

Expanding the intellectual capabilities of OLAP technology using neural networks

Gulnara Nabibayova

Institute of Information Technology, B. Vahabzade str., 9A, AZ1141, Baku, Azerbaijan

gnabibayova@gmail.com

orcid.org/0000-0001-8743-7579

ARTICLE INFO

<http://doi.org/10.25045/jpis.v15.i2.05>

Article history:

Received 09 January 2024

Received in revised form

10 March 2024

Accepted 15 May 2024

Keywords:

Data warehouse

Olap

Artificial intelligence

Machine learning

Neural network

Forecasting

Clustering

Classification

ABSTRACT

The article highlights the main characteristics, features and structure of Online Analytical Processing systems based on the same technology that perform online analytical processing of data. This technology allows analysts to explore and navigate a multidimensional indicator structure called an online analytical processing cube (data cube). Indicators (measures) of data cube play an important role in the decision-making process. To solve certain problems, these measures often need to be classified or grouped. Moreover, empty measures are common in data cube. This fact negatively affects strategic decision making. Unfortunately, online analytical processing itself is not well suited for classifying, clustering, and predicting empty measures of data cube in the presence of large data. In this regard, today there is a need to use new technologies to solve such problems. Such technologies include neural networks. The article discusses the problem of integrating online analytical processing and a neural network, showing the possibilities and advantages of such integration. It mentions that in the case of big data, the integration of OLAP and neural networks is very effective in solving problems of classification, clustering and empty measure prediction of data cube. An architectural and technological model for the integration of online analytical processing and neural networks is presented.

1. Introduction

People have always been eager to create all the necessary conveniences for everyday life for themselves. As a result of their efforts, proper technologies and machines appeared to assist them. One of the areas of such technologies is Artificial Intelligence (AI) (Iskanderova, 2023). The term "Artificial Intelligence" was first coined in history in 1956 by John McCarthy. The term refers to any system capable of performing creative functions and solving problems that typically require human intelligence, that is, which are traditionally performed by humans. In effect, AI imitates human intelligence in machines that are programmed to think and act like humans. AI plays a critical role in today's world by enabling automation, improving decision making, increasing efficiency and productivity. It opens up

new opportunities for innovation and growth in a variety of industries, including healthcare, finance, manufacturing, transportation, e-commerce, education, and more.

AI is frequently discussed and explored in parallel with machine learning (ML). ML is a branch of AI. The idea of ML is that machines should be able to learn and adapt through experience, making predictions based on statistical data collected by computers. Thus, the concept of AI is a broader concept compared to the concept of ML.

One of the areas of artificial intelligence is neural networks (NN). NNs are used to recognize hidden patterns in raw data, for clustering (Yunusova and Magsumova, 2019) and classification (Artikova and Rasulova, 2021), as well as for solving tasks in the field of AI.

Recently, both government agencies and private ones have widely used OLAP systems based on OLAP (Online Analytical Processing) technology

(What is Online Analytical Processing (OLAP)?). They are applied, for example, in banking, medicine, industry, telecommunications, trade, etc. (Alguliyev et al., 2019) describes the use of OLAP in the terminology environment, namely in the terminology system to expand its capabilities and for more efficient functioning.

NN and OLAP are essential tools for quickly and efficiently discovering valuable, non-obvious information from a large collection of data.

The goal of this paper is to study the possibility of integrating OLAP and NN, to identify the benefits of this integration and to develop a model for integrating NN and OLAP.

The second section reviews related studies. The third section describes the main characteristics of OLAP systems as methods and materials used. The fourth section provides information about NS: historical background, types of NS and their functions. The fifth section is devoted to the problem of integrating NN and OLAP and presents an architectural and technological model for integrating OLAP and NN. Then a discussion on this topic is presented. Finally, the Conclusion section presents the final conclusions of this paper.

2. Related work

In recent years, there has been a need to add new capabilities to OLAP. First of all, this is due to the fact that when solving some tasks, problems arise due to the presence of large data entering the systems from different sources. Big data refers to huge volumes of heterogeneous and rapidly generated digital information that cannot be processed with traditional tools. The use of big data is critical in various fields including healthcare, finance, education, government and many others.

On the other hand, it should be noted that today there is a need to solve various kinds of intelligent problems. Intelligent tasks based on big data are most effectively solved using NN. Using NN, big data can be analyzed efficiently. It is very important that the advantage of NN, such as the detection of hidden patterns that are invisible to humans, also works well on big data. The integration of OLAP and NN provides these benefits.

When there is a large amount of data in intelligent systems using OLAP NNs are effective for solving clustering and classification tasks. Note that clustering, unlike classification, does not have predefined categories into which all data should be grouped. In this case, the NN itself generates clusters based on common features of the data.

Clustering is one of the most important methods of data analysis. The article (Du, 2010) provides a comprehensive overview of clustering methods such as the self-organizing Kohonen map, as well as clustering algorithms such as k-means, fuzzy means algorithms, etc.

In (Kohonen neural network, self-organizing maps, learning, 2022) the Kohonen neural network is also noted as one of the NN classes primarily used to solve clustering problems.

k-means is the most popular and simplest clustering method. Its main disadvantages are: the number of clusters need to be known in advance; very sensitive to the choice of initial cluster centers.

In the decision-making process, you can encounter many fuzzy tasks. Therefore, the queries to the DW that the analyst is trying to formulate may often contain uncertainties. Clustering applied to the dimensions of an OLAP cube using NN produces linguistic variables. This will make it possible to solve fuzzy tasks as well (Kumar K., Krishna and Kumar, 2005).

For data mining, the task of data classification plays an important role. Currently, a large number of different types of classifiers have been developed, including those built on machine learning. These include NN. Although the classification task for NNs is not the main one, their use has a number of advantages:

- NNs are self-learning models, the operation of which requires almost no user intervention;
- NNs are universal approximators that allow you to approximate any continuous function with suitable accuracy;
- NNs are nonlinear models. This allows you to effectively solve classification tasks even in the absence of linear separability of classes (Fig. 1) (Oreshko, 2021).

By means of NN, forecasting problems that are of great importance in the production, economic and financial spheres are also solved. Forecasting in OLAP is important because when looking at the contents of a cube, it can often be sparse, meaning it is missing some measures, and may also be missing dimensions. This happens due to missing information or input errors. The absence of any measures and measurements is undesirable and can lead to incorrect analysis when making strategic decisions.

In OLAP systems, NN can be used in parallel with OLAP, i.e., OLAP cubes are created on historical DW data, and NN forecasting work is based on the same historical data (Gulesian, 2008).

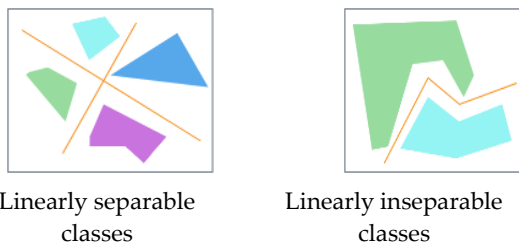


Fig. 1. Options for linear separability of classes

The disadvantage of this approach is that there are no training data sets.

The approach proposed in (Abdelbaki et al., 2012) includes two stages. First, principal components are analyzed to reduce the dimensionality of the data cube and special training sets are created. Then, in the second stage, a new OLAP-oriented multi-layer perceptron network (MLP) architecture is proposed whereby training is implemented on each training set and predicted dimensions are generated.

In (Abdelbaki et al., 2015), the possibilities proposed in (Abdelbaki et al., 2012) are expanded. First, the authors introduce a generalized framework, i.e., Multi-perspectives Cube Exploration Framework (MCEF), for applying the classical data mining algorithm to OLAP cubes. Secondly, the authors refer to modular NNs that apply a neural approach to predicting multidimensional cubes (NAP-NN). Modular NNs are a collection of several different networks that operate independently and contribute to the result. Each NN has its own set of input data. These networks do not interact with each other during task execution. The main advantage of modular NN is that the huge computational process can be divided into several subprocesses. This reduces computational complexity and increases computational speed. But ultimately, the processing time will depend on the number of neurons and their participation in calculating the results. Note that NAP-NN includes a preprocessing step. In this step, principal component analysis (PCA) is performed to reduce the size of the OLAP cube of the proposed method. Modular neural networks work effectively in cases where several directions of the system are simultaneously processed. Note that the article presents experimental results showing the effectiveness of NN.

Thus, it can be argued that the integration of two technologies such as OLAP and NN is very useful and important, as it enriches each of them: OLAP is the ability to navigate the multidimensional structure of indicators, and NN is the ability to intelligently solve tasks on large amounts of data.

3. Methods and materials

3.1. Main characteristics of OLAP systems

An OLAP system is an information and analytical data processing system developed based on OLAP technologies. The popularity of OLAP is explained by the fact that it is possible to solve many problems with its help, namely: to implement operational processing of information, including issuing information in various sections and dynamic report generation and its analysis based on the data obtained, to perform monitoring and forecasting (Codd et al., 1993). The OLAP system is designed for generating reports, constructing predictive scenarios and performing statistical calculations based on large collection of data with a complex structure (Main characteristics of OLAP systems, 2020). The key components of an OLAP system are a data warehouse (DW), an OLAP server and applications.

A DW is a source of processed information accumulated from already existing systems of geographically distributed units. A DW is a domain-specific, non-volatile, integrated, time-varying set of data for decision support (Inmon, 2005). OLAP is an element of the DW and takes advantage of its information.

The OLAP server is the core of the system, with the help of which multidimensional data structures are processed and communication between the DW and system users is ensured.

Applications are used for user work. They formulate queries and visualize the responses received. OLAP applications are used to store DW analysis contexts in multidimensional data structures, i.e., in OLAP cubes. OLAP cubes enable analysts to explore information and report through interactive, easy-to-use dashboards. It is OLAP cubes that contain indicators (measures) used for analysis and management decision-making. One of the important goals of OLAP is to make decisions based on historical data.

Note that OLAP provides any analytical report within a few seconds due to its wide visualization functionality.

4. Neural network as a method in artificial intelligence

The basic principles of NN operation were described back in 1943 by Warren McCulloch and Walter Pitts (McCulloch and Pitts, 1943). In 1957, neuroscientist Frank Rosenblatt developed the first

NN. He was the author of the first paper on perceptrons (Rosenblatt, 1958).

Below the main types of neural networks are considered to determine which problems it is more effective to use each of them for.

Each NN includes a first layer of neurons called the input layer. This layer does not perform any transformations or calculations; it receives and distributes input signals to other neurons. This layer is the only one that is common to all NN types.

Next, we consider the types of neural networks in order to determine which of them and for what purpose will be used in the developed model.

The main types of NN are as follows:

Perceptron. A perceptron is a fully connected artificial NN in which each neuron is connected to all neurons of the previous layer. Each connection has its own personal weighting coefficient. A perceptron is a single-layer or multi-layer feed-forward NN with binary or analog outputs, supervised learning.

Single layer neural network. It is a structure for the interaction of neurons, in which signals from the input layer are immediately sent to the output layer (Fig. 2).

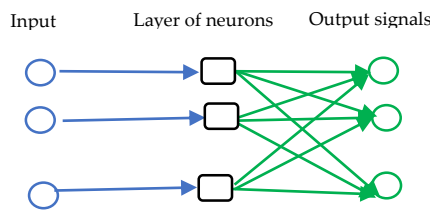


Fig. 2. Example of single layer neural network

Multilayer neural network. This NN, in addition to the output and input layers, has several hidden intermediate layers (Fig. 3).

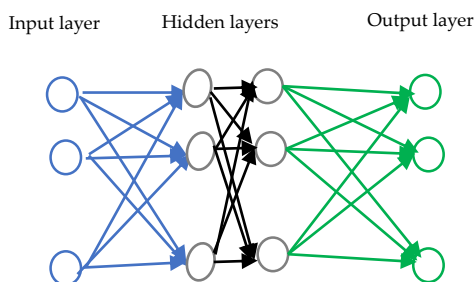


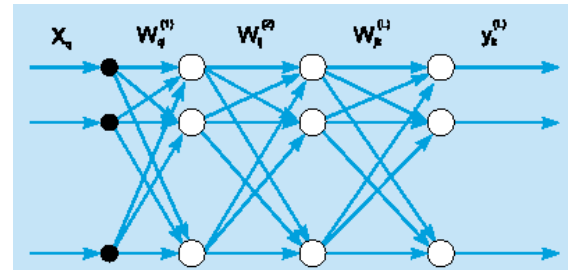
Fig. 3. Example of a multilayer neural network

The number of these layers depends on the complexity of the NN.

NN can be classified not only for the number of layers, but also according to the direction of information distribution along synapses (connections) between neurons:

Feed-forward neural network (unidirectional).

In this structure (Fig. 4.) the signal moves strictly in the direction from the input layer to the output layer.



Input layer Hidden layers Output layer

Fig. 4. A three-layer feed-forward neural network

Recurrent neural networks (with feedback) (RNN).

Here the signal moves both forward and backward. As a result, the output result can be returned to the input. Types of RNN:

- One to one
- One to many
- Many to one
- many to many

Self-organizing maps. These include self-organizing Kohonen maps. They are a powerful, self-learning clustering engine: the results are displayed in compact and easy-to-interpret two-dimensional maps. Thus, Kohonen maps combine two important data analysis paradigms - clustering and projection, that is, visualization of multidimensional data on a plane. The Kohonen map is used for exploratory data analysis. It is able to recognize clusters in data and also establish class proximity. The Kohonen network has only two layers: input and output, composed of radial neurons of an ordered structure. The output layer neurons are located at the nodes of a two-dimensional grid with rectangular or hexagonal cells. The number of neurons in the grid determines the degree of detail of the algorithm's result, and, ultimately, the accuracy of the generalizing ability of the map depends on it.

In addition, the Kohonen card is able to predict client behavior. If it is built a Kohonen map containing clusters for each group of clients according to their degree of loyalty, then with its help the expected behavior of the client can be predicted and applied appropriate marketing policies to them.

The Kohonen map is also capable of detecting anomalies. It distinguishes clusters in the training data and assigns all data to one cluster or another.

If after this the map encounters a data set that is unlike any of the known samples, then it will not be able to classify such a set and thereby reveal its anomaly (Kohonen map).

There are also other criteria for NN classification:

- depending on the types of neurons: homogeneous and hybrid;
- depending on the NN learning method: supervised learning, unsupervised learning, reinforcement learning;
- according to the type of input information, NNs are: analogous (use information in the form of real numbers), binary (operate with information presented in binary form); figurative (operate with information presented in the form of images, signs, hieroglyphs, symbols);
- according to the nature of synapse setup: with fixed connections (NN weight coefficients are selected immediately based on the conditions of the problem, with $dW/dt=0$, where W denotes NN weight coefficients); with dynamic connections (when the learning process is in progress in the settings of synaptic connections, that is, $dW/dt \neq 0$, where W denotes NN weight coefficients).

5. OLAP and neural networks integration model

Fig. 5 illustrates the architectural and technological model for integrating OLAP and NN. According to the figure, data from various sources,

before entering the DW or Data Mart (DM), first goes through ETL technology. During the ETL process, data is cleared of duplication, contradictions and typos and brought into a common format. OLAP cubes are built based on DW (or DM) data. The figure shows the integration of NN with an OLAP cube to classify and cluster OLAP cube data and predict empty measures. The integration of NN and OLAP is that the NN environment is built into OLAP applications that operate on a multidimensional structure and a large volume of data cubes.

Note that at this stage, the execution of processes characteristic of traditional OLAP is also ensured, namely: analytical queries are implemented on data for their rapid viewing and analysis, reports are issued based on the OLAP cube data, which can be with either intermediate or final results. It is also possible to view the same data from different angles.

Depending on the task set, the most appropriate NN is selected from the above types. For example,

- single-layer and multi-layer perceptrons are used for classification;
- single-layer and multi-layer perceptrons are used for classification;
- single-layer or multilayer perceptrons and Kohonen map are used for forecasting (Kohonen map, Neural networks, perceptron).

Finally, the results obtained will serve to make management decisions.

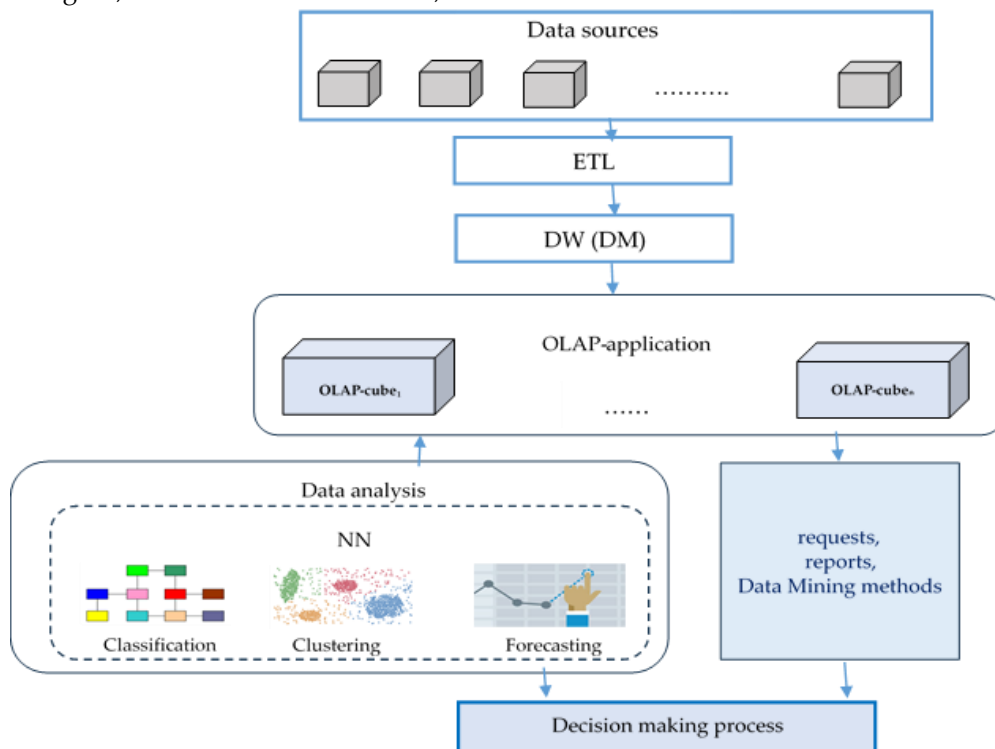


Fig. 5. Architectural and technological model of integration OLAP and NN

6. Discussion

The article examines the problem of integrating OLAP and NN. It is shown that the use of NN in OLAP systems expands the capabilities of OLAP in the context of big data and multiple data sources.

The article presents the developed conceptual architectural and technological model for the integration of OLAP and NN. It is noted that the integration of NN and OLAP is based on the fact that the NN environment is built into OLAP applications.

Future research will be devoted to the development and application of methods for introducing the NN environment into OLAP applications that operate with a multidimensional structure and a large volume of data cubes.

7. Conclusion

Currently, in the era of Industry 4.0, there is a dramatic increase in the flow of data. This creates great complications when solving some problems related, for example, to classification, data clustering, forecasting, and even in some cases makes solving these problems impossible. AI can provide quick and effective solutions to such problems. AI refers to any system capable of performing creative functions and solving problems that would typically require human intelligence. AI contributes to development and innovation in various industries, such as healthcare, finance, manufacturing, transport, e-commerce, education, etc. NNs are one of the areas of AI. NNs, being implemented into systems, can solve important tasks. Such systems include OLAP systems based on OLAP technology. Recently, OLAP systems have been widely used both in government agencies and in private ones.

The article analyzed related works, explored the possibility of integrating OLAP and NN, and showed the advantages of such integration.

The article presented the developed architectural and technological model for integrating NN and OLAP. According to this model, integration was carried out by embedding the NN framework into OLAP applications running on a multidimensional structure and a large volume of data cubes. As an example, the three most important problems were given, the solution of which on big data was most effective using NN.

References

- Abdelbaki, W., Messaoud, R., Yahia, S. (2012). Neural-Based Approach for Extending OLAP to Prediction, in: Proceedings of science conference Data Warehousing and Knowledge Discovery (DaWaK 2012), Springer-Verlag Berlin Heidelberg, 117-129.
- Abdelbaki, W., Yahia, S., Messaoud, R. (2015). Modular Neural Networks for Extending OLAP to Prediction. Book Transactions on Large-Scale Data- and Knowledge-Centered Systems XXI, Springer Nature, 73-93.
- Alguliyev, R. M., Nabibayova, G. Ch., Gurbanova, A. M. (2019). Development of a Decision Support System with the use of OLAP-Technologies in the National Terminological Information Environment, in: International Journal of Modern Education and Computer Science (IJMECS). 6, 43-52.
- Artikova, M., Rasulova, Sh. (2021). Data classification using neural networks, in: Scientific Collection «InterConf»: Scientific Goals and Purposes in XXI Century, (78), 403-409. <https://ojs.ukrlogos.in.ua/index.php/interconf/article/view/15124> [in Russian].
- Codd, E.F., Codd, S.B., Salley, C.T. (1993). Providing OLAP (on-line Analytical Processing) to User-analysts: An IT Mandate. http://www.estgv.ipv.pt/paginaspeessoais/jloureiro/esi_aid2_007_2008/fichas/codd.pdf.
- Du, K.-L. (2010). Clustering: A neural network approach, in: Neural Networks, vol. 23, issue 1, 89-107.
- Gulesian, M. (2008). Using Neural Networks and OLAP Tools to Make Business Decisions, <https://www.developer.com/database/using-neural-networks-and-olap-tools-to-make-business-decisions/>
- Inmon, W.H. (2005). Building the Data Warehouse, John Wiley & Sons, pp. 596.
- Iskanderova, Sh. (2023). The impact of artificial intelligence on the modern world, in: Science and Education Scientific Journal, Impact Factor 3.848. 4, 564-570. [in Russian]
- Kohonen map. <https://basegroup.ru/deductor/function/algorithm/kohonen> [in Russian].
- Kohonen neural network, self-organizing maps, learning, (2022). <https://microtechnics.ru/nejronnaya-set-kohonena-samoorganizuyushhiesya-karty-obuchenie/> [in Russian].
- Kumar, K., Krishna, R., Kumar, De S. (2005). Fuzzy OLAP Cube for Qualitative Analysis / Proceedings of the 3rd International Conference on Intelligent Sensing and Information Processing (ICISIP), 290-295, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1529464&tag=1
- Main characteristics of OLAP systems, (2020). <https://hsbi.hse.ru/articles/osnovnye-kharakteristiki-olap-sistem/> [in Russian].
- McCulloch, W., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, in: Bulletin of Mathematical Biology, 5, 115-133.
- Neural networks, perceptron. https://neerc.ifmo.ru/wiki/index.php?title=Нейронные_сети_и_перцептрон#cite_note-3 [in Russian].
- Oreshko, V. (2021). Data classification using neural networks. <https://loginom.ru/blog/neural-classification> [in Russian].
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain, Psychological Review, 65(6), 386-408.
- What is Online Analytical Processing (OLAP)? <https://aws.amazon.com/ru/what-is/olap/>. [in Russian].
- Yunusova, L., Magsumova, A. (2019). Clustering using neural networks and searching for dependencies, <https://cyberleninka.ru/article/n/klasterizatsiya-s-pomoschyu-neyronnyh-setey-i-poisk-zavisimostey-1>. [in Russian].