Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Data Mining in Gravity Field by Utilizing Clustering by Self-organizing Maps (Case Study in Southern Part of Iran)

Reza Shahnavehsi, and Farnusch Hajizadeh*

Department of Mining Engineering Faculty of Engineering, Urmia University, Urmia, Iran

Article Info

Received 4 August 2024
Received in Revised form 24 January 2025
Accepted 25 February 2025
Published online 25 February 2025

DOI: 10.22044/jme.2025.14879.2830

Keywords

Clustering, Gravity
Clustering
Gravity
Self-organizing maps
Numerical analysis

Abstract

The present work is mainly about a method for illustrating the relation between the raw data in the same time; clustering is a key procedure to solve the problem of data division; also illustrating the connection among the elements of the research area simultaneously is important. Therefore, we propose a novel kind of clustering for data mining in the gravity field to reach the presenting connection among all elements in the same time. For this research work, 867 gravity surveying points were collected in the southern part of Iran (near diapir of Larestan) with a range of absolute gravity from 978579.672 to 978981.186. In this paper, clustering by self-organizing- maps, by utilizing scatter plot matrix is utilized for detecting the relation between the easting, northing, elevation, and absolute gravity simultaneously. In the proposed method, the relations between arrays, two by two, are defined, and like matrix, each raw and column has different i and j values, which represent elements of the studied area, instead of number; for example, array A23 is data division between i = 2 or raw two (in our case northing) and j = 3 or column, three (in our case elevation). In this algorithm, firstly, by using self-organizing maps, clustering is done, and this processing is generated to all arrays by scatter plot matrix, and in all arrays, three clusters are proposed; the result of this clustering shows that in arrays A12, A13, A14, A21, A23, A24, A31, A32, A41, A42, clustering is performed perfectly, and the relationship between the parameters of the studied area near Larestan salt, diaper, can be useful in notifying the properties of this salt diapir.

1. Introduction

Geophysical modeling is a type of model connection among the data of geophysics, and the properties of sub-surface of earth [19], and specifying and describing the densities of the anomaly, which are caused by the gravity of sub-surface structures efficacy is the purpose of gravity surveying [14]. This method trails for identifying the anomalous mass within the interior bodies inside of the earth [20]. Therefore, it is popular that modeling of gravity anomaly in most cases is not unique [9]. One of the main steps for promoting of using of gravity method is integrating of knowledge, calculated out of several methods [26], and due to the significance of datamining is rising [13], and in the area of data mining and classification, clustering is the main field [12], and because datemining (one the area of machine learning) has a wide range of studies with so many applications in geophysical modeling (especially, in gravity modeling), for creating unique models; these methods build automated maps out of raw data, and clustering by self-organizing maps is manner, which is utilized broadly in gravity field [5], but in the previous studies, in gravity modeling, presentation relation of all elements of raw data in the same time can be found seldom, which is solved by the proposed model.

Clustering is significant in the area of 1- Pattern recognition, 2- classification 3- Model reduction 4- optimization, moreover. This type of numerical analysis possesses has many uses in so many fields [23]. By the basis of neural fields, self-organizing maps is introduced by Kohonen, and this manner, by the basis of the neural network, which causes by

* Corresponding author: farnooshhajizade@yahoo.com (F. Hajizadeh)

self-organizing process, this algorithm is done; in addition, this type of neural network, by using non-supervised algorithm, cause that the self-organizing maps are mapping dimensional from high space to lower space [8].

Several researchers promote the algorithm of clustering. For example, various elements are re-sampled to the same cluster. In another study, the graph division procedure is utilized, for solving the clustering performance; in addition, in another study, blended type of clustering is applied, and in some researches, in clustering algorithm, by classifying the element to transportation element and core element clustering algorithm is created [11]. In the area of gravity [24], introduced clustering methods for putting elements of joint gravity data points in the same clusters; also in the mentioned papers, only two elements are distributed to different clusters, and [25] introduced the clustering methods for putting elements of gravity data points in the same clusters, and [2] introduced the clustering method. In the mentioned papers, Despite of the mentioned papers, in the present paper, we use self-organizing maps as a tool of clustering. Other approaches in clustering were used variously, and digital image processing is one of the several uses of a self-organizing map, due to two reasons, the self-organizing maps are utilized significantly in clustering in comparison of other methods of clustering:

1. Organizing by the system, and retaining of topology are features of this approach; also onto the grid, after the training phase. inside the input data area, the self-organizing maps project the near date onto near prototype vectors, and due to this property, the input vectors with much similarity are more likely to place in one cluster.
2. This method visualizes the data prominently [27].

In this study, self-organizing maps are applied as an effective method for clustering, and for visualizing the relation between the parameters at the same time, a scatter plot matrix is used, and the following steps in the below structure are considered in this paper for reaching a novel type of clustering;

Firstly, we call self-organizing maps by the basic algorithm. Secondly, scatter plot matrix is defined according to position of arrays, and axes between two elements of each array. Thirdly, this clustering procedure (self-organizing maps clustering type) is applied to all arrays, by utilizing the scatter plot matrix algorithm. Self-organizing maps are classified according to the defined variables of each array to different clusters. In the last step, name of x-axis and name of y-axis, and title of each array are added. In the area of geophysics, especially in the gravity field, rare can study can be found that singular illustration can show all the elements of the studied area. In the same time, this unique method applies self-organizing maps to create the clustering method, and scatter plot matrix is generalized; this kind of clustering to all elements for reaching the purpose of illustrating all datasets simultaneously. Thus, this paper, with the approach of clustering by using self-organizing maps create novel type of clustering in the field of gravity for presenting data points in the same time, which is rarely found in the previous studies.

2. Method

2.1. Geology and location of studied area

The physiographic-tectonic zones with salt daipirs in Iran are: 1- Zagros mountain 2- basin of Persian Gulf 3-central part of Iranian basin, and Figure1 shows the main dipirs of Iran.

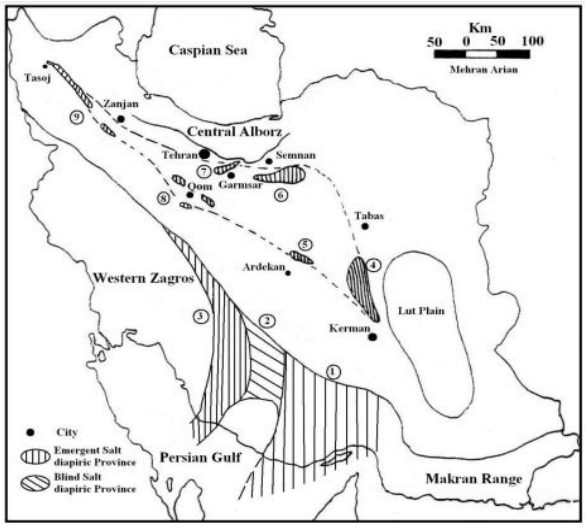

Figure 1. Main diapirs of Iran [3].

The diapir of studied area is placed in the basin of Persian Gulf. The trend of this structure is NW-SE; this diapir includes the south margin of non-complex folded Zagros's belt, which deposited on the norther area of passive margin of Arabian plate. In addition, Hormuz salt formed in the Persian Gulf in Late Precambrian to Early Cambrian.

The name of this area is taken out of the Iranian Hormuz island. In southern area of Zagros, in the Hormuzghan province, consists of 84 emergent diapirs; this zone can be assumed for the most important salt diapir in Zagros's salt basin. Most of

salt diapir in this area are pre-orogenic, but some of them are formed in post-orogenic [3].

The calculated data is placed in the southern part of Iran, and in norther part of the Hormuzghan province, near Larestan city, which is shown in Figure 2.

The main structure in the orogenic system in the larestan is salt diapirs. The Hormoz complex is constructed of 1- salt 2- brown and red shale 3- interbedded with black, white, and gray dolomites, 4- very fine siltstones, 5-sandstones, 6- gypsum 7- pyrite 8- iron oxide 9- igneous rocks. In the salt diapirs of Larestan region, the main body of many sedimentary rocks is the emergence of dolostones. They are frequently found as tiny as interbeds inside gypsum, marl sequences, and sometimes in gypsum-tuff parts observe in dolostones, and geological map of Larestan is illustrated in Figure 3.

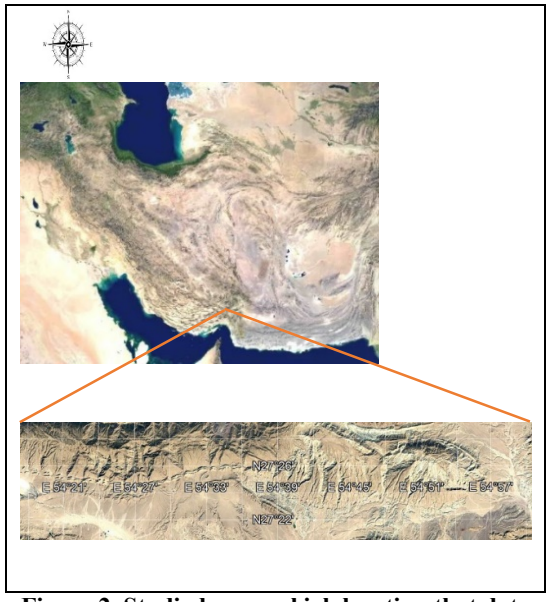

Figure 2. Studied area, which location that data collection was done.

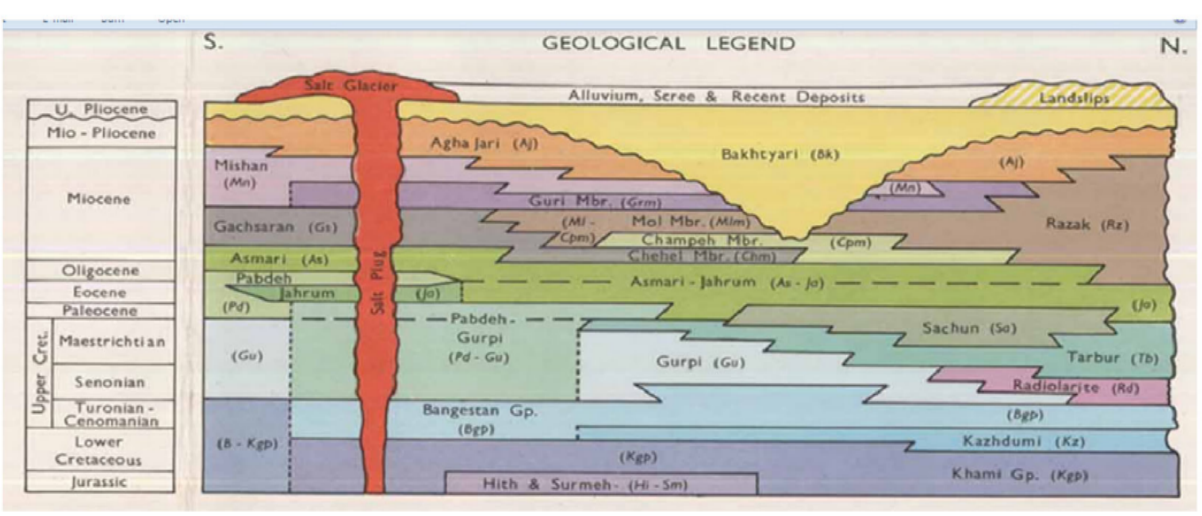

Figure 3. Geological map of Larestan with description of structure of the studied area and age of each structure [18].

2.2. Properties of method

In the clustering approach, the dataset to clusters of the same properties is distributed by clustering method [1]; also in the immense diversity of applied arrangement, the clustering method has the capability to name as principal learning method, and the usual goal of this approach is applying the method for dividing collected objects to groups [10]. Figure 4 shows the main structure of the most clustering method.

Kohonen introduced an unsupervised learning procedure in the area of artificial neural networks, which is called self-organizing maps [6]. Two layers of this system are 1- input layer 2- output layer.

The external information of each neuron is placed in the output layer via weight vectors, which is collected by Neurons onto the input layer, and the input layer of this system involves the same frame of back-propagation neural network (ANN), and the amount of neurons and sample dimensions are identical, and the output layer is placed in competing layer [28].

Tuning weight vectors till the map illustrates a portrait of the input data set is the fundamental axiom of self-organizing maps (SOM), and this algorithm continues up until the dataset amount of nodes of illustration becomes less than the amount

of items in data sets, and it is not possible to represent each input section out of the data space onto the map. Even, the goal of this type of neural network is to earn a configuration, in which the system reflects the distributed data, while the significant metric connection is retained [4]. Structure of the self-organizing map is shown in Figure 5.

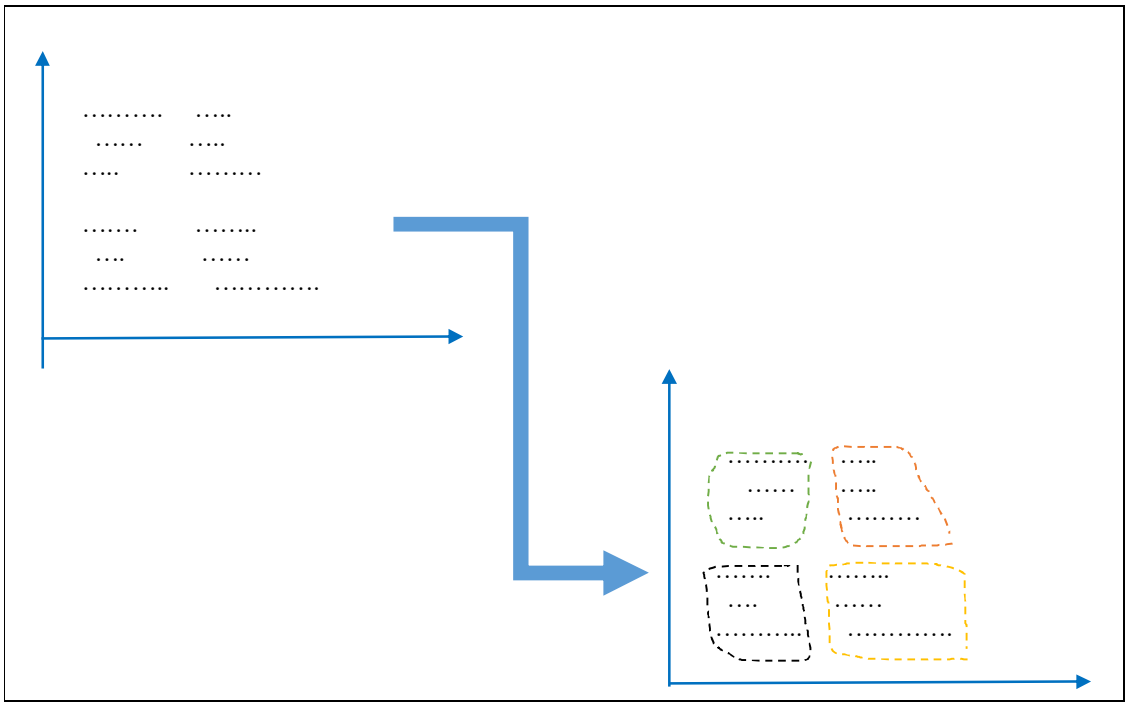

Figure 4. General structure of clustering.

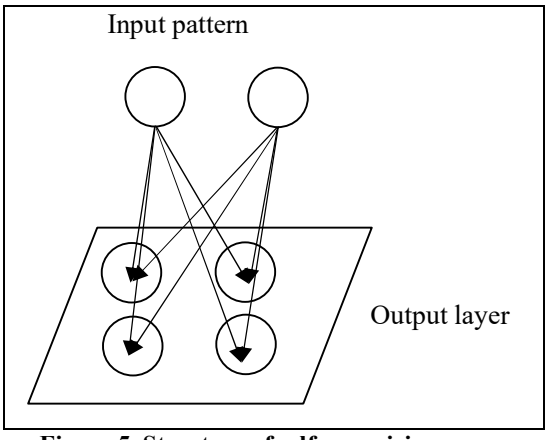

Figure 5. Structure of self-organizing maps.

Collecting of pattern (usually as measurements' representation's vector or a point into a multi-dimensional space) to similar basis cluster is the main structure of clustering analysis [17]. The optimal function for vector quantization is SOM. Clustering and classification, and visualizing of data are applications of SOM; however, it involves many other applications, and for high dimensional data, SOM as a clustering and projection function has been used. The competitive stage and the cooperative stage are the stages of SOM function [16]. Figure 6 shows clustering based on self-organizing maps.

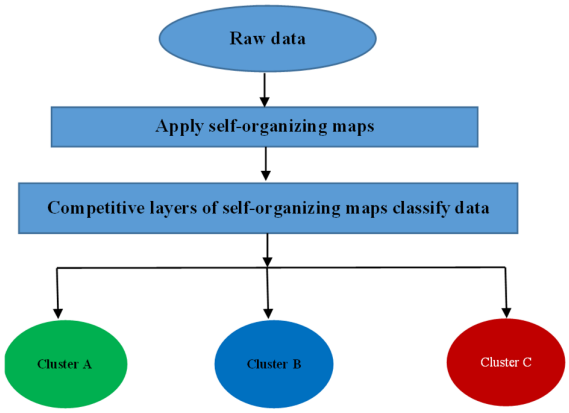

Figure 6. Clustering based on self-organizing maps with clustering by competitive layers of this type of neural network.

The overall structure of clustering by self-organizing maps is In the first step, we import defined variables. Secondly, by applying the related parameters of the competitive layer (this function is trained by SOM and separate data to different cluster, where each cluster is involved the

same properties, and has difference from other clusters) such as the number of clusters and learning rate to variables, the system is trained. After this steps, by utilizing the scatter plot matrix, this approach is generalized to all elements, and it causes that all variables, two by two, are divided to different cluster. In the last step, name of x-axis and name of y-axis and title of each array are defined; in the last step, addition plot in main diagonal (arrays A_{11}, A_{22}, A_{33}, A_{44}) are added; also title of mentioned array are added, and Figure 7 shows this procedure.

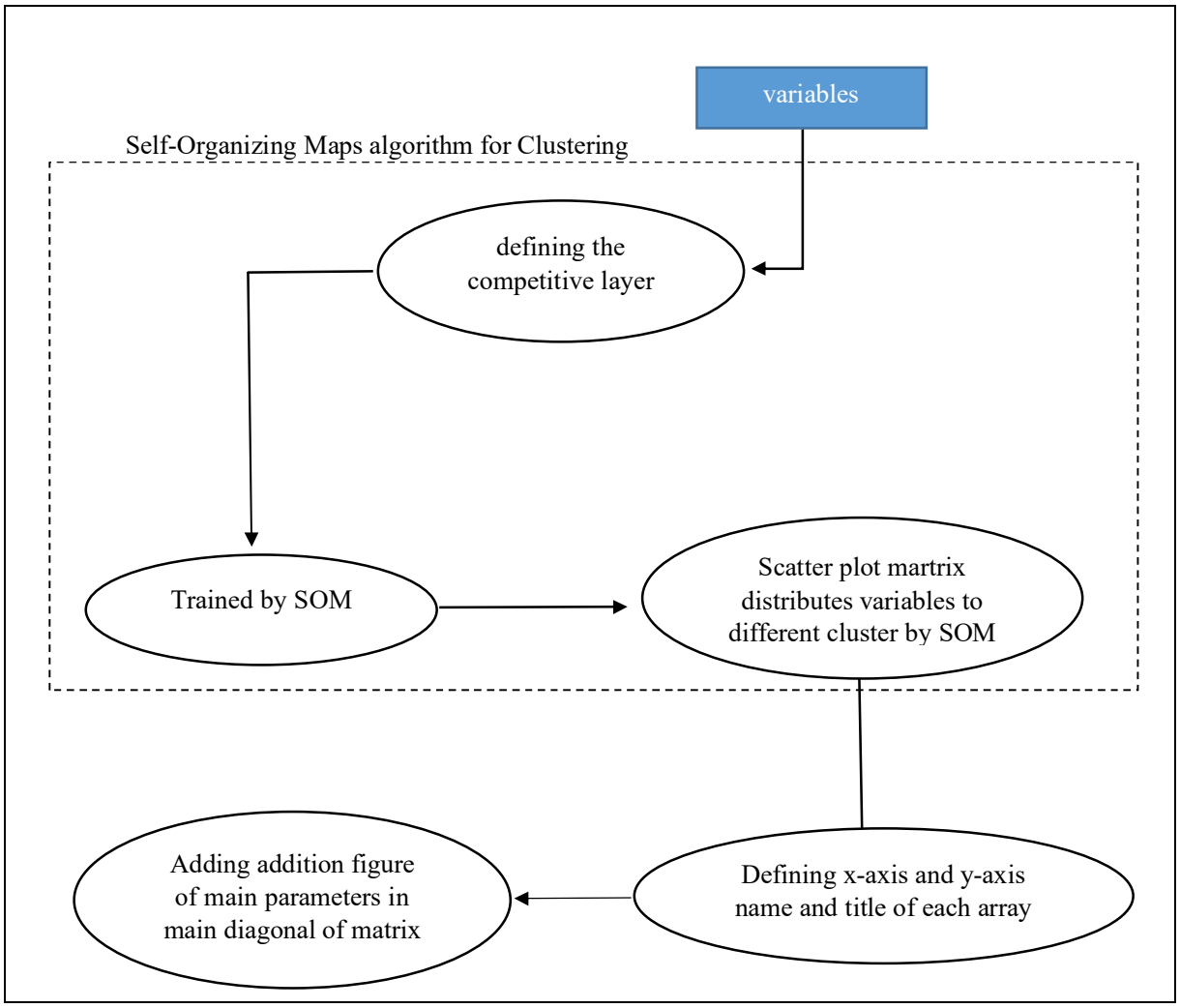

Figure 7. General function of clustering by self-organizing maps, which is disturbed by scatter matrix, which is illustrated by the main procedure of this clustering and adding name, and title and addition figure of main parameters to scatter plot matrix.

In the scatter plot matrix, instead of numerical arrays of the matrix, figures become arrays of the matrix. In this algorithm, the input data place in the main diagonal. Then, two by two, the raw data points are distributed to arrays of the matrix (place of the figure in scatter plot matrix shows the distributed data, as an example; array A_{34} in matrix $A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$ is defined clustering between element of raw three and the element of column four) then self-organizing maps is dividing data to different clusters, and Figure 8 shows the detail of this type of clustering for array A_{34}.

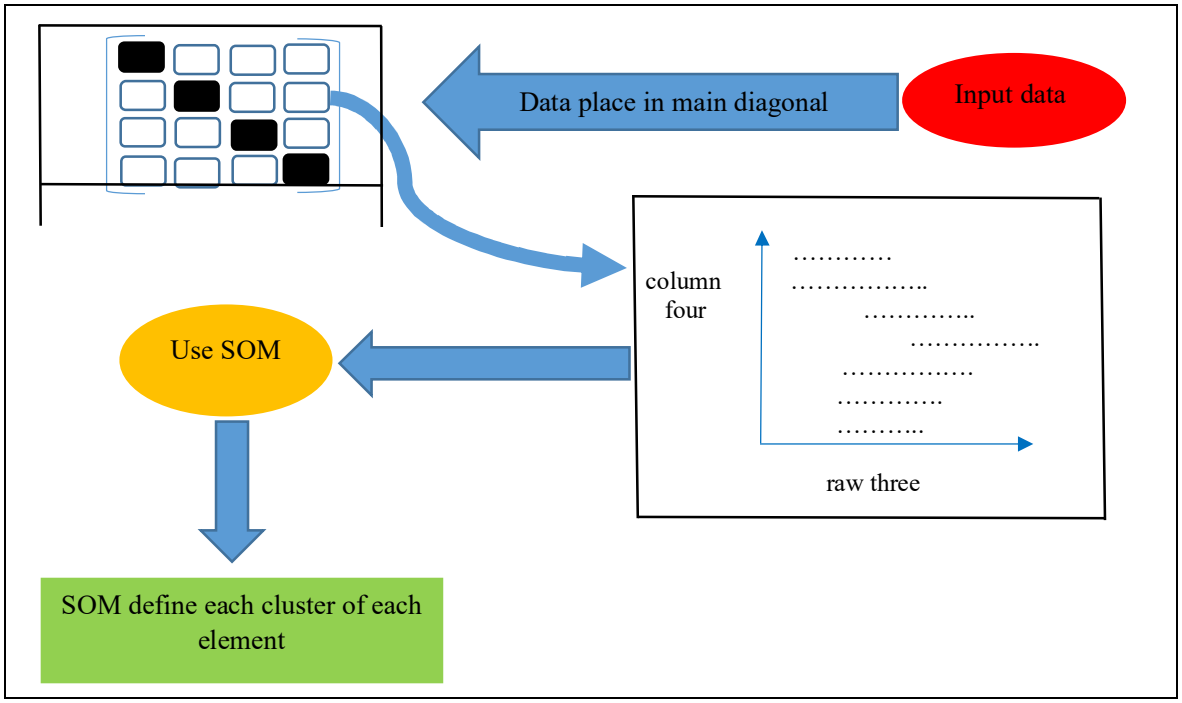

Figure 8. Detail of clustering with example of element A_{34}.

3. Data of studied area

We calculated the gravity data by CG3, and after some main corrections (like free air and Bouguer correction and other corrections) absolute gravity rate is computed, and absolute gravity is the main gravity of this paper, and we calculated the topography data by GPS. The number of data points are 867, and the area of the studied area is 880 km^2, and the distance between two station of prospecting is 500 meter, and the profile of the study area is north-south profile, and 51 profile lines was proposed for the studied area. The grid of the studied area shows in Figure 9, which shows the relation between easting and northing and gravity rate (absolute gravity rate) of the studied area, and the easting is placed in x-axis and northing is placed in y-axis.

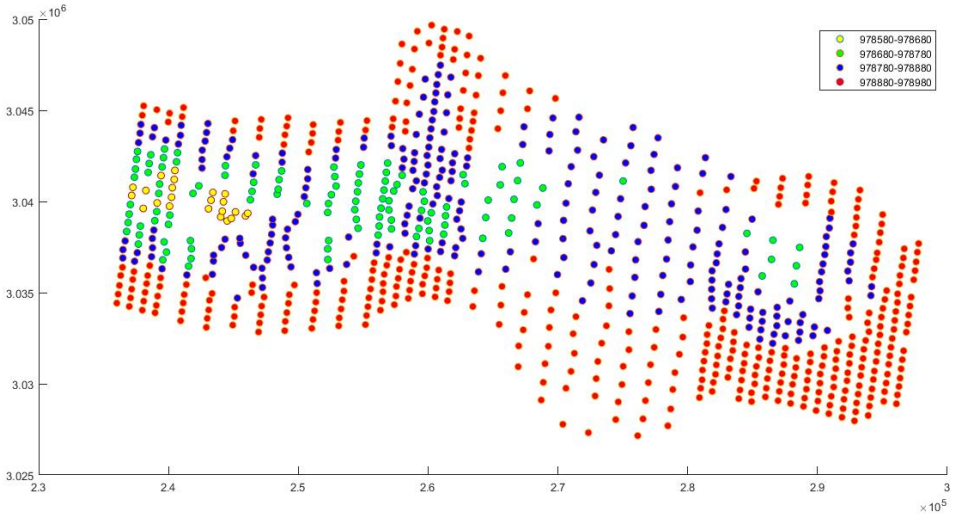

Figure 9. Grid of studied area with easting and northing (m) and absolute gravity (mgal).

Figure 9 shows absolute gravity, which ranges from 978579.672 to 978981.186, and the central part to western part of the studied area has a higher amount of absolute gravity rate, in comparison of other part of the studied area, and the western part

in the middle side has a maximum rate of absolute gravity rate.

4. Interpretation of algorithm by MATLAB code

It is better for more interpretation of the procedure we introduce the method by the MATLAB code. For a better comprehending of this algorithm, before describing the procedure of code, it is better for being algorithm more applicable. We introduce some prerequisite for algorithm; in this code, axis of each array due to properties of matrix are divided to two axes. In the x-axis, they are defined to i = 1,2,3,4, and in the y-axis, they are defined to j = 1,2,3,4 (number of i and j become the same until square matrix is created, and the number of i and j depend on the number of inputs, and become the same as the number of inputs). If we put name of A to matrix, the matrix for this procedure is $A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$; the main procedure of this code is:

Firstly, like most code, the raw data is imported; secondly, by using self-organizing maps, the raw data is trained, and Figure 10 shows this part of the code.

```
SelfOrganizingMaps=competlayer(number_of_clusters,...
Learning_rate_for_Kohonen_weights,Learning_rate_for_conscience_bias);
SelfOrganizingMaps.trainparam.epochs=300;
SelfOrganizingMaps=train(SelfOrganizingMaps,variables);
%% perform network
result=SelfOrganizingMaps(variables);
tableofclass=vec2ind(result);
```

Figure 10. Code of utilizing self-organizing maps, as the clustering method.

Figure 10 shows that by using compete layer algorithm, basic inputs of procedure of self-organizing maps are defined. In the next step, by utilizing the training function, this algorithm turns to self-organizing maps neural network; in the performed phase, by applying vec2ind, algorithm convert vectors to the class indices.

In the second section of algorithm, the main loop of algorithm is defined, and this section is divided to three parts; in the first part of this section, plotting of scatter plot matrix is programmed, and Figure 11 shows this part of the algorithm.

```
numeration=numeration+1;
if i~=j
    subplot(4,4,numeration);
    for category=1:number_of_clusters
    plot(variables(i,tableofclass==category),...
    variables(j,tableofclass==category),'.','color',colors(category,:));
```

Figure 11. Main loop of code, which define scatter plot matrix that causes date is distributed to different arrays.

In this part, if x-axis and y-axis do not have the same value, then plots of the mentioned arrays with specified properties of self-organizing maps are plotted (arrays of A_{12}, A_{13}, A_{14}, A_{21}, A_{23}, A_{24}, A_{31}, A_{32}, A_{34}, A_{41}, A_{42}, A_{43} are the mentioned arrays).

In the second part of this phase, name of x-axis and y-axis are defined; Figure 12 shows this part for putting name in x-axis.

In Figure 12, for putting axis name in x-axis, by using the properties of matrix-like axes raw (i), we divide the matrix to four arrays (number of elements of inputs are the same as the number of arrays, and in the mentioned matrix, if i = 1, we put the chosen name to x-axis of mentioned array (in our study, the name of axis is easting), and if i = 2, we put the chosen name to axis x-axis of mentioned array (in our study, the name of axis is northing), and if i = 3, we put the chosen name to x-axis of the mentioned array (in our study, the name of axis is elevation), and if i = 4, we put the

chosen name to x-axis of mentioned array (in our study, the name of axis is gravity).

```
if i==1
   xlabel('easting');
elseif i==2
   xlabel('northing');
elseif i==3
    xlabel('elevation');
elseif i==4
   xlabel('gravity');
end
```

Figure 12. Defining name for x-axis.

Figure 13. shows the algorithm for putting name in y-axis.

```
if j==1
   ylabel('easting');
elseif j==2
   ylabel('northing');
elseif j==3
   ylabel('elevation');
elseif j==4
   ylabel('gravity');
end
```

Figure 13. Defining name for y-axis.

In Figure 13, for putting the axis name in y-axis, by using the properties of matrix-like axes column, (j) we divide the matrix to four arrays (number of elements of the inputs are the same as the number of arrays), and in the mentioned matrix, in Fig. 14, if j = 1, we put the chosen name to that axis (in our study; the name of axis is easting), and if j = 2, we put the chosen name to that axis (in our study, the name of axis is northing), and if j = 3, we put the chosen name to that axis (in our study, the name of axis is elevation), and if j = 4, we put the chosen name to that axis (in our study, the name of axis is gravity).

In the third part of this section, the name of each array is programmed. Figure 14 shows the programming of this part.

If in the same time, i = 1 and j = 2, in this case, we put name A_{12} to this array, and if in the same time, i = 1 and j=3, in this case, we put name A_{13} to this array, and if in the same time, i = 1 and j = 4, in this case, we put name A_{14} to this array, and if in the same time, i = 2 and j = 1, in this case, we put name A_{21} to this array, and if in the same time, i = 2 and j = 3, in this case, we put name A_{23} to this array, and if in the same time, i = 2 and j = 4, in this case, we put name A_{24} to this array, and if in the same time, i = 3 and j = 1, in this case, we put name A_{31} to this array, and if in the same time, i = 3 and j = 2, in this case, we put name A_{32} to this array, and if in the same time, i = 3 and j = 4, in this case, we put name A_{34} to this array, and if in the same time, i = 4 and j = 1, in this case, we put name as A_{41} to this array, and if in the same time, i = 4 and j = 2, in this case, we put name A_{42} to this array, and if in the same time, i = 4 and j = 3, in this case, we put name A_{43} to this array.

```
if  i==1 && j==2
    title('A_1_2');
elseif i==1 && j==3
    title('A_1_3');
elseif i==1 && j==4
    title('A_1_4');
elseif i==2 && j==1
    title('A_2_1');
elseif i==2 && j==3
    title('A_2_3');
elseif i==2 && j==4
    title('A_2_4');

elseif i==3 && j==1
    title('A_3_1');
elseif i==3 && j==2
    title('A_3_2');
elseif i==3 && j==4
    title('A_3_4');
elseif i==4 && j==1
    title('A_4_1');
elseif i==4 && j==2
    title('A_4_2');
elseif i==4 && j==3
    title('A_4_3');
end
```

Figure 14. Defining code for naming of each arrays.

In the last section of this algorithm, the addition plots are added to the main scatter plot matrix, and Figure 15 shows this section.

5. Comparison with related works

hierarchical Clustering is one of the methods, which is capable of comparing with Clustering by Self-organizing maps. high-dimensional datasets are illustrated by self-organizing maps, also self-

organizing maps is acceptable method for non-linear and non-simple datasets. This manner permits to dynamic processing in non-simple datasets. This method is tending to visualize two-dimensional model meaningfully for inputs, and Self-organizing maps by overlapping of all other variables is trained [22].

For high connection, utilizing of hierarchical type is recommend, and in sparse dataset Hierarchical clustering method has more accurate prediction [15], But hierarchical can't handling fewer references which is the capability of self-organizing maps [21].

```
subplot(4,4,1);
histfit(easting);
title('A_1_1');
subplot(4,4,6);
histfit(northing);
title('A_2_2');
subplot(4,4,11);
histfit(elevation);
title('A_3_3');
subplot(4,4,16);
histfit(gravity);
title('A_4_4');
```

Figure15. Adding addition plot to algorithm.

Figure 15 shows programming for main diagonal of this algorithm, in this section, plots and titles of arrays A_{11}, A_{22}, A_{33}, A_{44} are added to algorithm.

6. Results and Discussion

Here, Self-Organizing Maps is utilized for clustering, and for illustrating the results simultaneously, a scatter plot matrix is used. With the basis of neighborhood relation, this type of neural network becomes one of the best methods for clustering. In these case, and because of the properties of this type of neural network, self-organizing maps cause better outcomes in clustering.

A Scatter plot matrix is performed for visualizing relevance between variables at the same time. This is a method to mitigate over plotting in density, and clustering is applied between easting, northing, elevation, and absolute gravity rate of the studied area. Separating to three clusters is appropriate for this analysis, and three clusters cause a perfect class identification. This method in the area of geophysics causes the visualization of contents become better outcomes in comparison of the [7] paper, because in this type of clustering, by the basis of similarity of properties regardless to distances of two or some points of the studied area, data points with similar properties (however, they are far from each other), can place in one cluster but in the applied method in the mentioned paper, vicinity is really important, and far data points are mostly divided to different clusters; otherwise, in our proposed method, the researchers with any previous knowledge about position of raw data and relationship among the variables, by considering this figure, have the ability to notice the features of the studied area completely, and have an overview about the connection of dataset, but in the mentioned paper, this goal can't available, and the researcher does not have the ability to consider all elements of the studied area in the same time in one figure.

Clustering is performed between two parameters in two dimensions, and the result of this experiment is shown in Figure 17, and the guide of Figure 17 is shown in Figure 16, which illustrates the relationship between the parameters, and notice that gravity rate in especial means absolute gravity rate.

The result figure in Figure 17 illustrates that the contribution of all variables of data set and self-organizing maps is created the method for clustering, which generates by scatter plot matrix. This method is establishing a way for illustrating the data set, and relationship of them in the same time. The various diversity of geophysical elements within this clustering type is suggested by utilizing range of the properties that maybe shows that, this kind of clustering is a transition clustering among other defined clustering, due to this method's ability for considering all data points without regarding to place of data points and vicinity of them.

Figure 16 is guidance of Figure 17, and has no especial application without Figure 17, and Figure 16 is designed only for better comprehending of Figure 17, and the main result of this study is Figure 17, and Figure 16 just defines Figure 17, and Figure 16 is only the addition description of Figure 17; due to this reason, it is better to describe both of them together until the outcome of this research become more applicable. Figure.17 shows a novel type of datamining which is look like matrix; this matrix type is scatter plot matrix, and in scatter plot matrix, instead of numerical arrays, algorithm uses figure in each arrays for more explanation, we call one scatter plot matrix with the name of A, which is 4*4 scatter plot matrix, and it looks like $A =$

$\begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix}$ and it is scatter plot matrix of studied of area, and instead of numerical arrays, figures are placed in each array, and it has one main diagonal. In this diagonal, the main variables of geophysical prospecting (eating, northing, elevation, absolute gravity) are placed, and in array A_{11}, easting is placed, and in array A_{22}, northing is placed, and in array A_{33}, elevation is placed, and in array A_{44}, gravity rate is placed. We put the titles A_{11}, A_{22}, A_{33}, A_{44} to the mentioned arrays, and this scatter plot matrix with the mentioned arrays (easting, northing, elevation, gravity (absolute gravity)) is:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and other arrays illustrate the clustering among the variables two by two.

Array A_{12} is clustering between easting and northing, which is using self-organizing maps for clustering, where easting is placed in x-axis, and northing is placed in y-axis, and clustering is performed among these two elements of studied area, we put title A_{21} to mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} = (easting\ in\ x-axis\ and\ northing\ in\ y-axis) & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

Array A_{13} is clustering between easting, and elevation which is using self-organizing maps for clustering, where easting is placed in x-axis, and elevation is placed in y-axis, and clustering is performed among these two elements of the study area, we put title A_{13} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} = (easting\ in\ x-axis\ and\ elevation\ in\ y-axis) & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{14} is clustering between easting and gravity rate, which is using self-organizing maps for clustering, where easting is placed in x-axis, and gravity rate is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{14} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} = (easting\ in\ x-axis\ and\ gravity\ in\ y-axis) \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{21} is clustering between northing and easting, which is using self-organizing maps for clustering, where northing is placed in x-axis, and easting is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{21} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} = (northing\ in\ x-axis\ and\ easting\ in\ y-axis & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{23} is clustering between northing and elevation, which is using self-organizing maps for clustering, where northing is placed in x-axis, and elevation is placed in y-axis, and clustering is performed among these two elements of the studied area; we put title A_{23} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} = (northing\ in\ x-axis\ and\ elevation\ in\ y-axis) & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{24} is clustering between northing and gravity rate, which is using self-organizing maps for clustering, where northing is placed in x-axis, and gravity rate is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{24} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} = (northing\ in\ x-axis\ and\ gravity\ in\ y-axis) \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{31} is clustering between elevation and easting, which is using self-organizing maps for clustering, where elevation is placed in x-axis, and easting is placed in y-axis, and clustering is performed among these two elements of the studied area; we put title A_{31} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} = (elvation\ in\ x-axis\ and\ easting\ in\ y-axis) & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(\ gravity) \end{bmatrix}$$

and array A_{32} is clustering between elevation and northing, which is using self-organizing maps for clustering, where elevation is placed in x-axis, and northing is placed in y-axis, and clustering is performed among these two elements of the studied area; we put title A_{32} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A = \begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} = (elevation\ in\ x-axis\ and\ northing\ in\ y-axis) & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{34} is clustering between elevation and gravity, which is using self-organizing maps for clustering, where elevation is placed in x-axis, and gravity rate is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{34} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A=\begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34}=(elevation\ in\ x-axis\ and\ gravity\ in\ y-axis) \\ A_{41} & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{41} is clustering between gravity rate and easting, which is using self-organizing maps for clustering, where gravity rate is placed in x-axis, and easting is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{41} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A=\begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41}=(gravity\ in\ x-axis\ and\ easting\ in\ y-axis) & A_{42} & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{42} is clustering between gravity rate and northing, which is using self-organizing maps for clustering, where gravity rate is placed in x-axis, and northing is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{42} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A=\begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42}=(gravity\ in\ x-axis\ and\ northing\ in\ y-axis & A_{43} & A_{44}(gravity) \end{bmatrix}$$

and array A_{43} is clustering between gravity rate and elevation, which is using self-organizing maps for clustering, where gravity rate is placed in x-axis, and elevation is placed in y-axis, and clustering is performed among the two elements of the studied area; we put title A_{43} to the mentioned array, and this array is placed in the mentioned area of scatter plot matrix A, in the following array:

$$A=\begin{bmatrix} A_{11}(easting) & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22}(northing) & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33}(elevation) & A_{34} \\ A_{41} & A_{42} & A_{43}=(gravity\ in\ x-axis\ and\ elevation\ y-axis) & A_{44}(gravity) \end{bmatrix}$$

Here, Figure 16. is guidance for Figure 17, and Figure 16 just explains position of elements of Figure 17, until Figure 17 becomes more comprehensive, and Figure 16 shows guidance of structure of clustering by self-organizing maps.

Figure 16. Guidance of clustering by self-organizing maps.

Figure 17 illustrates that data set of easting and northing in array A_{12}, which is divided to three clusters: cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area, data set of easting and elevation. Data points in array A_{13} is divided to three clusters-: cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area. Data set of easting, and gravity in array A_{14} is divided to three clusters: cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area. Data set of northing and easting in array A_{21} -are divided to three clusters-: Cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area. Data set of elevation and eastern in array A_{31} is divided to three clusters: cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area. Data set of gravity rate, and easting in array A_{41} -are divided to three clusters-: cluster 1 is in the western part of the studied area, and cluster 2 is in the middle side of the studied area, and cluster 3 is in the eastern part of the studied area, and in arrays A_{23}, A_{24}, A_{32}, A_{34}, A_{42},

A_{43}, data set is divided to all part of the studied area.

Figure.17 illustrate that in array A_{12}, A_{13}, A_{14}, A_{21}, A_{23}, A_{24}, A_{31}, A_{32}, A_{41}, A_{42}, clustering is performed greatly, and reason of this great clustering in mentioned arrays is digits in x-axis and y-axis are near, but in arrays A_{34}, A_{43}, because digits of x-axis and y-axis have so much difference, self-organizing maps can't separate data set to different clusters perfectly and the visualization of clusters in mentioned array are not well.

In Larestan salt diapir, by using figure 17 we can have significant knowledge about this diapir like position of this diapir or elevation of this diapir, by using array A_{14} and A_{41} we can notice relationship among eating and gravity rate which is useful in detecting easting side of salt diapir also by using array A_{24} and A_{42} we can notice the relationship among northing and gravity rate which is useful in detecting northing side of salt diapir, also by using array A_{34} and A_{43} we can notice relation among elevation and gravity rate which is useful in detecting elevation of salt diapir.

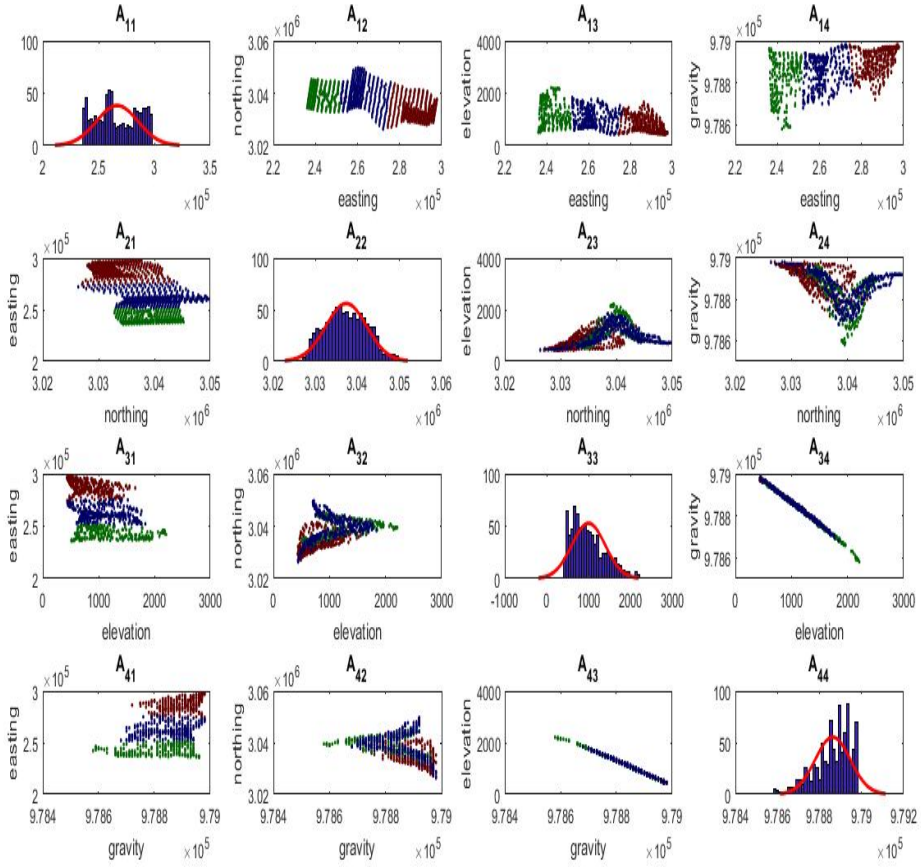

Figure 17. Scatter plot matrix of clustering by using self-organizing maps, in this figure easting (m) is placed in array A_{11} and northing(m) is placed in array A_{22} and elevation(m) is placed in array A_{33} and absolute gravity (mgal)is placed in array A_{44} and array A_{12} is clustering between easting and northing and array A_{13} is clustering between easting and elevation and array A_{14} is clustering between easting and absolute gravity and array A_{21} is clustering between northing and easting and array A_{23} is clustering between northing and elevation and array A_{24} is clustering between northing and absolute gravity and array A_{31} is clustering between elevation and easting and array A_{32} is clustering between elevation and northing and array A_{34} is clustering between elevation and absolute gravity and array A_{41}, is clustering between absolute gravity and easting and array A_{42} is clustering between absolute gravity and northing and array A_{43} is clustering between absolute gravity and elevation.

Finding the way for illustrating raw data in same time is really important and in the area of exploration gravity it can be found rarely, and most of current studies by using number of figures show relation between elements, and utilizing several figure for comprehending of relation between elements become time-consuming and in this paper, we present method for solving this subject. by using single figure maybe reliability of interpretation is increasing because only single

figure describes main finding of study area, and problem of multi-figure for interpreting the results is solved.

Comparing clustering by self-organizing maps with other type of clustering such as the hierarchical clustering method (HCM) is available;

Both of this methods by using scatter plot matrix can generate to different arrays but they have difference like;

In the hierarchical clustering method, the place of the elements is important, and the first element is compared with the second element, and the third element is compared with the fourth element, and it will continue until the last element, and after these steps, clusters are determined or in another type of hierarchical clustering method, all data points are divided to two clusters and that two clusters are divided to other clusters, and it is extended, until all data points are divided to appropriate clusters, and in this clustering methods, near elements in raw data place in one cluster and far elements although having same properties maybe does not site in one cluster but in Clustering by self-organizing maps because of features of competitive layer, separation of data to different clusters become better, and the place of elements in raw data does not effective and far elements can place in the same cluster and in this paper, if instead of SOM method, hierarchical clustering method (HCM) would be utilized, elements A_{12} and A_{13} and A_{14} and A_{21} and A_{31} and A_{41} maybe show the proper results because data points with same properties are near, and far data points in raw data points has less or no same properties and they are sited in different clusters, in this arrays vicinity are not so much important and most of the data with same properties are placed in the same area by using hierarchical clustering method instead of self-organizing maps, but in elements A_{23} and A_{24} and A_{32} and A_{34} and A_{42} and A_{43}, hierarchical clustering method does not have that much ability to separate data points appropriately (due to far data points have same properties, and they can place in the same cluster) and self-organizing maps can place far data points in the same cluster and in mentioned arrays self-organizing maps has better ability of separation, but in mentioned arrays, hierarchical clustering method can't separate them properly and hierarchical clustering method put near data points in same cluster and far data points are placed in different cluster, but self-organizing maps due to properties of competitive layers put far data points in same cluster and cause better clustering.

This clustering approach, open new way of clustering, and it has numerous usages. The importance of this study is: in the future study, other researchers without requiring to have any previous data set related to the studied area, have ability to consider study area. Also researchers can have more research in study area because they have enough knowledge about raw data of study area, and reason of this outcome is, they have all data as single figure in same time. This method causes that data interpretation becomes more easier and faster, and single figure provides all required data of study area. Utilizing self-organizing maps as a tool for clustering has so many advantages like place of data points doesn't have significant effect on clustering and far data points can place in same cluster. Second advantages of this approach is high-dimensional data can be clustering by this approach. third advantageous of this approach is clustering with high-accuracy is created by this type of clustering and reason of this high accuracy is properties of data points is important and vicinity isn't main reason of being data points in same cluster. The main deficiency of clustering by self-organizing maps is that if two axes have significantly difference in digits value, self-organizing maps can't separate data points to different clusters appropriately. Conversely, when in axes with near value of digits, self-organizing maps can become great type of clustering.

7. Conclusions

In most datamining method, relation among different elements of study area is presented two by two, and different figures shows data division among elements of study area. Seldom study can be found that single figure shows relation among all elements, the current paper presents the creative clustering manner for illustrating relation between elements of raw data of exploration gravity in same time, for reaching this aim self-organizing maps is utilized for visualizing the relationship between parameters of the study area, and scatter plot matrix generates this clustering to all elements, and main finding of this study are;

1. Clustering clearly shows that SOM produces an efficient platform due to features of the neighborhood of neurons and competitive layers, and other advantageous of this approach makes this method one of the best choices for Clustering in this type of data set.
2. This approach causes that the datamining of raw data and discovering of the best cluster is gained appropriately, and by utilizing scatter plot matrix

relation among elements are presented in same time, and different parameters of exploration data points separate two by two and visualize simultaneously

3. The proposed manner creates a simple procedure to display and specify the relationship among elements of exploration gravity data jointly for identifying relation between easting, northing, elevation, and absolute gravity clearly.

4. This method by considering all data points, make a better division for Clustering, and by utilizing of this method, the analysis of raw data points in the area of exploration gravity method is available with high accuracy, and in future consideration in study area, researchers by investigating one figure have ability to know most properties of study area.

5. this procedure doesn't have any restriction in number of inputs and depend on type of researches it can be change in the range of number of inputs.

6. By using clustering of self-organizing maps near Larestan salt diapir, we have capability to notify relationship among main parameters (easting, northing, elevation, and absolute gravity) of study area which is applicable in defining properties of mentioned salt diapir.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1]. Abubaker, A., Baharum, A., & Alrefaei, M. (2015). Automatic clustering using multi-objective particle swarm and simulated annealing. *PloS one*, *10*(7), e0130995.

[2]. Arian, M. (2012). Clustering of diapiric provinces in the Central Iran Basin. *Carbonates and Evaporites*, *27*, 9-18.

[3]. Arian, M., & Noroozpour, H. (2015). Tectonic geomorphology of Iran's salt structures. *Open Journal of Geology*, *5*(2), 61-72.

[4]. Behbahani, S., & Nasrabadiv, A. M. (2009). Application of SOM neural network in clustering. *Journal of Biomedical Science and Engineering*, *2*(8), 637-643.

[5]. Carter-McAuslan, A., & Farquharson, C. (2020). Application of SOMs and k-means clustering to geophysical mapping: Lessons learned. In *SEG Technical Program Expanded Abstracts 2020* (pp. 3843-3846). Society of Exploration Geophysicists.

[6]. Chaudhary, V., Bhatia, R. S., & Ahlawat, A. K. (2014). A novel Self-Organizing Map (SOM) learning algorithm with nearest and farthest neurons Alexandria Eng.

[7]. Czecze, B., & Bondár, I. (2019). Hierarchical cluster analysis and multiple event relocation of seismic event clusters in Hungary between 2000 and 2016. *Journal of Seismology*, *23*(6), 1313-1326.

[8]. Detorakis, G., Chaillet, A., & Rougier, N. P. (2020). Stability analysis of a neural field self-organizing map. *The Journal of Mathematical Neuroscience*, *10*(1), 20.

[9]. Florio, G. (2020). The estimation of depth to basement under sedimentary basins from gravity data: Review of approaches and the ITRESC method, with an application to the Yucca Flat Basin (Nevada). *Surveys in Geophysics*, *41*(5), 935-961.

[10]. Gerber, S., & Horenko, I. (2015). Improving clustering by imposing network information. *Science advances*, *1*(7), e1500163.

[11]. Hua, W., & Mo, L. (2020). Clustering Ensemble Model Based on Self-Organizing Map Network. *Computational Intelligence and Neuroscience*, *2020*(1), 2971565.

[12]. Jin, R., Kou, C., Liu, R., & Li, Y. (2013). Efficient parallel spectral clustering algorithm design for large data sets under cloud computing environment. *Journal of Cloud Computing: Advances, Systems and Applications*, *2*, 1-10.

[13]. Kim, J., Lee, W., Song, J. J., & Lee, S. B. (2017). Optimized combinatorial clustering for stochastic processes. *Cluster Computing*, *20*, 1135-1148.

[14]. Lowrie, W. (2007) Fundamentals of Geophysics. 2nd Edition, Cambridge University Press, Cambridge, 381.

[15]. Martinsson, J., & Törnman, W. (2020). Modelling the dynamic relationship between mining induced seismic activity and production rates, depth and size: a mine-wide hierarchical model. *Pure and Applied Geophysics*, *177*(6), 2619-2639.

[16]. Mokarram, M., & Najafi, M. Using Self-Organizing Maps for Determination of Soil Fertility. *Case Study: Shiraz Plain*, *3*(9).

[17]. Mythili, S., & Madhiya, E. (2014). An analysis on clustering algorithms in data mining. *International Journal of Computer Science and Mobile Computing*, *3*(1), 334-340.

[18]. Nokhbatolfoghahaei, A., Nezafati, N., Ghorbani, M., & Abdolabadi, B. E. (2019). Evidence for origin and alteration in the dolomites of salt diapirs, Larestan, Southern Iran. *Carbonates and Evaporites*, *34*, 389-403.

[19]. Pace, F., Santilano, A., & Godio, A. (2021). A review of geophysical modeling based on particle swarm optimization. *Surveys in Geophysics*, *42*(3), 505-549.

[20]. Ren, Z., Zhong, Y., Chen, C., Tang, J., Kalscheuer, T., Maurer, H., & Li, Y. (2018). Gravity gradient tensor of arbitrary 3D polyhedral bodies with up to third-order polynomial horizontal and vertical mass contrasts. *Surveys in Geophysics*, *39*, 901-935.

[21]. Riese, F. M., Keller, S., & Hinz, S. (2019). Supervised and semi-supervised self-organizing maps for regression and classification focusing on hyperspectral data. *Remote Sensing*, *12*(1), 7.

[22]. Russo, T., Scardi, M., & Cataudella, S. (2014). Applications of self-organizing maps for ecomorphological investigations through early ontogeny of fish. *PLoS One*, *9*(1), e86646.

[23]. Shafi, I., Ahmad, J., Shah, S., Ikram, A., Ahmad Khan, A., & Bashir, S. (2010). Validity-guided fuzzy clustering evaluation for neural network-based time-frequency reassignment. *EURASIP Journal on Advances in Signal Processing*, *2010*, 1-14.

[24]. Sun, J., & Li, Y. (2012). Joint inversion of seismic traveltimes and gravity data using petrophysical constraints with application to lithology differentiation. *ASEG Extended Abstracts*, *2012*(1), 1-4.

[25]. Sun, S., Yin, C., & Gao, X. (2021). 3d gravity inversion on unstructured grids. *Applied Sciences*, *11*(2), 722.

[26]. Vadoodi, R., & Rasmussen, T. M. (2022). Joint Interpretation of Magnetotelluric and Potential Field Data From North-Eastern Norrbotten, Sweden. *Pure and Applied Geophysics*, *179*(3), 1069-1088.

[27]. Yang, L., Ouyang, Z., & Shi, Y. (2012). A modified clustering method based on self-organizing maps and its applications. *Procedia Computer Science*, *9*, 1371-1379.

[28]. Yin, C., Zhang, S., & Kim, K. J. (2017). Mobile Anomaly Detection Based on Improved Self-Organizing Maps. *Mobile Information Systems*, *2017*(1), 5674086.

نشریه مهندسی معدن و محیط زیست
نشانی نشریه: www.jme.shahroodut.ac.ir
انجمن مهندسی معدن ایران
دانشگاه صنعتی شاهرود

داده کاوی در مطالعات گراویمتری با استفاده از ساختار نقشه‌های خود سازمانده(مطالعه موردی در جنوب ایران)

رضا شاه نوه سی و فرنوش حاجی‌زاده*

دانشکده فنی و مهندسی، دانشگاه ارومیه، ارومیه، ایران

اطلاعات مقاله

تاریخ ارسال: ۲۰۲۴/۰۸/۰۴
تاریخ داوری: ۲۰۲۵/۰۱/۲۴
تاریخ پذیرش: ۲۰۲۵/۰۲/۲۵

DOI: 10.22044/jme.2025.14879.2830

کلمات کلیدی

خوشه بندی
گراویمتری
نقشه‌های خود سازمانده
آنالیز عددی
هوش عددی

چکیده

مطالعه انجام شده عمدتاً در مورد روشی جهت نمایش دادن داده‌های خام به صورت همزمان می‌باشد. خوشه‌بندی یک روش کلیدی برای حل مشکل تقسیم‌بندی داده‌ها است. همچنین نشان دادن ارتباط بین پارامترهای تحقیق به طور همزمان حائز اهمیت است. بنابراین، ما نوع جدیدی از خوشه‌بندی را برای داده‌کاوی در گراویمتری پیشنهاد می‌کنیم تا به نمایش داده‌ها به صورت همزمان دست یابیم. برای این کار تحقیقاتی، ۸۶۷ نقطه برداشت گراویمتری در جنوب ایران (نزدیک دیاپیر لارستان) با دامنه گراویمتری از ۹۷۸۵۷۹.۶۷۲ تا ۹۷۸۹۸۱.۱۸۶ برداشت کرده‌ایم. در این مقاله، خوشه‌بندی با نقشه‌های خودسازماندهی، با استفاده از ماتریس نمودار پراکندگی برای تشخیص رابطه بین شرق، شمال، ارتفاع و گراویمتری به طور همزمان مورد استفاده قرار گرفته است. در روش پیشنهادی، روابط بین آرایه ها دو به دو تعریف شده است و مانند ماتریس، هر ستون و سطر دارای **i** و **j** های متفاوتی می‌باشد که پارامترهای منطقه مورد مطالعه را به جای اعداد در این ماتریس قرار داده می‌شود. در این الگوریتم ابتدا با استفاده از نقشه‌های خودسازماندهی، خوشه‌بندی انجام می‌شود و این پردازش با ماتریس نمودار پراکندگی به همه آرایه‌ها تعمیم داده می‌شود و در همه آرایه‌ها سه بازه جهت خوشه بندی پیشنهاد می‌شود. نتیجه این خوشه‌بندی نشان می‌دهد که در آرایه‌های A12, A13, A14, A21, A23, A24, A31, A32, A41, A42, خوشه‌بندی به نحو احسن انجام می‌شود و ارتباط بین پارامترهای منطقه مورد مطالعه نزدیک دیاپیر لارستان می‌تواند در دانستن خواص دیاپیر مفید باشد.

نویسنده مسئول مکاتبات: farnooshhajizade@yahoo.com (ف. حاجی‌زاده)