ORIGINAL ARTICLE

The experimental study of the effectiveness of Kohonen maps and autoassociative neural networks in the qualitative analysis of multidimensional data by the example of real data describing coal susceptibility to fluidal gasification

Dariusz Jamróz[1] (ID)

Abstract

The qualitative analysis of multidimensional data using their visualization allows to observe some characteristics of data in a way which is the most natural for a human, through the sense of sight. Thanks to such an approach, some characteristics of the analyzed data are simply visible. This allows to avoid using often complex algorithms allowing to examine specific data properties. Visualization of multidimensional data consists in using the representation transforming a multidimensional space into a two-dimensional space representing a computer screen. The important information which can be obtained in this way is the possibility to separate points belonging to different classes in the multidimensional space. Such information can be directly obtained if images of points belonging to different classes occupy other areas of the picture presenting these data. The paper presents the effectiveness of the qualitative analysis of multidimensional data conducted in this way through their visualization with the application of Kohonen maps and autoassociative neural networks. The obtained results were compared with results obtained using the perspective-based observational tunnels method, PCA, multidimensional scaling and relevance maps. Effectiveness tests of the above methods were performed using real seven-dimensional data describing coal samples in terms of their susceptibility to fluidal gasification. The methods' effectiveness was compared using the criterion for the readability of the multidimensional visualization results, introduced in earlier papers.

Keywords Multidimensional visualization · Multidimensional data analysis · Data mining · Self-organized neural network · Autoassociative neural network · Kohonen maps

1 Introduction

Methods utilizing neural networks for analyzing multidimensional data through their visualization are widely used in practice [1–5]. Visualization of multidimensional data consists in using the representation transforming a multidimensional space into a two-dimensional space representing a computer screen. This representation should preserve properties of these data crucial for the conducted analysis. Neural networks are well suited for different kinds of representations [6–9], so they can also be used for this type of representation. The important information which can be obtained in this way is the possibility to separate points belonging to different classes in the multidimensional space. Such information can be directly obtained if images of points belonging to different classes occupy other areas of the picture presenting these data. The paper presents the effectiveness of the qualitative analysis of multidimensional data conducted in this way through their visualization with the application of Kohonen maps and autoassociative neural networks. The obtained results were compared with results obtained using the perspective-based observational tunnels method, PCA, multidimensional scaling and relevance maps. The comparison of the

✉ Dariusz Jamróz
   jamroz@agh.edu.pl

1 Department of Applied Computer Science, AGH University
   of Science and Technology, al. Mickiewicza 30,
   30-059 Kraków, Poland

above methods was performed using real seven-dimensional data describing coal samples in terms of their susceptibility to fluidal gasification. The qualitative analysis of the presented data using each of the methods was conducted for this purpose. The purpose of the analysis was to state whether coal samples with different susceptibility to gasification occupy separate subareas of the multidimensional space of characteristics. This in turn allows to state whether selected characteristics are sufficient for the correct differentiation of samples well and poorly susceptible to fluidal gasification. The methods' effectiveness was compared using the criterion for the readability of the multidimensional visualization results introduced in earlier papers [1, 10]. This paper constitutes the experimental study of the effectiveness of Kohonen maps and autoassociative neural networks in the qualitative analysis of multidimensional data by the example of real data describing coal susceptibility to fluidal gasification. Real seven-dimensional data describing coal samples in terms of their susceptibility to fluidal gasification was used for the first time in the paper for the analysis of the effectiveness of methods utilizing neural networks. However, they were already utilized for the evaluation of other visualization methods' effectiveness [11–14]. Methods analyzed in the paper participated in the ranking of various methods of qualitative analysis of multidimensional data through its visualization which was developed in previous papers. This ranking [1] was created as a result of the analysis of completely different data describing different energy classes of coal. In practice, apart from neural networks, also other methods are used for the qualitative analysis of multidimensional data through their visualization. The perspective-based observational tunnels method [10, 13, 15] which constitutes the parallel projection with local orthogonal projection utilizing perspective is used. The PCA method [11, 16–20] constitutes a projection onto two eigenvectors corresponding to two eigenvalues of the dataset covariance matrix largest in terms of module. Multidimensional scaling [12, 21–23] constitutes such a representation that distance between each two images of points in the two-dimensional output space representing the screen is as close as possible to the distance between points in the input space corresponding to them. In the method of relevance maps [14, 24, 25], special points representing axes of the coordinate system are additionally used. These points and points representing vectors belonging to the analyzed set are distributed on the plane in such a way that the distance of each point representing a data vector to a point representing a given axis of the coordinate system is as close as possible to the value of this coordinate of a given data vector. The method of parallel coordinates [26–29] is also used to visualize multidimensional data. In this method, n coordinate axes are distributed in parallel next to each other. Each point is represented by a polyline going through each of the axes in a point corresponding to the coordinate value. A similar method is star graphs [30], in which all axes go radially outward from one point.

2 Visualization using autoassociative neural networks

The autoassociative neural network used for the visualization of multidimensional data has n inputs, one of the interlayers used for the visualization consisting of two neurons and n outputs [3, 4]. The number of network inputs and outputs is determined as equal to the number of dimensions of the analyzed data. It is multilayer feedforward neural network, which is trained by the method of error back-propagation. Autoassociative neural networks are indicative of learning criterion they depend on. The learning criterion of these networks is such that signals appearing for each data vector at outputs are the same as those provided at inputs. Thanks to this, the trained network performs the compression of n network inputs into two outputs of the interlayer used for the visualization and then decompression to n network outputs. It follows from this that if such a network is trained, then the whole information allowing to reconstruct n-dimensional data is sent by two outputs of the interlayer used for the visualization. Figure 1 presents the operating diagram of such a network.

Training such a network consists in counting all weights attributed to all neurons. At first, the input data should be scaled in such a way that it is within the range defined by the network outputs. Because a hyperbolic tangent function was assumed to calculate the value of neuron output, output values are contained within range $(-1, 1)$. Coordinates of dataset vectors were scaled thus to range $(-0.9, 0.9)$. Before the network's learning starts, drawing all weights of all neurons should be conducted. Each weight was attributed a random value from range $(-0.5, 0.5)$. Then, points
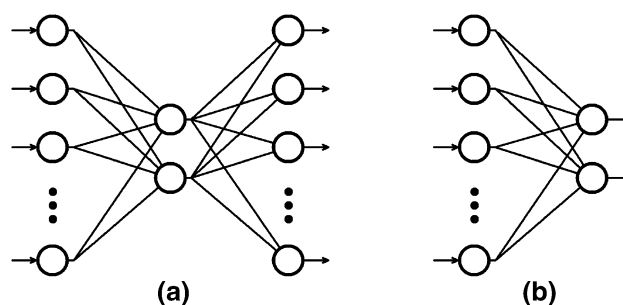

Fig. 1 A diagram of autoassociative neural network used for the visualization of multidimensional data. **a** Presents the diagram of the network being trained. **b** Presents a fragment of the previously trained network used during the visualization

1–5 are realized for each input data vector (such learning can be repeated multiple times):

1. For the next wth data vector, we calculate the output value of all neurons from the first layer:

$$y_{1,j} = g\left(w_{1,j,0} + \sum_{k=1}^{n} w_{1,j,k}x_{k,w}\right) \tag{1}$$

where g denotes the assumed nonlinear function (hyperbolic tangent was used in the conducted experiments), n is the number of network inputs, $y_{i,j}$ is the output value of a neuron placed in the ith network layer at the jth position (for neurons from the first layer i it is equal to 1), $w_{i,j,k}$ denotes the weight of the kth input of a neuron placed in the ith network layer at the jth position (the weight for input number 0 denotes the additional constant component), $x_{k,w}$ denotes the kth coordinate of the wth data vector.

For g function, a hyperbolic tangent was used as it is nonlinear, differentiable, increasing function and its value for the argument approaching infinity aims to 1 and for the argument approaching negative infinity aims to -1. The values are exactly within boundaries set for the operation of created network $(-1, 1)$. What is more, its derivative, which is used in the network learning process, is easy to calculate. It is important that the use of nonlinear g function makes it possible to increase capabilities of the created neural network to a large extent.

2. We calculate output values of all neurons located in the next network layers. The calculation of outputs of a given layer's neurons can occur after the calculation of output values of the previous layer's neurons:

$$y_{i,j} = g\left(w_{i,j,0} + \sum_{k=1}^{\text{size}(i-1)} w_{i,j,k}y_{i-1,k}\right) \tag{2}$$

where $\text{size}(i-1)$ denotes the number of neurons in layer number $i-1$, $y_{i,j}$ is the output value of a neuron placed in the ith network layer at the jth position, $w_{i,j,k}$ denotes the weight of the kth input of a neuron placed in the ith network layer at the jth position (the weight for input number 0 denotes the additional constant component), g is a nonlinear function, the same as in formula 1.

3. We count errors of the network outlet that is the difference between values at the network inputs and values we obtained at outputs of the last network layer. We multiply this difference by the derivative of the assumed function g, that is by the derivative of the hyperbolic tangent function:

$$\delta_{i,j} = \left(1 - y_{i,j}^2\right)\left(x_{j,w} - y_{i,j}\right) \tag{3}$$

where $\delta_{i,j}$ is the value of the error of the output of a neuron placed in the ith network layer at the jth position (in this formula i denotes the number of the last network layer), $y_{i,j}$ is the output value of a neuron placed in the ith network layer at the jth position, $x_{j,w}$ denotes the jth coordinate of the wth data vector.

4. We calculate errors of outputs of neurons from the remaining network layers, in the order from the penultimate layer to the first layer:

$$\delta_{i,j} = \left(1 - y_{i,j}^2\right) \sum_{k=1}^{\text{size}(i+1)} \left(\delta_{i+1,k}w_{i+1,k,j}\right) \tag{4}$$

where $\delta_{i,j}$ is the value of the calculated error of the output of a neuron placed in the ith network layer at the jth position, $w_{i+1,k,j}$ denotes the weight of the jth input of the kth neuron from layer $i + 1$, $\text{size}(i + 1)$ denotes the number of neurons in layer number $i + 1$, $y_{i,j}$ is the output value of a neuron placed in the ith network layer at the jth position.

5. Based on the previously calculated errors, we modify weights of all network neurons. We can do this by applying:

$$\widetilde{w}_{i,j,k} = w_{i,j,k} + \eta\delta_{i,j}y_{i-1,k} \tag{5}$$

where $w_{i,j,k}$ denotes the weight of the kth input of the jth neuron from the ith layer, $\widetilde{w}_{i,j,k}$ denotes the weight $w_{i,j,k}$ after the change, $\delta_{i,j}$ is the value of the error of the output of the jth neuron from the ith layer, $y_{i-1,k}$ is the output value of the kth neuron from layer $i - 1$, η is the parameter specifying the learning rate. Parameter η takes on a fixed value higher than zero.

After the completion of learning, visualization of each wth input data vector consists in calculating outputs of the next layers of neurons up to the moment of calculating the value of two outputs of the interlayer neurons used for the visualization. Two values obtained in this way constitute directly two coordinates of the screen specifying the location in which the image of the wth data vector should be drawn. In this way, we obtain the image of signals corresponding to separate multidimensional data vectors.

3 Visualization using Kohonen maps

Kohonen maps are one-layer neural networks with competitive learning principles [2, 5, 31, 32]. In these networks, the notion of neighborhood is additionally introduced. All network inputs reach every neuron. The number of network inputs is determined as equal to the number of dimensions

of the analyzed data. The network's learning proceeds in such a way that weights of the winner neuron, whose answer to a given set vector is the largest, are modified. Additionally, weights of neurons adjacent to the winner, that is neurons which are at some distance from the winner neuron, are modified. The modification of weights proceeds in such a way that the answer of the winner neuron and its neighbors (to a lesser extent) to a given set vector is even larger. Assuming the two-dimensional neighborhood, it is possible to arrange neurons within a grid of rows and columns. Then, the value of neuron output located in the ith row and jth column can be displayed on the screen as a point with coordinates (i, j). Figure 2 presents a simple example of such a network with three inputs, comprising two rows and three columns of neurons.

Training such a network consists in counting all weights attributed to all neurons. At first, the input data should be scaled in such a way that the length of each data vector is 1. For this purpose, we change each of n values of each wth input data vector:

$$\widetilde{x}_{k,w} = \frac{x_{k,w}}{\sqrt{\sum_{i=1}^{n}(x_{i,w})^2}} \tag{6}$$

where n is the number of data dimensions, $x_{k,w}$ is the kth coordinate of the wth input dataset vector, $\widetilde{x}_{k,w}$ denotes $x_{k,w}$ after the change.

Before the network's learning starts, drawing all weights of all neurons should be conducted. Each weight was attributed to a random value from range (0, 0.5). Then, points 1–3 are realized for each input data vector (such learning can be repeated multiple times):

1. For the next wth data vector, we calculate the output value of all neurons:
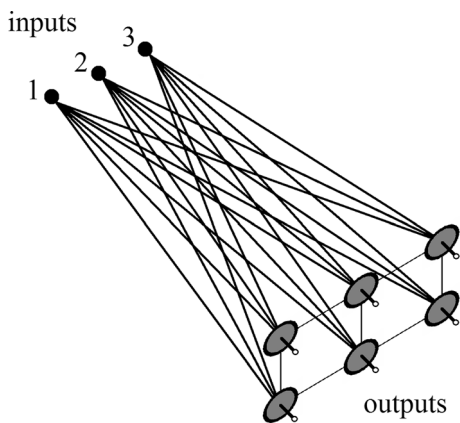

inputs

outputs

Fig. 2 The example of simple Kohonen maps, comprising two rows and three columns of neurons. The exemplary network has three inputs which reach all neurons

$$y_{i,j} = \sum_{k=1}^{n} w_{i,j,k} x_{k,w} \tag{7}$$

where n is the number of network inputs equal to the number of data dimensions, $y_{i,j}$ is the output value of neuron number (i, j) that is placed in the ith row and the jth column of the network, $w_{i,j,k}$ is the weight of the kth input of neuron number (i, j), $x_{k,w}$ is the kth coordinate of the wth dataset vector.

Based on the obtained results, we determine the neuron which is the winner, that is, the one at whose output the largest value appeared.

2. The modification of weights of the winner neuron and the winner's neighbors:

$$\widetilde{w}_{i,j,k} = w_{i,j,k} + \eta(x_{k,w} - w_{i,j,k}) \tag{8}$$

where

$$\eta = \begin{cases} \dfrac{0.01}{\text{dist} + 1} & \text{for dist} < \text{MAX\_DISTANCE} \\ 0 & \text{else} \end{cases} \tag{9}$$

where $dist$ is a distance using the Euclidean metrics of neuron number (i, j) from the winner neuron, MAX_DISTANCE is a parameter specifying the maximum distance of neurons treated as neighbors, $w_{i,j,k}$ is the weight of the kth input of neuron number (i, j), $\widetilde{w}_{i,j,k}$ denotes $w_{i,j,k}$ after the change, $x_{k,w}$ is the kth coordinate of the wth dataset vector.

It follows from the above that weights of neurons at a distance shorter than MAX_DISTANCE from the winner neuron are subject to the modification. Additionally, these modifications decrease hyperbolically along with an increase in the distance from the winner neuron. The above-assumed parameter η has been used before [5].

3. Vectors of weights of all neurons which were subject to changes are standardized:

$$\widetilde{w}_{i,j,k} = \frac{w_{i,j,k}}{\sqrt{\sum_{p=1}^{n}(w_{i,j,p})^2}} \tag{10}$$

where $w_{i,j,k}$ is the weight of the kth input of neuron number (i, j), $\widetilde{w}_{i,j,k}$ denotes $w_{i,j,k}$ after the change.

After the completion of learning, the visualization of each wth input data vector consists in calculating values of all neurons' outputs from Formula 7. Based on the obtained results, we determine the position of the neuron which is the winner, that is, the one at whose output the largest value appeared. If the winner neuron is in the uth row and vth column, we check whether a symbol representing a class other than the class of the

*w*th vector was earlier drawn in the location on the screen with coordinates (*u*, *v*):

- If yes, this means that the neuron with number (*u*, *v*) is at the same time the winner for vectors representing different classes. It follows from this that the network is unable to differentiate at least two data vectors belonging to different classes. In turn, it follows from this that the obtained view is not satisfying. Thus, we train the network again changing the number of learning repetitions or other parameters.
- If not, then in the location with coordinates (*u*, *v*), we draw a symbol representing the class to which the *w*th vector belongs.

Thanks to proceeding in that way for all data vectors, we obtain the image of neurons representing separate data classes on the computer screen.

4 Possibilities of visualizations based on autoassociative neural networks and Kohonen maps

Both autoassociative neural networks and Kohonen maps can create nonlinear mappings. On the one hand, this can be treated as a disadvantage, because it leads to the distortion of multidimensional data view. However, in the case of many types of analyses this is of no relevance—for example with the analysis conducted in this paper. Simultaneously, it must be noted that some topological dependencies are usually preserved in a view distorted in this manner. Moreover, in the case of many types of data, nonlinear mapping can have beneficial effect on the possibility to obtain views enabling to observe some significant features. An example of this type of data is situation in which one dataset surrounds another dataset from all sides and at the same time we want to obtain information on the possibility to separate these sets from each other. In order to present the analyzed situation in detail, artificial seven-dimensional data were prepared using a random number generator. These data consist of two subsets, one of which occupies the area of a seven-dimensional sphere and another occupies the area of a sphere of some thickness surrounding the first one. Both subsets contain 1000 points. Figures 3 and 4 present views of data obtained using autoassociative neural networks and Kohonen maps prepared in this manner. As seen, both these views allow to conclude that the possibility to separate subsets of analyzed data exists. This means that methods using autoassociative neural networks and Kohonen maps are effective even in
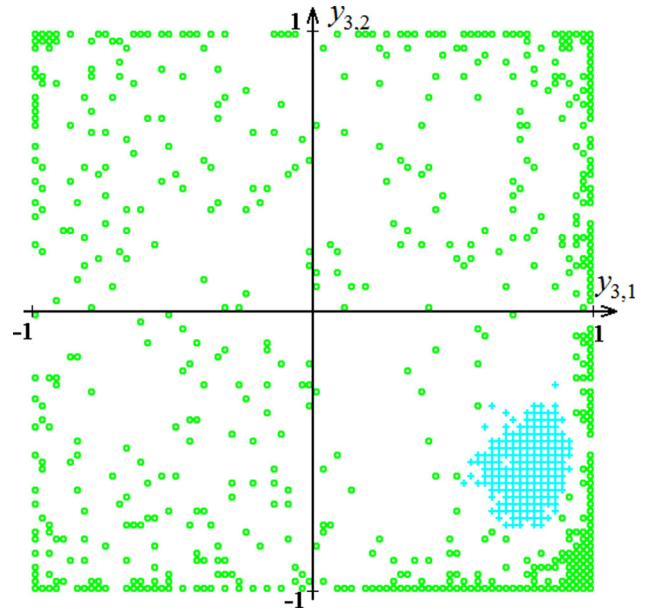

Fig. 3 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using the autoassociative neural network. Signals representing different subsets are marked with different symbols
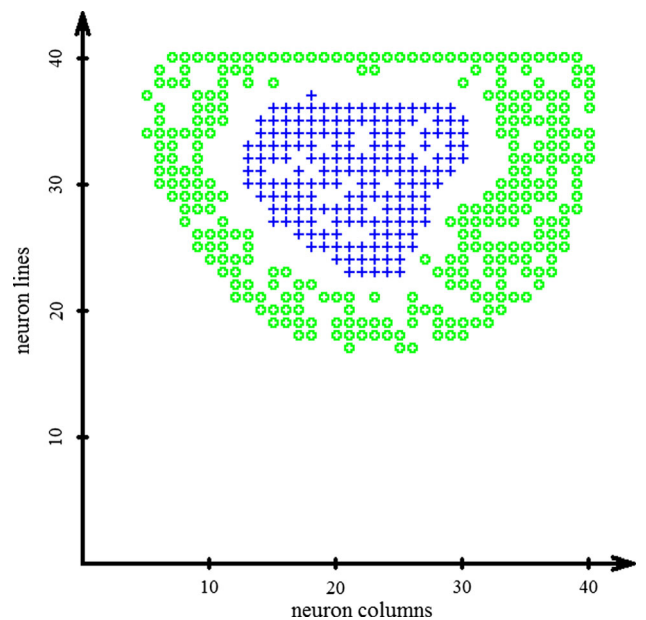

Fig. 4 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using Kohonen maps. Neurons representing different subsets are marked with different symbols

such an extreme case of surrounding one subset from all sides by another one.

For comparison, Figs. 5, 6, 7 and 8 present views of the same data obtained using other visualization methods. Figure 5 presents the view obtained using PCA. As seen, this view does not allow to obtain information about the
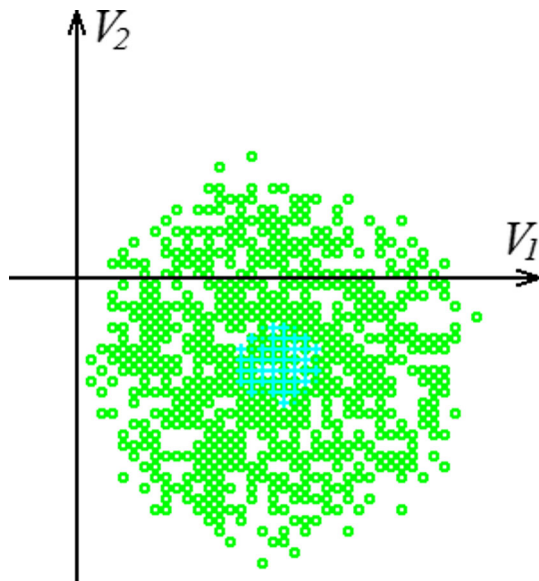
Fig. 5 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using PCA. Areas occupied by different subsets are marked with different symbols
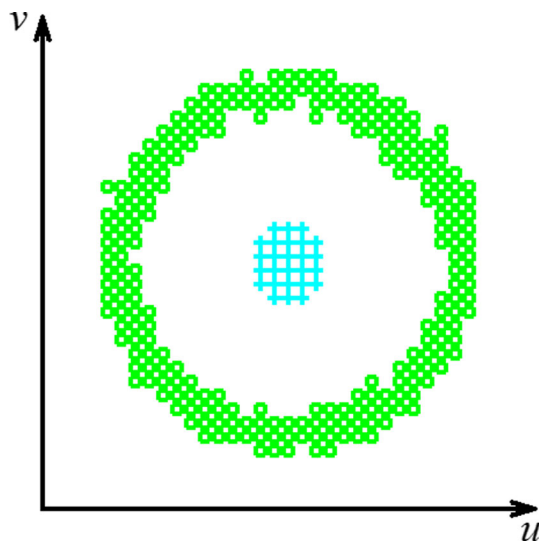

Fig. 7 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using relevance maps. The view obtained after 20,000 improvement cycles occurring after random initial values. Images occupied by different subsets are marked with different symbols

Fig. 6 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using multidimensional scaling. Images occupied by different subsets are marked with different symbols

Fig. 8 The view of artificially generated seven-dimensional data in which one subset surrounds another subset from all sides, obtained using relevance maps. The view obtained after 20,000 improvement cycles occurring after specially prepared initial values. Images occupied by different subsets are marked with different symbols

possibility to separate the analyzed subsets from each other. In the case of the presented artificially generated data, this method has a serious problem, because in this case principal components do not exist. It is caused by the fact that in the analyzed case, the analyzed data covariance matrix contains practically the same values on the main diagonal and values close to zero beyond the main diagonal. All eigenvalues are thus close to each other and are equal to each other for the number of points generated in
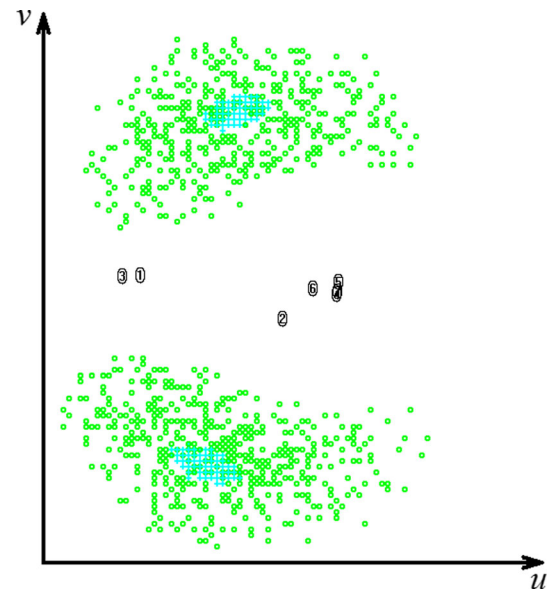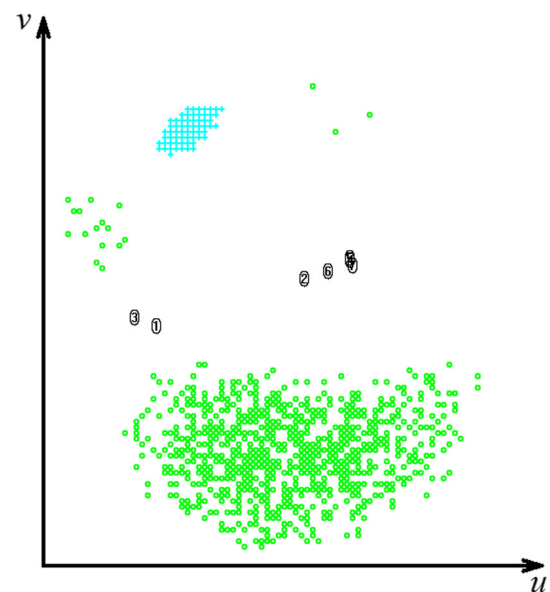
the described manner going to infinity—so there is no possibility to select those which are the largest in terms of module. It turns out that by selecting any eigenvectors for the analyzed data we obtain a view close to this in Fig. 5. This method is the example of a linear method, so it does not distort views. However, as can be observed, it is not

effective in the described case. But, its indisputable advantage is comfort of its use resulting from the fact that there is no need to select any parameters while using it. However, both autoassociative neural networks and Kohonen maps require determining many parameters. In particular, it is the number of learning repetitions. Sometimes, it is necessary to even change the assumed network topology. For example, Fig. 3 is obtained by a network in which all layers, apart from the third one consisting of two neurons (used for visualization) and the last one consisting of seven neurons, consist of 100 neurons. When initially the network in which all layers apart from the third one consisted of seven neurons was used, obtaining readable views allowing to determine the possibility to separate the analyzed subsets was not successful. Effective application of methods utilizing neural networks requires considerable experience. Moreover, the lack of obtaining readable views from the perspective of the conducted analysis does not have to mean that such views are not possible to obtain using these methods. It may turn out that readable views can be obtained assuming such parameters, network topology and such a number of learning repetitions which were not assumed in a given analysis.

Figure 6 presents the view obtained using multidimensional scaling which is a nonlinear method. The view obtained by it presents the possibility to separate the analyzed subsets from each other in a readable manner. This method is not so successful in the case of less regular data. An interesting result was obtained using the relevance maps method which is also a nonlinear method. Figure 7 presents the view obtained with a standard, random initial distribution of images of data points and relevance points. This view does not allow to obtain information about the possibility to separate the analyzed subsets from each other. It was not possible to obtain a better view at different combinations of random initial values. On the other hand, a readable view allowing to determine the possibility to separate subsets was obtained only with specially set initial values, in which relevance points are on a straight line and points belonging to different subsets are placed on the opposite sides of this straight line. Figure 8 presents the stabilized view obtained with such determined initial values of 20,000 repetitions of improvement cycles each.

As shown above on the artificially generated seven-dimensional data, methods utilizing neural networks are perfect even with the analysis of such data in which some subsets obscure other ones in a complicated manner. The remaining mentioned nonlinear methods, that is multidimensional scaling and relevance maps also allowed to obtain readable results, whereby relevance maps—only with specially prepared initial values. The mentioned linear method of PCA did not allow to obtain readable results.

A huge advantage of visualization methods based on neural networks is their capability to extract topological ordering from data and mutual relations between data categories. Topologically similar elements of data are gathered close to each other in the obtained view, while different ones are separated in remote clusters. Additionally, these methods can reflect data category density, that is, less frequently occurring items in data are represented by smaller clusters and those occurring more frequently— by larger clusters. Thanks to such an ability to maintain the topology, visualization methods utilizing neural networks, and in particular modifications of these methods specially created for the analysis of this type of data, are perfect also for visual analysis of sparse datasets [33–38]. Sparse data are characterized by that although vectors occurring in such data have very large dimensions, only a small percent of coordinates of these vectors is nonzero. Data of this type occur only in very specific areas, e.g., in the WWW analysis and text datasets and is not related to the subject of this article.

5 Experiments' results

To compare the effectiveness of methods presented in the paper, real seven-dimensional data describing samples of coal in terms of their susceptibility to fluidal gasification was used. These data were obtained thanks to the conducted physicochemical processes on 99 samples coming from two hard coal mines. In this way, seven features describing a given sample were obtained for each sample: total sulfur content, hydrogen content, nitrogen content, chlorine content, total carbon content, heat of combustion and ash content. Therefore, each sample can be represented by a vector in the seven-dimensional space of features. The whole dataset was published earlier [39]. A system written in the C++ programming language, specially created for the conducted experiments, was used. It was created based on theories presented in Sects. 2 and 3. During the research, the effectiveness of Kohonen maps and autoassociative neural networks in the qualitative analysis of multidimensional data was verified. The qualitative analysis of the presented data using each of the methods was conducted for this purpose. The purpose of the analysis was to state whether coal samples with different susceptibility to gasification occupy separate subareas of the multidimensional space of characteristics. This in turn allowed to state whether selected characteristics are sufficient for the correct differentiation of samples well and poorly susceptible to fluidal gasification. The criterion for the readability of the multidimensional visualization results introduced in earlier papers [1, 10] was used for comparing the effectiveness of methods. It consists in drawing a curve

separating images of points belonging to different classes in a figure. The more complicated this curve is, the less readable the view allowing to indicate the possibility to separate subsets of points from each other is. It was assumed that the curve consists of arcs and the more complicated it is, the more inflection points it has. Inflection points are points joining arcs turning into different sides.

During tests, in the case of autoassociative neural networks, the most readable views were obtained with the network consisting of six layers in total. Three layers were used for the change of seven-dimensional input space into two outputs $(y_{3,1}, y_{3,2})$ of the interlayer used for the visualization. Then, three layers were used to change these two outputs into seven network outputs. Figure 9 presents the view obtained using the autoassociative neural network. It is visible in the figure that signals being the response to data representing samples of coal with different susceptibility to gasification accumulate in aggregations. These aggregations can easily be separated from each other. We can conclude from this that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation in the susceptibility to fluidal gasification. Areas occupied by signals representing different susceptibility to gasification are separated by a curve in the figure. This curve does not have any inflection points. This means

that the obtained view constitutes the most readable result from the perspective of the assumed criterion for the readability of the visualization results. This view was obtained with ITER = 340, which means that network learning was repeated 340 times for each sample. It must be noted that the autoassociative neural network learning proceeds without any information on belonging of data vectors to specific classes. In this connection, the way in which signals of the interlayer used for the visualization are grouped depends only on some features of these data observed by the network.

Figure 10 presents the view obtained using the autoassociative neural network for data describing samples of coal with a different degree of susceptibility to fluidal gasification, but with the omission of the chlorine content effect. Such an approach causes that significantly more samples are marked as well susceptible to gasification. The consequence of using such samples for gasification is only a small increase in the level of contamination. It is also visible in the figure that signals being the response to data representing samples of coal with different susceptibility to gasification accumulate in aggregations. These aggregations can easily be separated from each other. We can conclude from this that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation in susceptibility to fluidal gasification also
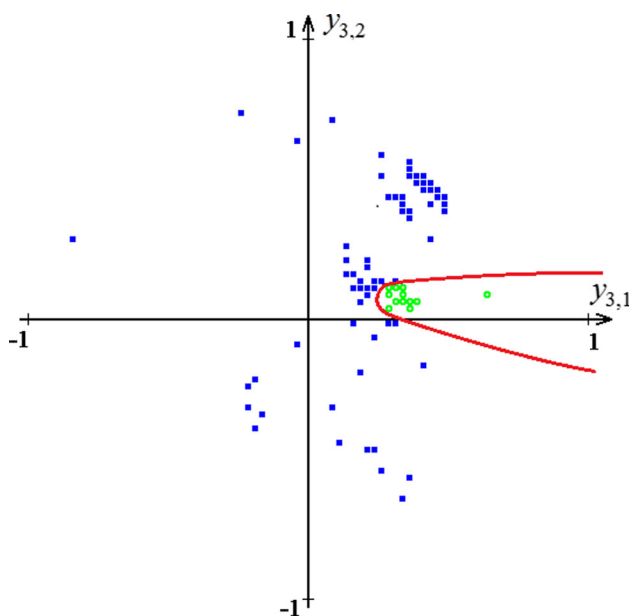

Fig. 9 The view obtained using the autoassociative neural network. Signals representing samples of coal poorly susceptible to gasification are marked with symbol filled square; signals representing samples of coal well susceptible to gasification are marked with a circle
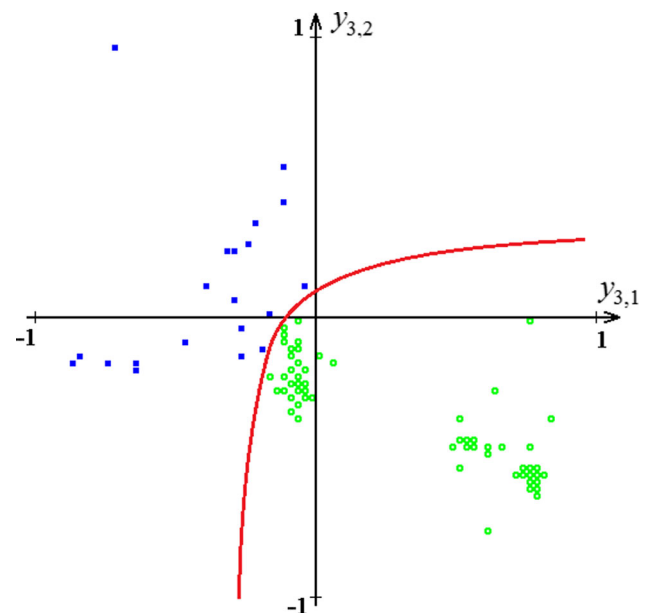

Fig. 10 The view obtained using the autoassociative neural network with the omission of a condition concerning the chlorine content. Signals representing samples of coal poorly susceptible to gasification are marked with symbol filled square; signals representing samples of coal well susceptible to gasification are marked with a circle

with the omission of the chlorine content effect. Areas occupied by signals representing different susceptibility to gasification are separated by a curve. This curve also does not have any inflection points. This view was obtained with ITER = 9000, which means that network learning was repeated 9000 times for each sample. Additionally, random initial values of weights other than for the previous figure were generated in order to obtain the above view.

Figures 11, 12, 13 and 14 present the results obtained using Kohonen maps. During the tests, the most readable views were obtained with the neural network consisting of 40 rows and 40 columns of neurons, thus of 1600 neurons. Figure 11 presents the view of a response of the neural network to one of input data vectors representing a sample well susceptible to fluidal gasification. The field brightness level denotes the values of a response of a neuron in a given position to a given sample. The brighter field denotes a greater value, and the darker field—a lower value at the neuron output. The winner neuron, that is, the neuron which obtained the largest value at the output, is marked with symbol 'x'. Figure 12 presents the view of a response of the neural network to one of input data vectors representing a sample poorly susceptible to fluidal gasification. Similarly as in the previous figure, the winner neuron, that is, the neuron which obtained the largest value at the output, is marked with symbol 'x'.

Figure 13 was obtained as a result of displaying winner neurons obtained in Figs. 11 and 12 along with winner
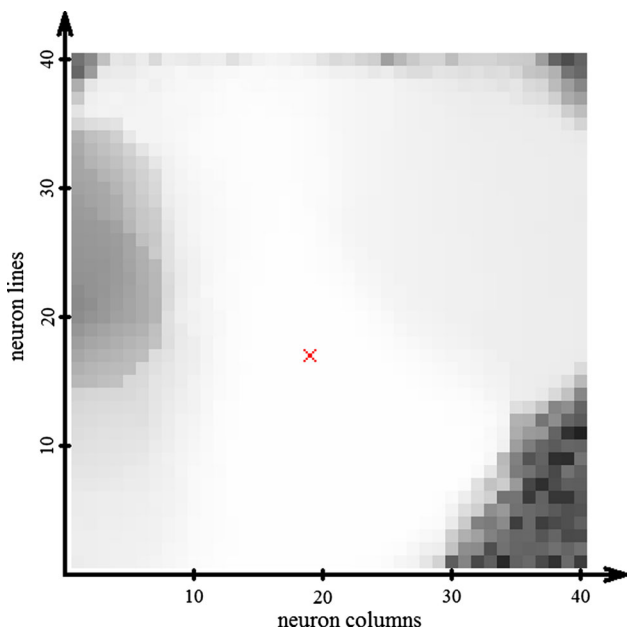

Fig. 12 The view of a response of Kohonen maps to one of input data vectors representing a sample poorly susceptible to fluidal gasification. The winner neuron is marked with symbol 'x'

Fig. 11 The view of a response of Kohonen maps to one of input data vectors representing a sample well susceptible to fluidal gasification. The brighter field denotes a greater value, and the darker field—a lower value at the neuron output. The winner neuron is marked with symbol 'x'
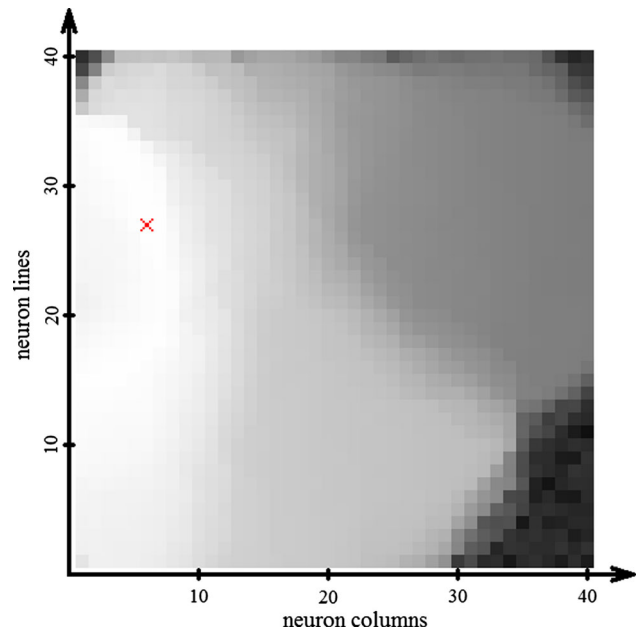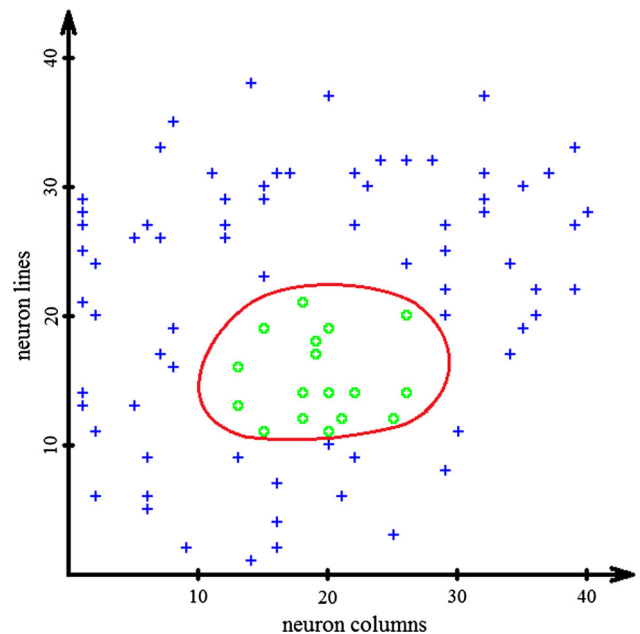

Fig. 13 The view obtained using Kohonen maps. Neurons representing samples of coal poorly susceptible to gasification are marked with symbol +; neurons representing samples of coal well susceptible to gasification are marked with a circle

neurons representing each of the remaining data vectors. It can be seen how neurons divided data representing a different degree of susceptibility to gasification between themselves. It is visible that neurons representing samples of coal with the same susceptibility to gasification accumulate in aggregations. These aggregations can easily be
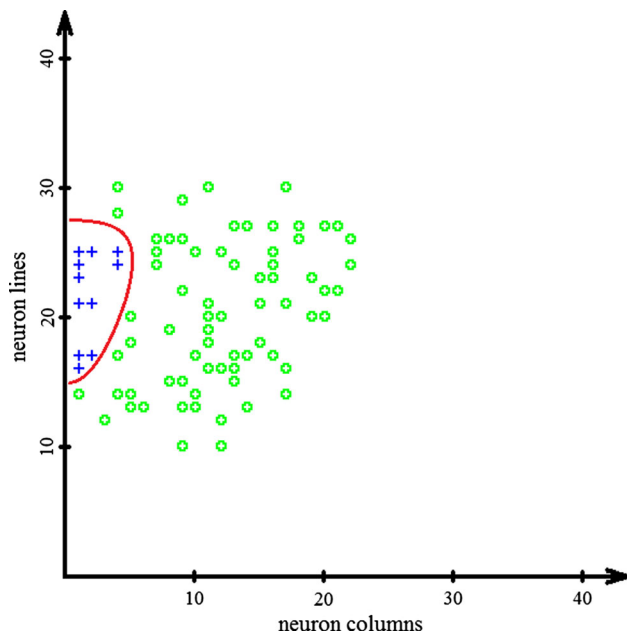
Fig. 14 The view obtained using Kohonen maps with the omission of a condition concerning the chlorine content. Neurons representing samples of coal poorly susceptible to gasification are marked with symbol +; neurons representing samples of coal well susceptible to gasification are marked with a circle

separated from each other. Based on the use of Kohonen maps, we can conclude that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation in susceptibility to fluidal gasification. Areas occupied by neurons representing different susceptibility to gasification are separated by a curve in the figure. This curve does not have any inflection points. This view was obtained with parameters MAX_DISTANCE = 7 and ITER = 1570. The assumed MAX_DISTANCE denotes that during the network self-organization, weights of neurons at the distance of less than 7 from the winner neuron were changed. The assumed ITER denotes the number of repetitions of the network self-learning performed for all data vectors. It must be noted that Kohonen maps self-learning proceeds without any information on belonging of data vectors to specific classes. In this connection, the way in which neurons are grouped depends only on some features of these data observed by the network.

Figure 14 presents the view obtained using Kohonen maps for data describing samples of coal with a different degree of susceptibility to fluidal gasification, but with the omission of the chlorine content effect. It can be seen how in this case neurons divided data representing a different degree of susceptibility to gasification between themselves. Also here, it is visible that neurons representing samples of

coal with the same susceptibility to gasification accumulate in aggregations. These aggregations can easily be separated from each other. Based on the use of Kohonen maps, we can conclude that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation of susceptibility to fluidal gasification also with the omission of the chlorine content effect. Areas occupied by neurons representing different susceptibility to gasification are separated by a curve in the figure. This curve also does not have any inflection points. This view was obtained with parameters MAX_DISTANCE = 4 and ITER = 820.

6 Discussion

It follows from the analysis presented above that both autoassociative neural networks and Kohonen maps allowed to obtain views in which images of samples representing different susceptibility to fluidal gasification can be separated by a curve without inflection points. The equally readable effect was obtained with the omission of the chlorine content effect. For comparison, Figs. 15, 16, 17, 18, 19, 20, 21 and 22 present the most readable views which were obtained during the conducted experiments for the analyzed seven-dimensional data describing coal samples in terms of their susceptibility to fluidal gasification using other visualization methods. Figures 15 and 16 present views obtained using the perspective-based observational tunnels method [10, 13, 15]. It can be observed in both figures that the curve separating areas occupied by samples with different susceptibility to fluidal gasification does not have any
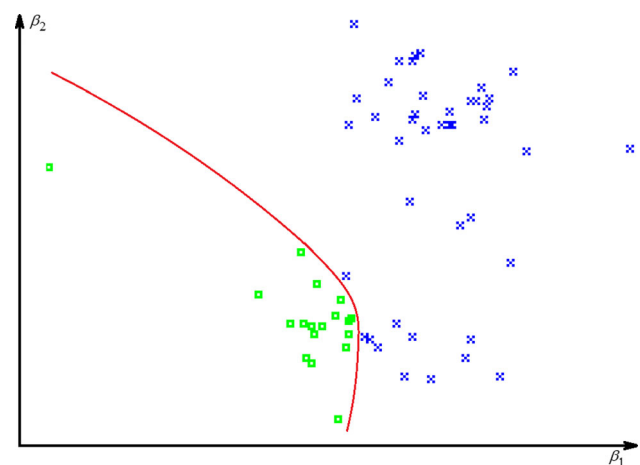

Fig. 15 The view obtained using the perspective-based observational tunnels method. Samples of coal poorly susceptible to gasification are marked with symbol x; samples of coal well susceptible to gasification are marked with a circle
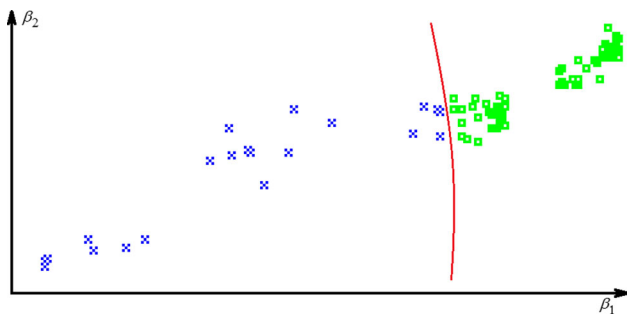
Fig. 16 The view obtained using the perspective-based observational tunnels method with the omission of the chlorine content effect. Samples of coal poorly susceptible to gasification are marked with symbol x; samples of coal well susceptible to gasification are marked with a circle
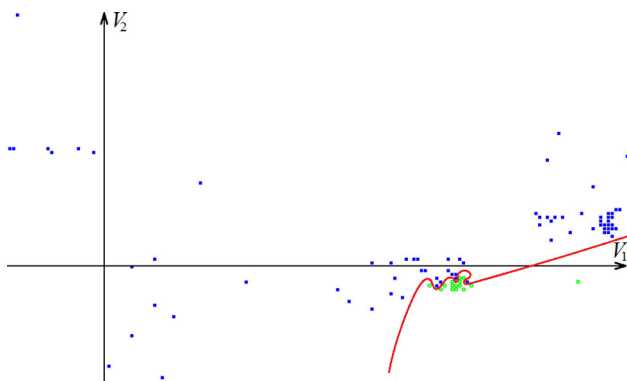

Fig. 17 The view obtained using the PCA method. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle. The curve separating areas has five inflection points
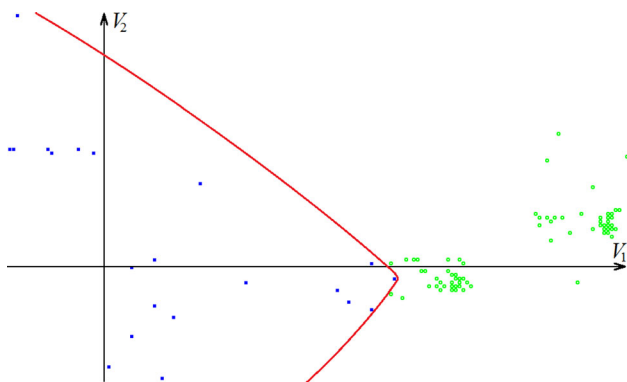

Fig. 18 The view obtained using the PCA method with the omission of the chlorine content effect. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle

inflection points. It follows that from the perspective of the assumed criterion, the perspective-based observational tunnels method provides just as readable views as Kohonen maps and autoassociative neural networks.
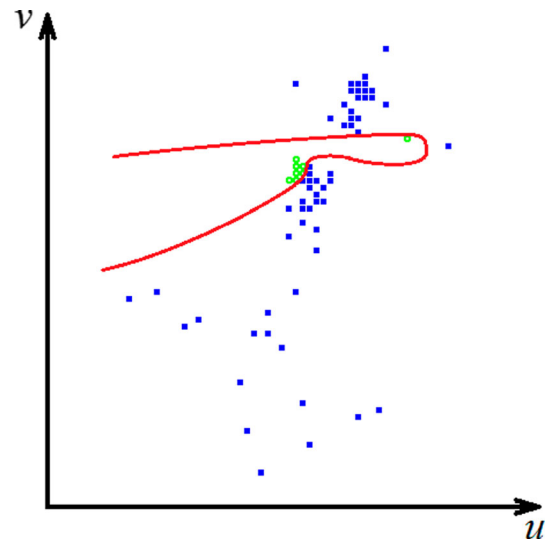

Fig. 19 The view obtained using multidimensional scaling. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle. The curve separating areas has two inflection points
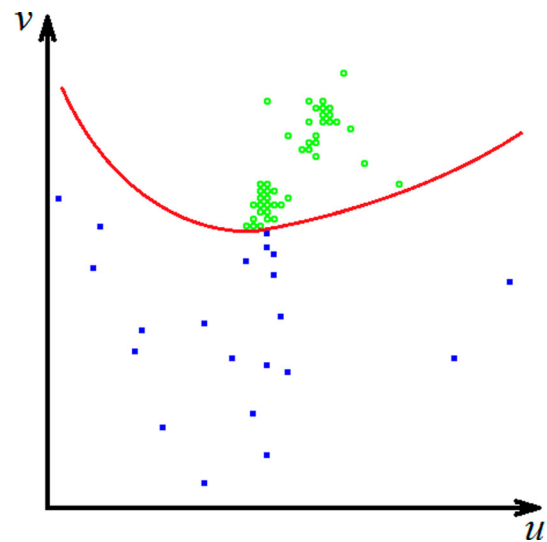

Fig. 20 The view obtained using multidimensional scaling with the omission of the chlorine content effect. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle

Figures 17 and 18 present views obtained using PCA [11, 16–20]. In order to obtain these views, the analyzed data covariance matrix was calculated and eigenvectors of this matrix (corresponding to two largest eigenvalues in terms of module) were obtained. The presented views constitute the orthogonal projection onto these vectors. Figure 17 shows that the curve separating areas occupied by samples with different susceptibility to fluidal gasification has five inflection points. It follows that from the
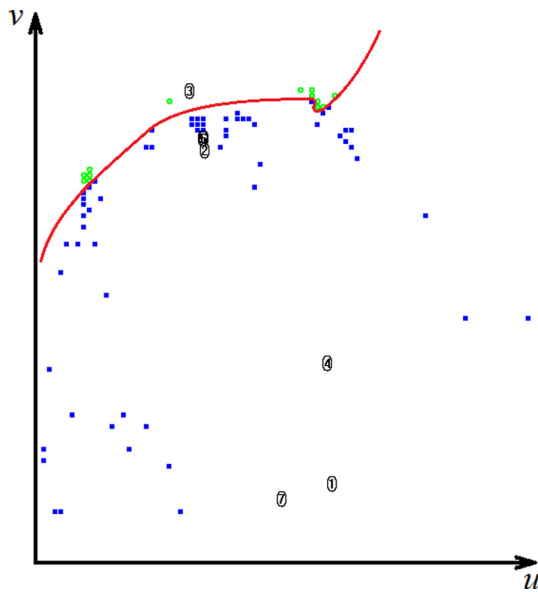
Fig. 21 The view obtained using relevance maps. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle. A digit with value *i* means a reference point representing the *i*th coordinate. The curve separating areas has one inflection point
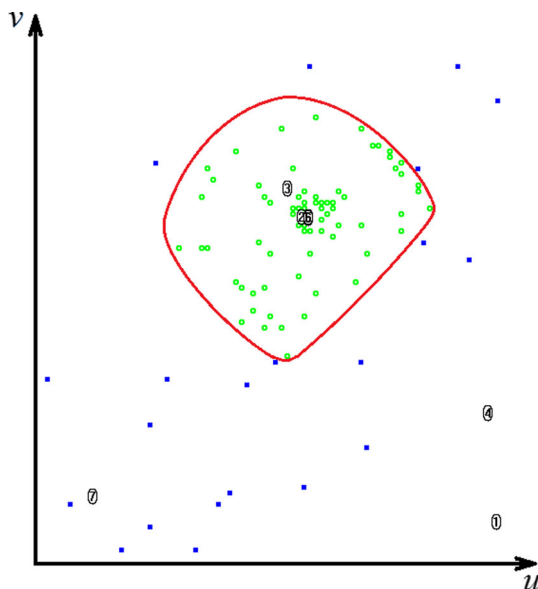

Fig. 22 The view obtained using relevance maps with the omission of the chlorine content effect. Samples of coal poorly susceptible to gasification are marked with symbol filled square; samples of coal well susceptible to gasification are marked with a circle. A digit with value *i* means a reference point representing the *i*th coordinate

perspective of the assumed criterion, this view obtained using PCA is significantly less readable than views obtained using Kohonen maps and autoassociative neural networks. However, the curve in Fig. 18 separating areas

with different susceptibility to gasification with the omission of the chlorine content effect has no inflection points. That is from the perspective of the assumed criterion, this view is just as readable as views obtained using Kohonen maps and autoassociative neural networks.

Figures 19 and 20 present views obtained using multidimensional scaling [12, 21–23]. Figure 19 shows that the curve separating areas occupied by samples with different susceptibility to fluidal gasification has two inflection points. However, the curve in Fig. 20 separating areas with different susceptibility to gasification with the omission of the chlorine content effect has no inflection points. It follows that from the perspective of the assumed criterion, the first of the views obtained using multidimensional scaling is significantly less readable and the second one is just as readable as views obtained using Kohonen maps and autoassociative neural networks.

Figures 21 and 22 present views obtained using the relevance maps method [14, 24, 25]. Figure 21 shows that the curve separating areas occupied by samples with different susceptibility to fluidal gasification has one inflection point. It follows that from the perspective of the assumed criterion, this view is less readable than views obtained using Kohonen maps and autoassociative neural networks. However, the curve in Fig. 22 separating areas with different susceptibility to gasification with the omission of the chlorine content effect has no inflection points. That is from the perspective of the assumed criterion, this view is just as readable as views obtained using Kohonen maps and autoassociative neural networks.

Table 1 presents a summary of the readability of results of visualization using Kohonen maps and autoassociative neural networks with methods of PCA, multidimensional scaling, relevance maps and perspective-based observational tunnels method. As seen, the smallest sum of inflection points equal to zero was obtained using Kohonen maps, autoassociative neural networks and perspective-based observational tunnels method. It follows that from the perspective of the assumed criterion these three methods from the tested methods allowed to obtain the most readable views, and in this way they simultaneously occupy the first position. The next were, according to the readability, relevance maps, multidimensional scaling and finally PCA. It should be noted that the above comparison of methods concerns solely the readability of views of real seven-dimensional data describing coal samples in terms of their susceptibility to fluidal gasification and the same data with the omission of the chlorine content effect. On the basis of the obtained results, it can also be stated that data with the omission of the chlorine content effect are significantly easier to analyze. This results from the fact that in the case of these data all of the analyzed methods obtained results in which the curve separating areas with

Table 1 Summary of the readability of visualization results

Location	Method	Inflection points	Inflection points (omission chlorine)	Sum
1	Autoassociative neural networks	0	0	0
1	Kohonen maps	0	0	0
1	Perspective-based observational tunnels	0	0	0
2	Relevance maps	1	0	1
3	Multidimensional scaling [MDS]	2	0	2
4	Principal component analysis [PCA]	5	0	5

The number of inflection points of the curve separating areas with different susceptibility to gasification and the number of inflection points of the curve separating areas with different susceptibility to gasification with the omission of the chlorine content effect are presented for each method. The last column presents the sum of inflection points

different susceptibility to gasification has no inflection points.

7 Conclusions

As a result of the conducted experiments, the qualitative analysis of seven-dimensional real data describing coal samples in terms of their susceptibility to fluidal gasification allowed to indicate:

1. In the case of autoassociative neural networks, the most readable views were obtained with the neural network consisting of six layers in total. Three layers were used for changing the seven-dimensional input space into two outputs of the interlayer used for the visualization. Then, three layers were used to change these two outputs into seven network outputs.
2. In the case of Kohonen maps, the most readable views were obtained with the neural network consisting of 40 rows and 40 columns of neurons, thus of 1600 neurons. During the analysis of data representing samples with different susceptibility to fluidal gasification, the most readable view was obtained with parameter MAX_DISTANCE = 7. During the analysis of the same data with the omission of the chlorine content effect, the most readable view was obtained with parameter MAX_DISTANCE = 4.
3. The qualitative analysis using autoassociative neural networks allowed to indicate that signals being the response to data representing samples of coal with different susceptibility to gasification accumulate in aggregations. These aggregations can easily be separated from each other. It can be concluded from this that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation of susceptibility to fluidal gasification.

4. The qualitative analysis using Kohonen maps allowed to indicate that neurons representing samples of coal with different susceptibility to gasification accumulate in aggregations. These aggregations can easily be separated from each other. It can be concluded from this that samples of coal well and poorly susceptible to gasification occupy separate subareas of the multidimensional space of features. This in turn allows to state that selected features are sufficient for the correct differentiation of susceptibility to fluidal gasification.
5. Both autoassociative neural networks and Kohonen maps were allowed to obtain views in which images of samples representing different susceptibility to fluidal gasification can be separated by a curve without inflection points. The equally readable effect was obtained during the data analysis with the omission of the chlorine content effect.
6. The smallest sum of inflection points equal to zero was obtained using Kohonen maps, autoassociative neural networks and perspective-based observational tunnels method. It follows that from the perspective of the assumed criterion these three methods from the tested methods allowed to obtain together the most readable views. The next were, according to the readability, relevance maps, multidimensional scaling and finally PCA.

Compliance with ethical standards

References

1. Jamroz D, Niedoba T (2015) Comparison of selected methods of multi-parameter data visualization used for classification of coals. Physicochem Probl Miner Process 51(2):769–784. https://doi.org/10.5277/ppmp150233

2. Kraaijveld MA, Mao J, Jain AK (1995) A nonlinear projection method based on Kohonen's topology preserving maps. IEEE Trans Neural Netw 6(3):548–559. https://doi.org/10.1109/72.377962

3. Aldrich C (1998) Visualization of transformed multivariate data sets with autoassociative neural networks. Pattern Recognit Lett 19(8):749–764. https://doi.org/10.1016/S0167-8655(98)00054-3

4. Jamroz D (2014) Application of multi-parameter data visualization by means of autoassociative neural networks to evaluate classification possibilities of various coal types. Physicochem Probl Miner Process 50(2):719–734. https://doi.org/10.5277/ppmp140224

5. Jamroz D, Niedoba T (2015) Application of multidimensional data visualization by means of self-organizing Kohonen maps to evaluate classification possibilities of various coal types. Arch Min Sci 60(1):39–50. https://doi.org/10.1515/amsc-2015-0003

6. Rubio JJ, Pan Y, Lughofer E, Chen M, Qiu J (2019) Fast learning of neural networks with application to big data processes. Neurocomputing. https://doi.org/10.1016/j.neucom.2019.10.057

7. Rubio JJ (2009) SOFMLS: online self-organizing fuzzy modified least-squares network. IEEE Trans Fuzzy Syst 17(6):1296–1309. https://doi.org/10.1109/TFUZZ.2009.2029569

8. Meda-Campana JA (2018) On the estimation and control of nonlinear systems with parametric uncertainties and noisy outputs. IEEE Access 6:31968–31973. https://doi.org/10.1109/ACCESS.2018.2846483

9. Rubio JJ, Garcia E, Ochoa G, Elias I, Cruz DR, Balcazar R, Lopez J, Novoa JF (2019) Unscented Kalman filter for learning of a solar dryer and a greenhouse. J Intell Fuzzy Syst 37(5):6731–6741. https://doi.org/10.3233/JIFS-190216

10. Jamroz D (2017) The perspective-based observational tunnels method: a new method of multidimensional data visualization. Inf Vis 16(4):346–360. https://doi.org/10.1177/1473871616686634

11. Jamroz D, Niedoba T, Surowiak A, Tumidajski T (2016) The use of the visualisation of multidimensional data using PCA to evaluate possibilities of the division of coal samples space due to their suitability for fluidised gasification. Arch Min Sci 61(3):523–535. https://doi.org/10.1515/amsc-2016-0038

12. Jamroz D, Niedoba T, Surowiak A, Tumidajski T, Szostek R, Gajer M (2017) Application of multi-parameter data visualization by means of multidimensional scaling to evaluate possibility of coal gasification. Arch Min Sci 62(3):445–457. https://doi.org/10.1515/amsc-2017-0034

13. Jamroz D (2018) The analysis of the effectiveness of the perspective-based observational tunnels method by the example of the evaluation of possibilities to divide the multidimensional space of coal samples. In: Computational science—ICCS 2018, lecture notes in computer science. Springer, Cham, vol 10862, pp 675–682. https://doi.org/10.1007/978-3-319-93713-7_64

14. Jamroz D, Niedoba T, Surowiak A (2016) Application of relevance maps method to evaluate the suitability of coal samples for fluidal gasification process. In: 1st International conference on the sustainable energy and environment development (SEED 2016), E3S web of conferences. vol 10, p 00065. https://doi.org/10.1051/e3sconf/20161000065

15. Jamroz D (2018) Application of perspective-based observational tunnels method to visualization of multidimensional fractals. In: Artificial intelligence and soft computing, ICAISC 2018, lecture notes in computer science. Springer, Cham, vol 10842, pp 364–375. https://doi.org/10.1007/978-3-319-91262-2_33

16. Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2(11):559–572. https://doi.org/10.1080/14786440109462720

17. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24(6):417-441–498-520. https://doi.org/10.1037/h0071325

18. Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York. https://doi.org/10.1007/b98835

19. Li W, Yue HH, Valle-Cervantes S, Qin SJ (2000) Recursive PCA for adaptive process monitoring. J Process Control 10(5):471–486. https://doi.org/10.1016/S0959-1524(00)00022-6

20. Niedoba T (2014) Multi-parameter data visualization by means of principal component analysis (PCA) in qualitative evaluation of various coal types. Physicochem Probl Miner Process 50(2):575–589. https://doi.org/10.5277/ppmp140213

21. Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29(1):1–27. https://doi.org/10.1007/BF02289565

22. Kim SS, Kwon S, Cook D (2000) Interactive visualization of hierarchical clusters using MDS and MST. Metrika 51:39–51. https://doi.org/10.1007/s001840000043

23. Jamroz D (2014) Application of multidimensional scaling to classification of various types of coal. Arch Min Sci 59(2):413–425. https://doi.org/10.2478/amsc-2014-0029

24. Assa J, Cohen-Or D, Milo T (1999) RMAP: a system for visualizing data in multidimensional relevance space. Vis Comput 15(5):217–234. https://doi.org/10.1007/s003710050174

25. Niedoba T (2015) Application of relevance maps in multidimensional classification of coal types. Arch Min Sci 60(1):93–106. https://doi.org/10.1515/amsc-2015-0007

26. Gennings C, Dawson KS, Carter WH, Myers RH (1990) Interpreting plots of a multidimensional dose-response surface in a parallel coordinate system. Biometrics 46(3):719–735. https://doi.org/10.2307/2532091

27. Chatterjee A, Das PP, Bhattacharya S (1993) Visualization in linear programming using parallel coordinates. Pattern Recognit 26(11):1725–1736. https://doi.org/10.1016/0031-3203(93)90027-T

28. Chou SY, Lin SW, Yeh CS (1999) Cluster identification with parallel coordinates. Pattern Recognit Lett 20(6):565–572. https://doi.org/10.1016/S0167-8655(99)00018-5

29. Inselberg A (2009) Parallel coordinates: visual multidimensional geometry and its applications. Springer, New York. https://doi.org/10.1007/978-0-387-68628-8

30. Sobol MG, Klein G (1989) New graphics as computerized displays for human information processing. IEEE Trans Syst Man Cybern 19(4):893–898. https://doi.org/10.1109/21.35357

31. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43(1):59–69. https://doi.org/10.1007/BF00337288

32. Kohonen T (1989) Self-organization and associative memory. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-88163-3

33. Melka J, Mariage JJ (2019) Adapting self-organizing map algorithm to sparse data. In: Sabourin C, Merelo J, Madani K,

Warwick K (eds) Computational intelligence. IJCCI 2017. Studies in computational intelligence. Springer, Cham, vol 829, pp 139–161. https://doi.org/10.1007/978-3-030-16469-0_8

34. Kaski S, Honkela T, Lagus K, Kohonen T (1998) WEBSOM-selforganizing maps of document collections. Neurocomputing 21(1–3):101–117

35. Lawrence RD, Almasi GS, Rushmeier HE (1999) A scalable parallel algorithm for self-organizing maps with applications to sparse data mining problems. Data Min Knowl Discov 3(2):171–195

36. Maiorana F (2008) Performance improvements of a Kohonen self organizing classification algorithm on sparse data sets. In: Proceedings of the 10th WSEAS international conference on mathematical methods, computational techniques and intelligent systems, MAMECTIS'08. World Scientific and Engineering Academy and Society (WSEAS), pp 347–352

37. Melka J, Mariage J (2017) Efficient implementation of self-organizing map for sparse input data. In: Proceedings of the 9th international joint conference on computational intelligence, IJCCI 2017. Funchal, Madeira, Portugal, pp 54–63

38. Olteanu M, Villa-Vialaneix N (2016) Sparse online self-organizing maps for large relational data. In advances in self-organizing maps and learning vector quantization. In: Proceedings of WSOM 2016. Advances in intelligent systems and computing. Springer, Houston, Texas, USA, vol 428, pp 27–37

39. Gawenda T, Krawczykowski D, Marciniak-Kowalska J (2014) Investigations of coal beneficiation by mechanical mineral processing, volume III: Investigation of the coal preparation process for terrestrial gasification in a fluidized bed gas generator with the application of mechanical processes of mineral engineering. Grafpol, Wroclaw