Set Pair Analysis based on Similar Forecast model of Civil car ownership

Sun Wei¹, Yuan Renteng², Su Shuaijie³ and Wang Zhongyu⁴,*

¹ Highway Institute, Chang’an University, Xi’an, Shanxi, 710064, China
² Highway Institute, Chang’an University, Xi’an, Shanxi, 710064, China
³ Highway Institute, Chang’an University, Xi’an, Shanxi, 710064, China
*Corresponding author’s e-mail: swdeyouxiang@163.com

Abstract: It shows the important meaning that the accurate prediction of civil car ownership. The most important factor of civil car ownership are gotten from set pair analysis and correlation coefficient. Set pair analysis method can be used to quantificationally compute the similarity between main physics vectors from the view of identical, different and opposite sides. According to the weighted average of relative membership degree, obtain the set pair analysis based on similarity forecast model of civil car ownership, named SPA-SF. The forecast result of civil car ownership of 2015-2017 show that SPA-SF has certain application value in the prediction of civil car ownership.

1. Introduction

With the rapid development of the national economy, the people's living standard is gradually improved. Among them, the cars are also fast entering every household typically. The increase of the quantity of the civil car provides more strict requirements for the transportation infrastructure, the management, the environment and so on. It is of great significance for the study of the level of social and economic development, the traffic planning and the comprehensive management of the environment.

The prediction of civil car ownership has been widely and deeply studied at home and abroad. The methods of grey model, neural network, multiple regression analysis and time clustering prediction are put forward. However, the set pair analysis based on similar forecast model has not been explored in civil car ownership[1]. In this paper, SPA-SF (Set Pair Analysis-based Similar Forecast model) is introduced into the prediction of civil car ownership.

2. Set pair analysis based on similar forecast model.

Similarity forecast method is a nonparametric prediction method, which is based on the principle that similar causes produce similar results, and finds one or more of the most similar samples from historical samples as prediction results[2]. According to the data of civil car ownership in the past, the samples with the highest degree of similarity are found to predict the future civil car ownership. The theory of set pair analysis is a systematic theory put forward by Zhao Keqin, a famous scholar in China, in 1989 to deal with the uncertainty caused by ambiguity, random intermediary and incomplete information by using $a + bi + cj$ of connection number. It is characterized by dialectical analysis and...
mathematical treatment of system determination and uncertainty as a system of similarities and differences, which is both definite and uncertain.

It takes some steps to establish the SPA-SF model:
- The main driving factors should be scientific and reasonable;
- Determine the connection degree between the driving factors of the historical samples and the driving factors of the predicted samples;
- The connection degree between driving factors and predictors;

2.1. Selection of drive factors

Based on a lot of previous research, a preliminary determination of a number of relevant factors: Regional GDP, per capita GDP, PCDI, UPDI, IIFA, Passenger traffic, Freight Traffic, population, Total Retail Sales of Consumer Goods, Transport vehicles in operation, Length of highways[3]. Because there is a certain trend in the increase of all the factors, the data are processed into the growth rate of each year compared with the previous year for the convenience of unified dimensional calculation. Let civil car ownership is \( Y^* = \{y_i^* | i = 0, 1, 2, \cdots, n \} \), The influencing factors are \( X^* = \{x_{ij}^* | j = 0, 1, 2, \cdots, m \} \) .

The annual growth rate of civil car ownership is \( Y = \{y_i | i = 0, 1, 2, \cdots, n \} \). Annual growth rate of each influence factor \( X = \{x_{ij} | i = 0, 1, \cdots, n, j = 1, 2, \cdots, m \} \). “m” and “n” are the number of influencing factors and the predictor of car ownership respectively[4].

2.2. the main influencing factors that affect the growth of civil car ownership determined by the theory of correlation coefficient hypothesis test and the theory of set-pair-analysis.

2.2.1. the theory of correlation coefficient hypothesis test

Correlation coefficients are statistical indicators of the degree of linear correlation between variables. The correlation coefficient formula of each influencing factor and the civil car ownership is as follows:

\[
r_j = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\left( \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)^{1/2}} \tag{2.1}
\]

According to the theory of sampling distribution, when the formula is satisfied, it can be determined that the amount of ownership is significantly related to this influencing factor:

\[
|r_j| > r_{min} = t_{\alpha/2} / \left( t_{\alpha/2}^2 + n - 2 \right)^{1/2} \tag{2.2}
\]

Let \( \alpha \) is significance level, \( t_{\alpha/2} \) is a degree of freedom \( n - 2 \), \( r_{min} \) is a Two-sided Constraint critical value.

2.2.2. Set pair analysis test

Pairing set \( H\{Y, X_j\} \) is set by the growth rate of civil car ownership \( Y \) and the growth rate of various influencing factors \( X_j \). The degree of identity, difference and opposite connection of the set pair is described by the connection degree determined by the exhaustive method. Let the number of all Characteristics of \( H\{Y, X_j\} \) as \( N \); the identity characteristics is recorded as \( S_j \); the difference characteristics are recorded as \( F_j \); the opposite characteristic is recorded as \( P_j \); Substitute formula[3]:

\[
\mu_{Y,X_j} = \frac{S_j}{Y} + \frac{F_j}{Y} 1 + \frac{P_j}{Y} J \tag{2.3}
\]
\[
S_j, F_j, P_j \quad \overline{Y}, \overline{Y}, \overline{Y}
\]
denote, respectively, identical degree, difference degree, opposites degree. I is Difference Degree coefficient, its value interval is \([-1,1]\). J is opposites degree coefficient, it is taken -1.

\[
\frac{S_j}{\overline{Y}} + \frac{F_j}{\overline{Y}} + \frac{P_j}{\overline{Y}} = 1
\]

In order to simplify the statistics of the civil ownership and the characteristics of each influencing factor, \(x_{i,j}\) and \(y_i\) are Symbolized\[6\].

\[
Y = Y_P + Y_S + Y_{ijj}
\]

In order to simplify the statistics of the civil ownership and the characteristics of each influencing factor, \(x_{i,j}\) and \(y_i\) are Symbolized\[6\].

\[
y_i = 0.44s_j y_i^0 + 0.44s_j y_i^0 + 0.44s_j y_i^0
\]

Standard deviation \(s_0\), \(s_j\) are the standard deviation of \(y_i\), \(x_{i,j}\), respectively. Contrast with \(Z_{i,j}\) if \([z_{i,j} - z_{i,0}] = 0\) then \(Z_{i,j}\), \(Z_{i,0}\) at the same class which include in \(S_j\); if \([z_{i,j} - z_{i,0}] = 1\) then \(Z_{i,j}\), \(Z_{i,0}\) at the adjacent class which include in \(F_j\), if \([z_{i,j} - z_{i,0}] = 2\), then \(Z_{i,j}\), \(Z_{i,0}\) at the opposite class which include in \(P\). So as the identical degree, difference degree, opposites degree \(a + bi + cj\) and \(\mu_{y-x}\). The influencing factors of satisfying the two conditions are the main influencing factors to predict the civil car ownership.

2.3. Determination of the connection degree between the main influencing factors of historical samples and the main influencing factors of forecast year

Set pair with forecast year and the previous year of growth rate of main influencing factors of civil car ownership, \(B_{n+1} = [z_{n+1,0} / = 1, \ldots, m] \quad B_{i} = [z_{i,0} / = 1, \ldots, m] \quad \mu_{B_{n+1} - B_{i}}\). Obtain the set pair \(H(B_{n+1}, B_{i})\), \(m\) is the number of main influencing factors. The degree of connection is also based on exhaustive method.

\[
\mu_{B_{n+1} - B_{i}} = a_i + b_i I + c_i J, \quad \mu_{B_{n+1} - B_{i}} \subseteq [-1,1]
\]

\(a_i, b_i, c_i\) are the identical degree, difference degree, opposites degree of two sets respectively. The closer to 0 after the calculation, the lower the correlation between the sample and the main influencing factors of the prediction year, the closer to 1, the higher the correlation.

2.4. Forecast of the future civil car ownership

Based on the connection degree between the predicted year sample and the historical samples, the number of samples is empirical, which can be selected by determining the lower limit of connection degree. \(k\) of the most similar samples were determined from \(n\) historical samples. Based on each relative membership degree from these \(k\) growth rate of the car ownership of the historical samples, after weighted average, obtain the forecast growth rate of the civil car ownership in the forecast year \(y_{n+1}\)[7].

\[
y_{n+1} = \sum_{i=1}^{k} \left[ \frac{v_{n+1,i} / \sum_{j=1}^{k} v_{n+1,j}}{y_i} \right] = \sum_{i=1}^{k} w_i y_i
\]

\(K\) is the number of samples with the highest similarity between the selected and predicted year sample, \(v_{n+1,i}\) is the relative membership degree of similitude of variable fuzzy sets \(B_{n+1}\) and \(B_{i}[8]\).

\[
v_{n+1,i} = 0.5 + 0.5u_{B_{n+1} - B_{i}}
\]
3. Application examples
Through the annual statistical yearbook of Shandong Province, the civil car ownership is obtained by the year 2000-2017 (see table 1).

| Regional GDP | per capita GDP | PCDI | UPDI | IIFA | Passenger traffic | Freight Traffic | Total Retail Sales of Consumer Goods | Transport vehicles in operation | Length of highways | Civil car ownership |
|--------------|---------------|------|------|------|-------------------|----------------|-------------------------------------|-------------------------------|-----------------|-------------------|
| 2001         | 1.1320        | 1.0952 | 1.0546 | 1.0941 | 1.1042           | 1.0703         | 1.0625                              | 1.0049                        | 1.1135          | 1.0927            |
| 2002         | 1.1180        | 1.1128 | 1.0569 | 1.0723 | 1.2498           | 1.0633         | 1.0998                              | 1.0045                        | 1.1220          | 1.0406            |
| 2003         | 1.1785        | 1.1731 | 1.0629 | 1.1032 | 1.5184           | 1.0158         | 1.0690                              | 1.0048                        | 1.1390          | 1.0589            |
| 2004         | 1.2456        | 1.2389 | 1.1133 | 1.1236 | 1.4318           | 1.1863         | 1.1146                              | 1.0060                        | 1.1390          | 1.0980            |
| 2005         | 1.1941        | 1.1861 | 1.1206 | 1.1385 | 1.3818           | 1.1055         | 1.1269                              | 1.0073                        | 1.1657          | 1.0599            |
| 2006         | 1.1936        | 1.1853 | 1.1114 | 1.1347 | 1.0564           | 1.1086         | 1.1353                              | 1.0067                        | 1.1703          | 1.0101            |
| 2007         | 1.1769        | 1.1694 | 1.1413 | 1.1700 | 1.1258           | 1.1356         | 1.1990                              | 1.0061                        | 1.1926          | 1.0423            |
| 2008         | 1.2013        | 1.1944 | 1.1316 | 1.1431 | 1.2312           | 1.7553         | 1.3150                              | 1.0053                        | 1.2383          | 1.0998            |
| 2009         | 1.0963        | 1.0904 | 1.0846 | 1.0923 | 1.2329           | 1.0933         | 1.1669                              | 1.0056                        | 1.1599          | 1.1673            |
| 2010         | 1.1564        | 1.1460 | 1.1424 | 1.1199 | 1.2231           | 1.0662         | 1.0508                              | 1.0115                        | 1.1826          | 1.1599            |
| 2011         | 1.1593        | 1.1492 | 1.1934 | 1.1427 | 1.1501           | 1.0059         | 1.0568                              | 1.0061                        | 1.1734          | 1.1399            |
| 2012         | 1.1036        | 1.0980 | 1.1324 | 1.1300 | 1.1676           | 1.0549         | 1.0622                              | 1.0050                        | 1.1455          | 1.0990            |
| 2013         | 1.1044        | 1.0989 | 1.1313 | 1.0438 | 1.1770           | 1.0142         | 1.0508                              | 1.0050                        | 1.1345          | 1.0825            |
| 2014         | 1.0761        | 1.0703 | 1.1119 | 1.0870 | 1.1551           | 0.2402         | 0.7377                              | 1.0058                        | 1.1263          | 0.9297            |
| 2015         | 1.0614        | 1.0552 | 1.0882 | 1.0795 | 1.1369           | 0.7568         | 0.9909                              | 1.0059                        | 1.1055          | 0.9164            |
| 2016         | 1.0637        | 1.0552 | 1.0792 | 1.0782 | 1.1037           | 1.0397         | 1.0957                              | 1.0102                        | 1.1039          | 1.0583            |
| 2017         | 1.0693        | 1.068  | 1.0834 | 1.0817 | 1.0353           | 1.0509         | 1.1534                              | 1.0059                        | 1.0980          | 1.0636            |

3.1. Determining the main influencing factors of civil car ownership
The main influencing factors were screened by correlation coefficient, and the lowest correlation coefficient $r_{min} = 0.4973$ was allowed when $\alpha = 0.05$ selected by hypothesis test$^9$. According to the formula (2.1), (2.2) and (2.3), the correlation coefficient and connection degree between each influencing factor and the civil car ownership can be obtained. Combined with the decision conditions, it can be seen that the main influencing factors of civil car ownership are regional GDP, per capita GDP and IIFA.

3.2. Establishment of similar forecast model
Set pair $H(B_{n+1}, B_i)$ could be obtained by setting pair with the main influencing factors of forecast year civil car ownership $B_{n+1}$ and that of historic year $B_i, i = 1,2,\cdots,16, i = 0.5, j = -1$. The connection degree is calculated by formula (2.4) (see table 2).

| year        | $a_i$ | $b_i$ | $c_i$ | $U_{B_{n+1}-B_i}$ |
|-------------|-------|-------|-------|-------------------|
| 2001        | 0.6666| 0.3333| 0.3333| 0.8332            |
| 2002        | 0     | 1     | 0     | 0.5               |
| 2003        | 0     | 0     | 1     | -1                |
| 2004        | 0     | 0     | 1     | -1                |
2005 0 0 1 -1
2006 0 0.3333 0.6667 -0.5000
2007 0.3333 0.3333 0.3333 0.1666
2008 0 0.3333 0.6667 -0.5000
2009 0.6667 0.3333 0 0.8333
2010 0 1 0 0.5000
2011 0.3333 0.6667 0 0.6666
2012 0.6667 0.3333 0 0.8333
2013 0.6667 0.3333 0 0.8333
2014 1 0 0 1
2015 1 0 0 1
2016 1 0 0 1

Four sets of samples most similar to the forecast year which selected from the historical year. In 2001, 2009, 2013, 2014, the connection degree was 0.8333, and the 2013 data were selected to participate in the forecast. The weight of most similar historical years are corresponding normalized value of relative membership degree. According to formula (2.8), the relative membership of the most similar year can be obtained. The predicted growth rate of civil car ownership in 2017 is 1.1172, the actual value is 1.1156, and the relative error is 0.15%.

In order to further verify the accuracy of the similar prediction model, the growth rate of civil car ownership in 2015 / 2016 is predicted by using the above method, and the results are shown in Table 3.

| year | real growth rate | Predicting growth rate | fractional error |
|------|----------------|------------------------|-----------------|
| 2015 | 1.1046         | 1.1456                 | 3.71%           |
| 2016 | 1.1268         | 1.1246                 | 0.19%           |
| 2017 | 1.1156         | 1.1172                 | 0.14%           |

4. Conclusion
Based on the set pair analysis theory, this paper establishes a similar forecast model of set pair analysis, determines the main influencing factors of civil car ownership through correlation coefficient hypothesis test and set pair analysis test, determines the relationship between forecast year and historical years, determines the most similar historical samples. By determining the growth rate of civil car ownership in the forecast year according to the relative membership degree weighted, the similar forecast model of set pair analysis is established.

The prediction of civil car ownership in Shandong Province in 2017 by SPA-SF shows that it is simple, convenient and accurate to use SPA-SF to predict car ownership, and the theory in this paper is complete, which is of great value in the prediction of civil car ownership.

References:
[1] Zhao, K.Q.(2000)Set Pair Analysis and its Preliminary Application.zhejiang Science and Technology Press,HangZhou.
[2] Wang, M.W., Jin, J.L., Zhou, Y.L.(2013)Set pair analysis based coupling methods and applications.Science Press,Beijing.
[3] Zhang, X.W., Chang, J.Y.(2012)Research on Urban Car Ownership Prediction Based on PCA-BP Neural Network.Computer Simulation,29(12):376-379.
[4] Zhou, R.X., Pan, Z.W., Jin, J.L.(2016)Application of the Method of Set Pair Analysis Based on Similarity Forecast Model in Water Consumption Prediction.Journal of North China University of Water Resources and Electric Power,37(6):67-71.
[5] Xu, Y.W., Li, Z.B., Wang, J.Y. (2015) Study on the similar forecast models of precipitation acidity and water quality based on set pair analysis. Environmental Pollution & Control, 37(2), 59-62.

[6] Jin, J.L., Wei, Y.M., Wang, W.S. (2009) Set pair analysis based on similarity forecast model of water resources. Journal of Hydroelectric Engineering, 28(1): 72-77.

[7] Shi, H.B. (2017) Study on groundwater level similarity prediction model under set pair analysis. Water Conservancy Science and Technology and Economy, 23(2): 37-40.

[8] Wu, W.Q., Xia, J. (2018) Forecasting China’s Car Ownership by Grey Neural Network with Simpson Formula. http://kns.cnki.net/kcms/detail/50.1190.U.20181108.1158.010.html. 2018-11-08.

[9] Zhu, C., Zhou, H.P., Zhong, B.Q. (2011) Forecasting on vehicle population based on combined model with mixed neural network. Journal of Changsha University of Science & Technology, 8(2), 13-16.