Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 1, 2023

Combination Forecasting Model of R&D Intensity in Anhui Province Based on IGOWA Operator

Weilun Xu, Yumei Wang

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233000, China

Abstract: Taking the R&D intensity data of Anhui Province from 2006 to 2020 as the sample, the grey prediction model, ARIMA model and Holt Winters non seasonal model were selected to fit and predict the R&D intensity of Anhui Province, and then the IGOWA operator variable weight coefficient combination prediction model based on the minimum criterion of the sum of squares of errors was constructed, and the combination prediction model corresponding to four special values of the operator parameters was taken, Establish an error evaluation index system to illustrate the effectiveness of the model, and analyze the sensitivity of the parameters. Through model evaluation, it can be found that the prediction effect of the combined prediction model in the sample period is significantly better than the three single prediction models, greatly improving the prediction accuracy. Finally, the combined forecasting model is used to predict the R&D intensity of Anhui Province from 2021-2025. The prediction results show that in 2021-2025, the R&D intensity of Anhui Province will continue to rise, and the growth rate is increasing year by year. The scientific research and innovation ability will continue to develop steadily.

Keywords: R&D intensity; IGOWA operator; Combination forecasting; Sum of squares of errors; Technological innovation.

1. Introduction

In today's world tide, the position of science and technology is increasingly prominent, and scientific and technological innovation is gradually becoming a key factor affecting the comprehensive national strength, which highlights the primary position of scientific and technological innovation, and also shows that in order to achieve rapid and stable development in the current society, we must adhere to the principle of independent research and development, and drive innovation. Since the 21st century, China's scientific research enthusiasm has continued to rise, and many scientific research achievements have emerged in combination with the corresponding scientific and technological development strategy and system construction. General Secretary pointed out in the report of the 20th National Congress of the Communist Party of China that innovation is the core driving force and plays a very important role in high-quality development. China has successfully entered the ranks of innovative countries. The "14th Five Year Plan" also clearly indicates five development directions, indicating that China is moving from the "quantitative expansion era" to the "quality expansion era".In order to enter the "quality expansion era" faster and better, it is necessary to promote efficiency through scientific and technological innovation and development. In 2020, the R&D expenditure of Anhui Province will reach 88.32 billion yuan, and the R&D intensity will reach 2.28%, 0.25 percentage points higher than that of the previous year, ranking 10th in the country and 2nd in the central region, successfully realizing the leap forward development from "a big province of science and education" to "a source of scientific and technological innovation".

The research of foreign scholars on R&D intensity is mainly based on the relationship between scientific and technological investment and the contribution rate of economic growth. Paul M. Romer et al. [1], Gene M. Grossman et al. [2], Paul S. Segerstrom et al. [3], all the studies had found that increasing investment in science and technology and R&D can boost the contribution rate of economic growth, and can better encourage enterprises to invest resources in R&D activities. At present, domestic scholars' research on R&D intensity mainly starts from the combination of R&D funds and GDP. Xuanwen Fang et al.[4] used the vector autoregression model and the grey prediction model GM(1,1), by selecting the data of China's R&D expenditure and GDP output value from 1990 to 2010, this paper forecasted the R&D expenditure at constant prices and current prices during the "12th Five Year Plan" period, and determined an equilibrium relationship with an elasticity of 0.6 in economic growth; Liyu Zhao et al.[5] selected the cointegration test of China's GDP and financial investment in science and technology from 1989 to 2007 showed that there was a long-term equilibrium relationship between economic growth and the total internal expenditure of R&D funds; Shi Chen et al.[6] believed that China's current R&D investment intensity was at the middle level of the world, but it maintained a good development trend. The government should increase relevant investment to further enhance China's R&D intensity; Jianwen He [7] compared China's R&D intensity data from 2006 to 2013 and the data of five developed countries verifies that China's R&D intensity is growing fast at present, and will not enter the slow growth stage immediately like most developed countries.

Scientific prediction requires integrating multiple prediction methods according to the history and reality of social and economic phenomena to minimize the loss of information and improve the prediction accuracy. Therefore, the concept of combined forecasting was first introduced by Bates and Granger [8]. It was proposed in 1969. Both theoretical and practical studies show that [9] in the case of different single prediction models and different data sources, the results derived from the combined prediction model may be better than any independent prediction value. The combined prediction model can reduce the systematic error of prediction, significantly improve the prediction effect, and reveal the development and change law of objective things.

To sum up, many scholars have analyzed and predicted the R&D intensity, but most of them still stay in the stage of using single prediction model to predict the data. Although these single prediction models can predict the time series data, they contain limited factors that can be considered and have a certain randomness, which may not make full use of all the effective information reflected in the sample data to provide prediction. Therefore, this paper selects the research and development intensity data of Anhui Province from 2006 to 2020, first analyzes and forecasts the data through the gray prediction model, ARIMA time series model, Holt-Winters non-seasonal prediction model, and then forms a combination prediction method with the help of the IGOWA (generalized induction weighted average) operator and the optimal criterion of minimizing the sum of squares of prediction errors [10]. The variable weight coefficient combination forecasting model based on IGOWA operator is constructed to make short-term prediction on the R&D intensity of Anhui Province in the next five years, in order to improve the prediction accuracy and increase the stability of the model. And according to the data predicted by the more stable model, it is compared with the actual data of research and development intensity of Anhui Province in 2021 to verify the effectiveness and statistical significance of the model.

2. Model Construction

2.1. Operator

2.1.1. OWA Operator

Set up f_ω by n-ary function: $R^n \to R$. There is a nonnegative weight vector $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ and meets $\sum_{i=1}^{n} \omega_i = 1, \omega_i \in [0,1]$.

$$f_\omega(a_1, a_2, \cdots a_n) = \sum_{i=1}^{n} \omega_i b_i \quad (1)$$

then f_ω is called n-dimension ordered weighted average operator, also known as OWA operator. b_i is the i-th largest number among $a_1, a_2, \cdots, a_n$, which indicates that the weight coefficient ω_i of the ordered weighted average operator is independent of a_i, but only related to the first i position after ordering n numbers in descending order.

2.1.2. IOWA Operator

Record two-dimensional array $< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >$. There is a nonnegative weight vector $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ and meets $\sum_{i=1}^{n} \omega_i = 1, \omega_i \in [0,1]$. Then

$$IOWA_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = \sum_{i=1}^{n} \omega_i a_{v-index(i)} \quad (2)$$

is called n-Dimension induced ordered weighted average operator, also known as IOWA operator. v_i is denoted as the induced value of the number a_i, and $v - index(i)$ is the subscript of the number $v_1, v_2, \cdots, v_n$ which is the greatest i in descending order [11]. This shows that the weighted averaging operator arranges n numbers according to their corresponding induced values in descending order, and then the weighted averaging cofficient ω_i has nothing to do with numerical value of a_i, but only relates to the i-th position after sorting according to the induced values.

2.1.3. IGOWA Operator

Set up f_ω by n-ary function: $R^n \to R$. Record two-dimensional array $< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >$. There is a nonnegative weight vector $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ and meets $\sum_{i=1}^{n} \omega_i = 1, \omega_i \in [0,1]$.

$$f_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = \left(\sum_{i=1}^{n} \omega_i (a_{v-index(i)})^\lambda\right)^{\frac{1}{\lambda}}. \quad (3)$$

f_ω is called generalized induced ordered weighted average operator and also called IGOWA operator where parameters $\lambda \in (-\infty, 0) \cup (0, \infty)$. v_i is denoted as the induced value of the number a_i, and $v - index(i)$ is the subscript of the number $v_1, v_2, \cdots, v_n$ which is the greatest i in descending order [11]. We can take any value within the effective value range of the parameter λ, so as to build different information aggregation operators. The following are several common operators when λ takes special values[12]:

When $\lambda = 1$, $f_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = \sum_{i=1}^{n} \omega_i a_{v-index(i)}$. The IGOWA operator at this time is the IOWA (induced ordered weighted average) operator;

When $\lambda = -1$, $f_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = \frac{1}{\sum_{i=1}^{n}\left(\frac{\omega_i}{a_{v-index(i)}}\right)}$. The IGOWA operator at this time is the IOWHA (induced ordered weighted harmonic average) operator;

When $\lambda \to 0$, $f_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = \prod_{i=1}^{n} a_{v-index(i)}{}^{\omega_i}$. The IGOWA operator at this time is the IOWGA (Induced Ordered Weighted Geometric Average) operator;

When $\lambda = 0.5$, $f_\omega(< v_1, a_1 >, < v_2, a_2 >, \cdots, < v_n, a_n >) = (\sum_{i=1}^{n} \omega_i \sqrt{x_{v-index(i)}})^2$. The IGOWA operator at this time is the IOWSA (Induced Ordered Weighted Square Average) operator.

2.2. IGOWA Operator Combination Prediction Model Based on the Criterion of Minimum Sum of Squares of Errors

Assume that the actual observation value of a required prediction phenomenon is $\{x_t, t = 1,2, \cdots, N\}$, n kinds of single prediction methods to predict this phenomenon, and x_{it} is the model prediction value of the first i kind of single prediction method in the t period. There are nonnegative weighting coefficients $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ and meet $\sum_{i=1}^{n} \omega_i = 1, \omega_i \in [0,1]$, then $W = (\omega_1, \omega_2, \cdots, \omega_n)^T$ constitutes the weight vector of all single forecasting methods in the combined forecasting model. The prediction precision is constructed by the absolute value of the prediction relative error of each single prediction at each time point to characterize the induced variable of the prediction value at each time point. Note

$$v_{it} = \begin{cases} 1 - \left|\frac{x_t - x_{it}}{x_t}\right|, & \left|\frac{x_t - x_{it}}{x_t}\right| < 1 \\ 0, & \left|\frac{x_t - x_{it}}{x_t}\right| \geq 1 \end{cases} \quad (4)$$

as the prediction accuracy of the i-th single prediction in the t period, then using v_{it} and x_{it} can form a two-dimensional array $(< v_{1t}, x_{1t} >, < v_{2t}, x_{2t} >, \cdots, < v_{nt}, x_{nt} >)$. Take it into the IGOWA operator for calculation, and the optimized combined predictive value is $\hat{x}_t = \left(\sum_{i=1}^{n} \omega_i (x_{v-index(it)})^\lambda\right)^{\frac{1}{\lambda}}$. According to this, it can be seen that the λ power prediction error and induced prediction error in the t period are $e_t(\lambda) = x_t{}^\lambda - \hat{x}_t{}^\lambda$ and $e_{v-index(it)} = x_t{}^\lambda - x_{v-index(it)}{}^\lambda$. In this paper, the four special values of λ listed in 2.1.3. are used for analysis and discussion, so the specific form of induced prediction error is as follows

$$e_{v-index(it)} = \begin{cases} x_t - x_{v-index(it)}, & \lambda = 1 \\ \frac{1}{x_t} - \frac{1}{x_{v-index(it)}}, & \lambda = -1 \\ \ln x_t - \ln x_{v-index(it)}, & \lambda \to 0 \\ \sqrt{x_t} - \sqrt{x_{v-index(it)}}, & \lambda = 0.5 \end{cases}. \quad (5)$$

From the induced prediction error, the n-order induced ordered weighted average prediction error matrix can be further denoted as $E = (E_{ij})_{n\times n} = (E_{ji})_{n\times n} = \left(\sum_{t=1}^{N} e_{v-index(it)} e_{v-index(jt)}\right)_{n\times n}; i, j = 1,2,\cdots,n$. It can be seen that the induced ordered weighted average prediction error matrix is a symmetric n-order invertible matrix. Therefore the sum of squares of the total forecast errors in N period is $Q^2 = \sum_{t=1}^{N}\left(\sum_{i=1}^{n} \omega_i e_{v-index(it)}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_i\omega_j \left(\sum_{t=1}^{N} e_{v-index(it)} e_{v-index(jt)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega_i\omega_j E_{ij}$, where $i, j = 1,2,\cdots,n$.

In conclusion, the IGOWA operator combination prediction model based on the optimal criterion of minimizing the sum of squares of prediction errors can be constructed as

$$\min Q^2 = W^T E W$$
$$s.t. \begin{cases} \sum_{i=1}^{n} \omega_i = 1; \\ \omega_i \geq 0, \quad i = 1,2,\cdots,n. \end{cases} \quad (6)$$

3. Empirical Analysis

3.1. Data Source and Descriptive Statistics

In recent years, R&D intensity has gradually become an important indicator reflecting the scientific and technological innovation strength and core competitiveness of a country or region. This article uses Jun Liu and others for reference [13]. The proposed R&D intensity measurement standard uses the proportion of R&D expenditure in GDP to measure the R&D intensity of a region. The research and development intensity index data of Anhui Province from 2006 to 2020 were collected by searching China Statistical Yearbook, China Science and Technology Statistical Yearbook and other data, and the trend line chart was drawn as shown in Figure 1. It can be seen from Figure 1 that the R&D intensity of Anhui Province has an obvious growth trend.

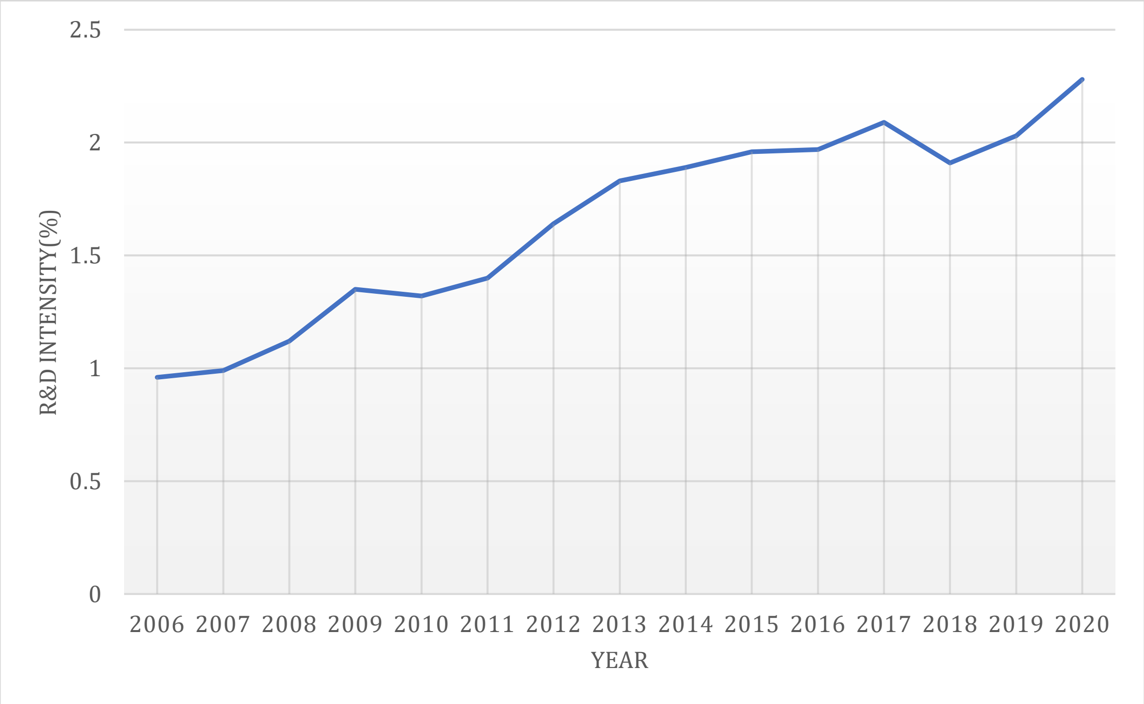

Figure 1. Trend chart of research and development intensity of Anhui Province from 2006 to 2020

3.2. Single Prediction Model

3.2.1. Grey Prediction Model GM (1,1)

In the early 1980s, Professor Julong Deng proposed an effective prediction method for small sample data or data with low integrity — grey system theory [14], which fully excavates the essence of data through differential equations, has a very significant effect. Before establishing the grey prediction model GM (1,1), it is necessary to carry out rank comparison test on the time series data. The purpose of this step is to verify whether the time series data has suitable regularity, so as to judge whether a satisfactory model can be obtained. According to the actual value of research and development intensity of Anhui Province from 2006 to 2020 as the original data of the model, the time series data $X = (x_1, x_2, \cdots, x_{15})$ is defined. According to the calculation formula of grade ratio

$$\phi(t) = \frac{x_{t-1}}{x_t}, t = 2,3,\cdots,15, \quad (7)$$

calculate all the grade ratios, if all the grade ratios are located in the interval $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ $(n = 15)$, the data is suitable for model construction. The calculation results of grade ratios are shown in Table 1. It can be seen from Table 1 that the grade ratios of the original data do not all fall within the allowable coverage of the interval, so the original data is shifted by three units to obtain a new sequence $Z = (z_1, z_2, \cdots, z_{15}) = (x_1 + 3, x_2 + 3, \cdots, x_{15} + 3)$. We can get that all the grade ratios of the translation-transformed series are located in the interval $(0.882, 1.133)$, indicating that the translation-transformated series are suitable for building the grey forecasting models.

The time series data $Z = (z_1, z_2, \cdots, z_{15})$ is accumulated once to generate a sequence $Y = (y_1, y_2, \cdots, y_{15})$, an accumulation matrix $B = \begin{bmatrix} -\frac{1}{2}(y_1+y_2) & 1 \\ -\frac{1}{2}(y_2+y_3) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(y_{14}+y_{15}) & 1 \end{bmatrix}$ and s constant vector $L = (z_2, z_3, z_4, \cdots, z_{15})^T$ are constructed. Use OLS method to calculate parameters

$$\hat{\alpha} = \begin{bmatrix} \alpha \\ \mu \end{bmatrix} = (B^T B)^{-1} B^T L. \quad (8)$$

By calculating the coefficient $\alpha = -0.019$, $\mu = 4.029$ with SPSSPRO software, the prediction formula of accumulating sequence Y is $\hat{y}_{t+1} = 216.013e^{0.019t} - 212.053$. According to which the model prediction value and prediction accuracy of grey prediction model GM (1,1) are shown in Table 4. It can be seen from Table 4 that the posterior difference ratio C of the model 0.082, and the average relative error is 6.347%, which means that the model has high precision and good fitting effect, and is suitable for forecasting by using this model.

Table 1. Grade ratio test results

Year	Original value	Grade ratio	Serial value after translation conversion	Grade ratio after translation conversion
2006	0.96	-	3.96	-
2007	0.99	0.970	3.99	0.992
2008	1.12	0.884	4.12	0.968
2009	1.35	0.830	4.35	0.947
2010	1.32	1.023	4.32	1.007
2011	1.40	0.943	4.40	0.982
2012	1.64	0.854	4.64	0.948
2013	1.83	0.896	4.83	0.961
2014	1.89	0.968	4.89	0.988
2015	1.96	0.964	4.96	0.986
2016	1.97	0.995	4.97	0.998
2017	2.09	0.943	5.09	0.976
2018	1.91	1.094	4.91	1.037
2019	2.03	0.941	5.03	0.976
2020	2.28	0.890	5.28	0.953

3.2.2. ARIMA Prediction Model

ARIMA model, also called autoregressive moving average model, is widely used to analyze various types of time series data and to model and predict. The ADF test results of the model are obtained through SPSS software processing, as shown in Table 2. When the difference is 0 order, the significance P value is 0.832, which does not show significance horizontally, so the original hypothesis that the series is unstable time series cannot be rejected. When the difference is 1 order, the significance P value is 0.012, showing significance at the 5% significance level, so the original hypothesis is rejected, and the series is a stable time series. When the difference is 2 order, the significance P value is 0.000, showing significance at the significance level of 1%, so the original hypothesis is rejected, and the series is a stable time series. Then judge whether the differential sequence is stable by looking at the differential sequence diagram. At the same time, the autocorrelation and partial autocorrelation analysis are carried out on the time series, and estimate the model's P and Q values according to the truncation and trailing of ACF and PACF diagrams. The ARIMA (1,1,2) prediction model is finally established by combining the model comparison of AIC information criteria.

ARIMA model requires that the model has pure randomness, that is, the residual error of the model is white noise. The residual error of the model can be tested by calculating the P value of the Q statistic of the model, and the P value of the statistic Q_6 obtained from the calculation result is equal to 0.845 and greater than 0.1, so the original hypothesis cannot be rejected, which means that the residual sequence of ARIMA (1,1,2) model does not have autocorrelation and meets the requirements of parameter diagnosis. Therefore, the ARIMA model established meets the requirements, and can be used for fitting and prediction. The specific parameter information is shown in Table 3.

According to the data in Table 3, the prediction formula of ARIMA (1,1,2) model built in this paper is

$$x_t = 0.065 + 0.299x_{t-1} - 0.596\varepsilon_{t-1} - 0.397\varepsilon_{t-2}. \quad (9)$$

The R&D intensity data of Anhui Province from 2006 to 2020 can be brought into formula (9) to obtain the prediction values and prediction accuracy of ARIMA (1,1,2) model in each period. The results are shown in Table 4.

Holt-Winters Non-Seasonal Exponential Smoothing Prediction Model

Holt-Winters non-seasonal exponential smoothing model is one of the exponential smoothing prediction models, which can be used to predict the type of R&D intensity data collected in this paper — data with no obvious seasonal change but with obvious temporal change trend. The prediction formula is $\hat{y}_{t+k} = a_t + b_t k, k > 0$, where $\hat{y}_t$ is the parameters a_t, b_t after three times from the original data sequence y_t, which is calculated by the following formula

$$\begin{cases} a_t = \alpha y_t + (1-\alpha)(a_{t-1} + b_{t-1}); \\ b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}. \end{cases} \quad (10)$$

In the formula, $\alpha, \beta \in (0,1)$ is called damping factor, which can be estimated by EViews software based on the principle of minimizing the sum of squares of errors and the minimum RMSE of root mean square error value [15][15]. In this way, the intercept a_t and slope b_t of the model can be calculated and brought into the prediction formula. Later, the predicetion results of R&D intensity data in Anhui Province are shown in Table 4.

Table 2. ADF Inspection Table

Variable	Difference order	T	P	AIC	Critical value		
					0.01	0.05	0.1
Research and development strength	0	0.754	0.832	-11.419	-4.012	-3.104	-2.691
	1	-3.369	0.012**	-7.084	-4.069	-3.127	-2.702
	2	-4.554	0.000***	-2.997	-4.138	-3.155	-2.714

Note: * * *, * * and * represent the significance levels of 1%, 5% and 10% respectively.

Table 3. ARIMA (1,1,2) model parameters

Term	Symbol	Coefficient	Standard error	Z-value	P-value	95% CI
Constant term	c	0.065	0.089	0.727	0.467	-0.110 ~ 0.240
AR parameters	α_1	0.299	1.116	0.268	0.788	-1.887 ~ 2.486
MA parameters	β_1	-0.596	23.374	-0.025	0.98	-46.407 ~ 45.216
	β_2	-0.397	8.16	-0.049	0.961	-16.391 ~ 15.596

AIC value: -14.225
BIC value: -11.030

Table 4. Predicted value and forecast precision of each single forecast model

Year	Actual value of R&D intensity (%)	Grey prediction model		ARIMA model		Holt-Winters model	
		Predicted value (%)	Prediction accuracy	Predicted value (%)	Prediction accuracy	Predicted value (%)	Prediction accuracy
2006	0.9600	0.9600	1.0000	0.9800	0.9792	0.9600	1.0000
2007	0.9900	1.1430	0.8455	1.0500	0.9394	0.9570	0.9667
2008	1.1200	1.2220	0.9089	1.0900	0.9732	0.9890	0.8830
2009	1.3500	1.3030	0.9652	1.2300	0.9111	1.1540	0.8548
2010	1.3200	1.3850	0.9508	1.4200	0.9242	1.4610	0.8932
2011	1.4000	1.4690	0.9507	1.3800	0.9857	1.4670	0.9521
2012	1.6400	1.5540	0.9476	1.5200	0.9268	1.5160	0.9244
2013	1.8300	1.6410	0.8967	1.7100	0.9344	1.7750	0.9699
2014	1.8900	1.7300	0.9153	1.8500	0.9788	2.0200	0.9312
2015	1.9600	1.8200	0.9286	1.9100	0.9745	2.0860	0.9357
2016	1.9700	1.9120	0.9706	2.0100	0.9797	2.1140	0.9269
2017	2.0900	2.0060	0.9598	2.0400	0.9761	2.0720	0.9914
2018	1.9100	2.1020	0.8995	2.1800	0.8586	2.1590	0.8696
2019	2.0300	2.1990	0.9167	2.0400	0.9951	1.9290	0.9502
2020	2.2800	2.2990	0.9917	2.2400	0.9825	1.9950	0.8750

3.3. Combined Forecasting Model

3.3.1. IGOWA Combined Forecasting Model for $\lambda = 1$

According to the model building method in 2.2., the third-order induced ordered weighted average prediction error matrix for calculating $\lambda = 1$ is:

$$E = \begin{bmatrix} 0.0626 & 0.0558 & 0.1057 \\ 0.0558 & 0.1841 & 0.0946 \\ 0.1057 & 0.0946 & 0.4108 \end{bmatrix}.$$

Then, the IGOWA operator combination prediction model based on the optimal criterion of minimizing the sum of squares of prediction errors can be constructed as:

$$minQ^2 = W^T EW = 0.0626\omega_1^2 + 0.1116\omega_1\omega_2 + 0.2114\omega_1\omega_3 + 0.1841\omega_2^2 + 0.1892\omega_2\omega_3 + 0.4108\omega_3^2 s.t. \begin{cases} \sum_{i=1}^{3} \omega_i = 1; \\ \omega_i \geq 0, \quad i = 1,2,3. \end{cases}$$

Using LINGO 18.0 software to solve, the weight coefficient obtained is:

$$\omega_1 = 0.9497, \ \omega_2 = 0.0503, \ \omega_3 = 0.$$

Then the corresponding IGOWA operator combination prediction model is

$$\hat{x}_t = 0.9497x_{v-index(1t)} + 0.0503x_{v-index(2t)}. \quad (11)$$

The predicted value and prediction precision of the combined prediction model calculated according to equation (11) are shown in Table 5. During the sample period, the sum of squares of prediction errors of the model is $Q^2 = 0.06226$.

3.3.2. IGOWA combined forecasting model for $\lambda = -1$

According to the model building method in 2.2., the third-order induced ordered weighted average prediction error matrix for calculating $\lambda = -1$ is:

$$E = \begin{bmatrix} 0.0081 & 0.0038 & 0.0110 \\ 0.0038 & 0.0287 & 0.0181 \\ 0.0110 & 0.0181 & 0.0766 \end{bmatrix}.$$

Then, the IGOWA operator combination prediction model based on the optimal criterion of minimizing the sum of squares of prediction errors can be constructed as:

$$minQ^2 = W^T EW = 0.0081\omega_1^2 + 0.0076\omega_1\omega_2 + 0.022\omega_1\omega_3 + 0.0287\omega_2^2 + 0.0362\omega_2\omega_3 + 0.0766\omega_3^2 s.t. \begin{cases} \sum_{i=1}^{3} \omega_i = 1 ; \\ \omega_i \geq 0, \quad i = 1,2,3. \end{cases}$$

Using LINGO 18.0 software to solve, the weight coefficient obtained is:

$$\omega_1 = 0.9852，\omega_2 = 0.0148，\omega_3 = 0.$$

Then the corresponding IGOWA operator combination prediction model is

$$\hat{x}_t = \frac{1}{\frac{0.9852}{x_{v-index(1t)}} + \frac{0.0148}{x_{v-index(2t)}}}. \quad (12)$$

The predicted value and prediction precision of the combined prediction model calculated according to equation (12) are shown in Table 5. During the sample period, the sum of squares of the reciprocal prediction errors of the model is $Q^2 = 0.00744$.

3.3.3. IGOWA Combined Forecasting Model for $\lambda \to 0$

According to the model building method in 2.2., the third-order induced ordered weighted average prediction error matrix for calculating $\lambda \to 0$ is:

$$E = \begin{bmatrix} 0.0204 & 0.0159 & 0.0333 \\ 0.0159 & 0.0653 & 0.0381 \\ 0.0333 & 0.0381 & 0.1535 \end{bmatrix}.$$

Then, the IGOWA operator combination prediction model based on the optimal criterion of minimizing the sum of squares of prediction errors can be constructed as:

$$minQ^2 = W^T EW = 0.0204\omega_1^2 + 0.0318\omega_1\omega_2 + 0.0666\omega_1\omega_3 + 0.0653\omega_2^2 + 0.0762\omega_2\omega_3 + 0.1535\omega_3^2 s.t. \begin{cases} \sum_{i=1}^{3} \omega_i = 1 ; \\ \omega_i \geq 0, \quad i = 1,2,3. \end{cases}$$

Using LINGO 18.0 software to solve, the weight coefficient obtained is:

$$\omega_1 = 0.9164，\omega_2 = 0.0836，\omega_3 = 0.$$

Then the corresponding IGOWA operator combination prediction model is

$$\hat{x}_t = x_{v-index(1t)}^{0.9164} x_{v-index(2t)}^{0.0836}. \quad (13)$$

The predicted value and prediction precision of the combined prediction model calculated according to equation (13) are shown in Table 5. During the sample period, the sum of squares of the prediction logarithmic error of the model is $Q^2 = 0.02002$.

3.3.4. IGOWA Combined Forecasting Model for $\lambda = 0.5$

Table 5. Predicted value and Prediction Precision of IGOWA operator combination prediction model

Year	Actual value of R&D intensity (%)	λ=1		λ=-1		λ→0		λ=0.5	
		Predicted value (%)	Prediction accuracy	Predicted value (%)	Prediction accuracy	Predicted value (%)	Prediction accuracy	Predicted value (%)	Prediction accuracy
2006	0.9600	0.9600	1.0000	0.9600	1.0000	0.9600	1.0000	0.9600	1.0000
2007	0.9900	0.9617	0.9714	0.9583	0.9679	0.9644	0.9742	0.9627	0.9724
2008	1.1200	1.0966	0.9791	1.0917	0.9748	1.1005	0.9826	1.0981	0.9804
2009	1.3500	1.2993	0.9625	1.3019	0.9643	1.2967	0.9605	1.2984	0.9617
2010	1.3200	1.3868	0.9494	1.3855	0.9504	1.3879	0.9486	1.3872	0.9491
2011	1.4000	1.3844	0.9888	1.3812	0.9866	1.3871	0.9908	1.3854	0.9896
2012	1.6400	1.5523	0.9465	1.5535	0.9472	1.5511	0.9458	1.5519	0.9463
2013	1.8300	1.7717	0.9682	1.7740	0.9694	1.7695	0.9669	1.7709	0.9677
2014	1.8900	1.8586	0.9834	1.8523	0.9801	1.8636	0.9861	1.8605	0.9844
2015	1.9600	1.9189	0.9790	1.9124	0.9757	1.9241	0.9817	1.9208	0.9800
2016	1.9700	2.0051	0.9822	2.0085	0.9805	2.0016	0.9840	2.0038	0.9829
2017	2.0900	2.0704	0.9906	2.0715	0.9912	2.0693	0.9901	2.0700	0.9904
2018	1.9100	2.1049	0.8980	2.1028	0.8990	2.1067	0.8970	2.1056	0.8976
2019	2.0300	2.0344	0.9978	2.0383	0.9959	2.0305	0.9998	2.0329	0.9986
2020	2.2800	2.2960	0.9930	2.2981	0.9921	2.2940	0.9939	2.2953	0.9933

According to the model building method in 2.2., the third-order induced ordered weighted average prediction error matrix for calculating is:

$$E = \begin{bmatrix} 0.0088 & 0.0075 & 0.0147 \\ 0.0075 & 0.0268 & 0.0147 \\ 0.0147 & 0.0147 & 0.0607 \end{bmatrix}.$$

Then, the IGOWA operator combination prediction model based on the optimal criterion of minimizing the sum of squares of prediction errors can be constructed as:

$$minQ^2 = W^T EW = 0.0088\omega_1^2 + 0.0015\omega_1\omega_2 + 0.0294\omega_1\omega_3 + 0.0268\omega_2^2 + 0.0294\omega_2\omega_3 + 0.0607\omega_3^2 s.t. \begin{cases} \sum_{i=1}^{3} \omega_i = 1 ; \\ \omega_i \geq 0, \quad i = 1,2,3. \end{cases}$$

Using LINGO 18.0 software to solve, the weight coefficient obtained is:

$$\omega_1 = 0.9372，\omega_2 = 0.0628，\omega_3 = 0.$$

Then the corresponding IGOWA operator combination prediction model is

$$\hat{x}_t = (0.9372\sqrt{x_{v-index(1t)}} + 0.0628\sqrt{x_{v-index(2t)}})^2. \quad (14)$$

The predicted value and prediction precision of the combined prediction model calculated according to equation (14) are shown in Table 5. During the sample period, the sum of squares of prediction errors of the model is $Q^2 = 0.00872$.

3.4. Evaluation of Model Effectiveness

In order to reflect the prediction effectiveness of different models, this paper selects the following five error indicators to build the effectiveness evaluation index system of IGOWA operator combination prediction model based on the principle of minimum sum of squares of errors [16]:

Sum of squares error:

$$SSE = \sum_{t=1}^{N}(x_t - \hat{x}_t)^2 ; \quad (15)$$

Mean square error:

$$MSE = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} ; \quad (16)$$

Square absolute error:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|x_t - \hat{x}_t\right| ; \quad (17)$$

Square absolute percentage error:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right| ; \quad (18)$$

Mean square percentage error:

$$MSPE = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(\frac{x_t - \hat{x}_t}{x_t})^2}. \quad (19)$$

According to the effectiveness evaluation index system, each error index under three kinds of single prediction models and four kinds of special values of combined prediction models is calculated and normalized [17]. The smaller the error index value is, the better the model can predict and fit the data in the sample period. The larger the error index value is, the worse the effect of the model on data prediction and fitting in the sample period is. The average prediction accuracy of various prediction methods in the sample period is supplemented, and the results are shown in Table 6.

Table 6. Prediction effect evaluation indicators of each model

Evaluating indicator	Grey prediction model	ARIMA model	Holt-Winters model	Combined forecasting model			
				λ=1	λ=-1	λ→0	λ=0.5
SSE	0.2095	0.1413	0.3067	0.0623	0.0625	0.0625	0.0623
SSE normalization	0.6831	0.4607	1.0000	0.2031	0.2038	0.2038	0.2031
MSE	0.0305	0.0251	0.0369	0.0166	0.0167	0.0167	0.0166
MSE normalization	0.8266	0.6802	1.0000	0.4499	0.4526	0.4526	0.4499
MAE	0.1022	0.0727	0.1200	0.0449	0.0464	0.0436	0.0444
MAE normalization	0.8517	0.6058	1.0000	0.3742	0.3867	0.3633	0.3700
MAPE	0.0635	0.0454	0.0717	0.0273	0.0283	0.0265	0.0270
MAPE normalization	0.8865	0.6332	1.0000	0.3808	0.3947	0.3696	0.3766
MSPE	0.0193	0.0150	0.0217	0.0096	0.0097	0.0096	0.0096
MSPE normalization	0.8894	0.6912	1.0000	0.4424	0.4470	0.4424	0.4424
Average precision	0.9365	0.9546	0.9283	0.9727	0.9717	0.9735	0.9730

From the data in Table 6, among the three single item prediction models, the prediction error of the time series model ARIMA (1,1,2) is relatively minimum, which indicates that the model has a good effect on single item prediction of data in the sample period. The Holt-Winters non-seasonal exponential smoothing prediction model has the largest prediction error index value, which indicates that the model performs poorly in single prediction fitting of data in the sample period. Then observe the prediction performance of the combined prediction model. No matter what parameters IGOWA operator takes or what error indicators are used, the prediction error index value of IGOWA operator is far less than the prediction error index value of various single prediction models. The average precision of the combined forecasting model is also higher than the average forecasting precision of each single forecast, which is higher than 97.17%. The high prediction accuracy indicates that the IGOWA operator combination prediction model has better prediction effect, more accurate fitting, and the results are closer to the true value, which has excellent prediction performance and is significantly superior to other single prediction models. Therefore, it is completely reasonable and effective to use the IGOWA operator combination prediction model based on the minimum sum of squares of errors to estimate and predict the sample data.

3.5. Model Sensitivity Analysis

When the parameter λ of IGOWA operator takes different values, the optimal weight coefficient of the combined forecasting model will also change. In order to show more intuitively the influence of parameter λ variation on weight coefficient ω and five prediction error evaluation indexes, the sensitivity of parameter λ is carried out [18]. The results are shown in Figure 2 and Figure 3 respectively.

It can be seen from Figure 2 that when λ gradually falls within the interval $[-1,0)$, the weight coefficient ω_1 decreases and ω_2 increases. When λ is gradually within the interval $(0,1]$, the weight coefficient ω_1 rises slightly and ω_2 decreases slightly. However, in the parameter interval, the weight changes little. ω_1 is always within $(0.9,1)$, ω_2 is always within $(0,0.1)$, ω_3 is always 0. This shows that the single prediction contribution degree ranks first in the prediction accuracy of each period when the combined prediction model is used in the sample period of this paper.

It can be seen from Figure 3 that, in the parameter interval, the MAE and MAPE prediction error indexes of λ change obviously. The variation of λ has little contribution to the variation of three prediction error indicators of SSE, MSE and MSPE. Among the five index values, the MSPE value of the mean square percentage error is the smallest, and all the error index values are below the 0.0625 horizontal line, which indicates that the prediction effect of the combined prediction model in this paper is very good, with high sensitivity, and meets the expectations.

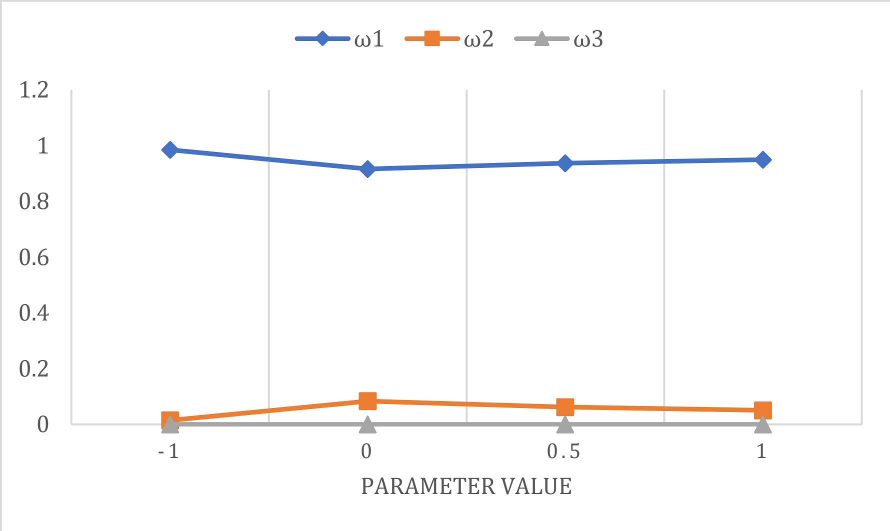

Figure 1. Influence of λ variation on optimal weight coefficient

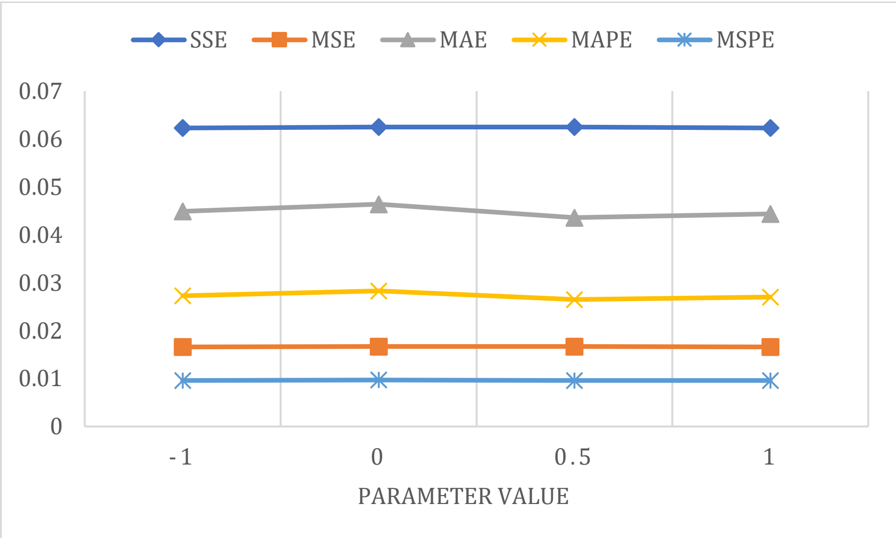

Figure 2. Influence of λ variation on five prediction error indexes

3.6. Prediction Results and Analysis

Based on the above analysis and test, it can be found that the forecasting effect of the combined forecasting model in the data sample period is significantly better than that of the grey forecasting model GM (1,1), time series model ARIMA (1,1,2), Holt-Winters non-seasonal exponential smoothing model. Therefore, it is naturally considered to use the IGOWA operator combination forecasting model based on the minimum sum of squares of error criterion to forecast the research and development intensity data of Anhui Province from 2021-2025, which will play a basic analysis and exploration role in the investment and planning of Anhui Province in scientific research and innovation in the next few years. In order to use the combination forecasting model to predict the R&D intensity data of the next five periods, we need to use the prediction accuracy of each single forecast in the same period to calculate. Since we cannot know the actual value of the R&D intensity of Anhui Province in the next several periods, we cannot know the prediction accuracy of each single forecast model in the next five periods, so we cannot determine how to assign the weight coefficient. Referring to the prediction precision processing method of prediction period proposed by Hongjun Yuan and Guiyuan Yang [19], based on the coherence principle of prediction, if we want to predict the future k periods in the prediction interval $[N+1, N+2, \cdots]$, we can use the fitting average accuracy $\frac{1}{k}\sum_{t=N-k+1}^{N} v_{it}$ of the latest k periods of the i single prediction model to reflect the prediction accuracy of $N+k$ periods in the prediction interval. Therefore, according to the prediction values of the next five periods of the three single prediction models in this paper, the prediction values of IGOWA operator combination prediction model with different values of parameter λ can be calculated respectively, as shown in Table 7.

It can be seen from Table 7 that no matter what parameter value is taken, the R&D intensity of Anhui Province will grow steadily from 2021-2025, and the scientific research and innovation ability will continue to improve. According to the 2022 Anhui Statistical Yearbook provided by the Anhui Provincial Bureau of Statistics, the actual value of Anhui's R&D intensity in 2021 is 2.34%, which is similar to the result predicted by the combination forecasting model, which proves the effectiveness of the model, and the model has statistical and forecasting significance.

Table 7. Prediction Results of Anhui R&D Intensity (Unit: %) from 2021-2025

Year	Single forecast model			Combined forecasting model			
	Grey prediction model	ARIMA model	Holt Winters model	λ=1	λ=-1	λ→0	λ=0.5
2021	2.4000	2.4020	2.3100	2.4019	2.4020	2.4018	2.4019
2022	2.5030	2.4870	2.4060	2.4878	2.4872	2.4883	2.4880
2023	2.6080	2.5770	2.5070	2.5786	2.5775	2.5796	2.5789
2024	2.7150	2.6690	2.6120	2.6713	2.6697	2.6728	2.6719
2025	2.8240	2.7620	2.7220	2.7651	2.7629	2.7671	2.7659

4. Conclusion

In this paper, the grey prediction model GM (1,1), time series ARIMA model and Holt Winters non-seasonal exponential smoothing prediction model are used to build and test the model, so that the three single prediction models are used to predict the R&D intensity data of Anhui Province from 2006 to 2020 and calculate the corresponding prediction accuracy. Then, the IGOWA operator combination prediction model based on the optimal criterion of the minimum sum of squares of errors is constructed. The prediction accuracy of three single prediction models is taken as the induced value in the combination prediction model with different λ values is established. Then the model error evaluation index system is constructed to test the effectiveness of the model and the sensitivity of parameter λ. By comparison, it can be found that the combined forecasting model optimizes the forecasting process and is more accurate and effective than each single forecasting method. Finally, the verified IGOWA operator combination prediction model is used to predict the R&D intensity of Anhui Province from 2021-2025 based on the data of the sample period. The research results show that the prediction results are in line with the actual situation, and in the next five years, the R&D intensity of Anhui Province will continue to rise, and the growth rate will increase year by year, and the scientific research and innovation ability will continue to develop steadily. In order to guarantee the investment in scientific research and innovation in Anhui Province, drive the synchronous development of R&D and society, and improve the quality of life of the people, the government needs to strengthen the dominant position of enterprise R&D investment, stabilize the R&D investment of colleges and universities, scientific research institutes, play the role of financial funds to guide, improve the level of incentive services, and gradually improve the social security system and relevant regulations. In turn, it will drive the sustained high-quality development of the province's R&D and innovation level, and lay a solid foundation for social transformation and upgrading.

Acknowledgements

The research is supported by the Natural Science Foundation of Anhui University of Finance and Economics (Item number: acszjyzd2022008).

References

[1] P.M. Romer: Endogenous technological change, Journal of Political Economy, Vol. 98 (1990) No.5, p.71-102.

[2] G.M. Grossman and E. Helpman: Quality ladders in the theory of growth, The Review of Economic Studies, Vol. 58 (1991) No. 1, p.43-61.

[3] P.S. Segerstrom, T.C.A. Anant and E. Dinopoulos: A schumpeterian model of the product life cycle, The American Economic Review, Vol. 80 (1990) No.5, p.1077-1091.

[4] X.W. Fang, J.H. Pan and C.L. Liu: Research on the prediction of R&D expenditure and intensity during the "Twelfth Five Year Plan" period in China, Science and Technology Progress and Policy, Vol. 28 (2011) No.24, p.144-147.

[5] L.Y. Zhao and P. Shi: An examination on relationship between the government’s R&D investment and economic growth — analysis on the data between 1989 and 2007, China Soft Science, (2010) No.2, p.53-58.

[6] S. Chen and W.J. Zhang: The international comparison and analysis of Chinese R&D/GDP, Studies in Science of Science, Vol. 31 (2013) No.7, p.1022-1031.

[7] J.W. He: Previous analysis on the trend of R&D intense growth in China, Value Engineering, Vol. 33 (2014) No.23, p.210-211.

[8] J.M. Bates and C.W.J. Granger: The Combination of Forecasts, OR, Vol. 20 (1969) No.4, p.451-468.

[9] H.Y. Chen: A kind of cooperative games method determining weights of combination forecasting, Forecasting, Vol. 22 (2003) No.1, p.75-77+32.

[10] G.Y. Yang and X.W. Tang: Research on optimization method of combination forecasting model with non-negative weight, Quantitative Economics and Technical Economics Research, (1998) No.3, p.56-60.

[11] L. Yang, H.Y. Chen and Y. Wang: Optimal combination forecasting model of induced generalized OWA operator based on closeness, Statistics and Decision, (2013) No.5, p.24-26.

[12] W.Y. Zhang, K. Wang and H.J. Yuan: Combination forecast of per capita disposable income of urban residents in Anhui Province, Journal of Chongqing Technology and Business University (Natural Science Edition), Vol. 39 (2022) No.5, p.93-104.

[13] J. Liu, J.W. Wang and Z.H. Cheng: Research on the influence of industrial agglomeration on the Collaborative Innovation Efficiency, China Soft Science, (2017) No.6, p.89-98.

[14] J.L. Deng: Grey Prediction and Grey Decision (Huazhong University of Science and Technology Press, China 2002), p.71-96.

[15] J.P. Xu: China’s agricultural energy consumption and CO2 emissions: current situation, trend and path to reduce emissions, Ecological Economy, Vol. 32 (2016) No.2, p.122-126.

[16] H.Y. Chen and C.L. Liu: A kind of combined forecasting method based on induced ordered weighted averaging (IOWA) operators, Forecasting, Vol. 22 (2003) No.6, p.61-65.

[17] H.Y. Yang, X, Y. Zhao and L. Wang: Review of data normalization methods, Computer Engineering and Applications, Vol. 59 (2023) No.3, p.13-22.

[18] H.J. Yuan, K. Du and L.Y. Hu: Interval-type combination forecasting model based on triangular fuzzy number similarity, Statistics and Decision, Vol. 35 (2019) No.6, p.24-29.

[19] H.J. Yuan and G.Y. Yang: The superior combination forecasting model based on an closeness degree and induced orderd weighted averaging (IOWA) operator, Statistics and Information Forum, Vol. 25 (2010) No.2, p.32-37.