Neural Network Forecasting Using Big Data

Sergey Ivanov[a], Nataliia Maksyshko[a] and Mykola Ivanov[a]

[a] *Zaporizhzhia National University, Zhukovsky str., 66, Zaporizhzhia, 69063, Ukraine*

Abstract

To increase the speed of data transfer while providing access to multidimensional data there is the use of Big Data as a tool. On the base of the MapReduce model you can use modern tools for working with big data. Therefore, in the paper researched Big Data as a single centralized source of information for the entire subject area. In addition, in this paper the structure of a neural network forecasting system is proposed, which include many databases, where transactions are processed in real time. For neural network forecasting of multidimensional data, a network in Matlab is considered and built. A matrix of input data and a matrix of target data, which determine the input statistical information, are used to teach the neural network.

The application of the Levenberg-Marquardt algorithm for training a neural network is considered. Also, the results of the training process of neural network in Matlab are presented. The obtained forecasting results are presented, which allows to conclude about the advantages of a neural network in multivariate forecasting.

Keywords

Big Data, MapReduce model, neural network, forecasting, matrix

1. Introduction

As Laurent Flores explained [1], the success of the use of information technology is determined by the success of digital economic, thus how they are measured and used. However, attention is paid to forecasting in the digital economy, taking into account intelligent systems.

Intelligent systems and the use of multidimensional communication determined the emergence of a new concept by the German economist Klaus Schwab, his economic forum in Davos [2]. According to this concept, it is argued that we are in the era of the fourth industrial revolution (Industry 4.0), when the virtual world is combined with the physical world using information technology. The fourth industrial revolution is characterized by a change in economic relations and the widespread use of intelligent technologies (general technologies, big data, artificial neural networks, and others).

It should be noted that with the use of digital technologies, D2C models have come to be used. The D2C (Direct to Consumer) model represents a direct selling system, where companies themselves manufacture, promote, sell and deliver their product without the involvement of intermediaries. So in the above paper (https://www.emarketer.com/content/nikes-d2c-sales-will-comprise-third-of-its-business) it is established that, in contrast to traditional promotion through retail chains, companies using D2C model develop their own distribution channels. Thanks to their good positioning, these companies not only have a competitive advantage in the market, but also have their own structure on the Internet. These companies have changed the producer-consumer relationship and are reducing the distance between them.

Today, any customer can contact the manufacturer directly, ask their question and make a purchase, avoiding extra charges and saving time. Renowned manufacturers have recognized the need to develop their own D2C strategies based on analytics. The authors of the paper acknowledge that

the use of D2C opens up additional opportunities for companies. According to the authors, Nike is a prime example, with D2C sales accounting for a third of total revenues by the end of 2020 based on its Consumer Direct strategy.

Given the widespread use of digital marketing, author Rimma Katz in paper [3] explored social commerce, which fostered development along the D2C model. Today, social commerce is used to increase the reach of consumers, those who may know about direct contact with the manufacturer.

So in the presented report by Jasmine Enberg (Updating social networks for the 4th quarter of 2020 - https://www.emarketer.com/content/social-media-pdate -q4-2020), a summary of the main events, their analysis for marketing management, solving problems strategic development of companies. Report author Jasmine Enberg determines that the global forecast for monthly social media users in 2020 has increased due to the effects of the pandemic. However, no platform will be able to maintain the growth it picked up at the beginning of the year. Therefore, in 2021, the growth rate will begin to normalize. Recent product launches, including Facebook live shopping and Instagram shopping tags, show that e-commerce continues to be a priority for the two platforms. Snapchat and Twitter focused more on effective marketing, namely the release of new sets of promotional offers with direct consumer response and others.

Therefore, the modern companies are based on direct selling (D2C) models. That allows you to get images of buyers and segment them. In addition to these tasks, it is necessary to solve forecasting the market, which changes every year.

It should be noted that worldwide targeted statistics for the entire sales system. This information is stored in cloud storage (Big Data). The information used includes data from the time of attracting a new consumer to the required resource information about the number, including repeated ones.

In this paper, we propose a neural network analytics method to forecast the main parameters for companies.

The innovation in the application of neural networks is the formation and application of neural network learning matrices. The peculiarity of these matrices is that their values have a random relationship with each other. For example, each family buys a conditional loaf of bread a day. This sales process has a pattern and can be attributed to a deterministic process. The next day, the family buys 0.5 white and 0.5 gray, and so on in different proportions. This sales process is random and belongs to a stochastic process.

2. Formal problem statement

Making forecasts is part of any firm's overall analytics. An important role in the forecasting method is paid by the multidimensionality of information and methods of their processing. The use of Big Data with OLAP technologies requires new approaches to processing and applying large amounts of data. This is due to the wide range of communication systems used in the novated companies.

Therefore, neural network forecasting based on Big Data is an urgent task, which is discussed in this paper.

3. Literature review

Modern companies use many available forecasting methods, they not only improve the quality of their products, but also get information about the needs of customers. Neural network forecasting models are a great way to predict customer preferences and apply new ways to stand out from the competition. Using practical forecasting models today is the best way to get the most effective and complete data to improve decisions. In this case, forecasting methods in digital economy should include not only customer surveys, their age, interests and price, but also the characteristics of the product, brand, logistics and others.

In the work of B. Yonathan [4], the problem of fitting mathematical models to numerical data was considered. Such a fit is often performed by the least squares method, regardless of previous knowledge of parameter values or the statistical nature of measurement errors.

According to the results of a study of the opinions of experts, A. Ashton [5] considered a scenario in which the results of the opinions of experts differ significantly from the polls of intentions. For this case, the author considered the problem of predicting market behavior.

Further development of the principle of intentions was found in the work of V. Morwitz [6]. The author proposed the principles of using intentions when solving the forecasting problem. In this paper is included research people behavior and how they would solve problems in different situations. For this, the method of polling the intentions of people was used. Intent surveys are widely used in marketing when sales data is unknown, for example, to forecast new products.

A continuation of these works found themselves in the work of J. S. Armstrong [7], where the role of a person as a dominant factor was considered. This task was solved as a role-playing game for making predictions of the behavior of people who interact with others. A key tenet of this approach is to provide realistic simulation of interactions.

This forecasting method is currently rarely used. G. Rowe and G.Wright [8] considered the application of the Delphi method as a procedure. The authors found that the accuracy of expert predictions can be improved through the use of Delphi structured methods. One of the principles of the method is that experts' forecasts should not depend on each other. Expert groups sometimes violate this principle; as a result, the data should not be used in forecasting.

D. Wittink and T. Bergestuen [9] considered the problem about "Intentions" in their work. This paper examines the intention as an indicator of a consumer to purchase a product under the influence of various factors. The consumer can declare his intentions to make a purchase of various goods. This method is based on the following principles, namely, using a new design to create an acceptable situation.

The formation of a digital marketing strategy was considered in the work of P. Mandal and N. Joshi [10]. The authors of the paper emphasize that digital technologies make marketing more effective, since it allows to identify individual consumer interests and how to make better manage campaigns and improve the product. In this paper, the authors propose a flowchart for developing marketing strategies.

M. Ivanov [11] discusses the analysis of the state of the digital economy and digital marketing in his paper. The author in the paper shows that the dynamics of processes in the economy is quite high and requires a quick analysis of multidimensional data. The author proposes a conceptual model and a method for assessing consumer demand in the target market, aimed at the prospective management of trading floors using Big Data.

4. The purpose of the paper

The paper is devoted to neural network forecasting based on the use of neural networks. The information base of neural network forecasting is Big Data. The process of constructing training matrices, training a neural network and making the predictions are presented in the result of paper.

5. Neural network forecasting modeling

The modern economy is characterized by the rapid dynamics of economic processes. Under these conditions, forecasting models of neural networks acquire new significance in the process of making managerial decisions. The process of the importance of making management decisions in digital marketing systems is shown in Figure 1.

In this Fig. 1 in the field of traditional marketing analytics (MTA), which is based on the use of classical approaches and forecasting methods. MTA is aimed at solving economic forecasting problems, where the time for processing and obtaining forecast results takes a long time. This time for solving forecasting problems can range from several days to a month or a year, which is incommensurate with the dynamic processes in the digital economy. Therefore, the forecast results become outdated, lose their relevance and can lead to negative results.

Therefore, decisions related to forecasting processes for a short period of time are an urgent task. The scope of this problem is depicted in Figure 1 as Marketing Actual Analytics (MRA). Marketing

Actual Analytics (MRA) and is aimed at solving the problem of predicting fast processes, namely from the moment information appears up to several hours.

Today the number of information sources of data in the world is growing rapidly.
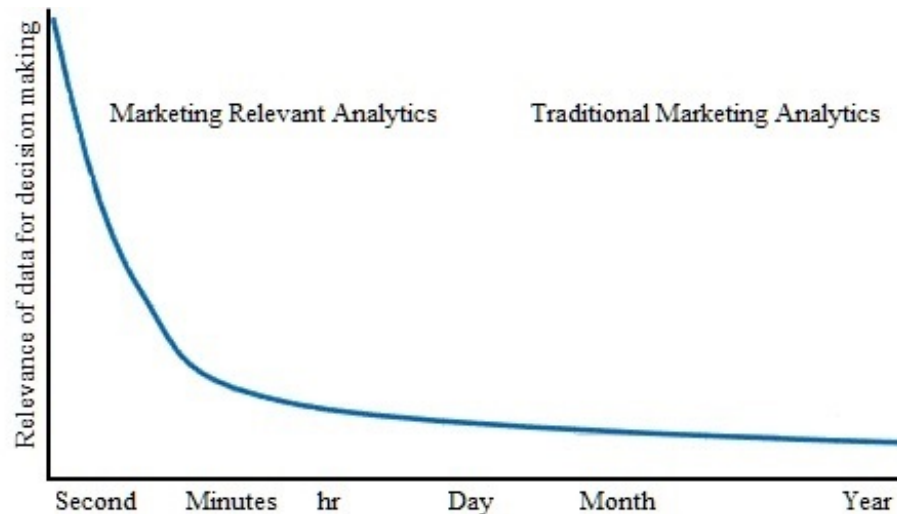

Figure 1: The importance of management decision making in digital marketing

Therefore, storage technologies and their processing of information are becoming more and more in demand. By storing information, one can single out the use of Big Data for which the basic principles of work can be formulated:

1. Horizontal scalability, which takes into account that the data can be arbitrarily large from any system. They have the ability to handle big data.

2. Tolerance to failures, which use the principles of horizontal scalability and apply methods of clustering systems; locality of data, which allows in large distributed systems to separate data from a large number of data centers. All modern tools for working with big data, one way or another, follow these three principles. The first principle is based on the MapReduce model. The MapReduce model provides for distributed data processing proposed by Google and is shown in Fig. 2.
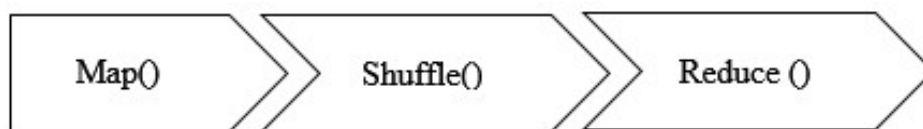

Figure 2: Data processing according to the MapReduce model

MapReduce provides that data is organized as relational or multidimensional data (OLAP). The data processing method takes place in three stages. The first stage is aimed at executing the Map () function. At this stage, the data is preprocessed using the Map () function, which is defined by the user. The work of this stage is to preprocess and filter the data.

The second stage of the model is performed by the Shuffle () function. This stage goes unnoticed by the user. At this stage, the Map () function performs the data immersion procedure similarly to the formation of data marts (Data Mart), that is, one Map () data output corresponds to each mart. In the future, these showcases will serve as an input for the Reduce () function.

The third stage of the model is aimed at executing the Reduce () function. Each data mart, which is formed in the second stage, transfers information to the input of the Reduce () function. The Reduce () function is user defined and calculates the result for individual storefronts. The set of all values returned by Reduce () is the result in this method. Therefore, Big Data technology is consider as a tool that allows you to increase the speed of data transfer while providing a large capacity of information carriers.

In addition, this technology can improve the availability of cloud applications and data services. Thus, digital marketing is shape around the mainstream e-commerce models. The interconnection of

the main models (B2B, B2A, D2C, C2A and C2C) of e-commerce systems based on the systems for collecting, storing and analyzing information in real time, which based on subsequent storage in historical data layers.

For the implementation of systems that perform Marketing Relevant Analytical tasks using data, OLAP data systems are used, which are structured according to the principle of multidimensional information presentation [11]. Reducing the cost of creating multidimensional warehouses can be achieve by using Data Mart. A data mart can only contain thematically aggregated data. Big Data is today a single, centralized source of information for the entire subject area. The structure of the marketing analytical system can be represent as follows (Fig. 3).

In a marketing analytical system, there are many databases. where transaction processing is done in real time. Therefore, online data source systems (ODS) provide information for processing in OLTP. OLTP systems provide storage and processing of information in real time. The processed data in OLTP is transferred to the Data Mart systems with the subsequent construction of multidimensional OLAP data cubes.

This multidimensional data is aimed at presenting information on thematic sections both on marketing information and other information from different areas of the economy.
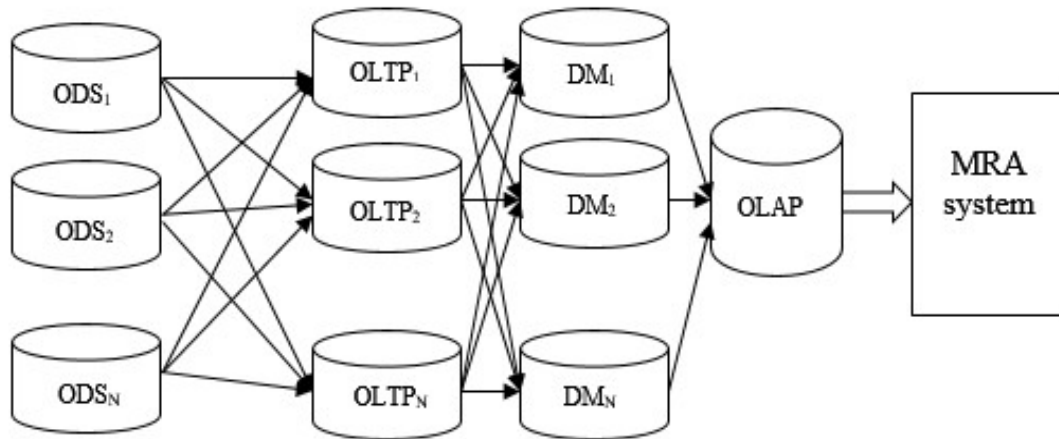

Figure 3: The structure of the marketing analytical system

The marketer has the ability to access multidimensional data in the repository, as well as complete economic information for conducting an MRA. The advantages of this approach are:

- simplicity of creating and filling OLAP, since filling comes from reliable sources of data marts;

- reducing the load on working with multidimensional data, namely, one multidimensional query processes multiple OLAP layers.

The data coming from the OID is transferred to the OLTP and the data marts are moved. OLAP stores data as multidimensional layers of measures and dimensions [11].

For neural network forecasting, a multidimensional query is formed to multidimensional data, which allows obtaining the following information input stream (Inflow) - formed by data from OLTP and DMN subsystems:

$$I = \{p_j = (g_j, in_j, mi_j)\}, \ j = \overline{1, N}, \tag{1}$$

where g_j is the product included in the analyzed sets of N — the object of research;

in_j — indicator of income j of the product;

mi_j — product indicators j.

Datasets from set I stored in OLAP and on demand allow you to obtain and conduct analysis with the subsequent storage of data, which are called transactions. Description of a transaction to set I as follows:

$$T = \{in_j | in_N \in I\}. \tag{2}$$

Such transactions for retail outlets on the Internet correspond to the nomenclature of goods that the consumer buys and the data stored in OLAP as multidimensional data cubes (OLAP).

Then solving the problem of neural network forecasting, data arrays of its training are formed. The forecasting technique using a neural network is formalized through the problem of pattern recognition. Data on the predicted economic indicators of a product for a certain period of time form an image, the class of which is determined by the values of the predicted indicators.

In the proposed methodology, the dimension of the multidimensional array will determine both the forecasting interval and the number of predicted indicators. Each next line of the array is formed as a result of a shift by one interval equal to the prediction interval.

The neural network is trained on the generated training array of product indicators and adjusts its weights accordingly. As a result, the neural network is trained to solve the forecasting problem for a certain forecasting horizon. It should be noted that two forecasting approaches are used: one-step and multi-step. One-step forecasting is used for short-term forecasts and multistep forecasting is used for long-term forecasting.

In general, the forecasting model can be represented as follows. Let the time interval $[t_0, t_k]$ be given, the indicators g_j, in_j, pr_j of the product are defined, where t_0 is the initial time value, t_k is the current time value. To find the predicted values on the prediction interval Δ, a method is proposed that includes the following stages:
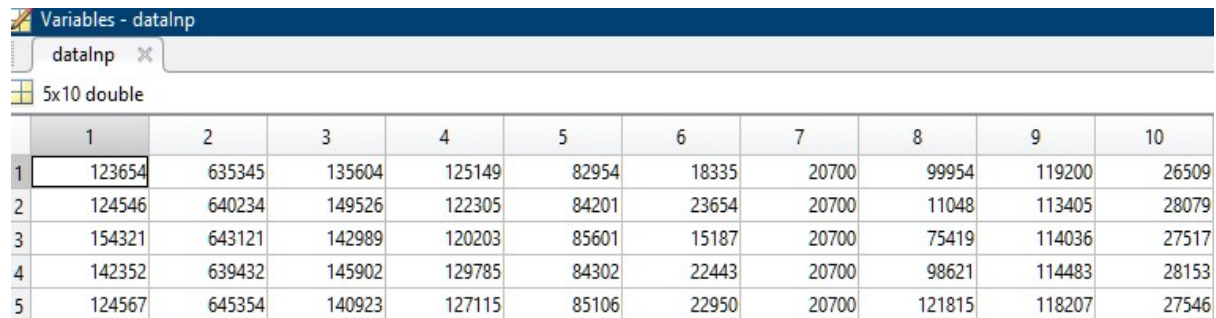
1. Analytical analysis of indicators and the formation of a learning matrix from selected values from historical slices of multidimensional databases (OLAP technology).

A learning matrix (ML - matrix learning) can be written as input data for representation in a neural network:

$$ML_{Input} = \begin{bmatrix} in_{01} = f_{01}(t_0) & in_{02} = f_{02}(t_0 + \Delta) \cdots & in_{0m} = f_{0m}(t_0 + (m-1)\Delta) \\ in_{11} = f_{11}(t_0) & in_{12} = f_{12}(t_0 + \Delta) \cdots & in_{1m} = f_{1m}(t_0 + (m-1)\Delta) \\ \cdots & \cdots & \cdots \\ in_{N1} = f_{N1}(t_0) & in_{N2} = f_{N2}(t_0 + \Delta) \cdots & in_{Nm} = f_{Nm}(t_0 + (m-1)\Delta) \end{bmatrix}, m = \overline{1, k}, \quad (3)$$

where Δ — is the horizon (time interval) of forecasting.

The input data in the neural network in Matlab is presented in the form of a 5x10 matrix, which represents static information on 10 products for five days of the week and has the following form (Fig. 4).

Variables - dataInp

dataInp ✕

5x10 double

	1	2	3	4	5	6	7	8	9	10
1	123654	635345	135604	125149	82954	18335	20700	99954	119200	26509
2	124546	640234	149526	122305	84201	23654	20700	11048	113405	28079
3	154321	643121	142989	120203	85601	15187	20700	75419	114036	27517
4	142352	639432	145902	129785	84302	22443	20700	98621	114483	28153
5	124567	645354	140923	127115	85106	22950	20700	121815	118207	27546

Figure 4: "BataInp" matrix view

Target data that determine the output statistical information for the neural output can be represented in the form of learning value matrix (TVM_{Output}):

$$TVM_{Output} = \begin{bmatrix} in_{01} = f_{01}(t_0 + \Delta) & in_{02} = f_{02}(t_0 + (k+1)\Delta) \cdots & in_{0m} = f_{0m}(t_0 + m\Delta) \\ in_{11} = f_{11}(t_0 + \Delta) & in_{12} = f_{12}(t_0 + (k+1)\Delta) \cdots & in_{1m} = f_{1m}(t_0 + m\Delta) \\ \cdots & \cdots & \cdots \\ in_{N1} = f_{N1}(t_0 + \Delta) & in_{N2} = f_{N2}(t_0 + (k+1)\Delta) \cdots & in_{Nm} = f_{Nm}(t_0 + m\Delta) \end{bmatrix}. \quad (4)$$

The target data, which determines the output statistical information in Matlab for neural output, is presented in Fig. 5.

	1	2	3	4	5	6	7	8	9	10
1	124546	640234	149526	122305	84201	23654	20700	11048	113405	28079
2	154321	643121	142989	120203	85601	15187	20700	75419	114036	27517
3	142352	639432	145902	129785	84302	22443	20700	98621	114483	28153
4	124567	645354	140923	127115	85106	22950	20700	121815	118207	27546
5	173876	647811	142989	121069	83241	23154	20650	97103	119905	27561

Figure 5: Target Output Matrix "DataOut"

2. NS training. The process of training a neural network is to match to each ML_{Input} element the value of the TVM_{Output} matrix corresponding to the mapping in the value of the elements of the weight matrix w_j:

$$w_j: ML_{Input} \rightarrow TVM_{Output} \ . \tag{5}$$

In the process of training the neural network, the task of minimizing the objective function is solved. With this approach, an algorithm is used for training, which is the most efficient not only in terms of errors, but also in time. The neural network in Matlab is trained using the Levenberg-Marquardt error backpropagation algorithm. The Levenberg-Marquardt algorithm uses a scalable conjugate gradient backpropagation.

Therefore, the training of the neural network is represented in time, and the network is tuned in accordance with its error. The magnitude parameter is used to measure the generalization of the neural network and stop learning when the generalization stops improving. The test score itself does not affect training and provides an independent assessment of the performance of the neural network during and after training.

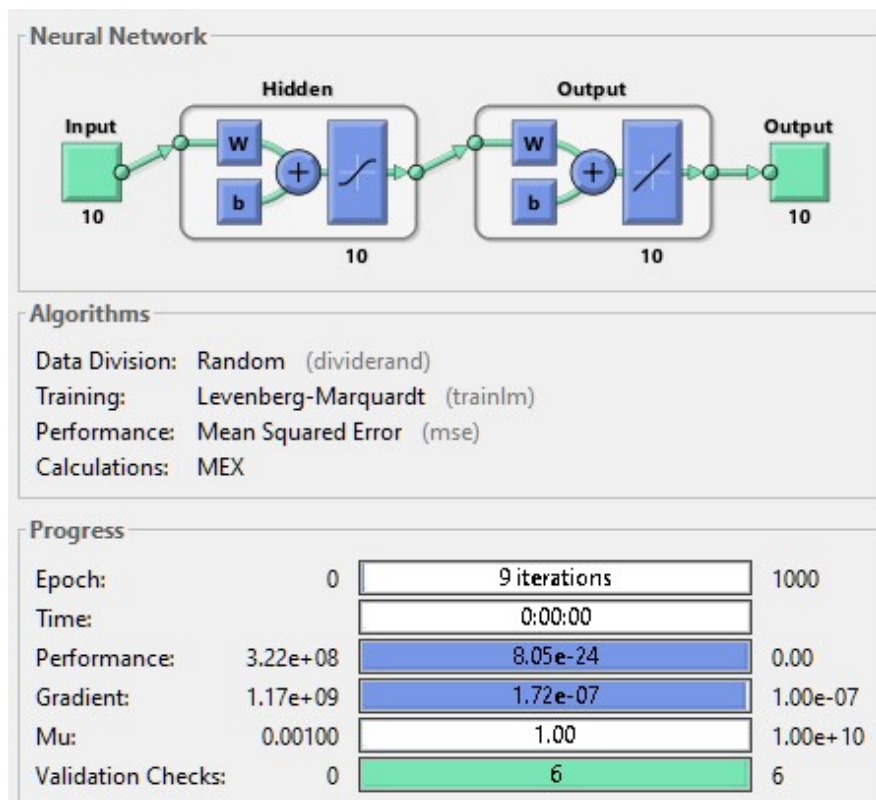The choice of the algorithm, as well as the learning process of the neural network, is shown in Fig. 6-7.

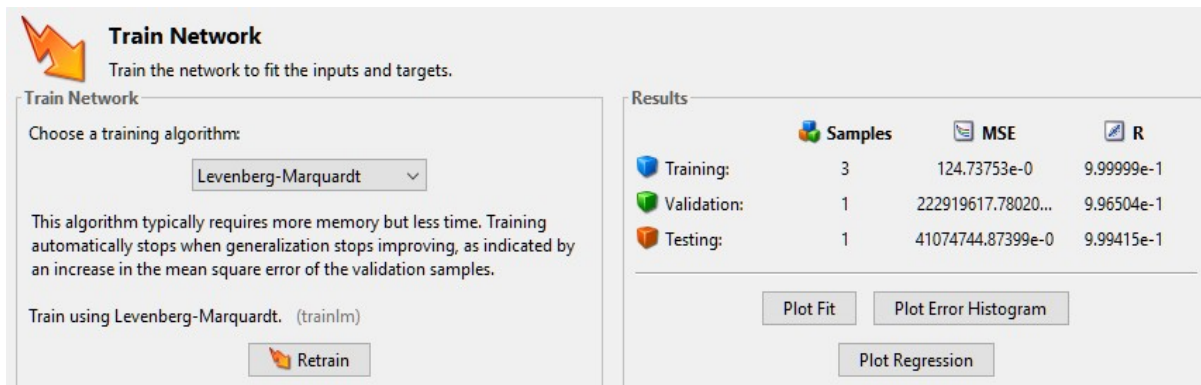
Figure 6: The process of learning the neural network

Figure 7: The choice of the learning algorithm

The number of neural network training epochs can be written as follows:

$$epochs = \left| \frac{t_k - t_0}{\Delta} \right|. \tag{6}$$

In the process of forecasting by a neural network, it is necessary to take into account the forecasting horizon. In the Matlab system, the sim (net,[;;]) function is implemented, which allows you to supply a variety of input values and get a solution at the output of a neural network. The forecast results for the sale of 10 goods are shown in Fig. 8.

Therefore, the created neural network does indeed make multiple predictive decisions. It allows you to solve the sales function of marketing and consider the dynamics of the sale of many products in real time. Neural network forecasting looks at the number of future periods that the forecast will cover. That is, you may need a forecast 7 days ahead, with data for every day. In this case, the period is a day, and the horizon is 7 days.

Finally, the prediction interval is the frequency with which a new prediction is made. Often the prediction interval coincides with the prediction period. The choice of the forecasting period and horizon is usually dictated by the conditions for making decisions.

Figure 8: The result of forecasting the sale of 10 products

Choosing these two parameters is one of the hardest parts of neural network forecasting. For forecasting to be meaningful, the forecasting horizon must be no less than the time required to implement the decision made on the basis of the forecast.

Thus, forecasting is highly dependent on the nature of the decision being made.

6. Conclusion

In this paper, the authors consider the solution to the problem of forecasting a neural network and the use of Big Data as a tool to increase the data transfer rate while providing access to multidimensional data (OLAP). The paper proposes the structure of a neural network for solving the forecasting problem, which uses the training matrices of this network. Matrix data can be built from data that provides information in OLAP.

The paper proposes the use of a neural network for predicting multidimensional data, which is built in the Matlab system. To solve the forecasting problem, the authors proposed the construction of the input and target data matrices for training a neural network. The paper also presents the construction procedures and the results of the neural network training process in Matlab.

The obtained forecasting results allow us to conclude about the advantages of a neural network for multivariate data prediction. Multidimensional data and their level of detail are important for solving the forecasting problem.

The wider use of digital systems makes it possible to apply the proposed approach to forecasting. It should be noted that the use of neural networks in forecasting could have the following disadvantages. So, in the process of forming training matrices, if there are no data in the initial information or data presented in an indistinct form (a linguistic variable, the appearance of a new product or there is no demand for a product), then an error may appear.

Therefore, this drawback in solving the forecasting problem can be eliminated if systems with fuzzy sets are used, which will allow to formalize fuzzy variables.

7. References

[1] L. Florès, How to Measure Digital Marketing, Springer, London, 2014.
[2] Kl. Schwab (Ed.), The Fourth Industrial Revolution, World Economic Forum, Cologny/Geneva, 2017.
[3] R. Kats, Mucinex Eyes Social Commerce to Help Bolster D2C Business, emarketer Web, 2020. URL: https://www.emarketer.com/content/mucinex-eyes-social-commerce-help-bolster-d2c-business/.
[4] B. Yonathan (Ed.), Nonlinear Parameter Estimation, New York, 1974.
[5] A.H. Ashton, Does consensus imply accuracy in accounting studies of decision making?, The Accounting Review, vol. 60, no. 2 (1985) 173–185. URL: www.jstor.org/stable/246784.
[6] V. Morwitz, Methods for forecasting from intentions data, Principles of Forecasting (2001) 33-56. doi:10.1007/978-0-306-47630-3_3
[7] J.S. Armstrong (ed.), Role Playing: A Method to Forecast Decisions, Springer, Boston, 2001, pp. 15-30. URL: https://repository.upenn.edu/marketing_papers/152.
[8] D.R. Wittink, T. Bergestuen, Forecasting with conjoint analysis, In: Armstrong J.S. (eds), Principles of Forecasting, International Series in Operations Research & Management Science, volume 30, Springer, Boston, MA, 2001, pp. 147-167. doi:10.1007/978-0-306-47630-3_8
[9] G. Rowe, G. Wright, Expert opinions in forecasting role of the Delphi technique, Springer, Boston, 2001, pp. 125-144. doi:10.1007/978-0-306-47630-3_7
[10] P. Mandal, N. Joshi, Understanding Digital Marketing Strategy, International Journal of Scientific Research and Management 5(6) (2017) 5428-5431. doi:10.18535/ijsrm/v5i6.11
[11] M. Ivanov, Cloud-based Digital Marketing, CEUR Workshop Proceedings, 2422, 2019.