Price Prediction of Traditional Chinese Medicine Based on ARIMA and Improved Elman Neural Network

Tao Fang¹, Xingliang Zhang¹, Chunlei Yang¹, Zhengzheng Huang¹, and Xiaodie Zhang¹
¹Hebei GEO University, Shijiazhuang 050000, China

Abstract. The price’s change of traditional Chinese medicine contains linear, non-linear and other miscellaneous factors. It is difficult for people to use a separate model such as neural network model to judge its price trend. Based on the background, a combined forecasting model is proposed in this paper, it consists of Autoregressive Integrated Moving Average model and Elman neural network which is improved by correlation analysis. The combined forecasting model can use its two algorithm model to deal with the linear and nonlinear factors. Meanwhile, the innovation of this paper is using correlation analysis to import extra additional parameters for the neural network, which can increase its accuracy. A large number of traditional Chinese medicine’s price data was collected to be training samples, the final results show that the combined forecasting model has advantages over stability and accuracy than ARIMA or Elman neural network.

1 Introduction

The price of traditional Chinese medicine changes faster than it of common agricultural products. According to the price data, ginseng, a common and traditional Chinese herbal medicine, changes a lot in seven years. From 2013 to 2019, its highest price is 1600 yuan/kg, and its lowest price is 520 yuan/kg, volatile prices often make it difficult for market to predict the price and respond the condition.

Traditional Chinese medicine’s price is a complex system with linear and nonlinear changes, it is difficult to define it as a simple linear model or a nonlinear model, which make it hard to predict it. However, price data is in chronological order, to some extent, it has obvious time-varying characteristics and can be regarded as time series data through some pretreatment. ARIMA¹ is widely used in the linear change system. However, there are many factors causing the price changing. So using combined model is necessary. Another major characteristic of traditional Chinese medicine is its relationship, it is obvious that a treatment regimen consists many herbs, herbs is also easy replaced by some other herbs. So there is a high degree of correlation between traditional Chinese medicines, which make the price change strangely occasionally, nevertheless, correlation analysis can greatly improve the situation. Artificial neural network²-³ has the ability to map any nonlinear functions, and it also has strong learning and prediction ability by deal with a large number
of data. In this context, the combined model can maximize its own advantages and avoid its inevitable weakness, so it is a more effective forecasting model for price forecasting.

2 Model establishment

Scholars at home and abroad generally use ARIMA model to predict the price\textsuperscript{[4]}, GDP, exchange rate and other economic aspects. The results show that ARIMA has excellent prediction ability in time series data, and traditional Chinese medicine can be regarded as time series under certain conditions, so it is easy to use ARIMA to deal with it. However, traditional Chinese medicine price contain linear and nonlinear systems, it is a complex system, so it is necessary to use calculus of differences to get the smooth data.

Elman neural network can map any nonlinear function by learning a mass of data. According to ‘no free lunch theorem’, there is no algorithm that can solve all problems perfectly. Different algorithms always have their advantages and disadvantages in different fields, the key to solve the problems is to find a suitable method, Elman neural network has an extra connection layer than other neural networks which make it has ability to solve the time series. Meanwhile, combining these two models can use their features to handle their own tasks that they are good at.

However, it is biased to predict the trend of a certain kind of medicine only through its own price. Traditional Chinese medicine has a strong correlation because of its medicinal properties, which make it possible to predict the medicine by using the price of relevant medicine.

2.1 ARIMA

ARIMA is putted forward through combining the advantages of AR (autoregressive) and MA(moving average), and ‘i’ means calculus of difference. ARMA can be determined as follows\textsuperscript{[5]}:

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}
\]  \hspace{1cm} (1)

where \( \phi \) and \( \theta \) is the undetermined coefficient that is not zero; \( e_{t-i} \) is error term.

However it can only be applied to stationary time series, so it’s necessary to use difference. After a series of calculations, ARIMA (6,1,6) is determined, taking honeysuckle for example, the predicted data is available.

![Predicted data by ARIMA.](image)
Figure 1 shows that ARIMA only predict the general trend, it’s hard for it to predict the small shifts in the short term, so it’s necessary to use other methods to make up for it and it also means that traditional Chinese medicine price is a complex system.

### 2.2 Coefficient of association

The complexity of objective things makes it difficult for people to know their causality and correlation from huge data. Establishing a reasonable mathematical model can greatly simplify the steps and improve the efficiency. Economic knowledge tells us that commodity demand $q$ is not only related to its price, but also related to the prices of substitutes and complementary products with similar functions. This is an example problem of multiple independent variables affecting a dependent variable. The difficulty is how to find out the data that affects the change from the vast amount of data. Meanwhile, correlation coefficient is one of the common correlation analyses\cite{6}, it’s efficient to use it to calculate the impact from many medicinal materials. It can be determined as follows:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$  \hspace{1cm} (2)

where $r_{xy}$ is sample coefficient, $s_{xy}$ is covariance, $s_x$ and $s_y$ are sample standard deviation, they are determined as follows:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$  \hspace{1cm} (3)

$$C = \begin{pmatrix}
    cov(x, x) & cov(x, y) & cov(x, z) \\
    cov(y, x) & cov(y, y) & cov(y, z) \\
    cov(z, x) & cov(z, y) & cov(z, z)
\end{pmatrix}$$  \hspace{1cm} (4)

By calculating the association from large amounts of data, the results are as follows:

**Table 1.** Tables of the coefficient of association.

<table>
<thead>
<tr>
<th>Drug name</th>
<th>honeysuckle</th>
<th>Rhizome atractylodis</th>
<th>Polygonatum</th>
</tr>
</thead>
<tbody>
<tr>
<td>honeysuckle</td>
<td>-</td>
<td>0.9366</td>
<td>0.9265</td>
</tr>
<tr>
<td>Rhizome atractylodis</td>
<td>0.9366</td>
<td>-</td>
<td>0.9042</td>
</tr>
<tr>
<td>Polygonatum</td>
<td>0.9265</td>
<td>0.9042</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1 shows that the correlation between the three herbs is very high, and they are all over 0.9. So, the prices of Rhizome atractylodis and Polygonatum are available to be factors of honeysuckle’s price.

### 3 Elman neural network

In 1990, J.L Elman proposed Elman neural network on the basis of BP neural network\cite{7}, which is a typical local regression neural network, characterized by local memory function.

260
and can feedback the values of some neurons to the next layer. It is composed of input layer, hidden layer, connection layer and output layer\cite{8}. The function of connection layer is to feed back the value part of hidden layer to input layer, which is the biggest difference between Elman neural network and BP neural network. Due to the existence of the connection layer, Elman neural network can dynamically remember the previous state and predict the current stage, while the price change is often greatly affected by the previous time period, so Elman neural network has greater advantages in the data prediction of time series such as price.

Figure 2. Structure of Elman neural network.

Be different from the normal Elman neural network, replacing the data to be forecasted with the data obtained in the previous section. The output of the improved Elman neural network is obtained, then combining it with the data from ARIMA. Final predict data is follow:

Figure 3. Predict data by Elman.
Figure 4. Predict data by combinatorial model.

Figure 3 shows that Elman neural network can predict the outcome accurately, especially in the short term, it can predict accurately and instantly, but its weakness is obvious, when prices are flat, its forecasts fluctuate. In order to deal with it, combining it with ARIMA, the results are shown in figure 4. It is clear that the predictions are very smooth and accurate.

3 Analysis of experimental results

In order to evaluate the results of different experiments intuitively, two indexes of mean relative error (MRE) and relative root mean square error (RRMSE) are used to compare and analyse the final experimental results, results are follow:

<table>
<thead>
<tr>
<th>Index evaluation table.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Elman</td>
</tr>
<tr>
<td>Combinatorial model</td>
</tr>
<tr>
<td>MRE</td>
</tr>
<tr>
<td>RRMSE</td>
</tr>
</tbody>
</table>

Figure 3, figure 4 and table 2 show that the combinatorial model sacrifice the accuracy of some periods to achieve the global stationarity and accuracy, it has the advantages of both and avoid their disadvantages.

4 Summary

It’s inaccurate to use a separate model to analysis a complex system, in this paper, the combined model can deal with this problem. It can use two models to deal with the linear and nonlinear parts respectively, which simplify the complexity of the problem, meanwhile, the correlation coefficient provides additional parameters for the neural network. From the results, the combined model has advantages over stability and accuracy than any single model.

The authors would like to acknowledge the science and technology planning project in Hebei Province (Grant no. 19450904D, 20550301D, 20557638D).

262
References


