Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 26, Issues 2, pp. 540–551, DOI 10.12694/scpe.v26i2.3934

FORECAST OF TOBACCO RAW MATERIAL DEMAND BASED ON COMBINATION PREDICTION MODEL

BIN CHEN*, JILAI ZHOU†, HAIYING FANG‡, RENJIE XU§ AND WEIYI QU¶

Abstract. In order to improve the prediction accuracy of tobacco raw material demand, this paper presented a combined prediction model. Combined prediction model first used Holt-winters exponential smoothing method and SARIMA model to forecast the demand of cigarette raw materials respectively, and then used BP neural network to aggregate the results of these two predictions to get the final prediction result. Holt-winters exponential smoothing method, SARIMA model and combined prediction model were used to forecast the demand data of tobacco raw materials, respectively. For the prediction of the same material, the error of the combined prediction model were all less than the other two models. The prediction accuracy of combined prediction model was higher.

Key words: Demand prediction; Holt-winters exponential smoothing method; SARIMA model; BP neural network; Combined prediction model.

1. Introduction. In the practical production process of tobacco firms, in order to meet the demand of production and realize the continuity of production, firms often appear the phenomenon of tobacco raw material replenishment, which consumes a lot of manpower and material resources. Therefore, accurate prediction of tobacco raw material demand is helpful for enterprises to make more reasonable production plans, realize rational warehouse planning, reduce enterprise resource consumption and improve enterprise work efficiency.

For raw material demand forecast, the traditional forecast methods mainly include trend extrapolation, wavelet analysis, regression analysis, grey prediction and time series prediction. The trend extrapolation method is usually used for relatively simple functional models, such as linear functions [1]. Although wavelet analysis can analyze and study each time component, it is difficult to restore the original time series because each time component has no practical significance [2].

Regression analysis determines future demand by fitting curves of past data into regression equations. However, when dealing with samples of uncertain factors, its performance is not satisfactory [3]. Grey prediction method can deal with the samples of uncertain factors well and requires a relatively small sample size [4]. The time series method fully considers the periodicity, seasonality, randomness and other significant characteristics involved in the time factor in the prediction, which can carry out dynamic analysis on the data and judge the trend of the data change over time. Based on the above advantages, time series method has been applied to air environment prediction, financial field, special material demand analysis and other fields [5][6].

ARIMA model and Holt-Winters exponential smoothing method have been widely used. ARIMA focuses on using difference method to analyze rules in time series [7]. Compared with ARIMA, seasonal ARIMA (SARIMA) adds seasonal factors in the forecasting process, therefore it is more suitable for time series with periodicity [8].

Exponential smoothing is a time series prediction method of weighted average movement [9]. Holt-Winters exponential smoothing method differs from ordinary exponential smoothing in that it can analyze the horizontal, trend, and seasonal components of historical data simultaneously [10]. Through the analysis and observation of the historical data of tobacco raw materials, it is found that it has obvious seasonality, so SARIMA and Holt-Winters exponential smoothing method are suitable for its prediction model.

*Hongyun Honghe Tobacco (Group) Co., Ltd., China.
†Hongyun Honghe Tobacco (Group) Co., Ltd., China.
‡Hongyun Honghe Tobacco (Group) Co., Ltd., China.
§Hongyun Honghe Tobacco (Group) Co., Ltd., China.
¶Business School of Hohai University, Nanjing, China (Corresponding author, weiyi_hh2198@163.com).

In the 21st century, machine learning has been applied to the field of prediction by many scholars. Common machine learning methods are Random Forest (RF), BP Neural Network, artificial neural network (ANN) [11][12]. Among them, BP Neural Network has good performance in nonlinear regression and has been widely used [13]. On the other hand, combinatorial forecasting methods have also attracted the attention of many scholars [14][15].

According to previous studies, BP neural network, Holt-winters exponential smoothing method and SARIMA model have been widely used in the field of prediction, but few studies have combined the three and applied them to the tobacco field. A combined demand prediction model combining BP neural network, Holt-winters exponential smoothing method and SARIMA model was constructed to fit the historical data of tobacco raw materials and achieve accurate prediction of tobacco raw material demand.

2. Model Construction.

2.1. SARIMA Model. The SARIMA model evolved from the basic ARMA model. The ARMA model is suitable for solving the prediction problem with stationary series. In this case, seasonal difference needs to be added to the sequence, and a more complex SARIMA model is needed to realize the prediction of time series.

The SARIMA model can be expressed by each different parameter as SARIMA(p, d, q) $(P, D, Q)_S$, whose modeling formula is shown in equations (2.1)-(2.5) [16]:

$$\phi_p(L)\Phi_P(L^S)\Delta^d\Delta_S^D y_t = \theta_q(L)\Theta_Q(L^S)u_t \tag{2.1}$$

$$\phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) \tag{2.2}$$

$$\Phi_P(L^S) = (1 - \Phi_1 L^S - \Phi_2 L^{2S} - \cdots - \Phi_p L^{PS}) \tag{2.3}$$

$$\theta_q(L) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \tag{2.4}$$

$$\Theta_Q(L^S) = (1 + \Theta_1 L^S + \Theta_2 L^{2S} + \cdots + \Theta_Q L^{QS}) \tag{2.5}$$

where p and P mean non-seasonal and seasonal autoregressive orders; ϕ_p and $\Phi_P(L^S)$ mean non-seasonal and seasonal autoregressive polynomials; d and D mean non-seasonal and seasonal difference orders; S means the seasonal cycle; Δ^d and Δ_S^D are difference and seasonal difference; y_t represents a time series; q and Q mean non-seasonal and seasonal moving average polynomials. $\theta_q(L)$ and $\Theta_Q(L^S)$ mean non-seasonal and seasonal moving average polynomials; u_t means the error.

This study adopted EViews software to construct SARIMA model, which could be divided into the following three steps:

Step 1: Stationarity test. In this paper, the stationarity of historical date series was determined by unit root test. If the series was non-stationary, it needs to make d-order difference; if the sequence had seasonality with period S, it needs to make D-order difference with step size S to make it into stationary sequence.

Step 2: Model selection. That is, the values of p, q, P, Q were certain. By observing the characteristics of the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) graph of the series after difference, the alternative combinations of each order of the model could be certain, and the optimal combinations were screened out by AIC, SC, HQC and other criteria.

Step 3: Model test. Each parameter of the model was determined in step 1 and step 2, and the residual sequence could be generated by running the model. In this paper, ADF (Augmented Dickey-Fuller) unit root test and Q test were used to test the residual error of SARIMA model.

Step 4: Model prediction. Made predictions according to the model selected in Step 3. By comparing the forecast data with the historical data, the prediction error was calculated. MAE, MAPE and RMSE were selected as the evaluation indexes. The three evaluation indexes showed that the lower the calculated value, the higher the prediction accuracy of the models. Each index formula is shown in formula respectively:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{y_i}| \tag{2.6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{y_i})^2} \tag{2.7}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{x_i - \hat{y_i}}{x_i}| * 100\% \tag{2.8}$$

Among them: x_i represents the historical data, $\hat{y_i}$ represents forecasts, n represents sample number.

2.2. Holt-Winters Exponential Smoothing Method. As a three-parameter smoothing method, exponential smoothing is mainly applied to time series with both linear and seasonal trends. The core idea of exponential smoothing method is to decompose time series into linear trend, seasonal trend and irregular change, then use exponential smoothing method to predict the above three trends or changes respectively, and finally establish addition or multiplication models for comprehensive prediction. The three parameters of Holt-winters exponential smoothing method include α, β and γ, which control the level, trend and season at time t, respectively. The value of each parameter ranges from 0 to 1. A larger value indicates a higher prediction ratio.

The addition model formula of Holt-winters exponential smoothing method is as follows:

$$u_t = \alpha(y_t - s_{t-T}) + (1-\alpha)(u_{t-1} + v_{t-1}) \tag{2.9}$$

$$v_t = \beta(u_t - u_{t-1}) + (1-\beta)v_{t-1} \tag{2.10}$$

$$s_t = \gamma(y_t - u_t) + (1-\gamma)s_{t-T} \tag{2.11}$$

The multiplication model formula of Holt-winters exponential smoothing method is as follows:

$$u_t = \alpha(\frac{y_t}{s_{t-T}}) + (1-\alpha)(u_{t-1} + v_{t-1}) \tag{2.12}$$

$$v_t = \beta(u_t - u_{t-1}) + (1-\beta)v_{t-1} \tag{2.13}$$

$$s_t = \gamma(\frac{y_t}{u_t}) + (1-\gamma)s_{t-T} \tag{2.14}$$

where, y_t is time series, u_t is linear trend, v_t represents the linear increasing rate of u_t, s_t is seasonal trend, T means seasonal cycle.

2.3. BP Neural Network. The structure of the BP neural network is shown in Figure 2.1. The original data is input layer, and then Data from the input layer pass to the hidden layer, hidden layer built-in transfer function used to process the input to the hidden layer of data, hidden layer will be processed data to the output layer, the output result z_1, z_2, $\cdots$, z_n. The BP neural network compares the output results with the original data and calculates the errors. If the errors meet the set requirements, the model can be determined. If the errors are greater than the set requirements, the errors will be reverse-transmitted to the hidden layer, and then the model parameters will be redetermined until the model errors meet the requirements.

There are three types of BP neural network data sets, namely training set, verification set and test set. In addition, in order to accelerate the training speed and improve the model performance and stability, BP neural network needs to normalize the original data. The formula is as follows:

$$S_t^{'} = \frac{S_t - min(S)}{max(S) - min(S)} \tag{2.15}$$

Among them: S means time sequence, S_t and $S_t^{'}$ means the value of the time series and normalized time series at time t, respectively. $max(S)$ and $min(S)$ mean the maximum and minimum value in the sequence.

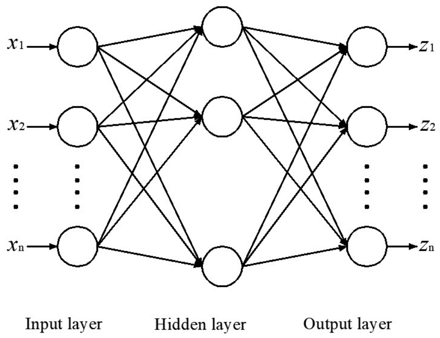


Fig. 2.1: Neural network structure diagram

2.4. Combinatorial Forecasting Model. Combinatorial forecasting techniques can be divided into two types: model grouping and result grouping. The model grouping method focuses on the analysis of the advantages and disadvantages of different models to put forward a combination method to realize the complementary advantages of the models. The common model grouping method includes the combination of qualitative prediction and quantitative prediction, linear prediction and nonlinear prediction, dynamic prediction and static prediction. The idea of result grouping method is to first obtain multiple prediction results by using different prediction models and methods, then appropriate methods are used to determine the weight of each model, and finally their weighted average is calculated and used as the predictive value. This paper chose the result grouping method and the expression of the weight model is as follows:

$$\hat{y_t} = w_{1t}y_{1t} + w_{2t}y_{2t} \tag{2.16}$$

In this formula: $\hat{y_t}$ is the output of BP neural network and represents prediction value of combination forecast model at time t. y_{1t} and y_{2t} are the input of BP neural network and represent the predicted values of the SARIMA model and Holt-winters exponential smoothing method at time t, respectively. w_{1t} and w_{2t} represent the respective weights of the predicted values of SARIMA model obtained by BP neural network and Holt-winters exponential smoothing method at time t.

The objective function of BP neural network is $L2$ norm loss function, and the formula is as follows:

$$L_2(\hat{y}, y) = \sum_{t=0}^{m} (y_t - \hat{y_t})^2 \tag{2.17}$$

In this formula: y_t represents value of the time series at time t, $\hat{y_t}$ represents prediction value of combination forecast model at time t.

There were three steps to implement the combined prediction model:

Step 1: SARIMA model and Holt-winters exponential smoothing method were respectively used to forecast the past tobacco raw material demand data, and the respective prediction results were obtained.

Step 2: Selected the appropriate proportion and divided the predicted results obtained in Step 1 and the corresponding real values into training set, verification set and test set. And take the actual demand as the model test standard. Using training set to train the BP neural network, and verification set to debug the parameters.

Step 3: Taking the training set and verification set as the input after parameter debugging, running the BP neural network, and then output prediction result. Calculated the prediction error and evaluated the prediction effect. The evaluation indicators were MAE, RMSE and MAPE.

It is worth mentioning that when the time series has 0 or the seasonal variation of the series is not significant, the SARIMA model will no longer be applicable, thus affecting the prediction accuracy of the combined model.

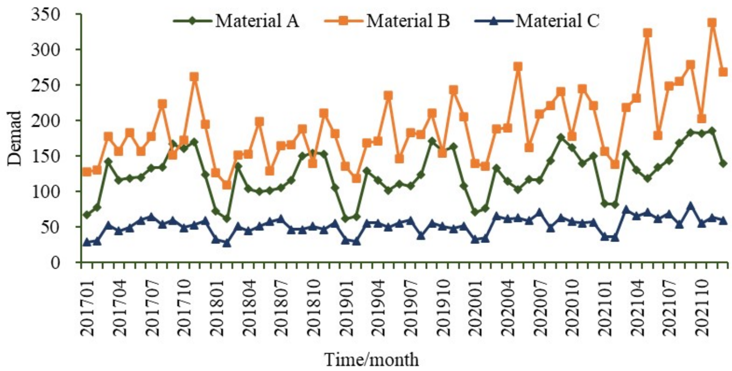


Fig. 3.1: Trend chart of historical data of materials demand

Table 3.1: Unit root text of DSX

		$t-Statistic$	$Prob$
ADF test statistic		-5.970426	0.0000
Test critical values	1%level	-3.615588	
	5%level	-2.941145	
	10%level	-2.609066	

3. Empirical Analysis.

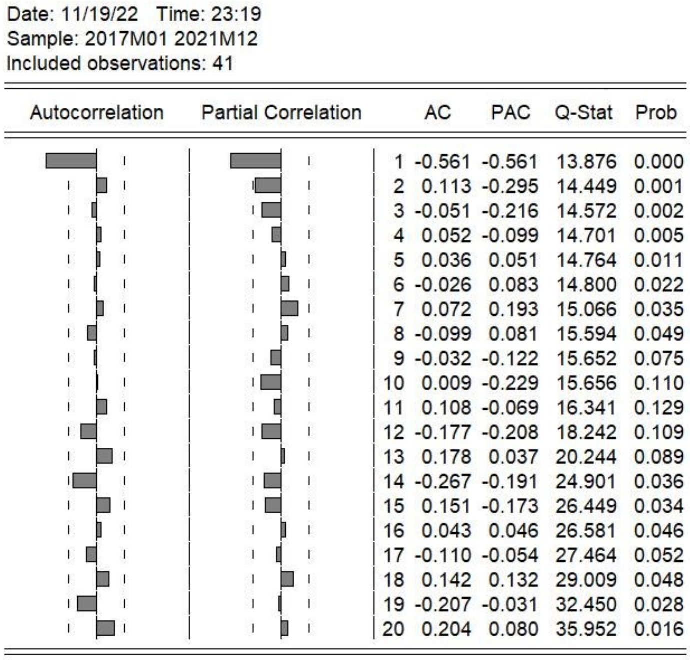
3.1. Data Sources and Description. In this paper, the materials historical demand data of material A, B and C of Y Tobacco Company from January 2017 to December 2021 were selected as samples, in which material C was the main raw material of Y Tobacco Company's high-end products, and material A and B were the raw material of common products. Here were 60 periods of historical data for different materials. In order to verify the prediction accuracy of the combined prediction model, the data from July to December 2021 were taken as the test set of the combined model. The historical data of each material are shown in Figure 3.1. Among them, the demand for material A and material B decreased slightly in 2018, and the demand for raw materials showed an overall rising trend from 2019 to 2021. As high-end raw materials, the overall demand for material C was less than that for material A and material B. The overall data of the three kinds of material showed non-stationary time series with obvious seasonality.

Due to space limitations, this paper only shows the process of predicting raw material demand of material A by using models. The prediction results of each model of all materials will be shown in Section 3.4.

3.2. Prediction of SARIMA Model. The first-order difference of $S = 12$ was applied to the historical data series of raw material demand of material A from January 2017 to June 2021, that was, SARIMA$(p, 1, q)$ $(P, 1, Q)_{12}$. The sequence before the difference was called X and the sequence after the difference was called DSX.

First, unit root test was performed on DSX. As could be seen from Table 3.1, T-test statistic was -5.970426, which was less than -2.941145 at 5% test level, indicating that residuals pass ADF test and DSX was stationary time series.

ACF and PACF of DSX were observed, as shown in Figure 3.2. PACF of DSX was significantly non-0 when lagging 1 order, so $p = 1$ could be considered; ACF was significantly non-0 when lagging 1 order, so $q = 1$ could be considered; in addition, since seasonal difference with period 12 was carried out in the time series, $P = 1$


Date: 11/19/22 Time: 23:19
Sample: 2017M01 2021M12
Included observations: 41

	AC	PAC	Q-Stat	Prob
1	-0.561	-0.561	13.876	0.000
2	0.113	-0.295	14.449	0.001
3	-0.051	-0.216	14.572	0.002
4	0.052	-0.099	14.701	0.005
5	0.036	0.051	14.764	0.011
6	-0.026	0.083	14.800	0.022
7	0.072	0.193	15.066	0.035
8	-0.099	0.081	15.594	0.049
9	-0.032	-0.122	15.652	0.075
10	0.009	-0.229	15.656	0.110
11	0.108	-0.069	16.341	0.129
12	-0.177	-0.208	18.242	0.109
13	0.178	0.037	20.244	0.089
14	-0.267	-0.191	24.901	0.036
15	0.151	-0.173	26.449	0.034
16	0.043	0.046	26.581	0.046
17	-0.110	-0.054	27.464	0.052
18	0.142	0.132	29.009	0.048
19	-0.207	-0.031	32.450	0.028
20	0.204	0.080	35.952	0.016

Fig. 3.2: ACF-PACF diagram of DSX

Table 3.2: Comparison results of SARIMA models

Model	Adjusted R-squared	AIC	SC	HQC
SARIMA$(1,1,1)(1,1,1)_{12}$	-	-	-	-
SARIMA$(1,1,1)(1,1,0)_{12}$	0.422709	7.864747	8.031925	7.925624
SARIMA$(1,1,1)(0,1,1)_{12}$	-	-	-	-
SARIMA$(1,1,0)(1,1,1)_{12}$	-	-	-	-
SARIMA$(0,1,1)(1,1,1)_{12}$	0.423823	7.882115	8.049293	7.942992
SARIMA$(1,1,1)(0,1,0)_{12}$	0.398515	7.856398	7.981781	7.902056
SARIMA$(0,1,0)(1,1,1)_{12}$	0.087883	8.316835	8.442218	8.362492
SARIMA$(1,1,0)(1,1,0)_{12}$	0.309010	7.994796	8.120179	8.040453
SARIMA$(0,1,1)(0,1,1)_{12}$	0.439722	7.833452	7.958836	7.879110
SARIMA$(0,1,1)(1,1,0)_{12}$	0.427561	7.840078	7.965461	7.885735
SARIMA$(1,1,0)(0,1,1)_{12}$	0.335584	7.980981	8.106364	8.026639
SARIMA$(1,1,0)(0,1,0)_{12}$	0.311724	7.959510	8.043099	7.989949
SARIMA$(0,1,1)(0,1,0)_{12}$	0.389070	7.847230	7.930819	7.877668
SARIMA$(0,1,0)(1,1,0)_{12}$	0.084441	8.278828	8.362416	8.309266
SARIMA$(0,1,0)(0,1,1)_{12}$	0.113267	8.268651	8.352240	8.299090

and $Q = 1$ could be considered.

Combined different values of P, p, Q, q into different SARIMA models, and calculate the Adjusted R-squared, AIC, SC and HQC values of each model. When the adjusted R-squared value was larger and the AIC, SC and HQC values were smaller, the prediction of the model was better. The comparison results of models could be seen in Table 3.2. Through comprehensive comparison and observation, SARIMA$(0,1,1)(0,1,1)_{12}$ performed best and selected as the model of material A.

The ADF test was carried out for the residual of SARIMA$(0,1,1)(0,1,1)_{12}$ model. The test results could

Table 3.3: ADF test of residual series

		$t-Statistic$	$Prob$
ADF test statistic		-5.354789	0.0001
Test critical values	1%level	-3.632900	
	5%level	-2.948404	
	10%level	-2.612874	

Table 3.4: Error of SARIMA$(0, 1, 1)(0, 1, 1)_{12}$ model

Time/month	Actual value	Prediction value	MAPE
2021.07	144	130.23	9.56%
2021.08	168	152.16	9.43%
2021.09	183	188.05	2.76%
2021.10	182	176.25	3.16%
2021.11	185	164.87	10.88%
2021.12	140	150.36	7.40%
MAE:11.82	RMSE:12.99	MAPE:7.20%	

be seen in Table 3.3. The T-test statistic was -5.354789, which was less than the -2.948404 at the test level of 5%, indicating that the residual has passed the ADF test. At the same time, the Q test with 20 order lag for the residual sequence was performed and $P = 0.294 > 0.05$ was obtained, indicating that the residual sequence passed the Q test and appeared as white noise sequence. If a sequence is a white noise sequence, it means that all information in the sequence has been extracted. Therefore, the SARIMA$(0, 1, 1)(0, 1, 1)_{12}$ model was valid and could be used to forecasting.

The SARIMA$(0, 1, 1)(0, 1, 1)_{12}$ model was used to fit and predicted the X series, the forecast data from July to December 2021 were used to compare with the real demand data, and calculated the prediction accuracy. Model errors were shown in Table 3.4, the MAPE was 7.20%, which was less than 10%. It was proved that the performance of the model was good, but it could also be seen that the stability of the model was poor for the prediction of different months.

3.3. Prediction by Holt-Winters Exponential Smoothing Method. EViews software was used for prediction, and the parameters in additive model and multiplicative model were the parameters with the highest fitting degree automatically calculated by the software. With the historical raw material demand data of A from January 2017 to December 2020 as the original data, the demand of material A from January 2021 to June 2021 was predicted. According to the size of the errors, judged which model of addition model or multiplication model had better prediction effect. Table 3.5 showed the prediction results of two models of Holt-winters. According to the table, compared with the multiplication model, most of the predicted values of the addition model were closer to the real values. But more specific calculations were needed, using errors to measure the performance of both models.

The evaluation indexes of the additive model and the multiplicative model were calculated and shown in Table 3.6. According to the evaluation results, each index of the addition model was superior to the multiplication model, and the addition model was selected as the prediction model of raw material demand for material A.

Taking the historical raw material demand data of A from January 2017 to June 2021 as the original data, the addition model was used to forecast the raw material demand from July 2021 to December 2021, and compared with the actual demand in the current month. The fitting error of the Holt-winters addition model could be seen in Table 3.7. The MAPE of addition model was 4.60% and less than 7.20% of SARIMA model, other indexes were also smaller. However, the maximal MAPE of different months was 10.60% and the

Table 3.5: Forecast result of Holt-winters models

Time/month	Actual value	model	
		Additive model	Multiplication model
2021.01	83	74.16	71.57
2021.02	82	76.41	73.58
2021.03	153	140.91	141.67
2021.04	130	118.66	118.19
2021.05	118	112.16	111.26
2021.06	134	118.66	118.05

Table 3.6: Fitting results of Holt-winters models

Model	Parameter			Fit evaluation		
	α	β	γ	MAE	$RMSE$	$MAPE$
Additive model	0.26	0	0	9.84	10.34	8.14%
Multiplication model	0.24	0	0	10.95	11.32	9.96%

Table 3.7: Fitting error of Holt-winters additive model

Time/month	Actual value	Prediction value	MAPE
2021.07	144	134.93	6.30%
2021.08	168	150.20	10.60%
2021.09	183	187.97	2.72%
2021.10	182	181.49	0.28%
2021.11	185	181.02	2.15%
2021.12	140	147.79	5.56%
MAE:9.13	RMSE:7.35	MAPE:4.60%	

minimum MAPE was 0.28%, the stability of the model was also not very well. In general, the fitting effect was better than that of SARIMA model.

3.4. Combination Model Prediction. The number of neurons in the BP neural network determines the complexity of the model, so as to adapt to the nonlinear mapping relationship of different complexity, but at the same time, the number of neurons also affects the computational complexity and training time. Therefore, the number of neurons of BP neural network needed to be confirmed. According to the empirical formula, the optimal number of neurons of hidden layers was between 2 and 8. Then, 35 groups of data from February 2018 to December 2020 were selected as the training set, the minimum error of the training target was set as 0.0001, the training times as 1000, and the learning rate as 0.01. Data from January to June 2021 were selected as the verification set. The number of each neuron was tested for 15 times, RMSE was used as the evaluation index, and the mean value of the 15 test results was taken. As can be seen from Figure 3.3, for the demand prediction of material A, the most appropriate number of hidden layer neurons was 2. As the number of neurons increased, the complexity of the model increased, and overfitting problems might occur, resulting in increased model errors.

After the number of neurons was determined, the combined prediction model could be constructed. The material A demand data from July 2021 to December 2021 were predicted. The prediction results and the values of each evaluation index were shown in Table 3.8. The MAPE of the combined prediction model decreased to

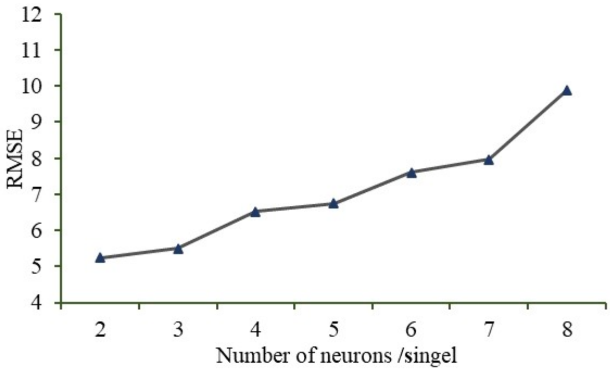


Fig. 3.3: Evaluation results of different neuron numbers

Table 3.8: Forecast results of combination forecasting model of material A

Time/month	Actual value	Prediction value	MAPE
2021.07	144	138.52	3.81%
2021.08	168	152.92	8.98%
2021.09	183	184.16	0.64%
2021.10	182	180.65	0.74%
2021.11	185	182.22	1.50%
2021.12	140	150.31	7.37%
MAE:7.90	RMSE:6.05	MAPE:3.84%	

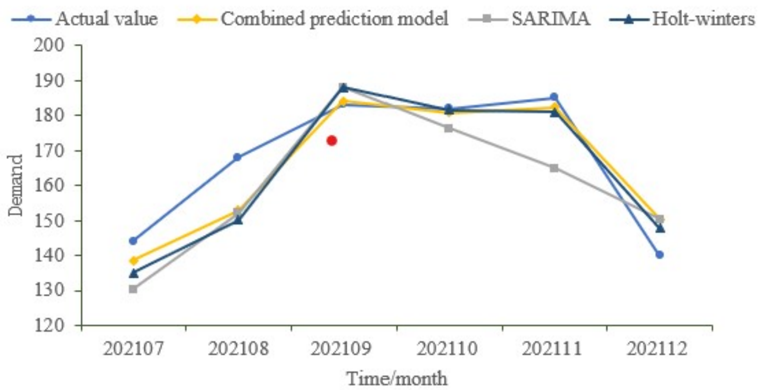


Fig. 3.4: Comparison of prediction results among three models of material A

3.84%.

The prediction results of the models for materials A, B and C were compared with the actual values of the current period to verify the validity of the combined model. The results were shown in Figures 3.4-3.6. By observing these figures, it could be known that the result of the combined prediction model was closer to the historical demand date, which represented it had good performance for prediction.

The accuracy indexes of different models of A, B and C were compared, and the results could be known in Table 3.9.

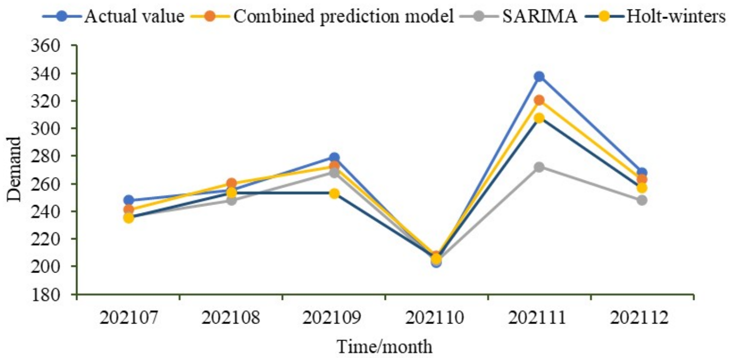


Fig. 3.5: Comparison of prediction results among three models of material B

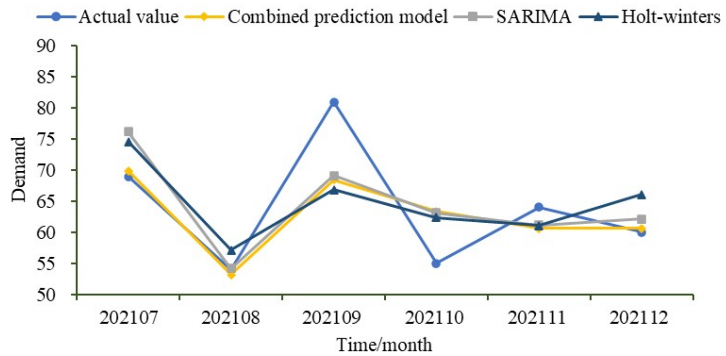


Fig. 3.6: Comparison of prediction results among three models of material C

Table 3.9: Comparison results of precision index among models of materials

Material	Model	MAE	RMSE	MAPE
Material A	Combined prediction model	6.03	7.90	3.84%
	SARIMA model	11.82	12.99	7.20%
	Holt-winters exponential smoothing method	7.35	9.13	4.60%
Material B	Combined prediction model	7.63	8.83	2.75%
	SARIMA model	19.47	29.04	6.49%
	Holt-winters exponential smoothing method	11.19	16.33	4.96%
Material C	Combined prediction model	4.44	6.33	6.64%
	SARIMA model	5.39	6.72	8.02%
	Holt-winters exponential smoothing method	6.53	7.54	9.90%

1. For different materials, the performance of three models of material C was worse than that of material A and B, which might be because the annual growth rate of overall demand for material C from 2019 to 2021 changes greatly. The growth rate of material A from 2019 to 2021 was 4.19%, 6.07% and 13.32%, respectively. Material B was 12.21%, 11.86% and 18.00%, while material C was 1.56%, 15.04% and 8.47%, respectively. Compared with material A and B, the demand growth rate of material C was

more volatile.

2. For different models, the MAPE of the three models was all less than 10%, which proved that the three models had better performance for the prediction of materials demand. Compared with the other two models, the prediction errors of the combined prediction model were reduced by 3.36% and 0.76% for material A, 3.74% and 2.21% for material B, and 1.38% and 3.26% for material C, respectively. It was proved that the overall fitting error of the combined prediction model was smaller for different order of magnitude of the material, and it had higher prediction accuracy.
3. For the prediction of material A, when the predicted values of SARIMA model and Holt-winters exponential smoothing method were directly averaged, the MAPE obtained was 4.60%, which was higher than 3.84% of the combined prediction model. It was proved that the weights of SARIMA model and Holt-winters exponential smoothing method were calculated by BP neural network, compared with direct average, the prediction accuracy can be improved effectively.

4. Conclusion. In this paper, a combination prediction model was proposed by combining BP neural network, Holt-winters exponential smoothing method and SARIMA model, and the effectiveness of the combination prediction model was analyzed by taking the historical data of tobacco raw material demand of Y enterprise as an example. The result grouping method and BP neural network were used to combine the predicted results of two models, and the combined predicted value was closer to the real value by calculating appropriate weight distribution.

1. Experimental results showed that the combined prediction model proposed in this paper had smaller prediction error and better fitting effect for different tobacco materials, and provided certain reference basis for cigarette factory to scientifically arrange production plan, inventory plan and cargo space allocation plan. In addition, the Holt-winters exponential smoothing method of BP neural network and the SARIMA model are often used in sales forecasting, financial forecasting and medical forecasting. Therefore, the combined forecasting model proposed in this paper may have strong applicability in these fields.
2. At the same time, it was found in the research process that the demand for cigarette raw materials in January and February was lower than that in other months due to the shutdown of cigarette factories during the Spring Festival, resulting in a large difference between the demand in January, February and other months. The Spring Festival and other holiday factors enhance the volatility of data. In addition, factors such as policy adjustments and market fluctuations will also have a great impact on the demand for cigarette raw materials, which may increase the prediction error. Therefore, in the subsequent research, factors such as holiday factors and policy adjustment factors could be introduced into the prediction model.

Acknowledgments. Thanks to the great support of the project team and Hohai University. Fundings from the Science and Technology Project (No. HYHH2020XX02) of Hongyun Honghe Tobacco (Group) Co., Ltd are gratefully acknowledged.

REFERENCES

[1] N. Y. Krakauer, M. D. Grossberg, I. Gladkova, and H. Aizenman, *Information content of seasonal forecasts in a changing climate*, Advances in Meteorology, vol. 2013, no. 1, Jan. 2013, pp. 1-12.

[2] S. Anil, K. Makoto, *Nexus between renewable energy certificates and electricity prices in India: evidence from wavelet coherence analysis*, Renewable Energy, Philadelphia, vol. 204, no. 0, Mar. 2023, pp. 836-847.

[3] K. Seungmi, K. Jaehwang, H. Ding, X. Xu, R. Chen, J. Guo and H. Fu, *Using multiple regression analysis to predict directionally solidified TiAl mechanical property*, Journal of Materials Science and Technology, vol. 104, no. 0, Mar. 2022, pp. 285-291.

[4] B. Men, Z. Wu, H. Liu, Z. Hu and Y. Li, *Improved grey prediction method for optimal allocation of water resources: a case study in Beijing in China*, Water Supply, vol. 19, no. 4, Jun. 2019, pp. 1044-1054.

[5] N. S. Muruganandam, U. Arumugam, *Dynamic ensemble multivariate time series forecasting model for PM2.5*, Computer Systems Science and Engineering, vol. 44, no. 2, Jun. 2022, pp. 979-989.

[6] L. Ana, H. P. Javier and M. Manuel, *A combined model based on recurrent neural networks and graph convolutional networks for financial time series forecasting*, Mathematics, vol. 11, no. 2, Jan. 2022, pp. 224-224.

[7] A. S. Ahmar, P. K. Singh, N. V. Thanh, N. V. Tinh and V. M. Hieu, *Prediction of BRIC stock price using ARIMA, SutteARIMA and Holt-Winters*, Computers, Materials and Continua, vol. 70, no. 1, Jan. 2022, pp. 525-534.

[8] O. Mahmod, I. Rachmah, S. A. Abubakar, Q. M. Bagus, and S. Rajalingam, *Model forecasting development for dengue fever incidence in Surabaya city using time series analysis*, Processes, vol. 10, no. 11, Nov. 2022, pp. 2454-2454.
[9] L. Rubio, A. J. Gutirrez-Rodriguez and M. G. Forero, *EBITDA index prediction using exponential smoothing and arima model*, Mathematics, vol. 9, no. 20, Oct. 2021, pp. 2538-2538.
[10] S. P. Kumar, P. A. Kumar, A. Sahil and K. Ravi, *Multiple forecasting approach: a prediction of CO2 emission from the paddy crop in India*, Environmental Science and Pollution Research, vol. 29, no. 17, Nov. 2021, pp. 25461-25472.
[11] P. Bajari, D. Nekipelov, S. P. Ryan, and M. Yang, *Machine learning methods for demand estimation*, The American Economic Review, vol. 105, no. 5, May 2015, pp. 481-485.
[12] H. Wumaier, J. Gao and J. Zhou, *Short-term forecasting method for dynamic traffic flow based on stochastic forest algorithm*, Journal of Intelligent and Fuzzy Systems, vol. 39, no. 2, Jun. 2020, pp. 1-13.
[13] F. O. John, S. Thokozani, *Path loss prediction in tropical regions using machine learning techniques: a case study*, Electronics, vol. 11, no. 17, Aug. 2022, pp. 2711-2711.
[14] C. Tarmanini, N. Sarma, C. Gezegin and O. Ozgonenel, *Short term load forecasting based on ARIMA and ANN approaches*, Energy Reports, vol. 9, no. S3, May 2023, pp. 550-557.
[15] S. Chatterjee, Y. Byun, *A synthetic data generation technique for enhancement of prediction accuracy of electric vehicles demand*, Sensors, vol. 23, no. 2, Jan. 2023, pp. 294-294.
[16] T. H. Noor, A. M. Almars, M. Alwateer, M. Almaliki, I. Gad and E. Atlam, *SARIMA: a seasonal autoregressive integrated moving average model for crime analysis in Saudi Arabia*, Electronics, vol. 11, no. 23, Nov. 2022, pp. 3986-3986.

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: May 14, 2024
Accepted: Aug 12, 2024