Multiple Traditional and Deep Learning-based Prediction Models for Quantity of Domestic Mail

Ruochen Xing *

Software Engineering, Harbin Institute of Technology (Weihai), Weihai, China

* Corresponding author email: 2201110306@stu.hit.edu.cn

Abstract. As the main role of domestic mail is gradually changing from the field of communication to the field of economy and culture, it is very important to provide relatively accurate volume prediction and feedback through the visualization process. This paper focuses on finding models with a high matching ability to achieve the purpose of prediction, and selects the best option by comparing multiple models. Eight different models ranging from simple to complex will be used in the following steps. More specifically, the predictive power of the models has steadily improved, from linear to curved, from constant to trend, from a single dimension to adding gradient factors to combining seasonal periods. The prediction principle is also gradually complicated. In the beginning, it focuses on the average, then weights the values in different stages to obtain the weighted average, and then makes a more refined prediction by means of single-layer convolution. To achieve the purpose of obtaining better results, it is of great necessity to introduce a self-learning prediction model, so Recurrent Neural Network (RNN) model and Long Short-term Memory (LSTM) model are chosen. In addition to obtaining the input data of this time, the RNN model will also combine the input data and output data of previous times to make the comprehensive prediction. The LSTM model further solves the situation that the RNN model cannot predict in some cases, by setting the limit to determine whether to keep the data better to fit the data with the trend. The data visualization results show that the performance of LSTM model is the most outstanding among all models, which can provide learners with more accurate prediction information.

Keywords: Domestic Letters; Data Prediction; Constant Prediction; Linear Prediction; RNN Model; LSTM Model.

1. Introduction

With the rapid development of the Internet, electronic communication has shown its convenient and efficient advantages in many fields, which makes paper media gradually withdraw from today's mainstream market. Sending letters is a completely antiquated form of communication, and few people would choose to write letters to connect with others without the emotional bond of a particular personal reason. According to the Statistical Report on the Development of the Internet in China released by the China Internet Network Information Center this year, by June 2022, the number of Chinese netizens reached 1.051 billion, and the Internet penetration rate reached 74.4% [1]. However, the paper media has not lost all space for existence. The meaning of sending letters in daily life is steadily diminishing, while it plays a more and more vital role in states like politics and business. Nowadays, many companies and individuals still choose to send important documents or agreements, including acceptance letters or contracts, through paper mail. Based on that, exploring the market trend of paper letter sending in recent years can provide valuable reference for merchants or individuals in need. At the same time, the development of data forecasting is in full swing, so the combination of the two is a valuable way to practice.

Data prediction has been talked about for a long time, which has provided a strong auxiliary role in the research of basic disciplines such as physics and chemistry, as well as urban planning, biochemical and other fields [2-4]. Nowadays, the important role of data prediction in the field of communication has been attracted a lot attention. The development of different branches of communication technology all need to make certain predictions on the basis of large amounts of data to gradually revise the research and development direction. As machine learning capabilities continue to improve, data prediction and machine learning are becoming more closely integrated and have

already provided significant assistance in related research. For example, in the field of telephone communications, data prediction is used to fit and solve the important issue of noisy instance in mobile phone data [5]. Similarly, in terms of network traffic, such kind of forecast is necessary in network protocol designing, network management and high-performance routers designing etc.

It can be seen that although data prediction has been widely used in the field of communication field, the research direction is almost emerging communication channels, while there are few researches on traditional letter quantity prediction, which leaves a gap in the process of letters transforming their own properties. In addition, in previous predictions, researchers chose ARIMA or FARIMA models for data fitting. For instance, Jiang M.et al used such model to fit the traffic prediction with the time scale [6]. Practice has proved that these two models can indeed reflect the timing changes of data to a certain extent, but the errors are still large. The author uses Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models, hoping to improve the fitting effect of the models.

In order to produce high quality forecasts, this study partnered with a postal station to obtain daily domestic letters volume data and then used a variety of methods to analyze and fit the same data set. After comparison, the advantages of LSTM model are obvious. Four kinds of models are mainly used as follows: (1) the first part is a constant prediction method with different average values as predictive values; (2) the second part is a linear prediction with trend added on the basis of the constant; (3) the third part adds periodic elements to the prediction model; (4) the last part is a highly fitting model represented by two deep-learning models. In the whole process, the complexity of the model increases step by step, and the simple model provides the theoretical basis for the use of the complex model, and finally achieves the result of almost complete fitting between the predicted value and the real value.

2. Methodology

2.1 Dataset Description and Preprocessing

The cooperative company provided the domestic letter quantity of each day for three and nine months from January 1, 2019 to September 30, 2022. According to the number changes, it can be intuitively observed that the data volume has similar fluctuations in every week, and there is a large gap between the peak value and the valley value. Therefore, this work first judged the time stamp of the data set, analyzed the day of the week of each date, divided the data set into seven groups based on this key word. After dividing each group data into two parts: training set and prediction set, the prediction analysis is carried out according to the group in the following research.

In this study, several prediction methods e.g., constant prediction, holt linear trend model and arima model are employed. More details of them are described as below.

2.2 Constant Prediction

In the prediction method where the predicted value is constant, the key is to find the average value of the different kinds. The naive prediction method focuses on the separation point between the training set and the prediction set and takes the average value of the last point of the training set and the first point of the prediction set as the prediction value. Moving Average Forecast sets an interval and takes the average of all the data within that interval as the predicted value applied to the forecast. Simple exponential averaging method is optimized on the basis of simple average method, in which more recent observations will be given more weight than those in earlier periods, while those farther back in time will have less weight. It calculates the predicted value from a weighted average, where the weight decreases exponentially as the observation changes from early to late.

2.3 Holt Linear Trend Model

Holt linear trend model pays much attention to the trend of the data set [7]. Although each of these methods can apply a trend, in this module, besides retaining the influence parameters in the horizontal

direction, the smoothing coefficient about the trend is increased, and the trend amplitude is adjusted while the horizontal direction variables are adjusted, so as to obtain the linear prediction results with the trend.

2.4 Holt-Winters Model

On the basis of the principle of the Holt model, Holt-Winters method introduces the seasonal item to the former model, which can be used to process the fluctuation behavior of fixed cycles in time series such as monthly data with cycle of 12, quarterly data with cycle of 4 and weekly data with cycle of 7. The model uses a similar calculation method to correct the data from three dimensions.

2.5 Arima Model

Arima model first converts non-stationary time series into stationary time series [8], and then regression is carried out according to the lag value and the random error term. Starting from the correlation of the time series value itself, moving average technology, correlation analysis technology and stationary technology (difference) are included in the model. To establish the regression equation between the time series values, the autocorrelation function and partial autocorrelation function are used to determine the two parameters, and the purpose of difference is achieved through layer-by-layer iteration, forming the establishment of the whole model.

2.6 RNN Model

The RNN is a class of networks that use sequence data as input [9], makes recursion in the progression direction of a sequence and connect in chain form. In the RNN model, each input value is passed into the hidden layer, and the value of the hidden layer is obtained by distributing the weight with the incoming data of the previous hidden layer, and then the output value is obtained by adding the weight. Finally, the output value is combined with the error, and the predicted value is formed after transformation by activation function. The whole process is as follows:

$$h_t = \emptyset(Ux_t + Wh_{t-1} + b) \quad (1)$$

$$o_t = Vh_t + c \quad (2)$$

$$y(t) = \sigma(o(t)) \quad (3)$$

x, o, L, b and y represent input, output, the loss function, biased and the label of the training set respectively. h is the hidden layer element. V, W and U are weights, and weights of the same type are connected with the same weight. ϕ () and σ () are the activation function, and the tanh function is the most likely function that can be chosen.

Appropriate step size should be selected for data set selection. If the step size is too short, the fitting effect cannot be good; if the step size is too long, the fitting efficiency will be greatly affected. After the step size is determined, the data set is divided, and the fitting is carried out according to the above formula based on the number of iterations and batch size to obtain the final result.

However, RNN model cannot handle long sequence problem, because RNN cycle multiplication results in gradient disappearing or gradient explosion, which can hardly be trained.

2.7 LSTM Model

LSTM is a supplement to the RNN model, mainly to solve the up two problems [10]. Simply conclusion, LSTM can perform better in longer sequences than ordinary RNN.

Whereas RNN has only one transmission state, LSTM has two transmission states, one cell state and one hidden state. At the same time, in the LSTM network, three gates are introduced to control the transmission of information, namely, the forgetting gate, the input gate and the output gate. The forgetting gate determines what information needs to be discarded by a real number between 0 and 1 indicating the degree of retention output by sigmoid function, where 0 means completely discarded and 1 means completely retained. The tanh function plays a vital role in the input gate to generate alternatives to update what needs to be updated. Finally, at the output gate deciding what value to

output, the two results gotten from the former two functions are multiplied to get the part we determined to output.

3. Result and Discussion

3.1 The Result of Different Models Used to Predict

Through the fitting of the above methods, a total of eight groups of different prediction figures were obtained. For the convenience of presentation, the views of Monday were taken as examples in each group of images. The first four groups are linear prediction, among which the first three chapters are constant prediction, and the last one is corresponding to the Holt model, which clearly reflects the trend characteristics. The last four are curve predictions, and the fitting curve is gradually closer to the original data set. Meanwhile, in the first six prediction images, the data collected from the training set is represented by the blue curve, the data collected from the valid set is represented by the orange curve and the green line is the prediction curve. In the images corresponding to the RNN and LSTM models, the blue curve is the data set, and the orange curve is the prediction curve. In the whole process, the epoch is set to 40 times, and the batch-size is 1. Finally, the error value of the RNN model is stable at about 0.007, and that of the LSTM model is stable at about 0.003.

Clearly, from constant predictions to LSTM models, each model is more complex and applied to the data set than the previous one. Constant prediction provides a horizontal line on which future data will float. To make the dividing line more accurate, tendency was added to the model, resulting in a linear model, and then periodicity was also added. ARIMA model has been preliminarily equipped with prediction ability, but limited by its own model, it can only predict the stable model or the model with stability after differentiation and cannot deal with the data with volatility. However, the letter delivery volume is affected by related activities, post office policies, protocol replacement and other reasons, resulting in the data itself is not stable, so the ARIMA model cannot be well fitted. Both RNN and LSTM models have the ability to learn and self-optimize based on known data. The LSTM model also solves the unpredictable phenomenon that may be generated by the RNN model during the prediction, so the LSTM model is the most suitable model for data fitting and short-term prediction.

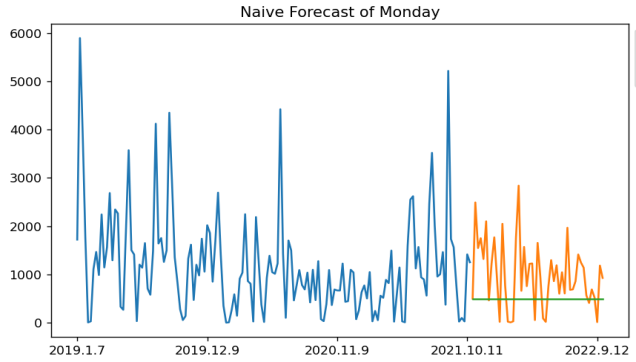


Fig 1. The naive forecast of Monday

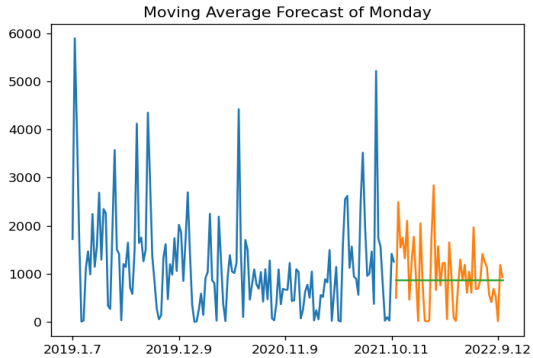


Fig 2. The moving average forecast of Monday.

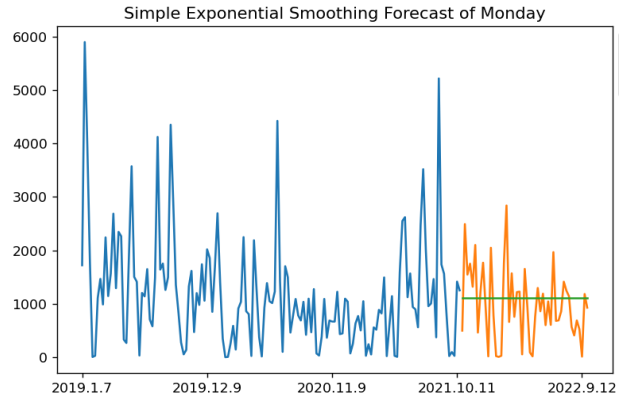


Fig 3. The simple exponential smoothing forecast of Monday.

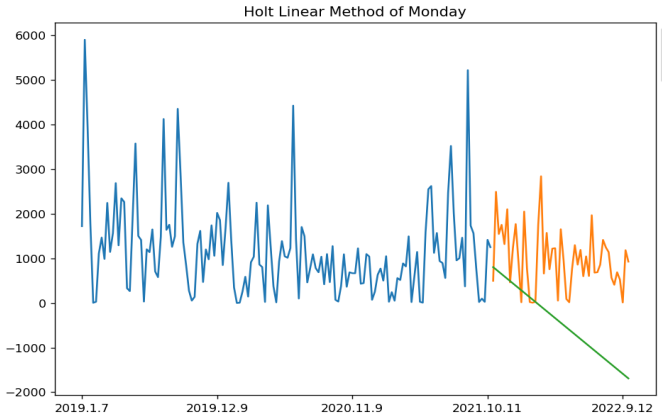


Fig 4. The Holt linear forecast of Monday.

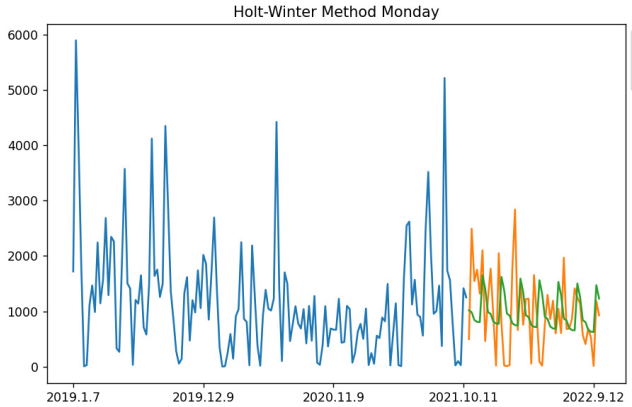


Fig 5. The Holt-Winter forecast of Monday

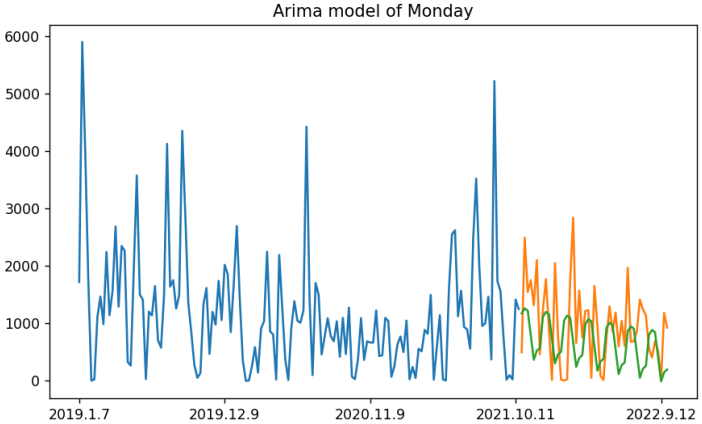


Fig 6. The Arima model forecast of Monday

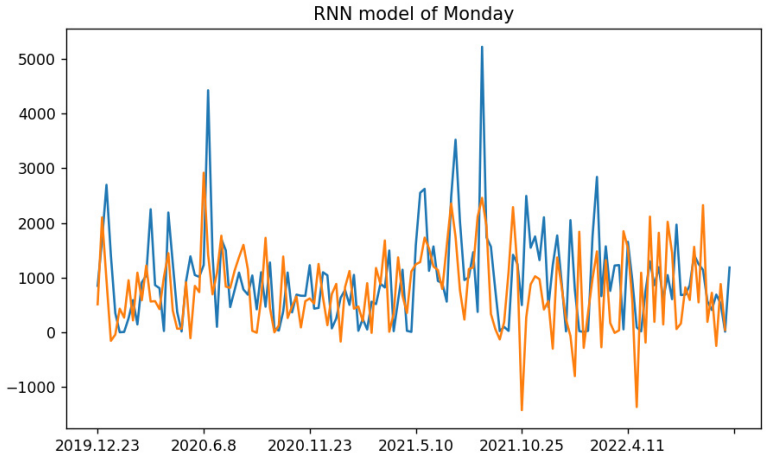


Fig 7. The RNN model of Monday

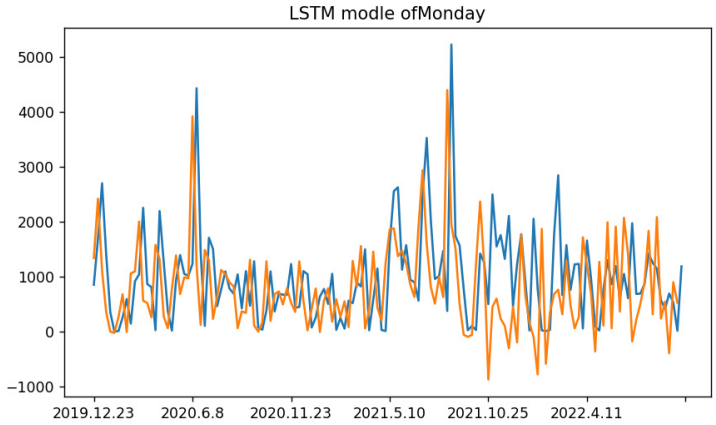


Fig 8. The LSTM model of Monday (x label is the date, y label is the quantity)

From Fig. 1 to Fig. 8, the horizontal and vertical coordinates of each chart correspond one-to-one with dates and quantities of the domestic letters, forming the x label and y label respectively, and represent different data sets according to the color distinctions described in the previous article

4. Conclusion

After obtaining the number of letters delivered in China, this paper starts from the data set itself and searches for the most suitable model for fitting prediction after simple processing. Starting from the naive prediction method, the study gradually complicates the fitting model, from the single dimension processing of the numbers to the two-dimension processing of adding linear trend, to the three-dimension processing of adding period cycle, and finally to the processing model of increasing self-learning and self-optimization ability. A total of eight different fitting schemes have been proposed and practiced. After all the results were obtained, this work made a comparative summary and proved that the LSTM model was highly applicable to the fitting of relevant data from the two aspects of the degree of image fitting and the amount of error. In the future, the research plans to study the variation trend of quantity change within a week, a quarter or even a year, and find out whether midweek, weekend, or different times at the beginning and end of the year also have a general influence on quantity, so as to provide better reference opinions and value.

References

[1] CNNIC. CNNIC releases the 50th Statistical Report on the Development of Internet in Chinahttp://www.cnnic.cn/n4/2022/0916/c38-10594.html, 2022.

[2] Piadeh F, Kourosh B, and Amir A. A critical review of real-time modelling of flood forecasting in urban drainage systems. Journal of Hydrology, 2022, 127476.

[3] Yu Q, Wang J et al. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 2022, 103323.

[4] Wu C, et al. Revisiting alloy design of low-modulus biomedical β-Ti alloys using an artificial neural network. Materialia 21, 2022, 101313.

[5] Sarker I H, Kabir M A, Colman A, et al. An effective call prediction model based on noisy mobile phone data. the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2017 ACM International Symposium on Wearable Computers. ACM, 2017.

[6] Jiang M, Wu C M, Zhang M, et al. Research on the Comparison of Time Series Models for Network Traffic Prediction. Acta Electronica Sinica, 2009, 37(11):2353-2358.

[7] Bujang M A, Tassha H, Adnan P, et al. Prediction number of deaths by occurrence in Malaysia: A comparison between simple linear regression model and holt's linear trend model, 2009.

[8] Contreras J, Espinola R, Nogales F J, et al. ARIMA Models to Predict Next-Day Electricity Prices. IEEE Transactions on Power Systems, 2003, 18(3):1014-1020.

[9] Lukosevicius M, Jaeger H. Reservoir computing approaches to recurrent neural network training. Computer Science Review, 2009, 3(3):127-149.

[10] Dyer C, Ballesteros M, Ling W, et al. Transition-Based Dependency Parsing with Stack Long Short-Term Memory. Computer Science, 2015, 37(2):321–332.