Complex Systems Informatics and Modeling Quarterly
CSIMQ, Issue 6, March/April 2016, Pages 1-11
Published online by RTU Press, https://csimq-journals.rtu.lv
http://dx.doi.org/10.7250/csimq.2016-6.01
ISSN: 2255-9922 online

Big Multidimensional Datasets Visualization Using Neural Networks – Efficient Decision Support

Gintautas Dzemyda, Viktor Medvedev, Audrone Lupeikiene,
Olga Kurasova and Albertas Caplinskas

Vilnius University, Institute of Mathematics and Informatics,
Akademijos str. 4, LT-08663 Vilnius, Lithuania

{Gintautas.Dzemyda, Viktor.Medvedev, Audrone.Lupeikiene,
Olga.Kurasova, Albertas.Caplinskas}@mii.vu.lt

Abstract. Nowadays business information systems are thought of as decision-oriented systems supported by different types of subsystems. Multidimensional data visualization is an essential part of such systems. As datasets tend to be increasingly large, more effective ways are required to display, analyze and interpret information they contain. Most of the classical visualization methods are unsuitable for large datasets. This paper focuses on the artificial neural networks-based methods for visualization of big multidimensional datasets; namely, on the approaches for the faster obtaining of visual results. The new strategy, which is identified by the decreased number of cycles of data reviews (passes of training data) up to the only one, when training neural networks, is proposed. To test this strategy, the results of experiments, using two unsupervised learning methods on benchmark data, are briefly presented.
Keywords: Data visualization, big multidimensional dataset, neural networks-based method, decision support.

1 Introduction

The features of today's world, such as globalization, dynamics and often unpredictable changes, huge amounts of data, are being observed on any of its entities. Even a generic philosophy concerning business information systems (BIS) is changing. Nowadays, they are thought of first of all as the decision-oriented systems supported by different types of subsystems. Managers are faced with a federated environment and the need to make time-critical decisions; consequently, data in BIS should be presented timely, in a meaningful manner and easily understandable form. Multidimensional data visualization is an essential constituent of such systems because it enables to discover knowledge hidden in big datasets. Dimensionality reduction is one of the basic operations in this context.

This paper focuses on visualization of big multidimensional datasets. As datasets tend to be increasingly large more effective ways are required to display, analyze and interpret the information contained within them. Given a large set of measured variables, the main task is to represent them with a smaller set of more "condensed" variables. Another reason for reducing the dimensionality is to decrease computational load for further data processing. However, the most of the conventional visualization methods are unsuitable for big multidimensional datasets.

Here the attention is directed to the artificial neural networks-based methods for visualization[1] of big multidimensional datasets. Two unsupervised learning methods are considered: SAMANN (a feed-forward neural network to learn Sammon's mapping) and SOM (self-organizing map). To cope with the data processing time problem, a new strategy is proposed to decrease the number of data passes (reviews) up to the only one when training neural networks. It is based on the assumption that a huge amount of data includes many similar objects, so even in one pass a neural network can access large amount of similar objects. After training, neural network can be used for decision-making support. In other words, any number of new objects can be converted to a meaningful form, i.e., presented as points on the plane to display interpretable results. This trained network can also be used to see the outliers. Regardless the fact that outliers are not representative, they may reveal potentially valuable information.

Empirical research has been carried out. The experiments using two unsupervised learning methods and on two sets of benchmark data have been conducted to test the hypothesis which explains the proposed strategy. The additional evaluation of the outcome has been done using the visualisation results of traditionally trained neural networks and the processing of outliers.

The structure of this paper is as follows. Section 2 discusses the context and positioning of BIS focuses on their main "duties" in the contemporary world. Section 3 considers visualization methods based on artificial neural networks and reviews the recent approaches to neural networks training by large data. Section 4 describes the new strategy of training through a single pass of data and performs an experimental investigation. Section 5 concludes the paper and comments unresolved problems.

2 Context: Efficient Decision Support

The concept of information system has significantly changed throughout its more than 50 year history. These changes reflect the system's role and importance in business enterprise and can be seen in the development approaches, methodologies, frameworks, architectural design decisions, and technologies.

Initially, management information systems (MIS) served business management. Their purpose was to cater to the information needs for planning, controlling and decision making. MIS is dependent on underlying transaction processing systems, but, in fact, can itself be thought of as a transaction processing system that can interact with a decision-support subsystem. From technological point of view, MIS was a set of applications centered on a database.

This philosophy has changed at the beginning of the 21th century when it was realized that an enterprise system should be developed as a whole [1], [2], [3]. Information system (IS) in this paper refers to a real world system which provides information services required to support business. IS is a component of an enterprise system and it should be aligned with the business's mission and goals, thus serving as critical success factor. Consequently, the whole enterprise system is viewed as a three-layered system: business systems, information systems, and supporting software.

Nowadays, information systems should be thought of first of all as a decision-oriented systems supported by different types of subsystems. This can be noticed when observing the relations between enterprise resource planning (ERP) and advanced planning and scheduling systems (APS) (ERP is one of the types of BIS). According to [4] and [5], planning and scheduling process is a primary aspect of decision making in manufacturing enterprises. APS system is not a part of ERP, but rather an entire planning and scheduling system within an enterprise supported by the ERP system (Figure 1).

One of the challengers in this context is the ability to process big amounts of data in near-real time to facilitate decision making.

[1] The term *visualization* means both *dimensionality reduction* and *displaying meaningful results* in this paper.

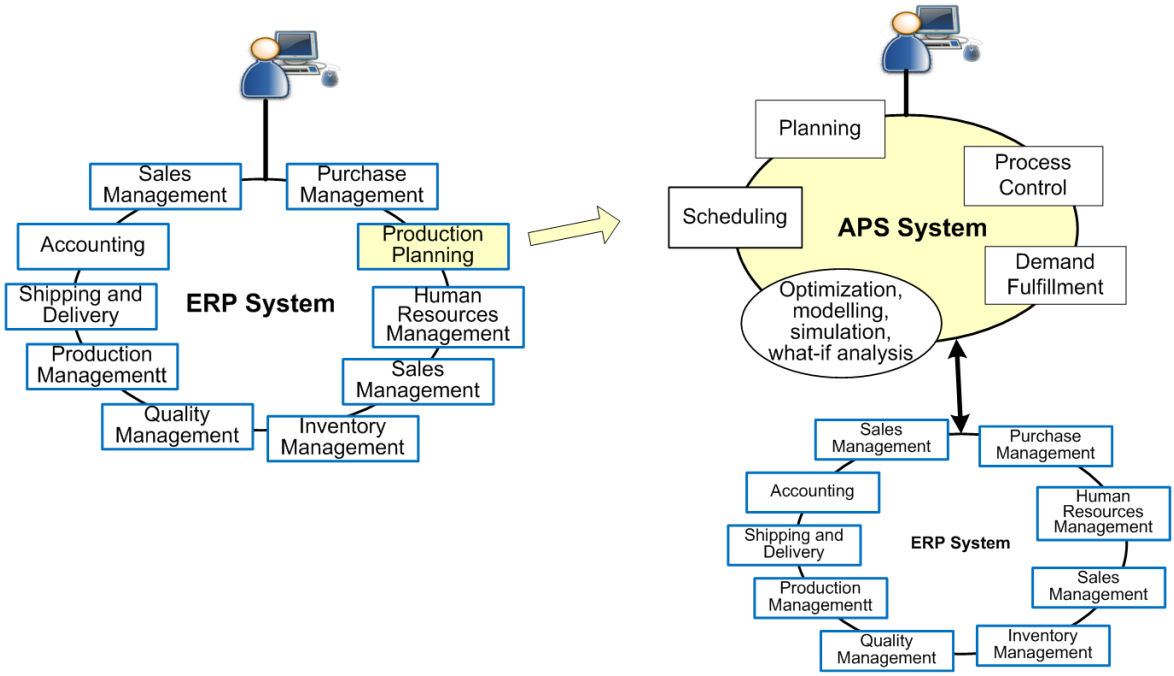


Figure 1. Decision support as a primary aspect of contemporary information systems

3 Artificial Neural Network-based Data Visualization

3.1 Multidimensional Data Visualization

Big multidimensional datasets bring new challenges to data analysis because large volumes and different varieties of data must be taken into account. In many cases, data are just being generated faster than they can be analyzed. To analyze big data, many data mining and machine learning algorithms have been developed. In this paper we focus on dimensionality reduction algorithms which reduce data dimensionality from original high dimension space to target dimension (2D in visualization case). Data visualization is the presentation of multidimensional data. It is very important to see analytical results presented visually, to find interdependencies / interrelationships among many objects.

Visualization is one of the basic operations in the toolbox of data analysis. Given a large set of some measured variables, the main idea is to represent them with a reduced set of more informative ones. Another reason for reducing the dimensionality is to decrease computational load for further data processing. Today's large multidimensional datasets contain a huge amount of data. It becomes almost impossible to analyze them manually and to extract valuable information. Therefore, more effective ways to display, analyze and interpret the information contained within them are required.

Data from the real world are frequently described by an array of variables $x_1, x_2, \dots, x_n$. Any variable may take some numerical values. A combination of values of all variables characterizes a particular data object $X_j = (x_{j1}, x_{j2}, \dots, x_{jn}), j \in \{1, \dots, m\}$, from the whole set $X_1, X_2, \dots, X_m$, where n is the number of variables, m is the number of the analyzed objects. If objects $X_1, X_2, \dots, X_m$ are described by more than one variable, the data are called multidimensional data. Often the objects are interpreted as points in the n-dimensional space R^n, where n defines the dimensionality of the space. In fact, we have a table of numerical data $\{x_{ji}, j = 1, \dots, m, i = 1, \dots, n\}$ for the analysis. An intuitive idea is to present multidimensional data, stored in such a table, in some visual form. It is a complicated problem that many

researchers seek to address in order to enable decision-makers to gain a deeper insight into the data, draw conclusions, and directly interact with the data. A type of multidimensional data visualization is based on dimensionality reduction. The goal of dimensionality reduction is to represent the input data in a lower-dimensional space so that certain properties (e.g., clusters, outliers) of the structure of this dataset are preserved as closely as possible.

An example of visual presentation of the data table (n = 6, m = 20) using multidimensional scaling method (MDS) is presented in [7] (see Figure 2). In this example, the dimensionality of data is reduced from 6 to 2. As we can see from the figure, objects X_4, X_6, X_8 and X_{19} form a separate cluster that can be clearly observed visually on a plane and that cannot be recognized directly from the table without a special analysis.

	x_1	x_2	x_3	x_4	x_5	x_6
1	0,57	0,11	0,47	0,54	0,84	0,53
2	0,65	0,14	0,53	0,55	0,89	0,54
3	0,61	0,13	0,49	0,52	0,92	0,57
4	0,44	0,16	0,45	0,42	0,99	0,45
5	0,55	0,11	0,53	0,52	0,92	0,55
6	0,44	0,16	0,43	0,41	0,98	0,43
7	0,49	0,12	0,56	0,46	0,82	0,48
8	0,41	0,17	0,42	0,43	0,98	0,42
9	0,54	0,12	0,59	0,46	0,85	0,48
10	0,69	0,16	0,63	0,56	0,91	0,49
11	0,66	0,16	0,51	0,48	0,87	0,55
12	0,70	0,13	0,52	0,48	0,81	0,52
13	0,65	0,14	0,63	0,52	0,89	0,49
14	0,59	0,11	0,51	0,57	0,89	0,52
15	0,53	0,11	0,52	0,57	0,89	0,50
16	0,52	0,13	0,51	0,61	0,92	0,51
17	0,58	0,12	0,55	0,61	0,95	0,52
18	0,45	0,10	0,52	0,53	0,92	0,51
19	0,42	0,17	0,41	0,42	0,98	0,40
20	0,53	0,11	0,42	0,57	0,91	0,57

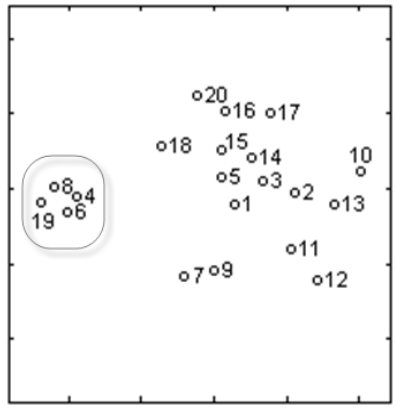


Figure 2. Example of data visualization

A comprehensive review of the dimensionality reduction methods is presented in [7], [8]. The Principal component analysis (PCA) [9] is one of the well-known dimensionality reduction methods. It can be used to display the data as a linear projection on such a subspace of original data space that best preserves the variance of the data. The PCA cannot preserve nonlinear structures, consisting of arbitrarily shaped clusters or curved manifolds since it describes the data in terms of a linear subspace. An alternative approach to dimensionality reduction is multidimensional scaling (MDS) [10]. MDS is a classical approach that maps an original high dimensional space to a lower dimensional one, but does so in an attempt to preserve the distances of corresponding data points. The starting state of MDS is a matrix consisting of the pairwise dissimilarities of data points.

The effectiveness of PCA is limited by its global linearity. The MDS method is nonlinear method, unsuitable for large datasets. It requires too much computational resources. Therefore, the combinations of different data visualization methods are under active development today because the combination of different methods can facilitate more efficient data analysis, while minimizing the shortcomings of individual methods.

3.2 Visualization Methods Based on Neural Networks

Artificial neural networks (ANNs) may also be used for dimensionality reduction and data visualization. The MDS were experimented with by the neural network researchers [11], [12]. As

a result, several neural network-based methods for the visualization of big multidimensional datasets have been proposed, including SAMANN [7], [8], [12], SOM [8], [13], etc. Most ANN based visualization methods are unsuitable for large datasets due to the demand of huge computational resources. One possible solution employs hardware, i.e., increased memory, parallel processing and cloud computing. The second solution is to go the other way and to develop a more mature neural network-based visualization theory. Therefore, the new strategies, approaches, and methods for training neural networks are required.

Visualization of the large dataset requires a huge amount of computational resources and time. The most of conventional visualization methods are unsuitable for large datasets. As has been shown in [7], SAMANN neural network can be successfully used for the large multidimensional dataset visualization, notwithstanding some limitations. It tries to optimize a projection error that describes how well the pairwise distances in a dataset are preserved and it is not suitable for large datasets. The whole mapping procedure has to be repeated when a new data point has to be mapped. The back propagation-like learning rule SAMANN [7], [8], [12] allows a feed-forward artificial neural network to learn MDS-based SAMANN's mapping in an unsupervised way. This neural network is able to project new points after the training. In each learning step, two objects are given to this neural network. The weights of neural network are updated according to the update rule using the error measure. One training iteration of the neural network is completed if all possible pairs of objects from the dataset are shown to the neural network. After training, the network is able to project previously unseen data using the obtained generalized mapping rule.

One of the ways to minimize the computational expenditure for the neural network training is working with a subset of the dataset. The results of the experiments showed [14] that it is possible to find such a subset of the dataset that, while training the SAMANN network with this subset, lower projection errors are obtained faster than by training with all the points of the dataset.

In Figure 3 the process of the visualization of the large multidimensional dataset is presented. At first, it is necessary to create the subset of the dataset. The subset can be created of dataset points chosen randomly (or using some deterministic way) from the dataset. The points of the subset are projected on a plane using SAMANN with optimal parameters and the weights of the network are calculated using iteration weights updating process. Then all the remaining points of the dataset are projected using the calculated weights of the neural network.

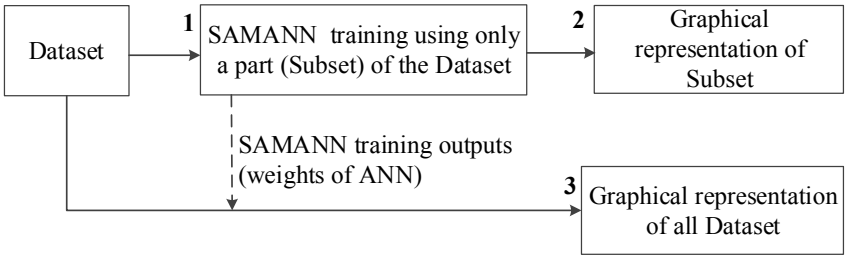


Figure 3. Scheme of the visualization process suitable for the large dataset: (1) selection of the subset of the primary dataset and projection of the subset using SAMANN; (2) graphical representation of the subset; (3) graphical representation of the dataset using the calculated weights

The training process of the SAMANN network depends on different parameters, such as learning rate, activation function, momentum term, the initial range of the network weights, etc. Heuristically searching for the optimal parameters of the network we can significantly speed-up the training of the network and apply it to the large dataset. Heuristic decisions are often applied when there is no possibility to get them theoretically [15].

Parameters that are associated with the neural network can be divided into two main groups. The first group involves control parameters of the neural network training: learning rate, momentum term. The second group involves parameters of neural network's architecture:

number of hidden layers, number of units in each hidden layer, weights, type and parameters of activation functions, etc. The dependence of the projection error on the learning rate and momentum term was investigated in [16].

The experiments have been performed in [7] to evaluate how the SAMANN network training process (visualizing the large dataset) depends on the initial range of the weights of the network and the slope parameter of the activation function of neurons.

The self-organizing map (SOM) is another class of neural networks that are trained in an unsupervised way using a competitive learning [8], [13]. A distinctive characteristic of this type of neural networks is that they can be used for both clustering and visualization of multidimensional data. SOM is a set of neurons, connected to one another via a rectangular or hexagonal topology. Each neuron is defined by the place in SOM and by the so-called codebook vectors. After SOM learning, the data are presented to SOM and winning neurons for each data point are found. In such a way, the data points are distributed on SOM and some data clusters can be observed. Moreover, according the position on the grid, the neurons are characterized by n-dimensional codebook vectors. An intuitive idea is to apply the dimensionality reduction methods to additional mapping of the codebook vectors of the winning neurons on the plane. MDS may be used for such a purpose. The scheme of the combination SOM and MDS (SOM-MDS) is presented in Figure 4. Additionally, the number of winning neurons is smaller than the number of data points, so a smaller dataset should be visualized by MDS than in those cases where the whole dataset is processed by MDS. This distinctive characteristic of SOM is very useful for big multidimensional datasets visualization.

Figure 4. Scheme of the visualization process by the combination of SOM and MDS

4 New Training Strategy: Single Pass of Data

To visualize big multidimensional datasets using SAMANN and SOM, a new strategy for training these networks has been proposed. The advantage of this strategy is that the network can be trained to visualize the multidimensional data through a single pass of training data. After training, the network can be used for visual presentation of the desirable number of multidimensional objects on the plane. The strategy is based on the assumption that a huge amount of data includes many similar objects so that even during one pass, the neural network can visualize a large amount of similar objects.

A new strategy of large multidimensional datasets visualization using SAMANN is presented in Figure 5: (1) training of SAMANN neural network through a single pass (only 1 iteration); calculating its weights; (2) graphical representation (visualization) of the dataset; (3) graphical representation of new previously unseen data points using calculated weights without additional neural network training.

A new strategy using the combination of SOM and MDS is presented in Figure 6: (1) training of the SOM neural network through a single pass; SOM winning neurons are calculated; (2) visualization of two-dimensional points that are two-dimensional representations of the codebook vectors of the winning neurons by MDS; (3) graphical representation of the dataset; (4) graphical representation of new previously unseen data points using the winning neurons without additional SOM training.

The ellipsoid dataset has been used to investigate the ability to visualize big multidimensional dataset using SAMANN and SOM-MDS. This ellipsoidal dataset consists of 7354 10-dimensional points from 10 overlapping ellipsoidal-type clusters. The dataset has been

obtained using the ellipsoidal cluster generator [17]. This generator creates ellipsoidal clusters with the major axis of an arbitrary orientation. The boundary of a cluster is defined by four parameters: the origin (which is also the first focus); the interfocal distance, uniformly distributed in the range [1.0, 3.0]; the orientation of the major axis, uniformly located amongst all orientations; the maximum sum of Euclidean distances to two foci, belonging to the range [1.05, 1.15] – equivalent to the eccentricity ranging from [0.870, 0.952]. For each cluster, data points are generated at a Gaussian-distributed distance from a uniformly random point on the major axis in a uniformly random direction and are rejected if they lie outside the boundary.

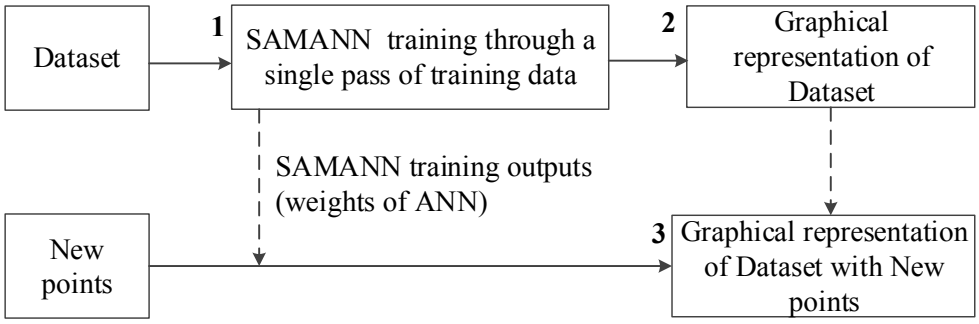


Figure 5. A new strategy of big multidimensional datasets visualization using SAMANN

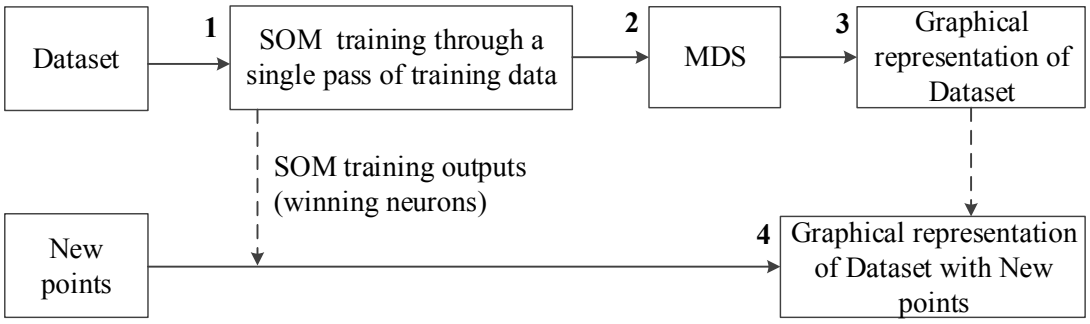


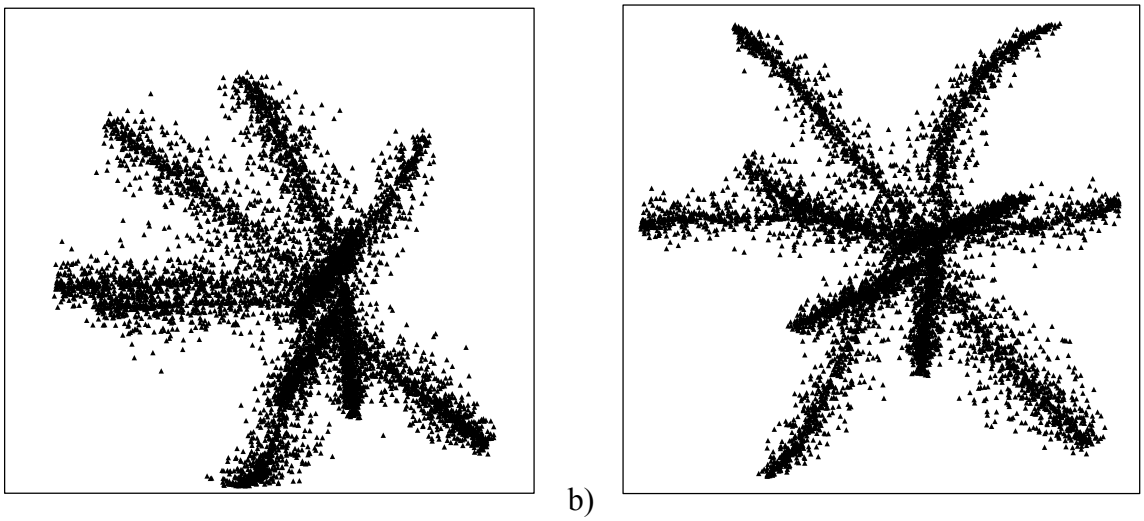
Figure 6. A new strategy of big multidimensional datasets visualization using SOM-MDS

In order to compare the visualization results obtained by standard SAMANN training and the proposed strategy through a single pass of training data, the experiment has been performed using the ellipsoidal dataset. The results of the experiment of multidimensional data visualization by SAMANN through a single pass of training data with optimal control parameters are presented in Figure 7a. The points of the dataset are marked by black triangles. Figure 7b presents the visualization results of the same dataset when neural network is trained using 10000 iterations. By comparing the results presented in Figure 7 (a and b), it can be concluded that the visualization results of the ellipsoid dataset using SAMANN through a single pass of training data are visually rather similar to the visualization results of the dataset using 10000 iterations. In both Figure 7a and Figure 7b, we can clearly observe 10 overlapping ellipsoidal-type clusters.

The additional evaluation of the strategy has been done in order to estimate the possibility to visualize new points without additional training of the neural network. Ten new points were generated without additional training. The visualization results of the dataset (obtained after the training of SAMANN) and new points (obtained without additional training, using the already calculated weights) are presented in Figure 8. The circles correspond to the new points that were not used for training.

The visualization of the dataset and outliers using single pass of training data strategy is presented in Figure 9. In this case, outliers are points (events or observations) which do not conform to an expected pattern or other items in a dataset. The outliers are marked by circles. On

the one hand, the experiment shows that the outliers can be properly identified; on the other hand, decision makers can easily recognise issues that need their attention. Such visualization of outliers can be useful for unsupervised anomaly detection.



a) b)

Figure 7. a) Visualization results of the ellipsoidal dataset using SAMANN through a single pass of training data; b) visualization results of the dataset using 10000 iterations

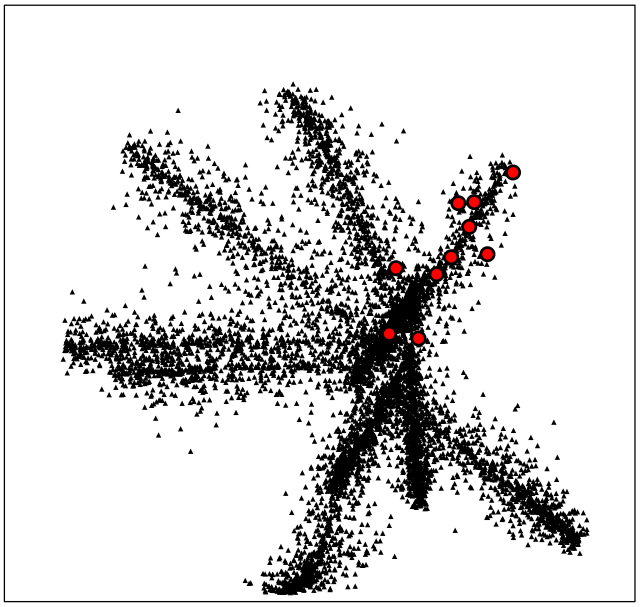


Figure 8. Visualization results of the ellipsoidal dataset and new points using SAMANN

The results of the experiment of multidimensional data visualization using SOM-MDS through a single pass of training data and using 100 iterations are presented in Figure 10. The visualization enables meaningful visual evaluation of the single pass and the traditional training strategies.

The visualization results of the ellipsoidal dataset and new points using the already calculated winning neurons without additional SOM training are presented in Figure 11. The circles correspond to the new points that were not used for training.

The visualization of the dataset and outliers using trained SOM by single pass of the training data strategy is presented in Figure 12. The circles correspond to the outliers.

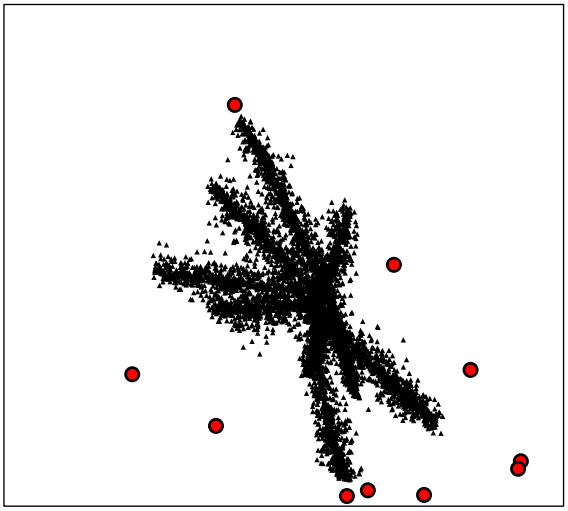


Figure 9. Visualization results of the ellipsoidal dataset and outliers using SAMANN

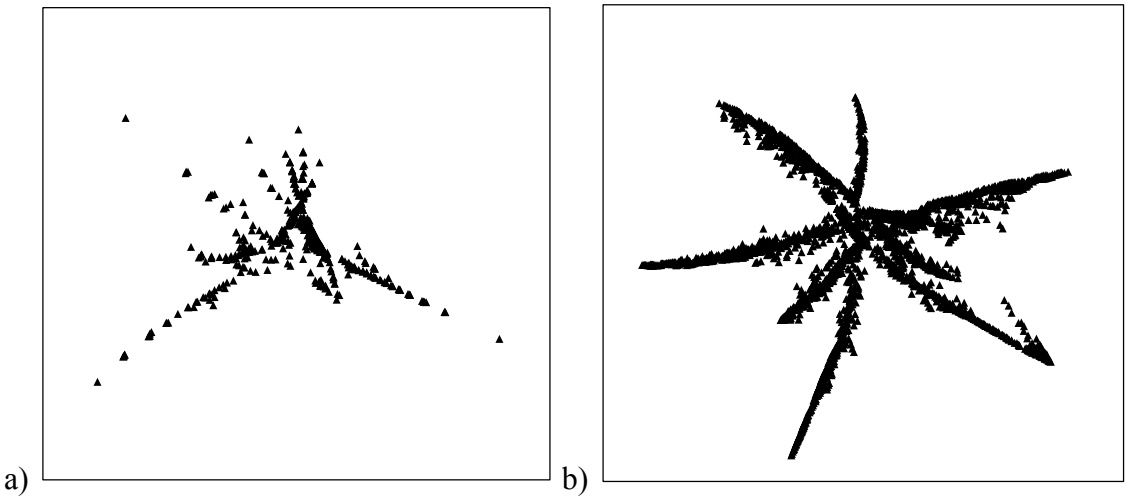


Figure 10. a) Visualization results of the ellipsoidal dataset using SOM-MDS through a single pass of training data; b) visualization results of the dataset using 100 iterations

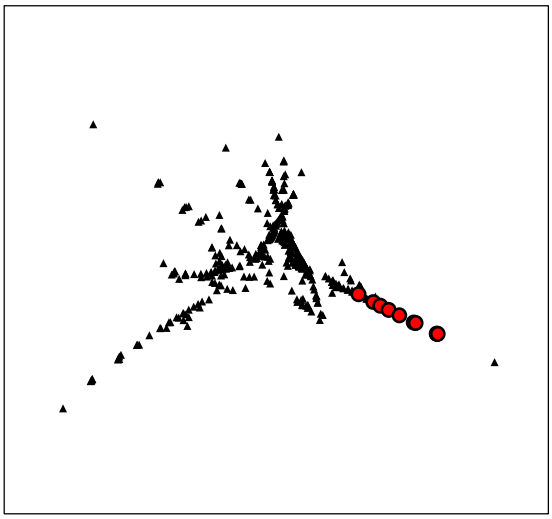


Figure 11. Visualization results of the ellipsoidal dataset and new points using SOM-MDS

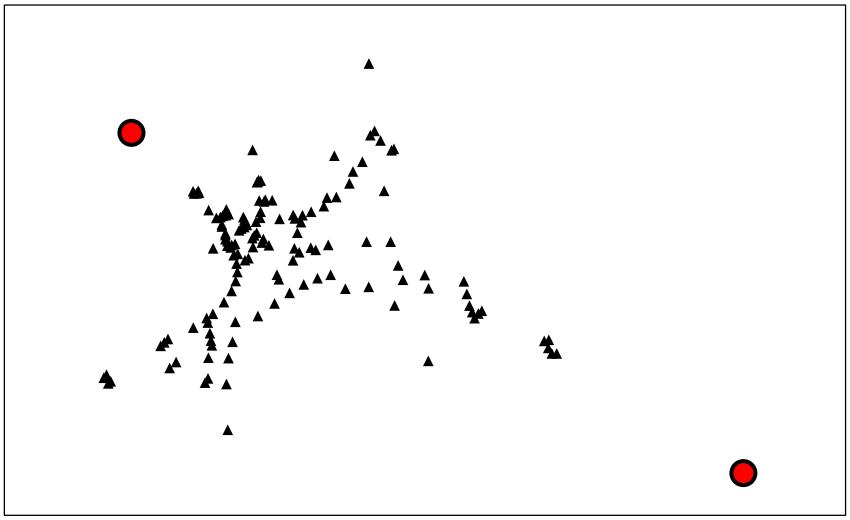


Figure 12. Visualization results of the ellipsoidal dataset and outliers using SOM-MDS

The analysis of the results using SAMANN and SOM-MDS shows that it is possible to get the suitable projections of the dataset through a single pass of training data and to visualize new points. The experiments show that even after one pass fairly reliable projections can be obtained and used to display meaningful results to support, at least, trend detection in predictive analytics.

5 Conclusion and Future Research

Big multidimensional datasets visualization is an essential constituent of today's business information systems, in the age of growing amounts of data to be interpreted and analyzed to support decision making. The challengers in this context include the ability to process huge amounts of data in near-real time, as well as to quickly understand the meaning of the data and to access outliers.

The new strategy for training SAMANN and SOM neural networks is proposed and examined. The characteristic feature of this strategy is that neural network can be trained to visualize data through a single pass of training data. The results of experiments on the two sets of benchmark data to demonstrate this strategy allow us to conclude that the unsupervised learning of SAMANN and SOM neural networks are effective in producing the visual projections of big multidimensional datasets, where we do not need any additional knowledge on the objects, i.e., the known numerical values of the variables are sufficient. The obtained visualization results are significant and computational expenses are acceptable if compared with the traditional learning when many iterations are required. However, the visualization results using SAMANN are more exact than using SOM. The experiments show that the strategy of a single pass of training data can be used to display meaningful results; at least, in decision support to detect trends in predictive analytics and to identify the issues that are visualised as outliers.

Further research should be focused on the theoretical background of the proposed single pass strategy, as well as on discovering new domains (e.g., streaming data analysis [18]) where big multidimensional datasets are required to be visualized when making proper human decisions. The additional experimental research must be carried out using much larger datasets of varying characteristics to evaluate the suitability of the proposed approach for big multidimensional datasets visualization. Moreover, the comprehensive evaluation of the proposed training strategy should be made using some quantitative measures. The next step of the research should involve the application of this strategy to big multidimensional datasets in real-world situations.

References

[1] R. Maes, D. Rijsenbrij, O. Truijens and H. Goedvolk, “Redefining Business – IT Alignment through a Unified Framework”, Report 2000-19, Amsterdam, Universiteit van Amsterdam, Department of Information Management , 2000.

[2] A. Caplinskas, A. Lupeikiene and O.Vasilecas, “Shared conceptualisation of business systems, information systems and supporting software”, in Databases and Information Systems II, H.-M. Haav, A. Kalja, Eds., Kluwer Academic Publishers, pp. 109-120, 2002. Available: http://dx.doi.org/10.1007/978-94-015-9978-8_9

[3] P. van Eck, H. Blanken and R. Wieringa, “Project GRAAL: towards operational architecture alignment”, in International Journal of Cooperative Information Systems, vol. 13, no. 3, pp. 235-255, 2004. Available: http://dx.doi.org/10.1142/S0218843004000961

[4] Y. Nishioka, “Object model for planning and scheduling integration in discrete manufacturing enterprises”, in Knowledge Sharing in the Integrated Enterprise: Interoperability Strategies for the Enterprise Architect, P. Bernus, M. Fox, Eds., IFIP 183, Springer, pp. 215-224, 2005. Available: http://dx.doi.org/10.1007/0-387-29766-9_18

[5] Y. Kristianto, P. Helo and A. Mian, “Value chain re-engineering by the application of advanced planning and scheduling”, in Handbook on Business Information Systems, A. Gunasekaran, M. Sandhu, Eds., World Scientific Publishing Company, pp. 147-187, 2010. Available: http://dx.doi.org/10.1142/9789812836069_0007

[6] A. Lupeikiene, G. Dzemyda, F. Kiss and A. Caplinskas, “Advanced planning and scheduling systems: modelling and implementation challenges”, in Informatica, vol. 25, no. 4, pp. 581-616, 2014. Available: http://dx.doi.org/10.15388/informatica.2014.31

[7] G. Dzemyda, O. Kurasova and V. Medvedev, “Dimension reduction and data visualization using neural networks”, in Emerging Artificial Intelligence Applications in Computer Engineering, I. Maglogiannis, K. Karpouzis, M. Wallace, J. Soldatos, Eds., Frontiers in Artificial Intelligence and Applications, vol. 160, IOS Press, pp. 25-49, 2007.

[8] G. Dzemyda, O. Kurasova and J. Žilinskas, “Multidimensional Data Visualization: Methods and Applications”, Springer, Heidelberg, 2013. Available: http://dx.doi.org/10.1007/978-1-4419-0236-8

[9] I. T. Jolliffe, “Principal Component Analysis”, Springer, Heidelberg, 2002. Available: http://dx.doi.org/10.1007/b98835

[10] I. Borg and P. Groenen, “Modern Multidimensional Scaling: Theory and Applications”, Springer, Heidelberg, 2005. Available: http://dx.doi.org/10.1007/0-387-28981-X

[11] D. Lowe and M. E. Tipping, “Feed-forward neural networks and topographic mappings for exploratory data analysis”, in Neural Computing and Applications, vol. 4, no. 2, pp. 83-95, 1996. Available: http://dx.doi.org/10.1007/BF01413744

[12] J. Mao and A. K. Jain, “Artificial neural networks for feature extraction and multivariate data projection”, in IEEE Transactions Neural Networks, vol. 6, no. 2, pp. 296-317, 1995. Available: http://dx.doi.org/10.1109/72.363467

[13] T. Kohonen, “Self-organizing Maps”, Springer, Heidelberg, 2001. Available: http://dx.doi.org/10.1007/978-3-642-56927-2

[14] S. Ivanikovas, V. Medvedev and G. Dzemyda, “Parallel realizations of the SAMANN algorithm”, in Adaptive and Natural Computing Algorithms, Lecture Notes in Computer Science, vol. 4432, Springer, pp. 179-188, 2007. Available: http://dx.doi.org/10.1007/978-3-540-71629-7_21

[15] G. Dzemyda and L. Sakalauskas, “Large-scale data analysis using heuristic methods”, in Informatica, vol. 22, no. 1, pp. 1-10, 2011.

[16] V. Medvedev and G. Dzemyda, “Optimization of the local search in the training for SAMANN neural network”, in Journal of Global Optimization, no. 4, vol. 35, pp. 607-623, 2006. Available: http://dx.doi.org/10.1007/s10898-005-5368-1

[17] J. Handl and J. Knowles, “Cluster Generators for Large High-dimensional Data Sets with Large Numbers of Clusters”. [Online]. Available: http://personalpages.manchester.ac.uk

[18] J. Bernatavičienė, G. Dzemyda, G. Bazilevičius, V. Medvedev, V. Marcinkevičius and P. Treigys, “Method for visual detection of similarities in medical streaming data”, in International Journal of Computers Communications & Control, vol. 10, no. 1, pp. 8-21, 2015. Available: http://dx.doi.org/10.15837/ijccc.2015.1.1310