Possibilities for predicting the state of usability web resources

I F Astachova 1, K A Makoviy 2 and Yu V Khitskova 3

1 Department of Applied Mathematics, Informatics and Mechanics, Voronezh State University, 1, University Square, Voronezh, 394018, Russia
2 Faculty of Economics, Management and Information Technology, Voronezh State Technical University, 84, 20th anniversary of October, Voronezh, 394006, Russia
3 Economics Department, Voronezh State University, 1, University Square, Voronezh, 394018, Russia

E-mail: prosvetovau@list.ru

Abstract. The article discusses the possibilities of predicting the state of the web resources usability. The usability testing procedure is quite costly from both financial and time points of view. Therefore, a system that reduces these costs is useful for modern organizations. Different approaches of forecasting the number of visitors: ARIMA model and Neural Networks are considered. An important time series property for ARIMA model being applicable is the stationarity of the series. It is shown that this model is not suitable enough for the investigated time series, some types of neural networks are also not suitable for various reasons. As a result, NARX networks are selected, which are successfully used for time series forecasting, providing an opportunity to use an exogenous variable.

1. Introduction

Internet marketing is growing very quickly, becoming more and more popular; website promotion on the Internet is one of the most important ways to develop a business.

Internet marketing tools provide users with various applications that help to track the information resources usability. But applications are required to simplify the usability testing procedure, by intellectualizing it. Such systems make it possible to automate the processes of assessing the company's resources, to understand how attractive the resource is for potential and real consumer goods and services. Before developing a usability testing system, it is necessary to propose criteria that must correspond to a web resource page. The main criteria are highlighted: page functionality, availability of the main page buttons, especially the contacts, pages which reflect the basic assortment of goods, the pages describing the conditions of purchase behind, payment and delivery. The page headers size, the main text font size, the type of headers on the page, the type of main font, the page color scheme are also important. The resulting indicator that characterizes the usability of a resource is page traffic. Website traffic depends on many external factors besides the usability of the resource: the time of the year, days of weeks and, holidays, etc. The most important external factor, which largely affects the number of visitors - it is the
day of the week. Therefore, it was decided to develop a forecast of website traffic based on the previous period. It is desirable to take into account the exogenous variable – the days of the week in this forecast.

2. Methods of system intellectualization

Different methods are used to intellectualize systems. These can be traditional methods: information retrieval, simulation modeling, situational analysis, regression methods. Some of these methods have been developed as part of artificial intelligence research.

The traditional forecasting methods are moving average, exponential smoothing, least square method and regression analysis [1, 2].

At present, the intellectualization of the system and forecasting can be carried out by rather new methods, in comparison with traditional methods: machine learning methods: ARIMA model, neural networks. We have considered various options for predicting the assessment of site usability.

The integrated model -Autoregressive Integrated Moving Average (ARIMA) -a model describing the time series and short-term forecasting. In general, ARIMA is described by the following equation (1), [3].

$$\phi_p(B)(1-B)^d z_t = \theta_q(B)a_t,$$

Where $\phi_p$ and $\theta_q$ – are unknown parameters; $z_t$ is a time series; $B$ – is the backward shift operator, $d$ - is the order of differentiation; and $a_t$ – is a row that represents white noise.

The next common method is artificial neural networks. After the development of learning algorithms, the resulting models began to be used for practical purposes: in forecasting problems, for pattern recognition, in control problems. The network processes input information and forms a set of output signals in the process of changing its state in time. The work of the network consists in transforming input signals in time. Because of this, the internal state of the network changes and output effects are formed. Typically, the NN operates with digital, not symbolic values [5].

There are several types of neural networks that can be used in deep machine learning. Each type has its own advantages and disadvantages. The expediency of using one or the other type depends on the subject area and application tasks. The following main types of neural networks are distinguished, the main ones are systematized in the works [1, 4, 5, 6].

1. Convolutional neural networks (CNN) contain five types of layers: input, convolution, subsampling (downsampling), fully connected and output. Each layer has a specific purpose, such as summing, connecting or activating Convolutional neural networks have gained popular image classification and object detection. However, CNN is also used in other areas such as natural language processing and forecasting.

Deep learning technology assumes the presence of a process of configurational complication of informative features in a sequence of neural layers. Starting with K. Fukushima, many options for implementing this idea have been proposed. One of the successful solutions is the architecture of convolutional neural networks [7, 8], which has shown high efficiency in solving various problems. A distinctive feature of this architecture is the presence in convolutional layers of several data processing channels (called maps or planes). In each plane, the convolution of the output image of the previous layer with a fixed low-dimensional kernel is performed. Convolutional layers alternate with pooling layers, which multiply the dimension of the feature space. Pooling layers are optional and there are options to remove them completely from the network architecture. Another distinctive feature of the architecture is the use of semi-linear activation functions that act as switching keys controlled by the values of the variables of the hidden layers.

2. Recurrent neural networks (RNN) use sequential information such as time-stamped data from a sensory device or an oral sentence consisting of a sequence of terms.
3. FNN - Feedforward neural networks. Feedforward neural networks, in which every perceptron in one layer is connected to every perceptron from the next layer. Information is transmitted from one level to the next only in the forward direction. There is no feedback loop in a neural network.

4. Neural networks NARX Nonlinear autoregressive exogenous model are considered as a class of generalized, nonlinear, nonparametric, data-driven statistical methods. Nonlinear autoregressive exogenous NN is a combination of all the above NN and is successfully used for forecasting time series, as well as predicting causal relationships.

5. Another well-known type of neural networks are autoencoder networks. Autoencoders take any input, convert it to some kind of compressed version, and use it to represent the output. So basically the input \( x \) goes to the hidden layer \( h, h = f(x) \) and is obtained as a reconstruction of \( r, r = g(h) \). An autoencoder is good when \( r \) is close to \( x \), or when the output looks like an input. Autoencoder neural networks are used to create abstractions called encoders, created from a given set of input data [9]. Although autoencoders similar to more traditional neural networks, they try to simulate the input data, and so the method is considered to be out of control. The premise of autoencoders is to desensitize irrelevant and sensitize appropriate. As layers are added, further abstractions are formulated at higher layers (the layers closest to the point at which the decoder layer is represented). Linear or non-linear classifiers can then use these abstractions.

Several types of neural networks use the gradient descent method. This is one of the most famous algorithms in machine learning [10]. Its main strength lies in its ability to bypass the problem of having a large amount of data. This problem affects systems, such as neural networks, with too many variables to be able to calculate their optimal values. However, gradient descent breaks down dimensional problems by increasing the local low point or local low of the multidimensional error or cost function. This helps the system determine a configurable value or weight to be assigned to each of the blocks in the network, bringing the accuracy back into alignment. This method is also used in NARX neural networks.

3. Using the considered models for forecasting time series

Consider the possibility of using the models to predict the number of site page visitors.

Usually consider two types of ARIMA model, stationary ARMA model and non-stationary ARIMA model. The stationarity of a time series is associated with the type of change in statistical temporal characteristics over time, so the probability distribution is constant over time. Stationarity or the so-called weak stationarity is defined as follows:

The expected value of the time series is independent of time.

The autocovariance function is a function \( k \), where for each \( k \) \( y(z) = \text{Cov}(z_t, z_{t+k}) \).

Considering the equation (1), we can distinguish two components - autoregressive (AR) and moving average (MA). Thus, the ARIMA model can be the AR (p) model or the MA (q) model, or a combination of both, that is, ARMA(p,q), see (2).

The autoregressive model AR (p) is described by the following equation:

\[
z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \alpha_t ,
\]  

(2)

Where \( \alpha_t \) is an error that does not correlate with itself in any way, where \( E[\alpha_t] = 0 \) and \( Var(\alpha_t) = \sigma^2 \). Coefficients \( \phi_i, i = 1, \ldots, p \) - parameters which are to be calculated.

A moving average model MA is a model that contains the «average» noise for the current period and for the previous period. The MA (q) model is described by equation (3).

\[
z_t = \alpha_t + \theta_1 \alpha_{t-1} + \theta_2 \alpha_{t-2} + \ldots + \theta_q \alpha_{t-q}
\]  

(3)
Coefficients $\theta_i, i=1,...,q$ is calculated parameters.
Consider the time series of site visitors, see (4).

$$y_1, ..., y_T \in \mathbb{R}$$  \hspace{1cm} (4)

This series is shown in figure 1. Feature Value - is the number of visitors of the main page, the time interval – daily, and sources data is the Google Analytics. The row is schematically shown in figure 1.

![Figure 1. Time series – the number of visitors to the main page of the site.](image)

Further we analyze this time series according to the stages shown in figure 2.

![Figure 2. Simulation steps.](image)

In the following paragraphs, an analysis of this time series according to the ARIMA and NARX models will be presented.

3.1. ARIMA model

The ARMA and ARIMA models have undoubted advantages. You can use several tools for their implementation: the Python, R-language, as well as the MATLAB computing environment [11]. Before compiling the model, it is necessary to consider the properties of the time series and the possibility of using the ARMA model.

ARMA (p, q) model contain AR and MA components. This model does not contain the "i" element because it is a model that is already stationary. The element in equation (1) is 0. The mathematical notation of the ARMA (p, q) model is as follows (5) [12].

$$z_t = c + \phi_1 z_{t-1} + ... + \phi_p z_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q} \hspace{1cm} (5)$$

As can be seen from equation (5), ARMA (p, q) is the sum of two components of the time series – AR (p) and MA (q).

An important property of the time series for using the ARMA model is the stationarity of the series. The series $y_1, ..., y_T$ stationary if $\forall s$ – distributions $y_t, ..., y_{t+s}$ does not depend on $t$, that is, its properties do not depend on time. The ARMA model is only used for the fully stationary row.
For time series that do not meet the requirements of criteria for stationary use differentiating. A series that can be modeled as a stationary ARMA \((p, q)\) after differentiating in time \(D\) times is denoted ARIMA \((p, D, q)\). Mathematically, the form ARIMA \((p, D, q)\) is written in the following form (6):

\[
\Delta^D z_t = c + \phi_1 \Delta^D z_{t-1} + \ldots + \phi_p \Delta^D z_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q},
\]

where \(\Delta^D z_t\) denotes \(D\) times the differentiated time series.

The autocorrelation value of the stationary series is near zero (any insignificant value). In our model, autocorrelation is shown in figure 3.

![Figure 3. Autocorrelation of the number of site page visitors, data for 7 months, by day.](image)

Figure 3 shows points of significant autocorrelation. Significance can be checked by the Ljung–Box \(Q\) test or Student’s \(t\)-test. The hypothesis of the Student’s criterion that autocorrelation is equal to zero is tested against some alternative, most likely against some two-sided. For example, autocorrelation is not equal to zero [13].

The figure shows that there are 8 points at which autocorrelation goes beyond what is possible in the case of using the ARIMA model. The stationarity of the series can be checked and improved by differentiating it. A cycle is observed in our series, but this does not mean that it is not stationary. An additional check is needed for an accurate determination [12].

Differentiate the series. After the first differentiation, we got the following, figure 4.
We can see that the series is still not stationary. The indicators of the time series are significantly displayed, there is an upward and downward trend. Further differentiation of the series makes forecasting difficult. We can conclude that the ARIMA model is not suitable for us to make a forecast. In addition, the ARIMA model does not allow the use of an exogenous parameter, which is the day of the week.

3.2. Neural networks
Let’s consider various types of neural networks as possible methods for predicting the state of usability of web resources [14]. The architecture convolutional networks disadvantage is the lack of theoretically based methods for choosing the network structure and parameters of convolution kernels. The choice of the structure of the convolutional network is a subject of art until now. Several perfectly working network configurations have been proposed to solve individual problems, but the general algorithm for their construction is unknown. The second drawback is related to the training time of convolutional networks. On a typical processor, times can vary from several hours to several days, so high-performance GPUs are often used to train networks. In our case, the site contains more than 120 pages (and this is not the largest site). Therefore, it is impossible to spend too much time on processing one time series. This type of neural networks is not suitable for our project.

RNN are used in time series forecasting, sentiment analysis, and other text-based applications. Recurrent NNs are essentially feed-forward NNs with a recurrent loop, so the outputs are fed back to the input. This is consistent with the purpose of defining usability. All inputs to a recurrent neural network are not independent of each other, and the output for each element depends on the calculations of its previous elements, unlike traditional neural networks. But the entrance to our network supposes the existence of independent variables. So this type of neural networks is also not suitable for the site visitors time series.

The template is required that is submitted as input for RNN networks to work, but there is no such template when predicting the usability of website pages. Therefore, this type of networks is not applicable in this case. Autoencoders are almost never used in forecasting.

4. Results and discussion
As noted above, the NARX network has been successfully used for time series forecasting. A conventional neural network consists of an input layer that receives external information, one or more
hidden layers which provide nonlinearity of the model, and an output layer that provides the target value. Each layer consists of one or more nodes. All layers are connected through an acyclic arc [15]. Each input node in the input layer is associated with a corresponding weight. Activation function is applied to the weighted sum of the inputs to calculate the output. The activation function is either an identity function or a sigmoidal function [16, 17]. In this type of neural networks, additional exogenous variables are used (in our case, it is a time factor). At the heart of NARX neural networks is the Levenberg-McGraft method, sometimes called the backpropagation algorithm, it was first used to minimize prediction errors. It is also currently used to determine the weight (synapse) of the NARX model. The algorithm combines the gradient descent method and Newton's method, and has fast convergence and stable performance [17]. The main advantage of NARX in our case is the possibility of introducing an exogenous parameter. It make possible to predict website traffic not only taking into account its own changes, but also taking into account the time factor, which is especially important. Because page traffic often depends on the season, especially on the day of the week (off it or workday) NARX neural networks are suitable for the prediction of our time series.

The analysis result carried out, we choose the NARX model as a model for predicting traffic on the website pages.

The configuration and operation stages of the neural network are shown in figures 5-6. The standard time series modeling network algorithm consists of two stages. At the first stage, the network is trained using the time series of the number of site visitors and two exogenous parameters as input: the day of the week and the binary series of the weekend − working day.

![Figure 5. Scheme of a network with an open feedback loop.](image)

At the second stage of the algorithm operation, the feedback loop is closed. When the feedback is open on the NARX network, it performs a one-step-forward prediction. This predicts the next \( y(t) \) value from the previous \( y(t) \) and \( x(t) \) values. Closed-loop, it can be used to perform multi-stage forward predictions. This is because the predictions \( y(t) \) will be used instead of the actual future values of \( y(t) \) figure 6.
Figure 6. Scheme of a network with a closed feedback loop.

A closed loop network allows predicting the values of a time series using the input data of the series $y(t)$ obtained as a result of the forecast. The dynamics of training a neural network for 9 epochs is shown in figure 7.

Figure 7. MSE during training.

The histogram of the time series modeling error at each stage of the neural network operation is shown in figure 8.
The figure shows that there is a slight discrepancy between the data and the forecast. So this model can be used. After obtaining the NARX network model, the feedback loop is closed, which makes it possible to predict the values of the time series $y(t)$. The graph of the predicted values obtained using the network is shown in figure 9.

Figure 8. Histogram of the simulated value error.

Figure 9. Forecasting website page traffic using NARX.
The blue trend shows the data values, the lilac trend highlights the time series obtained as a result of the neural network operation at all three stages, the predicted values are displayed in black in the figure. Because of the network operation, five values of the time series were obtained, which can be used, for example, in Internet marketing to predict various indicators that depend on this data. The model is quite universal, it can be used for other information resources.

5. Conclusion
The ability to evaluate and forecast a website traffic is a key factor for successful internet marketing in a modern world. The proposed approach to forecasting bases on a NARX model that gives an opportunity to consider an exogenous parameter and thus take into account fluctuations depending on business activity concerning type of a day, e.g. public holidays.

References

[1] Mitrea C A, Lee C. K. M., Wu Z. A. 2009 *Comparison between neural networks and traditional forecasting methods: A case study* International journal of engineering business management. V. 1 XI

[2] Hubbard E A, Venkatramani K, Anderson D P, Adiga A K, Hewgill G D, & Lawson J A 2011 *Dynamic coordination and control of network connected devices for large-scale network site testing and associated architectures*. U.S. Patent Application No. 12/462,600.

[3] Tun C C, Majid N 2018 *Comparison between artificial neural network and autoregressive integrated moving average model in bitcoin price forecasting* Journal of Quality Measurement and Analysis JQMA. V 14.2 XLV-LIII

[4] Khajouei R, Zahiri Esfahani M, Jahani Y 2017 *Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems* Journal of the American Medical Informatics Association. V 24 eLV-eLX.

[5] Zaentsev VI 1999 *Neural networks: basic models* Voronezh Voronezh State University, LXXVII.

[6] Haykin S 2008 *Neural Networks: A Comprehensive Foundation Second Edition* Moscow Williams Publishing House, MCIV.

[7] Dorogov A Yu 2019 *Deep learning neural networks with controlled switching of neural planes* Distance educational technologies CCLXXXIV - CCXCVI.

[8] Domingos . 2012 *Several useful things to know about machine learning* ACM Messages V 55 LXXVIII – LXXXVII.

[9] Dong G, Liao G, Liu H, Kuang G 2018 *A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images*. IEEE Geoscience and Remote Sensing Magazine, V 6(3), XLIV- LXVIII.

[10] Zhang Y D, Pan C, Chen X, Wang F 2018 *Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling* Journal of computational science, V 27 LVII – LVII.

[11] Hussain A, Mkpojiogu E O C, Kamal F M 2016 *A systematic review on usability evaluation methods for m-commerce apps* Journal of Telecommunication, Electronic and Computer Engineering (JTEC) V 8.10 XXIX XXXIV.

[12] Cadenas E, Rivera W, Campos-Amezquita, R, Heard C 2016 *Wind speed prediction using a univariate ARIMA model and a multivariate NARX model* Energies V 9(2) CIX.

[13] Tun C C, Majid N 2018 *Comparison between artificial neural network and autoregressive integrated moving average model in bitcoin price forecasting* Journal of Quality Measurement and Analysis JQMA V 18.2 XLV- LIII.
[14] Galanov A E, Selyukova G P 2019 *Neural networks and neural technologies* Actual problems of science and economy: new challenges and solutions CCCXCIX-CDV.
[15] Demuth H B, Beale M H, De Jess O, & Hagan M T 2014 *Neural network design* Martin Hagan.
[16] Andrienko M P, Yudin P A, Vishnevetskaya E Yu 2020 *Understanding of repetitive neural networks: a preferred neural network for time series data* Actual problems of modern science and education XCVI–XCVIII.
[17] Paul R K, Sinha K 2016 *Forecasting crop yield: ARIMAX and NARX model* RASHI V 1.1 LXXVII – LXXXV.