Research on sales Forecast based on XGBoost-LSTM algorithm Model

He Wei¹, QingTao Zeng²*

¹Beijing Institute of Graphic Communication, BeiJing, China
²Beijing Institute of Graphic Communication, BeiJing, China
*Corresponding author’s e-mail: zengqingtao@bigc.edu.cn

Abstract. Reasonable sales forecast is very important for enterprises. The short-term and long-term sales changes of a product are helpful for enterprises to make marketing strategies and sales decisions. On the basis of in-depth analysis of the characteristics of a certain algorithm model and long and short memory neural network, and according to the data set provided by a supermarket chain in kaggle competition, a XGBoost-LSTM neural network combination model for sales forecasting and a classical time series prediction model are constructed to compare the experimental results. The experimental results show that the XGBoost-LSTM neural network prediction model has higher accuracy than the time series prediction model, which can provide an important scientific basis for the supermarket chain to make sales forecast.

1. Introduction
Business forecast refers to the prediction of the future sales volume of the enterprise according to the past sales information of the enterprise. So that enterprises can better grasp the market demand, so as to formulate a better marketing strategy.

With the rise of artificial intelligence technology, the enterprise market competition is becoming more and more fierce. In the operation of many enterprises, the rational use of enterprise sales data to predict will greatly increase the core competitiveness of enterprises. At present, the most commonly used forecasting method of sales forecasting is time series modeling. By establishing an appropriate model, the traditional time series forecasting relies on too simple data, easy to fit the data, and often faces the problem of lag. XGBoost and LSTM models have their own advantages. In practical application, the single algorithm model of sales forecast can not carry out time prediction and analysis very well. Therefore, when in-depth analysis of the characteristics of time series data, many scholars will use different methods to integrate several models to better optimize the prediction effect of the model. Among them, the prediction and comprehensive analysis model based on autoregressive moving average model, support vector regression model, Adaboost and improved particle swarm optimization algorithm and BP neural network have achieved good results in different fields.

This paper works on the sales time series data of an enterprise, and puts forward a forecasting method of XGBoost-LSTM combination model based on weighted set. The sales time series data of the enterprise in the past two years are modeled by using XGBoost and LSTM neural network models,

¹ Beijing science and technology innovation service capability construction project (PXM2016_014223_000025)
respectively. The proposed XGBoost-LSTM combination forecasting model is better than the single forecasting model in the time series forecasting of sales volume, and has better forecasting ability.

2. Research theories and methods

2.1. LSTM neural network model

Long Short Term Memory [1] (LSTM) is a common cyclic neural network. In LSTM, the main purpose is to solve the problem of gradient disappearance and gradient explosion in the process of long sequence training (see fig 1).

Fig.1 LSTM input and output

LSTM still has an update door \( \Gamma_u = \sigma ( W_u [a^{<t-1b>}, x^{<t>}] + b_u) \). A new feature of LSTM is that there is not only one update door to control, but also a new door called forgetting door, To represent with \( \Gamma_f \), \( \Gamma_f = \sigma ( W_f [a^{<t-1b>}, x^{<t>}] + b_f) \). Then a new output door was added, \( \Gamma_o = \sigma ( W_o [a^{<t-1b>}, x^{<t>}] + b_o) \). The updated value of memory cells has been changed to \( c^{<t>} = \Gamma_u \cdot c^{<t-1>} + \Gamma_f \cdot c^{<t-1>} + b_c \). The complete formula of LSTM is as follows,

\[
\begin{align*}
\Gamma_u &= \sigma ( W_u [a^{<t-1b>}, x^{<t>}] + b_u) \\
\Gamma_f &= \sigma ( W_f [a^{<t-1b>}, x^{<t>}] + b_f) \\
\Gamma_o &= \sigma ( W_o [a^{<t-1b>}, x^{<t>}] + b_o) \\
c^{<t>} &= \Gamma_u \cdot c^{<t-1>} + \Gamma_f \cdot c^{<t-1>} + b_c \\
a^{<t>} &= \Gamma_o \cdot c^{<t>}
\end{align*}
\]

Fig.2 Hidden layer cell structure diagram
The related components are described as follows:
Input Gate: Controls whether information flows into Memory cell, marked as it.
Forget Gate: Controls whether the information in the previous time Memory cell is accumulated into the current time Memory cell, recorded as ft.
Output Gate: Controls whether the information in the current time Memory cell flows into the current hidden state ht, recorded as Ot.
Cell: The memory unit, which represents the memory of the state of the neuron, makes the LSTM unit have the ability to save, read, reset and update long-distance historical information, recorded as ct.

In the training process of LSTM neural network, firstly, the data characteristics of t time are input to the input layer, and the results are output through the excitation function (see fig 3).

![Fig.3 Inside structure of LSTM](image)

### 2.2. XGBoost model
Gradient lifting is the model optimization idea of Boosting algorithms in ensemble learning. Ensemble learning is one of the hottest research fields in machine learning. Its basic idea is to combine many weak learners to form a powerful model that can predict accurately. Ensemble learning is not only a simple superposition of multiple classifiers, but also realizes the optimal combination of weak learners by training the aggregated model, which is much more accurate than the results predicted by a single model.

Extreme gradient lifting tree [5] (XGBoost) is an integrated learning algorithm, it is a tool for massively parallel integrated decision tree, and it is the fastest and best integrated decision tree algorithm at present. It is a joint decision made by multiple associated decision trees, that is, the input sample of the next decision tree will be related to the training and prediction results of the previous decision tree. At the beginning of the model training, the number of decision trees is 0, and with the training iteration, the decision tree is added, that is, the decision function is added. As one of the ensemble learning methods, the XGBoost prediction model can be expressed as:

\[
\hat{y} = \sum_{k=1}^{k} f_k(x_i)
\]

Where k is the total number of trees; represents the k-th tree; represents the prediction result of the sample.

The objective function can be expressed as:

\[
Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k)
\]

\[l(y_i, \hat{y}_i)\] Training error of sample, \[\sum_{k=1}^{k} \Omega(f_k)\] Represents the regular term of the k-th tree to prevent over-fitting of the model.
3. Construction and Forecast of XGBoost and LSTM Model

3.1. Data preprocessing
This paper selects the daily sales data of a supermarket chain in kaggle competition in Beijing for three consecutive years. The brand has a total of 78 retail stores in Beijing, and 78 groups of 1096 sales time series are obtained. One set of sales data and data format are listed in Table 1. According to the isometric series sorted from January 1, 2017 to December 31, 2019, the data in the following table can be used as random time series for different stores.

|       | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | ... |
|-------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|-----|
| 2017  | 4 | 6 | 8 | 7 | 3 | 4 | 6 | 5 | 7 | 3  | 10 | 5  | 9  | 6  | ... |
| 2018  | 13| 11| 10| 7 | 3 | 4 | 8 | 5 | 11| 3  | 10 | 3  | 6  | ... |
| 2019  | 10| 14| 4 | 6 | 3 | 7 | 14| 6 | 3 | 11 | 8  | 4  | 7  | ... |

In order to more intuitively observe the data characteristics of product sales in the supermarket, the sales data show a relatively peaceful trend, volatility is not very large, and there is randomness, there are large fluctuations in some parts (see fig 4).

3.2. XGBoost and LSTM combined forecasting model
In order to make full use of the advantages of XGBoost model and LSTM neural network model, an optimal combination forecasting model based on XGBoost and LSTM neural network is proposed for time series prediction (see fig 5).
4. Experiments and Conclusion

4.1. Experiments
In this paper, in the commodity sales of supermarket chains from 2017 to 2019, 78 groups of 1096 sales time series were selected to construct the data set, and 70% positive samples and 70% negative samples were randomly selected to form the training set. the remaining 30% positive samples and 30% negative samples constitute the test set, and 10% cross-validation is used to find the optimal parameters of the model. The training results obtained from the LSTM model are re-added to the new data set as a new feature, and the data set is used to train the XGBoost model(see fig 6).

---

**Fig. 5 Model flow chart**

**Fig. 6 Sales per month**
4.2. Conclusion
This paper analyzes the rules and characteristics of sales data, and the performance of using XGBoost-LSTM model to deal with time series data is much higher than that of the original XGBoost single model, which maximizes the advantages of the two prediction models, and can help enterprises to formulate marketing strategies, which is of great commercial value to enterprises. In addition, the parameters of LSTM neural network can be optimized to find the best prediction model.

Acknowledgments
(1) Major Science and Technology Projects of Guangdong Province in 2019, No. 190826175545233
(2) Beijing science and technology innovation service capability construction project (PXM2016_014223_000025)
(3) BIGC Project (Ec202007)

References
[1] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: A Search Space Odyssey [J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 28 (10): 2222-2232.
[2] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization [J]. Computer Science, 2014.
[3] Li Y, Yang L, Yang B, et al. Application of interpretable machine learning models for the intelligent decision [J]. Neurocomputing (S0925-2312), 2018.
[4] Tadesse M M, Lin H, Xu B, et al. Personality predictions based on user behavior on the Facebook social media platform [J]. IEEE Access (S2169-3536), 2018: 1-1.
[5] Pang L, Wang J, Zhao L, et al. A novel protein subcellular localization method with CNN-XGBoost model for Alzheimer’s disease [J]. Frontiers in Genetics (S1664-8021), 2019, 9: 751.