Forecasting time series using neural networks on the example of primary sales of a pharmaceutical company

Dmitrii Galkin, *Tatiana* Dudkina, and *Natalia* Mamedova [*]

Plekhanov Russian University of Economics, Basic Department of Digital Economy, 117997 Moscow, Stremyanny lane, 36, Russia

Abstract. This article is devoted to neural network forecasting of primary sales of a pharmaceutical company. The research includes an analysis of the main forecasting methods, various types of neural network architectures used to predict time series, as well as an analysis of neural network training methods. The primary sales of the pharmaceutical company AKRIKHIN JSC were considered as the applied basis of the study. As part of the study, an analysis of the revenue of the pharmaceutical company AKRIKHIN JSC from 2016 to 2021 was carried out, the selection of the main hyperparameters of the neural network was carried out and a forecast of primary sales for 4 weeks ahead was obtained. The practical significance of the research results is due to the fact that forecasting primary sales gives management an understanding of which direction the company is moving in and whether additional management decisions are needed to meet the planned values. This study can be used as an alternative method of sales forecasting, so the results will be in demand among analysts of pharmaceutical companies in sales planning departments.

1 Introduction

Forecasting is one of the most popular sales management tools today. It allows you to gain an understanding of future sales for inventory management, which makes it possible to relocate the plant's capacity, plan expenses for the future period, increase the manageability of the business process, reduce external risks. This study was conducted using the neural network modeling method. As an example, data on the activities of the pharmaceutical company AKRIKHIN JSC are used.

AKRIKHIN JSC is one of the leading Russian pharmaceutical companies. The company's product portfolio includes more than 200 drugs. Almost half of them are included in essential drug list [1]. The company's development strategy focuses on the development and release of the first generics, new combinations of molecules, doses and release forms.

We formulated the research request as the development of a solution for the selection of parameters and metrics that allow predicting primary sales in the selected subject area with the greatest accuracy using the example of the data of AKRIKHIN JSC. The received forecast

[*] Corresponding author: nmamedova@bk.ru

is intended to determine and timely take measures to reduce the consequences of risks in the company's activities. A correctly interpreted forecast will help the management to make the right decisions, the results of which will be to maintain sales dynamics at the same level and maintain the strategic guidelines for the company's development.

The specificity of decision-making by companies operating in the pharmaceutical market lies in the existence of different states of demand and heterogeneity of the assortment of goods. The demand for medicines is highly variable due to seasonality, as well as the tendency to purchase drugs "for the future" among consumers.

Considering the specifics of the selected market, among the many existing forecasting methods, we selected methods based on neural network training. Among the many well - known neural networks, the following options were subject to testing:
- Single-layer and multi-layer perceptron.
- Recurrent networks.
- LSTM networkss.
- Convolutional neural network.

2 Methodology of research

Many direct sales indicators are presented in the form of a discrete time process – a time series. Time series forecasting in general is a complex type of forecasting task. Therefore, among the possible options, we chose a special type of neural networks – LSTM (Long short-term memory) networks because of their ability to process large sequences of values. The rationale for choosing the type of network to train the model is given below.

It is known that all recurrent neural networks (RNNs) are built according to the type of chain of repeating modules (Fig. 1).

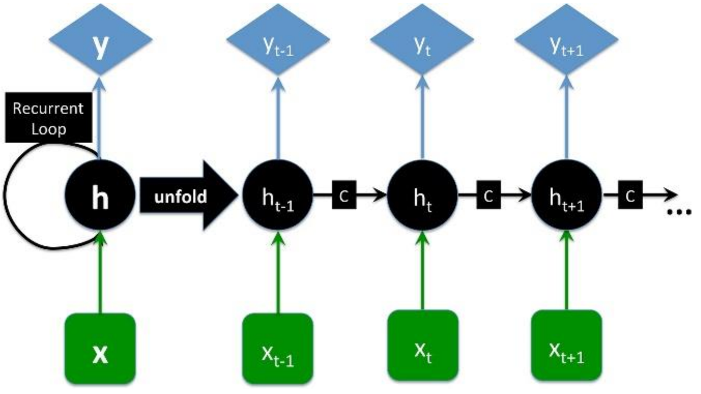

Fig. 1. Scheme of a recurrent neural network with one hidden layer [2]

The structure of the standard RNS module includes one neural layer of hyperbolic tangent. In turn, LSTM networks, as a special type of RNS, have a similar structure, but at the same time four layers interact with each other. This makes it possible to use LTSM networks for learning long-term dependencies [3]. The data moves between the cells (cells) of the chain with small linear transformations, and in the later training cycles without any changes at all. The LSTM network is controlled by recurrent gates, which are called "forgetting" gates. The state of the cell (cell) is indicated by a horizontal line running along the upper part of the circuit. LSTM networks can delete or add information to the cellular state. At the same time, the entire process is controlled by valves that selectively pass data.

There are also other variations of LSTM networks [4]. In one of the variations, "eye connections" are added, which means that the valves can "spy" on the cellular state. The other uses paired forgetters and input gates, which allows you to simultaneously make decisions about what to forget and where to add new information.

A likely option was to use the gate mechanism for recurrent neural networks (Gated Recurrent Unit, GRU). In this mechanism, both the inlet valve and the forgetting valve are combined in one updating valve. As a consequence, the cellular state merges with the hidden layer [5]. The resulting model is simpler than the usual LSTM model, however, when predicting time series, GRU networks predict the trend without highlighting the features of the series. And this limitation does not allow us to use GRU when building our model.

Summarizing all of the above, we can conclude that the LSTM network is the optimal type of neural networks for solving the task.

After choosing the type of neural network to predict, the network was trained. The main stages of the network training process for forecasting are shown in Figure 2.

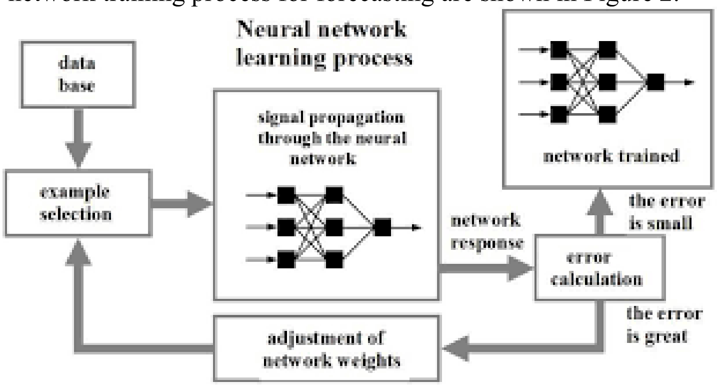

Fig. 2. The structural diagram of the learning process of a neural network [6].

There are several types of neural network training, among which it was necessary to choose the most appropriate type for the chosen subject area. The types of training are presented below:

- with a teacher;
- without a teacher;

When deciding on the type of network training, the main preference was given to such a characteristic as the presence of a test sample, since the existing large time series of direct sales values allowed using not only a training sample. In the case of teaching with a teacher, each input data vector (training sample) has a target vector (test sample) representing the required output. Together they form a training pair. In the learning process, there is a purposeful modification of the synaptic connections (weights) of the neural network, in order to achieve the greatest correspondence between the output of the network with the target vector. Teaching without a teacher does not imply the presence of a test sample. The neural network generates output signals only on the basis of a training sample, which is the most plausible learning model in a biological system.

The error back propagation algorithm is one of the most common algorithms for training multilayer neural networks with a teacher [7]. This is an iterative gradient learning algorithm that uses the gradient method to find the minimum of the error function. In this case, error signals are recorded in the direction from the output of the neural network to the input, that is, in the opposite direction to the direct propagation of signals.

In general, the algorithm does not allow to achieve the global minimum of the error function. This is not its obvious disadvantage, since for many problems the global minimum point is not considered the optimal solution to the problem. The optimal solution of the learning problem involves achieving such a network error in which the neural network has a given generalizing ability. Due to the generalization property, the neural network is able to produce correct results for a variety of new data, including those not used in training [8]. To predict direct sales, obtaining the optimal solution to the training problem is crucial for success. Summarizing the above conditions, it was decided to use the method of teaching with a teacher to train a neural network and using the method of error back propagation.

Next, popular metrics are presented, which we considered as a working option for evaluating the quality of the obtained models. The choice was made in favor of the metrics listed below, since they are all related to the estimation of the prediction error. We were guided by the principle: the smaller the prediction error, the better the model (the model is more accurate) and vice versa, if the prediction error is large, then the model is less qualitative [9]. The selected metrics for determining the most qualitatively constructed network are presented below:

- MAPE (Mean Absolute Percentage Error) – shows the average percentage absolute error as a percentage and is calculated using the formula:

$$MAPE = \frac{1}{n} * \frac{\sum_1^N |\widehat{y_1} - y_i|}{|y_i|} * 100\% \tag{1}$$

- MSE (Mean Squared Error) – the mean-square error is measured in units that are the square of the target variable, calculated by the formula:

$$MSE = \frac{\sum_1^N \left(\widehat{y_i} - y_i\right)^2}{n} \tag{2}$$

- RMSE (Root Mean Squared Error) – the square root of MSE - the root-mean-square error of prediction is measured in the same units as the target variable, calculated by the formula:

$$RMSE = \sqrt{\frac{\sum_1^N (\widehat{y_i} - y_i)^2}{n}} \tag{3}$$

MAE shows how much difference there is on average from the simulated and reference values. MAPE is expressed as a percentage and tells you how many percent the model is wrong in forecasting. The MSE has no physical meaning, however, when comparing, the model with less MSE is selected.

During neural network training, RMSE is most often minimized. To correctly compare models by metrics, it is necessary to compare those models that were trained on the same row. This is due to the fact that the considered metrics are not comparable on different time series due to the different range of values and behavior of the series. In our case, the time series is a highly volatile series of weekly initial sales, each value of which is more than 100 million.

3 Research results

In general, the problem of forecasting primary sales can be formulated as a problem of forecasting a time series. To train and test the neural network, we divided the available time series of initial sales (small and large-scale launches) into training and test data.

Guided by general approaches to the construction of recurrent models [2, 10], test data were not considered during RNN training, and RMSE and MAE values were used to test the model. In particular, low values of RMSE and MAE in the test sample give higher accuracy compared to randomly selected parameter values from the basic distribution of the experimental ensemble. In addition, this approach also makes it easy to detect outliers.

Most often, 80% of the series is taken as a training sample, the remaining values are a test sample. However, since in our case the time series are less variable, it was decided to use 80 to 90 percent of the available values for the training sample.

Then the time series was normalized, that is, the scale of the data from the original range was changed for the convenience of visual analysis. In this case, the first 255 values (~90%) of the values were taken as a training sample. Such a large volume of the training sample was determined by us in order to present the neural network with more features of the behavior of the series.

The schedule of initial sales of the AKRIKHIN company is shown in Figure 3.

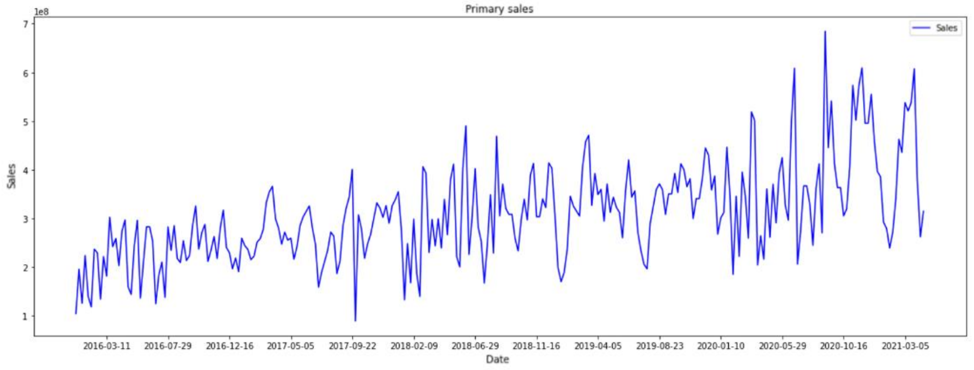

Fig. 3. Primary sales chart.

Source: compiled by the authors.

After preparing the data needed to feed the LSTM network, it is necessary to build a neural network by determining the number of LSTM modules and the number of neurons in each layer.

As a result of building a neural network, it was expected to obtain a network capable of predicting 4 values ahead with an optimal MAPE value (less than 10%). In our opinion, forecasting exactly 4 values is sufficient, based on the volume of the available time series of primary sales values, since the error increases for each subsequent forecast value.

To solve this problem, 5 LSTM networks were built, differing in:

- the size of the window (the time interval on the basis of which the forecast is based, in classical NS, from 15 to 30 values are used);
- optimizer (used in network training, stochastic gradient descent with a descent step of 0.001 or more);
- the number of LSTM modules (for simple series, 1 or 2 modules are used, for highly volatile ones from 4);
- the number of neurons in the module (selected for better learning, increases if there is a problem of under-learning in the model);
- the number of epochs (the parameter responsible for the number of neural network iterations).

4 Discussion of the results

Table 1 shows the results of all the constructed networks, which we compared using previously selected metrics calculated using formulas 1-3.

Table 1. Results of building neural networks.

Neural network	Neural Network Characteristics	MAE	RMSE	MAPE
1	Window – 30 Optimizer – Adam Number of LSTM modules – 4 The number of neurons in the module is 50 Numbero of epochs – 500	118 082 460	145 063 650	27,14%
2	Window – 30 Optimizer – Adam Number of LSTM modules – 4 The number of neurons in the module is 50 Numbero of epochs – 1000	97 269 859	121 847 344	22,49%
3	Window – 30 Optimizer – Adam Number of LSTM modules – 6 The number of neurons in the module is 50 Numbero of epochs – 1000	64 178 853	81 266 683	15,58%
4	Window – 15 Optimizer – Adam Number of LSTM modules – 6 The number of neurons in the module is 90 Numbero of epochs – 1000	75 328 267	98 997 343	18,36%
5	Window – 30 Optimizer – Adam Number of LSTM modules – 6 The number of neurons in the module is 50 Numbero of epochs – 500	22 685 630	28 690 515	5,37%

The fifth network has shown sufficient results due to its power. Compared to the first model, the number of layers was increased (from four to six), the number of neurons in each LSTM module was increased (from fifty to ninety). The selection of metrics solved the problem of under-learning (the problem in which the model poorly repeats the values in the training sample and poorly predicts the values in the test sample) by increasing the epochs. This allowed not only to train the network more efficiently, but also to avoid retraining (a problem in which the model explains well only examples from the training sample, but cannot predict values on the test sample). All our results for MAE and RMSE have shown that retraining problems can be largely ignored.

With each change in the set of parameters, they led to an improvement in the metrics of neural networks. MAE decreased from 118 to 23 million. The greatest change in metrics for the better was observed when the window was increased from 15 to 30 and the neurons in the module increased to 90 (4 and 5 neural networks).

For all the metrics considered, the fifth network shows the best results. The average absolute error was 22 million (6 times less compared to the first network), and the average absolute percentage error was 5.37%. Given the huge range of values (from 140 to 650 million), this can be considered a normal deviation.

The forecast is based on 4 values (weeks) ahead. The result of training and forecasting can be seen in Figure 4, in which a series of data with actual values of initial sales is indicated in blue, and a series of forecast values superimposed on actual data is indicated in red. A number of forecast values repeat the sales dynamics with high accuracy, which indicates a sufficient amount of values included in the training sample.

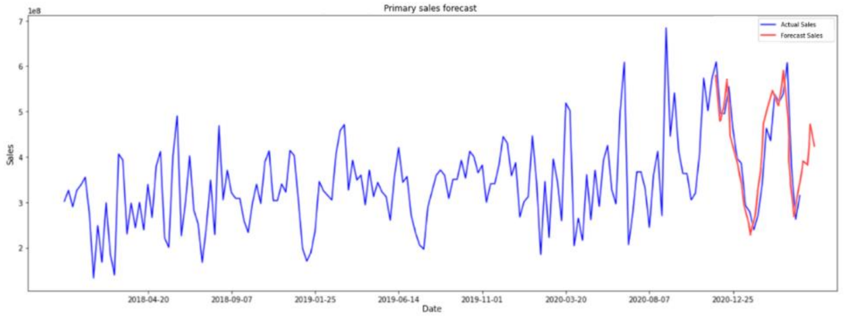

Fig. 4. Forecast for 4 values of the fifth network

Source: compiled by the authors.

5 Conclusion

Summarizing the results of the research process and the results obtained, the following should be pointed out. The type of neural network and the learning algorithm were chosen correctly.

The constructed neural network gives adequate predictive values, correctly guesses the direction of the trend, given the strong variability of the observed process. The result of the conducted research is the choice of the optimal type of neural network forecasting, considering the peculiarities of the pharmaceutical market and forecasting primary sales for 4 values ahead, which allows us to develop the right strategy for the company's sales plan and its consistent development.

The considered method of sales forecasting is not common in pharmaceutical companies, where more preference is given to classical methods (regression analysis, expert assessments, etc.) and can be offered as an alternative method of sales forecasting.

References

1. About AKRIKHIN Company. Retrieved from https://www.akrikhin.ru/cooperation/Informaciya_ob_Akrihine_2020.pdf (2020)
2. J. Smiatek, Ch. Clemens, L. Montano Herrera, S. Arnold, B. Knapp, B. Presser, A. Jung, Biotechnology Reports **31**, e00640 (2021). https://doi.org/10.1016/j.btre.2021.e00640
3. Y. Bengio, P. Simard, P. Frasconi, IEEE Transactions on Neural Networks **5(2)**, 157 (1994)doi: 10.1109/72.279181
4. R. DiPietro, G. D. Hager, *Chapter 21 - Deep learning: RNNs and LSTM*, Handbook of Medical Image Computing and Computer Assisted Intervention, Academic Press, 503 (2020) https://doi.org/10.1016/B978-0-12-816176-0.00026-0
5. F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru, A. Dairi, *Chapter 7 - Unsupervised recurrent deep learning scheme for process monitoring, Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches*, Elsevier, 225 (2021). https://doi.org/10.1016/B978-0-12-819365-5.00013-9
6. M. Mazurov, A. Mikryukov, M. Gorbatih, T. Bergaliev, N. Shchukina, CEUR Workshop Proceedings **2919**, 161 (2021)
7. S. H. Oh, Neurocomputing, **74(6),** 1058 (2011). https://doi.org/10.1016/j.neucom.2010.11.024

8. M. Hamamoto, J. Kamruzzaman, Yu. Kumagai, *Artificial Neural Networks*, North-Holland, 1067 (1992). https://doi.org/10.1016/B978-0-444-89488-5.50050-6
9. A. Tealab, Future Computing and Informatics J. **3(2)**, 334 (2018). https://doi.org/10.1016/j.fcij.2018.10.003
10. S. Chakraborty, A. Preece, M. Alzantot, T. Xing, D. Braines, M. Srivastava, 20th International Conference on Information Fusion (Fusion), 1 (2017). doi: 10.23919/ICIF.2017.8009785