Comparative Analysis of Prophet and LSTM Model in Drug Sales Forecasting

Jie Meng, Xiao Yang*, Chengwei Yang and Yang Liu
School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, China
Email: yangxiao@sdufe.edu.cn

Abstract. With the rapid development of China's pharmaceutical industry, accurate prediction of drug sales has become the key to enterprises' competitiveness. Sales forecasting research has very important value for strategic decisions and improvement measures made by enterprises. We studied mainly two machine learning methods of pharmaceutical sales prediction in this paper, analyzed deeply the prophet model and LSTM, and carried out a comparative experiment on these two methods using real sales data. The experimental results show that the LSTM model is more accurate than the Prophet model on drug sales forecasting.

Keywords. Pharmaceutical industry, Sales forecasting, Prophet model, LSTM.

1. Introduction
Nowadays, medical demand in China presents the increasing trend year by year, the state for the pharmaceutical industry's emphasis is also increasing. Facing the increasing market demand, enterprises must do a certain degree of demand management to survive in the competitive market environment, in which forecasting is an important part of demand management. For the enterprises, accurate prediction of the sales trend of products can help them more accurately grasp the changes in the market, so as to provide an important reference for enterprise inventory management and marketing. Nowadays, most pharmacies in China generally lack of scientific forecasting methods, and most of them judge the demand and inventory of drugs according to the subjective experience of salesmen [1]. This, to some extent, reduces the efficiency and accuracy of the prediction, thus making it impossible to make accurate prediction.

Among a large number of sales forecasting methods, time series modeling is a commonly used method with good prediction effect. It can predict the future trend according to the past changes of the market, and the data it relies on is relatively simple, so it is easy to get the prediction results. Depending on the complexity of the system, artificial intelligence neural network can simulate the human brain neural network and carry out information processing by adjusting the inter-connected relations among a large number of internal nodes. Moreover, it has a good ability of self-learning and self-adaptation, and has the function of nonlinear mapping in prediction. Through comparative analysis, many scholars used these two methods to forecast sales volume, and obtained pretty good results.

Based on the above considerations, this paper selected the time series model Prophet [2] and LSTM neural network model [3] as prediction model using the real sales volume data of S enterprise from January 2017 to December 2019. The experimental results showed that the error of LSTM model was smaller than Prophet model.
2. The Prophet Model and LSTM model

2.1. The Prophet Model

Prophet [2] is an open-source time series analysis framework developed by facebook, which analyzes a variety of time series features: periodicity, trends, partial outliers and holiday effects to make predictions. Compared with some traditional time series methods, Prophet is more simple and easy to use. Since it encapsulates Python and R language interfaces, we can easily build a Python predictive analysis environment for time series analysis [4].

The basic form of the time series model is:

\[ y(t) = g(t) + s(t) + h(t) + \epsilon_t \]  \hspace{1cm} (1)

The model divides the time series into the summation of three parts, where the growth function of \( g(t) \) is represented to fit the aperiodic changes. \( s(t) \) is used to represent periodic changes, such as weekly, monthly, etc., \( h(t) \) represents changes caused by special reasons such as holidays, festivals etc., and finally \( \epsilon_t \) is the noise term, which is explained as:

\[ g(t) = \frac{c}{1 + e^{-k(t-t_0)}} \]  \hspace{1cm} (2)

\( g(t) \) is a logical function, where \( c \) represents the capacity of the model, \( k \) represents the growth rate, and \( b \) represents the offset.

\( s(t) \) is a periodic term, which use Fourier series to approximate the periodic component. It can be expressed as follows:

\[ s(t) = \sum_{n=1}^{2n} (a_n \cos \left( \frac{2\pi n}{T} t \right) + b_n \sin \left( \frac{2\pi n}{T} t \right)) \]  \hspace{1cm} (3)

where \( T \) is the period, and \( 2n \) is the number of cycles expected to be used in the model. \( h(t) \) is the way to handle a holiday is to set a dummy variable for the same holiday in the past and future. The model is expressed as follows:

\[ h(t) = \sum_{i=1}^{K} K_i \mathbb{1}(t \in D_i) \]  \hspace{1cm} (4)

\[ Z(t) = [I(t \in D_H) \land I(t \in D_i)] \]  \hspace{1cm} (5)

\[ h(t) = Z(t) \kappa, \kappa \sim \text{Normal}(0, \nu) \]  \hspace{1cm} (6)

where \( K_i \) represents the influence of holidays in the window period on the predicted value, \( D_i \) represents the fourth dummy variable. If the time variable \( t \) belongs to the dummy variable, the value of the dummy variable is 1; otherwise, it is 0. \( i \) represents the holiday, \( D_i \) represents the time \( t \) included in the window, and \( \epsilon(t) \) represents the error term.

2.2. LSTM Neural Network Model

Long Short-Term Memory Network (LSTM) [3] is a Recurrent Neural Network (RNN) specific pattern that studies long-term dependence. LSTM adopt subtle gate control to combine short-term memory and long-term memory, solved the problem which RNN cannot deal with the issue of long-distance dependence. It can keep the error along the time in the other direction, and layer can keep the error in the more constant level. Let the recursive network learning can be more than one-time step, so as to establish causal relationship over a long distance. LSTM differs from RNN mainly in that it adds a "processor" to the algorithm to determine whether the information is useful or not. This processor acts on a structure called a cell. Three gates are placed in a cell, namely the input
gate layer, the forget gate layer and the output gate layer. A message enters the LSTM network and can be judged according to the rules whether it is useful or not. Only the information that conforms to the algorithm authentication will be left, and the information that does not conform will be forgotten through the forgetting gate. This principle can solve the big problem of neural network by repeated operation. At present, it has been proved that LSTM is an effective technology to solve the long-term dependence problem of neural network, and this technology has been widely used because of its high universality. Its structure is shown in figure 1.

![Figure 1. Internal structure of LSTM network neurons.](image)

In the figure 1, the forget gate determines how much of the unit state of the last moment \((C_{t-1})\) is retained to the current moment \((C_t)\), which is denoted as \(f_t\); Input gate determines how much network input \((X_t)\) is saved to the cell state \((C_t)\) at the current time. Output gate controls how many outputs of the unit state are sent to the current output value \((h_t)\) of LSTM, denoted as \(O_t\).

Cell is a memory unit which represents the memory of the state of the neuron, so that the LSTM unit has the ability to save, read, reset and update the long-distance historical information, which is denoted as \(C_t[5]\).

The specific formula of LSTM is expressed as follows:

\[
f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (7)
\]

\[
i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)
\]

\[
o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)
\]

\[
c_t^\prime = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)
\]

\[
c_t = f_t \cdot c_{t-1} + i_t \cdot c_t^\prime \quad (11)
\]

\[
h_t = o_t \cdot \tanh(c_t) \quad (12)
\]

Here, the recursive connection weight represents its corresponding threshold, \textit{sigmoid} function and \textit{tanh} function are two activation functions respectively.
3. The Experimental Result

3.1. Prophet Model Construction

The data needed by Prophet includes two columns, \( ds \) and \( y \). The \( ds \) column needs to contain the date or specific time point, while the \( y \) column represents the sales volume we hope to predict. This paper first preprocesses the input data \( y \), where the sales data are mapped to the interval of \((0,1)\) by using deviation standardization, and the main holidays in China are set as shown in the Table 1 below.

| Major holidays in China | 2017-1-1 | 2018-1-1 | 2019-1-1 | 2020-1-1 |
|------------------------|----------|----------|----------|----------|
| New Year's Day         |          |          |          |          |
| Spring Festival        | 2017-1-28| 2018-2-16| 2019-2-5 | 2020-1-25|
| Qingming Festival      | 2017-4-5 | 2018-4-5 | 2019-4-5 | 2020-4-5 |
| International Workers' Day | 2017-5-1 | 2018-5-1 | 2019-5-1 | 2020-5-1 |
| Dragon Boat Festival   | 2017-5-30| 2018-6-18| 2019-6-7 | 2020-6-25|
| Mid-Autumn Festival    | 2017-10-4| 2018-9-24| 2019-9-13| 2020-10-1|

Initialize Prophet model and set \( \text{changepoint}=0.15 \) according to experience to make the growth trend more sensitive to change [5]. Specify a forecast interval of one year in the future. Finally, the following analysis results come out.

![Figure 2. Time series trend chart.](image)

We break down the trend in the figure 2 to obtain a separate analysis result. The picture from top to bottom shows the growth trend, holidays, weekly trend and annual trend of sales data broken down by the model. As can be seen from the figure 3 the sales volume of enterprise S is gradually increasing, and the sales volume reaches its peak on weekends.
The Prophet model includes the time series cross-validation function, which is used to measure the prediction error using historical data. This is accomplished by selecting the cut-off point in the historical record, and finally the predicted value is compared with the actual value. The cross-validation process can use the cross validation function to automatically complete the history truncation, specify the horizon, and then select the initial training period (initial) size and the interval between truncations. In this paper, initial = "900 days", period = "55 days" and horizon = "110 days" were selected for verification. RMSE is used to measure the predictive effect of the prophet model. The final effect is shown in the figure 4 below, and the final RMSE is 0.05.
3.2. LSTM Model Construction

The LSTM model is constructed as follows. The original data is transformed linearly by using deviation standardization, and the values of the original data are mapped between [0,1]. The resulting image of the conversion result is as follows figure 5:

![Figure 5. Adopt the sales volume chart after deviation standardization.](image)

(1) Divide the training set and the test set. In this paper, 80% of the original data is used as the training data, and the remaining 20% is used as the test data.

(2) LSTM neural network model was initialized, random seed was set to be 7, and LSTM model was established: the input layer had 1 input, the hidden layer had 40 neurons, and the output layer was to predict a value. The activation function was sigmoid, iterated for 50 times, and the batch size was 80.

(3) Make the prediction and convert the predicted data into the same unit before error analysis.

(4) In this paper, RMSE is used for error analysis, and the final figure 6 is as follows. Finally, the rmse of the test set is 0.0472.

![Figure 6. Training result diagram.](image)
3.3. Experimental Error Comparison
According to Table 2, the error of LSTM is smaller than Prophet.

Table 2. Experimental error comparison.

|        | RMSE  |
|--------|-------|
| Prophet| 0.050 |
| LSTM   | 0.047 |

4. Conclusion
In this paper, based on the real sales data provided by S enterprise, A Prophet model and a deep learning model LSTM are respectively used to predict and evaluate the sales volume. The results show that the accuracy of LSTM experiment is higher than that of Prophet model, which proves the superiority of LSTM model. In the future, more models can be used for comparison experiments, and factors affecting drug sales can be further analyzed, so as to obtain more accurate results. This is also the research direction of the follow-up work in this paper.

Acknowledgments
This work is supported by Key R & D project of Shandong Province (2019GSF108222), the Project of Shandong Province Higher Educational Science and Technology Program (Project No. J15LN56).

References
[1] Ma Y L 2017 Study on Inventory Management of Z Pharmaceutical Company (Beijing: Capital University of Economics and Business)
[2] Taylor S J and Lethem B 2018 Forecasting at Scale the American Statistician 72 (1) 37-45
[3] Greffk R, Srivastava K, Koutnik J 2015 LSTM: A search space odyssey IEEE Transactions on Neural Net-works & Learning Systems 28 (10) 2222-2232
[4] Liu P 2018 Research On Sales Forecasting Method Based On Time Series Analysis (Dalian: Dalian University of Technology)
[5] Ge N, Sun L Y, Shi X D and Zhao P 2019 Research on sales forecasting of prophet-LSTM combination model Computer Science 46 (S1) 446-451