УДК 519.2:004.9

Zinkevich I.E., Kirichenko L.O., Radivilova T.A.

COMPARATIVE ANALYSIS OF PREDICTION METHODS FOR WEAKLY CORRELATED TIME SERIES

Annotation. Prediction for weakly correlated time series by using methods of exponential smoothing, neural network and decision tree using real data of online store as example are considered in the paper. The advantages and disadvantages of each method are described.
Keywords: time series, forecasting, exponential smoothing, decision tree, long-term memory network.

Introduction and goal

In today's world, there is often a need for analysis and prediction of time series (TS). TS is a common and important form of data description as possible to observe the whole history of value changes interested to us. This gives opportunity to judge about "typical" behavior and values of deviations from such behavior. One of the challenging and interesting areas of TS analysis is the area of e-commerce.

E-commerce is in process of development, which is facilitated by new technologies, services and tactical tools [1]. To "survive" and stand out among the many online stores, it is important to understand the user's behavior from the moment of the first arrival on the site: to track his movements, to know what products he looked at, put in the basket, where he clicked, what saw, the time he left, how and when he returned. Web analytics will help in this, which involves ongoing collection, analysis and interpretation of data about visitors, work with basic metrics.

Quality analyst online store always starts at the visitor's path, which he held before making a purchase. Let’s describe a suspended user's path, which came from the social network: visit the page; visitor pays attention to the post; clicks on it; goes to the landing page to which the link leads from the post; studies the characteristics of product; examines methods of payment and delivery; adds product to “shopping cart”; places an order; makes a purchase.

At each stage the user can stop and terminate the process without making a purchase. The percentage of conversion is calculated as the ratio of the number of visitors to buyers. For example, if a site was visited by 100 people, but bought - 2, then the conversion is equal to 2%. This measurement is the mail in web analytics for all commercial sites. The increase of percentage depends on many factors: from the design of the page to its functionality. Monitoring conversion allows to understand in time that the e-shop needs to be improved.

Analyzing and predicting TS daily values of conversion percentage is critical to optimizing the effectiveness of online business. However, it should be noted that practically all of the classical methods of TS analysis are based on calculating the correlation between TS values [2]. In the case of weakly correlated TS, and also in the case when TS has a discharged zero value, which is typical for many electronic sales sites, these methods are not appropriate or have a large error.

Neural network approach has been widely used to solve forecasting problems. Neural networks allow you to model complex relationships between data as a result of learning by example. However, the prediction of TS using neural networks has its disadvantages. Firstly, long lengths TS required for neural network training. Secondly, the result essentially depends on the choice of the architecture of the network, as well as the input and output data. Third, neural networks require pre-preparation of the data or preprocessing. Preprocessing is one of the key elements of prediction: the quality of prediction of a neural network can depend crucially on the form in which information is presented for its learning. The general goal of preprocessing is to increase the information content of inputs and outputs. An overview of the methods for selecting input variables and preprocessing is contained in [3, 4].

Recently, for the analysis of the TS regularities the methods of Data Mining and machine learning [5] have been increasingly used to detect various patterns in the time series. In this case, logical methods are of particular value in the detection of such patterns. These methods allow to find logical if-then rules. They are suitable for analyzing and predicting both numerical and symbolic sequences, and their results have a transparent interpretation

The goal of this work is to carry out a comparative analysis of time series prediction, based on classical prediction methods and machine learning methods on the example of a real online shop data.

Methods of research

E-commerce is in process of development, which is facilitated by new technologies, services and tactical tools. Suppliers, the range of buyers, the range of products change regularly, which leads to a rapid obsolescence of information. Therefore, methods that require large arrays of TS, such as, for example, autoregressive and moving average models, work bad.

Methods of exponential smoothing. The basis of the exponential smoothing (ES) is the idea of a permanent revision of predicted values as the actual receipt. ES model assigns exponentially decreasing weights to observations as they age. Thus, the latest available observations have a greater effect on the predicted value than the older observations.

The ES model has the form

$$Z(t) = S(t) + \varepsilon_t, \quad (1)$$

$$S(t) = \alpha \cdot Z(t-1) + (1-\alpha) \cdot S(t-1),$$

where α – Smoothing ratio; $0 < \alpha < 1$; $Z(t)$ – predicted TS; $S(t)$ – smoothed TS; initial conditions are defined as $S(1) = Z(0)$. In this model, each subsequent smoothed value $S(t)$ is the weighted average between the previous value of the time series $Z(t)$ and the previous smoothed value $S(t-1)$.

Machine learning methods - extremely broad and dynamically developing area of research, using a great number of theoretical and practical methods. One of the methods of machine learning is the decision tree method. The decision tree is a decision support tool used in statistics and data analysis for predictive models.

The structure of the tree is "leaves" and "branches". On the edges ("branches") of the decision tree, attributes are recorded, on which the objective function depends, in the "leaves" the values of the objective function are recorded, and in the remaining nodes - the attributes for which the cases differ. To classify a new case, it is necessary to go down the tree to the leaves and give the appropriate value. Similar decision trees are widely used in intellectual data analysis. The goal is to create a model that predicts the value of the target variable based on several variables at the input.

Each leaf represents a target value of a variable modified in the course of movement from the root of the leaf. Each internal node corresponds to one of the input variables. The tree can also be "learned" by dividing the original sets of variables into subsets based on testing attribute values.

This process is repeated for each of the resulting subsets. Recursion ends when the subset at the node has the same values of the target variable, so it does not add value to the predictions. In data mining, decision trees can be used as a mathematical and computational methods to help describe, classify and summarize a set of data that can be written as follows: $(x, Y) = (x_1, x_2, x_3, \ldots, x_k, Y)$. The dependent variable Y is the target variable that needs to be analyzed, classified and generalized. A vector x consists of input variables x_1, x_2, x_3, etc., which are used to perform this task [5].

Long Short Term Memory networks (LSTM) are a special type of recurrent neural networks that are capable of learning long-term dependencies. They were proposed by Hochreiter and Schmidhuber and finalized and popularized by others in subsequent work. They work incredibly well on a wide variety of problems and are currently widely used [3, 4].

LSTM is specifically designed to avoid the problem of long-term dependencies. Remembering information for a long period of time is practically their default behavior, not something that they are just trying to do. All recurrent neural networks have the form of a repeating module circuit of a neural network. In standard recurrent neural networks these repeating modules will have a very simple structure.

The forecast errors. To obtain quantitative characteristics of the comparative analysis of the models, the following characteristics of forecast errors were chosen. The Mean Absolute Derivation (MAD) measures the accuracy of the forecast by averaging the values of the forecast errors. Using MAD is most useful when the analyst needs to measure the forecast error in the same units as the original series. This error is calculated as follows:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|X(t) - \hat{X}(t)\right| .$$

Mean squared error (MSE) is another way of estimating prediction method. Since each deviation value is squared, this method emphasizes large forecast errors. The MSE error is calculated as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(X(t) - \hat{X}(t))^2 .$$

The average Absolute Percentage Error (MAPE) is calculated by finding the absolute error at each time moment and dividing it by the actual observed value, followed by averaging the absolute percent errors obtained. This approach is useful when the size or quantity of predicted values are important in assessing the prediction accuracy. MAPE accentuates magnitude forecast errors compared to actual values of series. This error is calculated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|X(t)-\hat{X}(t)\right|}{X(t)}$$

Input data

Input data in the work were daily data from the online sales site, which included the number of clicks on the site from social networks, the number of purchases and the corresponding conversion rate. In addition, there was information about which language the buyer used, from which country order was made and other information.

Figure 1 shows typical TS clicks, orders and percent conversion. The series of orders and, accordingly, conversions, are characterized by zero values, which significantly complicates the prediction for the next day. The correlation function for a series of conversions is shown in Figure 1 on the right. Obviously, there is almost no correlation between TS values.

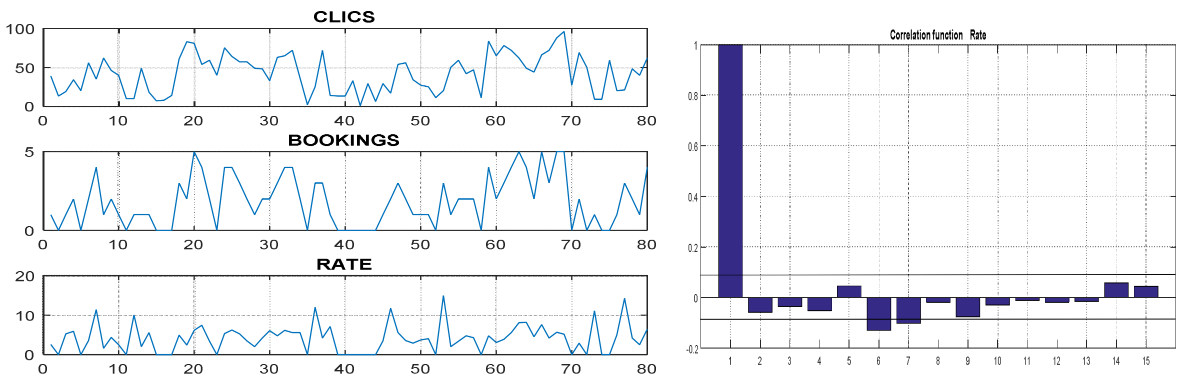

Figure 1. - Time series of clicks, orders, and percent conversion (left); Correlation function for a number of percent conversion (right).

Results of research

For building models and a neural network, the Python language was used with libraries that implement machine learning methods. For the prediction the TS have been divided into two parts, where the first is used to train the model, and the second - to assess its plausibility. The models were trained on the S last values (S was chosen equal to 20).

Checking the models for predicting *m* values was carried out as follows: take the last 20 values of first row and make a prediction for one value ahead. Next, let's shift our window one value forward, including in window a prediction for new value, and make a bet on a single value forward again, and repeat *m* times.

In Fig. 2 presents the results of each model prediction to 7 values forward. The solid line shows the real values, line 1 - values obtained by the method of exponential smoothing, 2 - based on the decision tree, 3 - using a neural network.

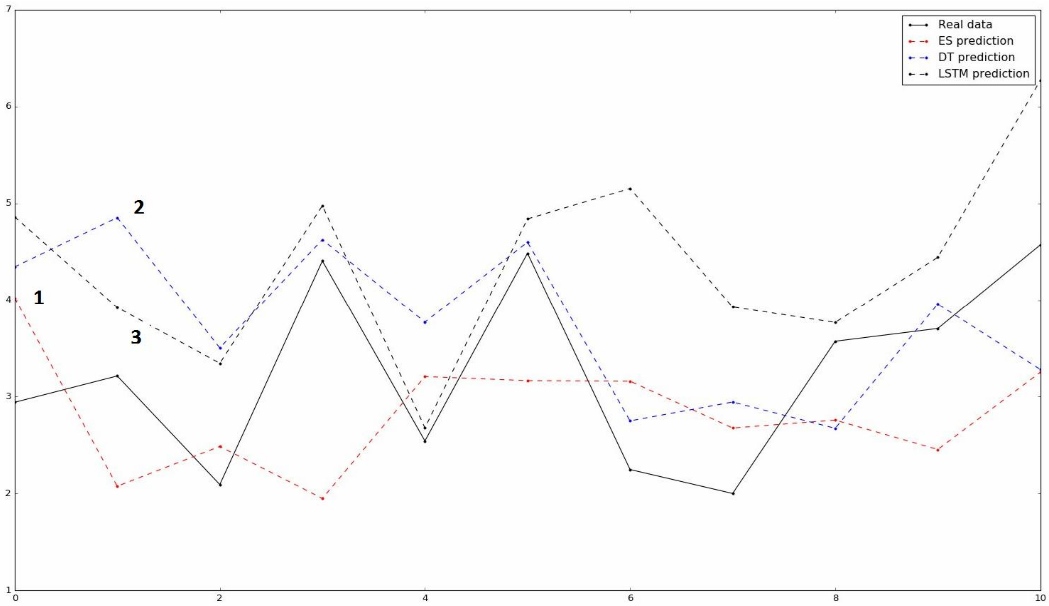

Figure 2. - Predicted values for different models

Predicted values were calculated for S = 20 and m = 1 (this choice of parameters determined by the requirements of an online store) by using 100 number values for the TS plurality percent conversion. The calculation results are typical for most of series are shown in Table 1

Table 1 - Method fault

Method	Exponential Smoothing	Decision tree	Neural network
MAD	0.013751	0.014218	0.002645
MSE	0.000369	0.000353	0.000012
MAPE	0.491	0.513	0.072

As a result of the analysis of prediction of different values of *S* and и *m* was established that the method of exponential smoothing, in spite of its simplicity and not exacting in the amount of data on which the prediction will be constructed, has smallest fault of predicted values in most cases, but at the same time, some predicted values are significantly removed from the real ones. The decision tree method has proved to be inconvenient in choice of parameters and has fault comparable with errors of exponential smoothing, but without strongly remote predicted values. The

LSTM neural network, which has a more complex structure and needs to be preliminarily trained on a rather large time series, has shown excellent results, as well as in the overall prediction fault, and in the remoteness of prediction from real time series values.

Conclusion

The results of a study of methods for predicting weakly correlated time series typical of e-commerce conversion series have shown that exponential smoothing is the simplest, fastest and most convenient predictive method, but in the case of complex or long-term dependencies, it does not apply. The decision tree method is fast in learning, not difficult to understand, but inconvenient in the choice of parameters and does not work well when learning on data that have many characteristics. The neural network is cumbersome, long in training, requires a lot of parameters to be selected, but it has very good predictors and an order of magnitude smaller errors.

Bibliography

1. [On-line resource]: http: // lpgenerator.ru/blog/2015/07/02/kakoj-dolzhna-byt-veb-analitika-internet-magazina
2. Khank D. Biznes-prognozirovaniye / Khank D. Izd. Dom «Vil'yams», 2003. - 656 s.
3. Guyon I. An Introduction to Variable and Feature Selection. J. Guyon Isabelle, Elisseeff Andre / Of Machine Learning Research 2003. –S.1157-1182.
4. Yezhov A.A Neyrokomp'yuting i yego primeneniya v ekonomike i biznese / Yezhov A.A., Shumskiy S.A. –M., 1998. –S.216.
5. V'yugin V.V. Matematicheskiye osnovy teorii mashinnogo obucheniya i prognozirovaniya. / V'yugin V.V.–M., 2013. –S.387.Ханк Д. Бизнес-прогнозирование / Ханк Д. Изд. Дом «Вильямс», 2003. -656 с.

УДК 519.2:004.9

Зинькевич И.Э., Кириченко Л.О., Радивилова Т.А. **Сравнительный анализ методов прогнозирования слабо коррелированных временных рядов //** Системные технологии. Региональный межвузовский сборник научных работ.- Выпуск?(??).- Днепропетровск, 2017. –С. ??-??.

В работе проведен сравнительный анализ прогнозирования слабо коррелированных временных рядов методами экспоненциального сглаживания, нейронной сети и дерева решений, на основе данных реального интернет-магазина. Рассмотрены преимущества и недостатки каждого метода.

Библ.5 , рис. 2, табл.1 .

УДК 519.2:004.9

Зінькевич І.Е., Кіріченко Л.О., Радівілова Т.А. **Порівняльний аналіз методів прогнозування слабо корельованих часових рядів** // Системні технології. Регіональний міжвузівський збірник наукових робіт.- Випуск?(??).-Дніпропетровськ, 2017. –С. ??-??.

У роботі проведено порівняльний аналіз прогнозування слабо корельованих часових рядів методами експоненціального згладжування, нейронної мережі та дерева рішень, на основі даних реального інтернет-магазину. Розглянуто переваги та недоліки кожного методу.

Бібл.5, рис. 2, табл.1.

UDC 519.2:004.9

Zinkevich I.E., Kirichenko L.O., Radivilova T.A. **Comparative analysis of predicting methods for weakly correlated time series //** System technologies.- N(??).-Dnipropetrovsk, 2017. –P. ??-??.

The article compares the forecasting of weakly correlated time series using exponential smoothing methods, a neural network, and a decision tree. Time series were data from a real online store. The advantages and disadvantages of each method are considered.

Ref.5, fig. 2, tab.1.