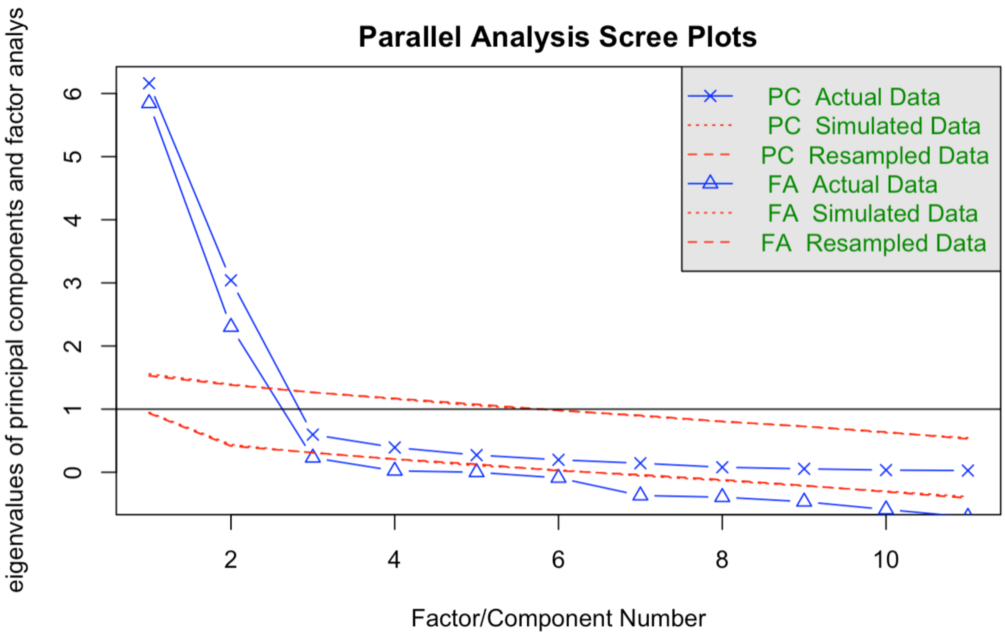

Figure 2. Scree plot to determine the component number (Photo credit: Original)

In Fig. 2, by observing the three lines of the FA actual data, FA simulated data, and FA resampled data related to factor analysis, we can conclude that three factors can well interpret the data set.

Based on the above discussion on factor rotation, we use the maximum likelihood estimation method to generate the following results.

```
Loadings:
                    Factor1 Factor2 Factor3
Total.Power          0.79    0.41   -0.41
Total.Torque         0.92
Battery.Capacity     0.63    0.48
Price                0.77    0.48   -0.31
Length               0.39    0.86
Width                0.44    0.83
Gross.Vehicle.Weight 0.55    0.75
Fastcharge.Speed     0.34    0.54   -0.58
Height                                0.93
Cargo.Volume                 0.49     0.64
Seats                                 0.84
```

Figure 3. Loadings via maximum likelihood estimation (Photo credit: Original)

Fig.3 shows that the first load factor mainly reflects the common characteristics of total power, total torque, and price, while the second mainly reflects the characteristics of length, width, and weight, and the third the characteristics of height and the number of seats.

```
               Factor1 Factor2 Factor3
SS loadings        3.3    3.22    2.81
Proportion Var     0.3    0.29    0.26
Cumulative Var     0.3    0.59    0.85

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 148.89 on 25 degrees of freedom.
The p-value is 1.35e-19
```

Figure 4. Contributions of factors (Photo credit: Original)

The cumulative variance contribution of these three factors in Fig.4 is 85%. In addition, assuming that the validation results give the p-value slightly less than 1.35×10^{-19} for the characteristics of these variables, it can be seen that these three factors are sufficiently indicative of these variables.

The Fig. 5 below is the factor load matrix A:

	[,1]	[,2]	[,3]
[1,]	0.79	-0.31	0.54
[2,]	0.41	0.39	-0.58
[3,]	-0.41	0.86	0.00
[4,]	0.92	0.00	0.00
[5,]	0.00	0.44	0.93
[6,]	0.00	0.83	0.00
[7,]	0.63	0.00	0.49
[8,]	0.48	0.55	0.64
[9,]	0.00	0.75	0.00
[10,]	0.77	0.00	0.00
[11,]	0.48	0.34	0.84

Figure 5. Factor load matrix (Photo credit: Original)

By combining equations 8 and 9 to get equation 10, and obtain the scaled factor score coefficients (See Appendix for details)

$$f = \begin{pmatrix} f_1 \\ \cdots \\ f_2 \end{pmatrix} = \begin{pmatrix} b_1 x \\ \cdots \\ b_2 x \end{pmatrix} = Bx = A'R^{-1}x \tag{10}$$

The factor score function is thus obtained as follows:

$$f_1 = 0.46 * total\ power - 0.20 * total\ torque - 0.45 * battery\ capacity + 0.88 * fast\ charge\ speed - 0.23 * length - 0.30 * width + 0.31 * height + 0.74 * gross\ weight - 0.25 * cargo\ volume + 0.04 * number\ of\ seats - 1.07 * price$$

$$f_2 = -1.15 * total\ power + 1.08 * total\ torque + 1.15 * battery\ capacity - 1.09 * fast\ charge\ speed - 0.87 * length + 1.12 * width - 1.12 * height - 1.14 * gross\ weight + 1.10 * cargo\ volume + 0.98 * number\ of\ seats + 0.17 * price$$

$$f_3 = 0.69 * total\ power - 0.89 * total\ torque - 0.70 * battery\ capacity + 0.21 * fast\ charge\ speed + 1.09 * length - 0.81 * width + 0.81 * height + 0.39 * gross\ weight - 0.84 * cargo\ volume - 1.02 * number\ of\ seats + 0.90 * price$$

Next, applying K-means clustering to analyze the sample. The optimal cluster number can be obtained through iteration.

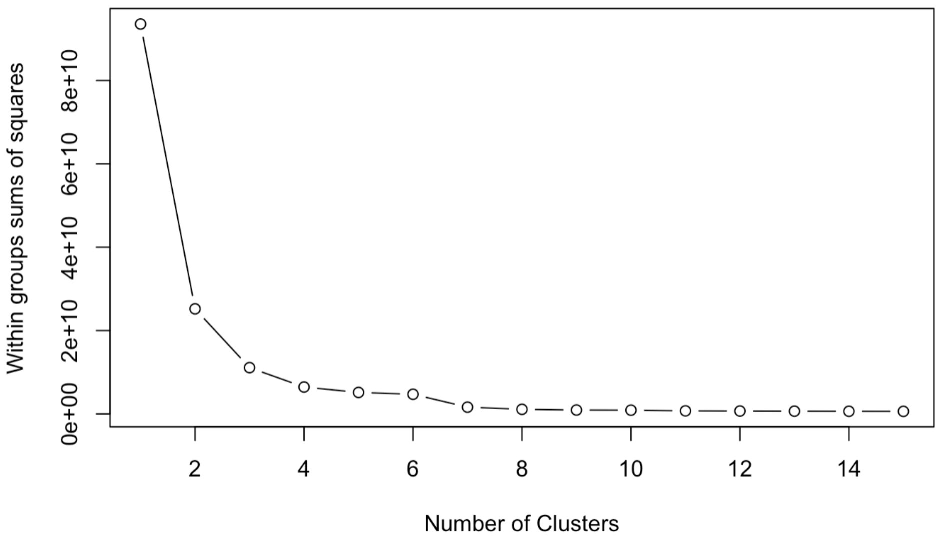

Figure 6. Scree plot for determining cluster number (Photo credit: Original)

According to Fig. 6, after multiple iterations, the optimal number of clustering is $k = 4$, Therefore randomly select three sample $x_8, x_{29}, x_{92}, x_{132}$ at the beginning of the algorithm as the initial mean vectors as

$$\mu_1 = (110, 210, 55.0, \dots, 4, 32970)$$
$$\mu_2 = (237, 365, 78.6, \dots, 4, 56482)$$
$$\mu_3 = (315, 680, 79.0, \dots, 5, 95839)$$
$$\mu_4 = (504, 945, 102.9, \dots, 5, 127109)$$

Sample $x_1 = (168, 350, 77.4, 1020, 4515, 1890, 1580, 2495, 432, 5, 47005)$, which is 18.47, 28.94, 38.34 and 56.23. away from the current mean vectors μ_1, μ_2, μ_3 respectively, so x_1 will be grouped into cluster C_1. Similarly, after reviewing all samples in the data set, the current cluster can be divided into C_1, C_2, C_3, C_4. Thus, a new mean vector $\mu'_1, \mu'_2, \mu'_3, \mu'_4$ can be obtained from C_1, C_2, C_3, C_4 respectively. After updating the current mean clustering, the above process is repeated continuously. The result of the fifth iteration is the same as that of the fourth iteration. The sizes of each cluster are 16, 25, 54, and 51 respectively and the centers of each cluster are

	Total.Power	Total.Torque	Battery.Capacity	Fastcharge.Speed	Length	Width	Height
1	485.3571	915.0000	99.18571	897.1429	5013.000	1957.571	1463.500
2	333.5000	627.6875	89.57500	806.8750	4913.688	1943.688	1477.500
3	205.4324	445.2432	74.67027	565.4054	4679.054	1878.486	1646.757
4	119.3143	284.7143	56.33143	378.0000	4376.429	1825.686	1606.686

	Gross.Vehicle.Weight	Cargo.Volume	Seats	Price
1	2980.000	436.5000	4.357143	123094.64
2	2888.250	443.5000	4.437500	81715.06
3	2669.838	528.7297	5.486486	53477.30
4	2235.514	443.9429	5.257143	34747.29

Figure 7. The result of the fifth iteration (Photo credit: Original)

A two-dimensional k-mean graph with length as x axis, price as y axis and the number of seats as the shape of the points are generated to test the clustering effect in Fig. 7. It can be seen from the figure that the clustering effect is relatively excellent, and the clustering boundary is obvious.

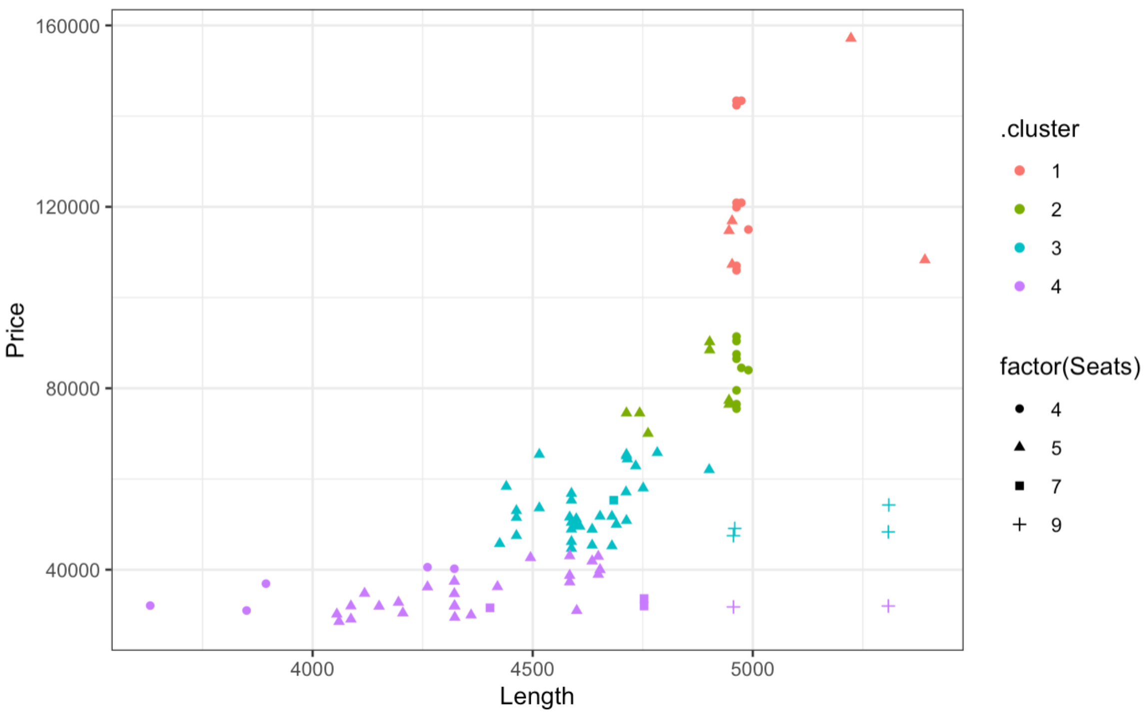

Figure 8. Two-dimensional K-means clustering (Photo credit: Original)

According to Fig. 7 and Fig. 8, electric vehicles in the data set can be divided into four categories according to K-means clustering. Here, the four clusters can be named high-end vehicles, sub-high-end vehicles, mid-end vehicles, and low-end vehicles respectively according to several variables with distinct characteristics, such as price, total power, and battery capacity.

Using two different models to analyze the data set, both models have their own defects. Although factor analysis can provide a specific functional model for the data set and has certain general applicability, as a matter of fact when designing an electric vehicle, manufacturers of electric vehicles must make certain choices on some parameters according to the positioning of different electric vehicles due to cost and other reasons. So, it may be too general to use a factor analysis model to generalize all the information contained in the data set. K-means clustering can well divide all-electric vehicles into four categories, however, it is not possible to get a specific functional model for a particular cluster.

Therefore, in order to solve the above problems and obtain a model with universal applicability and specificity, we can combine the two models. Firstly, K-means clustering is used to obtain four clusters, and then maximum likelihood estimation factor analysis is carried out on each cluster respectively. In this way, corresponding function models can be obtained according to the positioning of different vehicle types.

```
Loadings:
                     Factor1 Factor2 Factor3
Battery.Capacity      0.84    0.45
Length                0.60
Cargo.Volume          0.92    0.35
Fastcharge.Speed             -0.95
Height                0.46    0.79
Gross.Vehicle.Weight  0.54    0.81
Seats                 0.61    0.67
Total.Power          -0.45            0.85
Total.Torque                          0.83
Price                                 0.83
Width                -0.42
```

Figure 9. Loadings for the high-end EV class (Photo credit: Original)

Fig. 9 indicates the load distribution of factor analysis for the high-end EV class, the characteristics of different factors are more clearly distinguished and have better interpretability. For example, factor 3 reflects the common characteristics of total power, total torque, and price. In real life, these three parameters are also the indicators that many potential consumers who intend to buy high-end electric vehicles concerning about.

```
               Factor1 Factor2 Factor3
SS loadings       3.34    3.13    2.25
Proportion Var    0.30    0.28    0.20
Cumulative Var    0.30    0.59    0.79

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 68.66 on 25 degrees of freedom.
The p-value is 6.08e-06
```

Figure 10. Contributions of factors for the high-end EV class (Photo credit: Original)

According to Fig. 10, the cumulative variance contribution rate of these third factors reaches 79%. In addition, the p-value is 6.08×10^{-6}, both indicating that these three factors are sufficient to explain these variables (Factor analysis data for other clusters are detailed in the appendix).

4. Conclusion

The combination of K-means clustering and factor analysis has a much better effect than using the two analysis methods separately. The final model can better indicate how factor analysis can interpret the most information with fewer variables under different EV positioning. However, the final model still has some shortcomings. Since there is no market positioning data of electric vehicle manufacturers in the data set, this paper did not set the training set and the test set respectively, and thus cannot judge the accuracy of the model. Moreover, since K-means algorithm assumes that data follows Gaussian distribution, the performance of data with non-Gaussian distribution may be poor. After being tested by qqplot, some variables, such as the number of seats and width, do not show obvious normality.

References

[1] Johnson, Richard A. (Richard Arnold), and Dean W. Wichern. Applied Multivariate Statistical Analysis. 6th ed., Pearson Prentice Hall, 2007.

[2] Kabacoff, Robert. R in Action: Data Analysis and Graphics with R. Manning, 2011.

[3] Ziegel, Eric R., and John Rice. “Mathematical Statistics and Data Analysis.” Technometrics, vol. 37, no. 1, 1995, p. 127 –, https://doi.org/10.2307/1269179.

[4] Qiao, Qinyu. Electric Vehicle Recycling in China: Economic and Environmental Benefits. Jan. 2019, https://www.sciencedirect.com/science/article/abs/pii/S0921344918303288#preview-section-abstract.

[5] Muratori, Matteo. “The Rise of Electric Vehicles—2020 Status and Future ... - Iopscience.” The Rise of Electric Vehicles—2020 Status and Future Expectations, IOPSCIENCE, 25 Mar. 2021, iopscience.iop.org/article/10.1088/2516-1083/abe0ad.

[6] Graham, John D., and Eva Brungard. “Affordable Electric Vehicles: Their Role in Meeting the U.S. Contribution to the Paris Climate Goals.” Frontiers, Frontiers, 5 Aug. 2022, www.frontiersin.org/articles/10.3389/fenvs.2022.962942/full.

[7] Milev, George, et al. “The Environmental and Financial Implications of Expanding the Use of Electric Cars - A Case Study of Scotland.” Energy and Built Environment, Elsevier, 12 Aug. 2020, www.sciencedirect.com/science/article/pii/S2666123320300799.

[8] Stephenson, William. "Technique of factor analysis." Nature 136.3434 (1935): 297 - 297.

[9] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." Journal of the royal statistical society. series c (applied statistics) 28.1 (1979): 100 - 108.

[10] Krishna, K., and M. Narasimha Murty. "Genetic K-means algorithm." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29.3 (1999): 433 - 439.

Appendix

1. Factor analysis result of sub-high-end EVs

```
Call:
factanal(x = ev12, factors = 3, scores = "regression", rotation = "varimax")

Uniquenesses:
        Total.Power        Total.Torque     Battery.Capacity     Fastcharge.Speed
               0.28                0.00                 0.85                 0.04
             Length               Width               Height Gross.Vehicle.Weight
               0.09                0.23                 0.07                 0.09
       Cargo.Volume               Seats                Price
               0.41                0.04                 0.22

Loadings:
                     Factor1 Factor2 Factor3
Total.Power           0.62   -0.34    0.47
Fastcharge.Speed      0.77   -0.49   -0.35
Width                 0.85
Price                 0.85
Height               -0.66    0.70
Gross.Vehicle.Weight          0.94
Cargo.Volume                  0.74
Seats                -0.64    0.74
Total.Torque                          0.97
Length                0.61    0.41   -0.61
Battery.Capacity              0.37

               Factor1 Factor2 Factor3
SS loadings       3.70    3.18    1.79
Proportion Var    0.34    0.29    0.16
Cumulative Var    0.34    0.63    0.79

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 82.88 on 25 degrees of freedom.
The p-value is 4.01e-08
```

2. Factor analysis result of mid-end EVs

```
Call:
factanal(x = ev13, factors = 3, scores = "regression", rotation = "varimax")

Uniquenesses:
        Total.Power        Total.Torque     Battery.Capacity     Fastcharge.Speed
               0.00                0.26                 0.07                 0.03
             Length               Width               Height Gross.Vehicle.Weight
               0.58                0.80                 0.13                 0.05
       Cargo.Volume               Seats                Price
               0.00                0.11                 0.26

Loadings:
                     Factor1 Factor2 Factor3
Battery.Capacity      0.84    0.45
Length                0.60
Cargo.Volume          0.92    0.35
Fastcharge.Speed             -0.95
Height                0.46    0.79
Gross.Vehicle.Weight  0.54    0.81
Seats                 0.61    0.67
Total.Power          -0.45            0.85
Total.Torque                          0.83
Price                                 0.83
Width                -0.42

               Factor1 Factor2 Factor3
SS loadings       3.34    3.13    2.25
Proportion Var    0.30    0.28    0.20
Cumulative Var    0.30    0.59    0.79

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 68.66 on 25 degrees of freedom.
The p-value is 6.08e-06
```

3. Factor analysis result of low-end EVs

```
Call:
factanal(x = ev14, factors = 3, scores = "regression", rotation = "varimax")

Uniquenesses:
        Total.Power        Total.Torque     Battery.Capacity     Fastcharge.Speed
               0.00                0.26                 0.07                 0.03
             Length               Width               Height Gross.Vehicle.Weight
               0.58                0.80                 0.13                 0.05
       Cargo.Volume               Seats                Price
               0.00                0.11                 0.26

Loadings:
                     Factor1 Factor2 Factor3
Battery.Capacity      0.84    0.45
Length                0.60
Cargo.Volume          0.92    0.35
Fastcharge.Speed             -0.95
Height                0.46    0.79
Gross.Vehicle.Weight  0.54    0.81
Seats                 0.61    0.67
Total.Power          -0.45            0.85
Total.Torque                          0.83
Price                                 0.83
Width                -0.42

               Factor1 Factor2 Factor3
SS loadings       3.34    3.13    2.25
Proportion Var    0.30    0.28    0.20
Cumulative Var    0.30    0.59    0.79

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 68.66 on 25 degrees of freedom.
The p-value is 6.08e-06
```