Proposal of Clustering Approach Based on Structural Mechanics: An Application of Multi-Dimensional Truss

Kazuyuki Hanahara and Yukio Tada

Graduate School of System Informatics
Kobe University
Kobe, Japan
Email: {hanahara, tada}@cs.kobe-u.ac.jp

***Abstract*—We deal with an approach of clustering based on structural mechanics. Rupture of multi-dimensional truss due to a universal repulsive force is adopted as the process of clustering. The structural behavior of multi-dimensional truss is formulated. The feasibility of the proposed approach is examined and demonstrated by a number of calculation examples.**

***Keywords*–Clustering; Multi-Dimension; Truss; Tidal force.**

I. INTRODUCTION

Clustering is one of the important processes for data management, especially in the case of pattern identification and recognition [1]. Several methods of clustering have been proposed [2][3]. There are typical approaches such as hierarchical algorithms e.g., the group average method and the Ward method and partitioning algorithms e.g., the k-means method and its families. Another approach such as based on PCA (Principal Compoment Analysis) has also been developed and studied [4]. These approaches having being developed so far are basically based on some mathematical or geometrical viewpoints. The authors see that these approaches are somewhat artificial, in the sense that the clustering processes are controlled by one or more mathematical parameters that are intentionally determined.

In the current study, we deal with an approach of clustering based on structural mechanics. Taking account of the mechanical characteristics of the target data set, a clustering of somewhat natural manner is considered to be possible. One of the significant problem is that structural systems dealt with in structural mechanics are two or three-dimensional entities, but the data set to be clustered can be an entity of higher dimensional space. In the case of truss structural system, however, it is possible to formulate the structural mechanical characteristics such as the stiffness matrix even in the case of a truss structure of four or higher dimensional space.

In this article, we regard the data elements to be clustered as the truss nodes. We develop the formulation of stiffness matrix of truss structure of general dimension. Rupture of the truss structure due to universal repulsive force is calculated; the obtained separated parts are recognized as the clusters.

In Section II, a general formulation of the nodal stiffness matrix of multi-dimensional truss is introduced. The developed clustering procedure is explained in Section III. Some preliminary example calculation results are demonstrated and discussed in Section IV and Section V gives the conclusion and future work.

II. MULTI-DIMENSIONAL TRUSS

A truss structure consists of a number of truss nodes and truss members connecting them. We denote the truss nodal positions vector as $\mathbf{X} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_N^T]^T$, where $\mathbf{x}_n$ is a D-dimensional vector corresponding to the nth element of the data set to be clustered. Truss member connections are denoted as $\mathcal{C} = \{c_1, \cdots, c_M\}$, where $c_m = \{c_{m0}, c_{m1}\}$ and c_{m0} and c_{m1} correspond to the two truss nodes connected by the mth truss member.

A. Geometrical Relation

We denote the truss member lengths vector as $\mathbf{L} = [l_1, \cdots, l_M]^T$. Each of the member lengths is given as the Euclidean distance between the corresponding nodes expressed as

$$l_m = \left[(\mathbf{x}_{c_{m1}} - \mathbf{x}_{c_{m0}})^T (\mathbf{x}_{c_{m1}} - \mathbf{x}_{c_{m0}})\right]^{(1/2)} \tag{1}$$

For all of the truss member lengths and the truss nodal positions, (1) can be collected and expressed in the following form:

$$\mathbf{L} = \mathbf{L}(\mathbf{X}) \tag{2}$$

The total differential of (2) is given as

$$d\mathbf{L} = \frac{\partial \mathbf{L}}{\partial \mathbf{X}} d\mathbf{X} \tag{3}$$

which is obtained as the collection of the total differential of (1) expressed as

$$dl_m = \frac{\mathbf{x}_{c_{m1}} - \mathbf{x}_{c_{m0}}}{l_m} d\mathbf{x}_{c_{m1}} - \frac{\mathbf{x}_{c_{m1}} - \mathbf{x}_{c_{m0}}}{l_m} d\mathbf{x}_{c_{m0}} \tag{4}$$

B. Stiffness Matrix

In the case of linear elastic model with small deformation, the strain energy U of the entire truss under deformation is expressed as

$$U = \sum_{m=1}^{M} \frac{1}{2} k_m r_m^2 = \frac{1}{2} \mathbf{R}^T \mathbf{K}_L \mathbf{R} \tag{5}$$

where k_m and r_m are the stiffness and the elastic change in length of the mth truss member, $\mathbf{R} = [r_1, \cdots, r_M]^T$ is the member deformation vector and $\mathbf{K}_L = \text{diag}[k_1, \cdots, k_M]$ is the member stiffness matrix. Since we deal with the case of small deformation, the nodal displacement vector $\mathbf{U} = [\mathbf{u}_1^T, \cdots, \mathbf{u}_N^T]^T$ and the member deformation vector have the following linear relation referring to (3):

$$\mathbf{R} = \frac{\partial \mathbf{L}}{\partial \mathbf{X}} \mathbf{U} \tag{6}$$

Substituting (6) for (5), we obtain

$$U = \frac{1}{2}\mathbf{U}^T \left(\frac{\partial \mathbf{L}}{\partial \mathbf{X}}\right)^T \mathbf{K}_L \frac{\partial \mathbf{L}}{\partial \mathbf{X}} \mathbf{U} = \frac{1}{2}\mathbf{U}^T \mathbf{K}_X \mathbf{U} \tag{7}$$

where

$$\mathbf{K}_X = \left(\frac{\partial \mathbf{L}}{\partial \mathbf{X}}\right)^T \mathbf{K}_L \frac{\partial \mathbf{L}}{\partial \mathbf{X}} \tag{8}$$

is the nodal stiffness matrix of the given truss structure.

III. Clustering Based on Rupture of Truss

The clustering is dealt with in terms of rupture of the truss structure corresponding to the given data elements, which experiences a kind of universal repulsive force.

A. Generating Member Connection

In the current study, we deal with two types of member connections among the truss nodes corresponding to the data elements. One is the full-connection type, where all of the combination of two nodes are connected by truss members. The other is a simplex-connection type, where truss members are connected to form appropriate simplices in the given dimensional space. Figure 1 shows such examples of member connections in 2D space. For the simplex-connection, we adopt the truss members in the order of length, from the shortest, among all the possible connections. The full-connection type is easy to generate; however, it is not natural from the viewpoint of truss structural system. On the other hand, the computational time to generate the simplex-type connection is not insignificant in the case of higher dimensional space; however, the obtained member connection is natural and more reasonable as a truss structural system.

Mechanical characteristics of a truss structure also depends on the stiffness of the members. We use the following relation to determine the stiffness of the truss members taking into account the weight values assigned to the data elements:

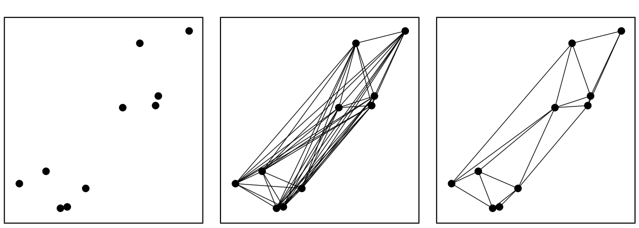
$$k_m = C_S \frac{1}{l_m^S} w_{c_{m0}} w_{c_{m1}} \tag{9}$$

where C_S is an adequate constant, S is the distance-evaluation parameter and $w_{c_{m0}}$ and $w_{c_{m1}}$ are the weight values of the data element nodes to be connected by the truss member m. The equation indicates that in the case of higher order of S, the connection strength of two data elements decreases rapidly in accordance with their distance.

B. Universal Repulsive Force

As the force to deform and rupture the truss structure, we introduce a universal repulsive force among the truss nodes corresponding to the data elements. The force between any two nodes is expressed as

$$\mathbf{f}_{ni} = C_R(\mathbf{x}_n - \mathbf{x}_i) d_{ni}^{R-1} w_n w_i, \quad d_{ni} = \|\mathbf{x}_n - \mathbf{x}_i\| \tag{10}$$



(a) Data elements (b) Full-type (c) Simplex-type

Figure 1. Two types of member connection. (2D case example)

where C_R is an adequate constant, w_n and w_i are the mass assigned to the two nodes corresponding to the weight values assigned to the data elements, and R is the parameter which denotes the nature of the repulsive force. On the basis of the introduced universal repulsive force between two nodes written as (10), the nodal force vector is obtained as follows:

$$\mathbf{F}_X = [\mathbf{f}_1^T, \cdots, \mathbf{f}_N^T]^T, \quad \mathbf{f}_n = \sum_{i=1}^{N} \mathbf{f}_{ni} \ (i \neq n) \tag{11}$$

It should be noted that the repulsive nodal force pattern for the case $R = 1$ is corresponding to the so-called tidal force, though the obtained force is not uni-directional but multi-directional.

C. Clustering Procedure

We deal with two-group clustering of the given data elements $\mathbf{x}_1, \cdots, \mathbf{x}_N$ having the weight values $w_1, \cdots, w_N$. The process is performed as follows:

Step 0 Generate truss member connections $\mathcal{C}$.

Step 1 Calculate the nodal stiffness $\mathbf{K}_X$ and the nodal force $\mathbf{F}_X$.

Step 2 Solve the stiffness equation

$$\mathbf{K}_X \mathbf{U} = \mathbf{F}_X \tag{12}$$

taking into account the condition of rigid body motion and obtain the nodal displacement $\mathbf{U}$.

Step 3 Calculate the member deformation $\mathbf{R}$ by (6) and obtain the magnitude of the member strain as follows:

$$\epsilon_m = |r_m / l_m| \quad (m = 1, \cdots, M) \tag{13}$$

Note that the obtained value immediately corresponds to the magnitude of the member stress, since we assume uniform structural material.

Step 4 Delete the truss member connections in the order of the magnitude of strain until the truss corresponding to the data set is separated into two parts.

Characteristic of the proposed clustering approach is determined by the type of member connection, the distance evaluation parameter S and the repulsive force parameter R. The constants C_S and C_R do not affect the clustering result.

IV. Example Calculations

Since this is a study still at a preliminary stage, we conduct example calculations in order to examine the feasibility of the proposed clustering approach. Influence on the clustering results of the types of truss member connection as well as the introduced two parameters is also discussed.

For each of the data sets to be clustered, the nth data element in D-dimensional space, $\mathbf{x}_n = [x_{n(1)}, \cdots, x_{n(D)}]^T$, is generated by the following equation for $i = 1, \cdots, D$ as

$$x_{n(i)} = \begin{cases} +\dfrac{D_G}{2} + D_R & (n = 1, \cdots, \dfrac{N}{2}) \\ -\dfrac{D_G}{2} + D_R & (n = \dfrac{N}{2} + 1, \cdots, N) \end{cases} \tag{14}$$

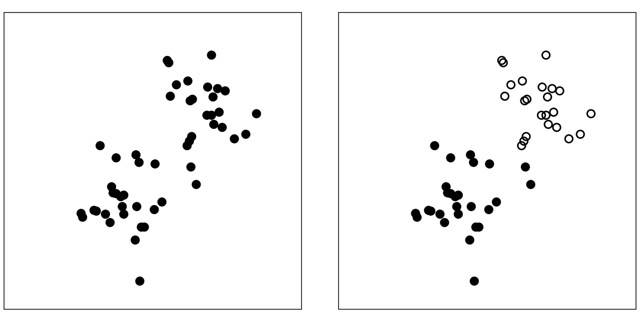
where N is the number of data elements, D_G is the assumed gap parameter between the two cluster centers and D_R is a random number. In the following calculation examples, the number of elements is $N = 50$, the gap parameter is $D_G = 0.6$ and the random number D_R is assumed to have the normal distribution of standard deviation 0.2.

A. *Evaluation of Member Connection Type*

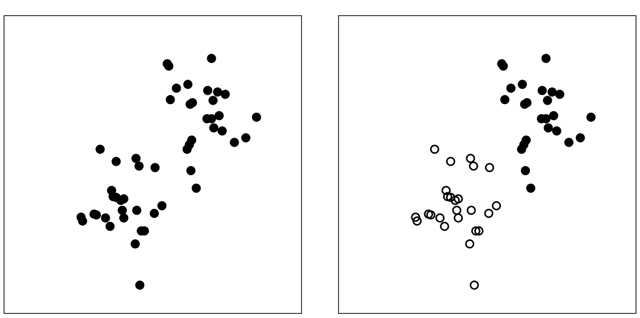
First, we examine the difference between the results based on the two types of member connections. We use $S = 2$ and $R = 1$ a priori as the two parameters in the following examples. The distance evaluation parameter $S = 2$ is selected from the clustering point of view, which indicates that the thickness of truss member connection between two data elements becomes thinner in accordance with their distance. The repulsive force parameter $R = 1$ is selected because the value corresponds to a really existing repulsive force, that is the tidal force, although this is not unidirectional.

Figure 2 shows typical clustering results based on the two types of member connections. Figures (a) and (b) are the examples respectively based on the full-connection and the simplex-connection of truss members. Figures (a-1) and (b-1) are the same data elements to be clustered and Figures (a-2) and (b-2) are the clustering results. The data elements of the obtained major cluster are depicted as a filled circle (•) and the others are depicted as an empty circle (○). Both obtained results shown in Figures (a-2) and (b-2) are similar and considered to be acceptable; however, small difference is observed with the two data elements at the right-hand side of the center.

Figure 3 shows another clustering results based on the data elements shown in Figure (a-1). On the basis of the full-connection type truss, the first clustering result and the succeeding second clustering result respectively shown in Figures (a-2) and (a-3) are considered insufficient. The succeeding third clustering result shown in Figure (a-4) does not seem natural. On the basis the simplex-connection type truss, the first clustering result shown in Figure (b-1) can also be regarded as insufficient; however, the result having a cluster of single data element is unacceptable from the clustering point of view. The succeeding second result shown in Figure (b-2) is considered to be reasonable.



(a-1) Data elements (a-2) Clustering result
(a) Based on full-connection truss



(b-1) Data elements (b-2) Clustering result
(b) Based on simplex-connection truss

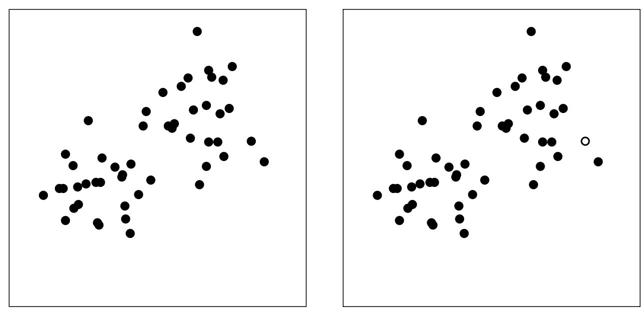
Figure 2. Evaluation of connection-type based on data set A. ($S = 2, R = 1$)

The examples shown in Figures 2 and 3 are typical results. Another clustering calculation examples also show similar tendency. In the following calculation examples, we use the simplex-connection type truss members for the clustering.

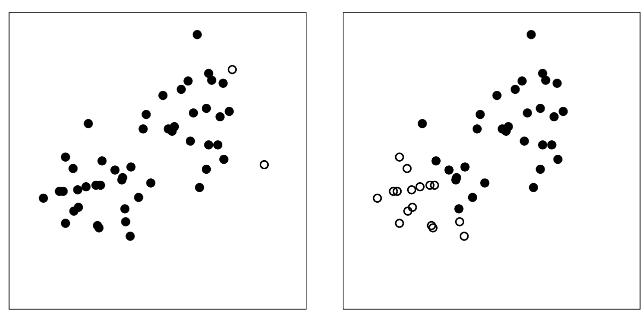
B. *Evaluation of Two Introduced Parameters*

We examine the influence of two parameters S and R. Another data set is adopted this time, since no significant difference with the parameters is observed in the clustering results based on the two data sets adopted in the previous examples. The case shown in Figure 4 is adopted as the reference. Figure (a) is the adopted data set for the parameter evaluation and Figure (b) is the clustering result based on $S = 2$ and $R = 1$. Clusters of this data set are comparably unclear; however, the clustering result shown in Figure (b) is considered to be reasonable.

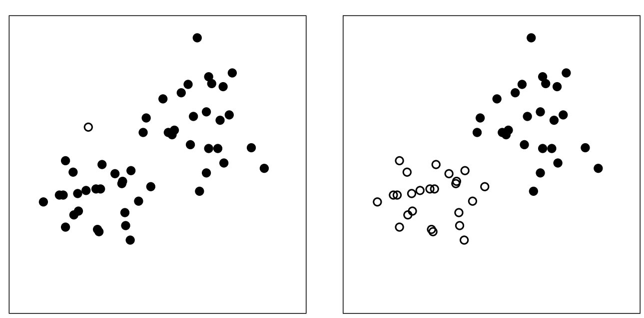
Figure 5 shows the clustering results based on different values of S in the case of $R = 1$. Figures (a), (b) and (c) respectively based on $S = 0$, $S = 1$ and $S = 3$ exhibit different results. It can be observed for all the cases that the



(a-1) Data elements (a-2) 1st clustering result

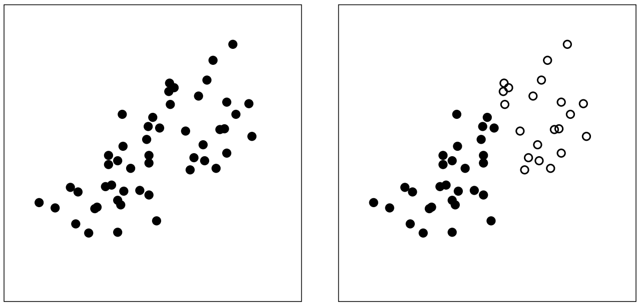


(a-3) 2nd clustering result (a-4) 3rd clustering result
(a) Based on full-connection truss



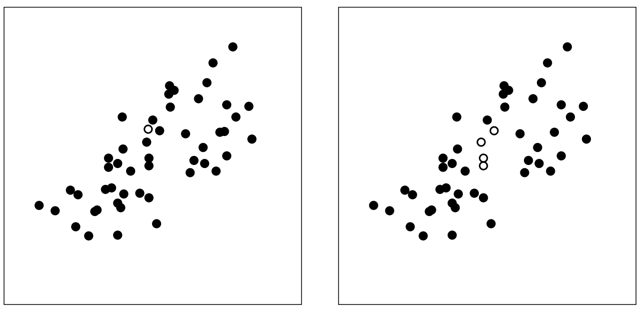
(b-1) 1st clustering result (b-2) 2nd clustering result
(b) Based on simplex-connection truss

Figure 3. Evaluation of connection-type based on data set B. ($S = 2, R = 1$)



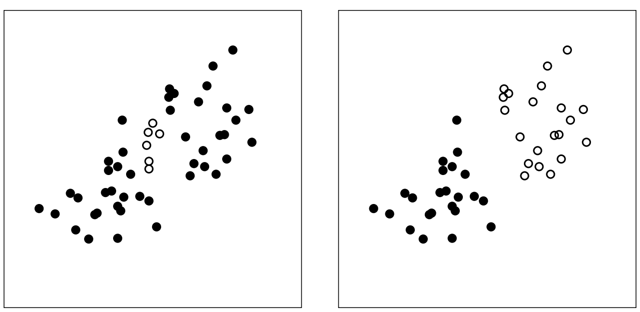
(a) Data elements (b) Clustering result

Figure 4. Reference clustering result based on data set C. ($S = 2, R = 1$)



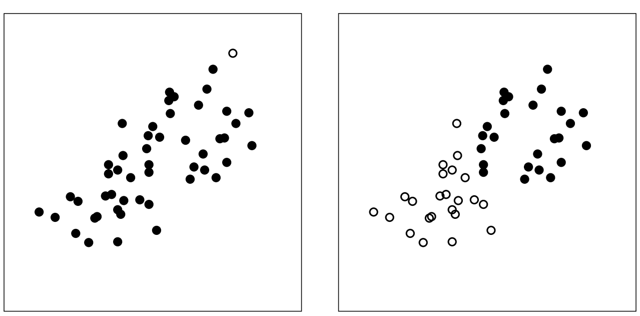
(a-1) 1st clustering result (a-2) 3rd clustering result

(a) $S = 0$



(b-1) 1st clustering result (b-2) 2nd clustering result

(b) $S = 1$

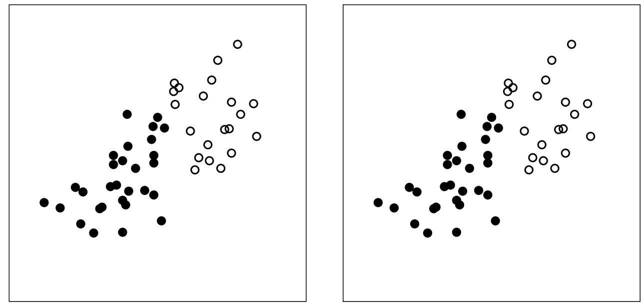


(c-1) 1st clustering result (c-2) 3rd clustering result

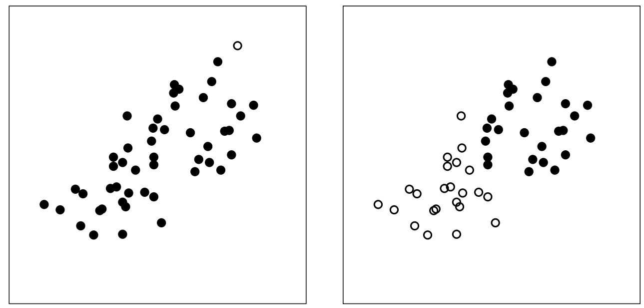
(c) $S = 3$

Figure 5. Evaluation of parameter S based on data set C. ($R = 1$)

first clustering results shown in Figures (a-1), (b-1) and (c-1) are insufficient. The second clustering result shown in Figure (b-2) in the case of $S = 1$ and the third clustering result shown in Figure (c-2) in the case of $S = 3$ are, however, considered to be reasonable results. The case $S = 0$ is clearly not acceptable



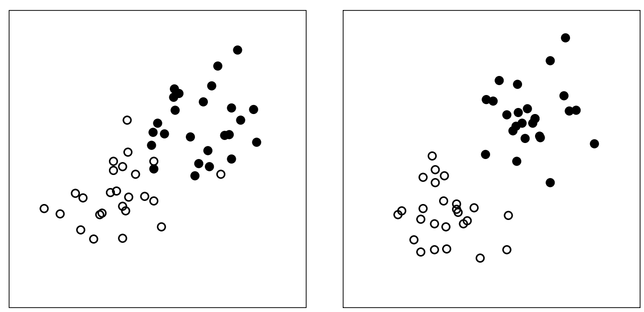
(a) $R = 0$ (b) $R = 2$



(c-1) 1st clustering result (c-2) 2nd clustering result

(c) $R = 3$

Figure 6. Evaluation of parameter R based on data set C. ($S = 2$)



(a) xy-plane (b) zw-plane

Figure 7. Four-dimensional example clustering result example based on expanded data set C. ($S = 2, R = 1$)

even for the third clustering result shown in Figure (a-2). Since all of the member stiffness values are assumed to be the same irrespective of their lengths in this case, the thickness of the assumed truss member becomes larger in accordance with the distance of the two data elements to be connected. This type of truss structural system is considered to be unreasonable from the clustering viewpoint.

Figure 6 shows the clustering results based on different values of R in the case of $S = 2$. As shown in the figure, the influence of different values of R is less significant than the case of S. The reference clustering result of $R = 1$ shown in Figure 4(b) is the same as the results of $R = 0$ and $R = 2$ respectively shown in (a) and (b) of Figure 6. Only the case of $R = 3$ shown in Figure 6(c) is slightly different.

Other calculation results that have been conducted so far exhibit similar tendencies. As a preliminary result, we conclude that the parameters determined a priori, $S = 2$ and $R = 1$, are considered to be appropriate, though the further

 ISBN: 978-1-61208-419-0

examination is required especially for the case of R.

C. Higher Dimensional Examples

Figure 7 shows an example clustering result of four dimensional data. The adopted data set in xy-plane is the same as the previous case shown in Figure 4(a), but it is expanded to z and w axes this time. In Figure 7, the clustering result plotted on xy-plane shown in (a) is slightly unreasonable but the result plotted on zw-plane shown in (b) demonstrates its adequateness.

V. Conclusion and Future Work

We proposed an approach of clustering based on structural mechanics, which is an application of multi-dimensional truss. The feasibility of the proposed approach was examined based on a number of calculation examples. As a preliminary result, we conclude that the clustering process based on the truss of simplex-type connection with the distance-evaluation parameter $S = 2$ and the repulsive force parameter $R = 1$ is considered to be adequate.

In the current study, the example artificial data sets are assumed to consist of only two clusters. In the case of a data set consisting of more clusters, iterative use of the proposed approach for the obtained clustering results is considered to be applicable. More detailed characteristics of the approach have to be studied with various patterns of data set examples. On the basis of the insights to be obtained, application of the approach to some practical problems has to be taken into consideration. These are considered to be part of the future work.

References

[1] C. M. Bishop, Pattern Recognition and Machine Learning. Springer-Verlag New York, 2006.

[2] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, Cluster Analysis, 5th Edition. John Wiley & Sons, 2011.

[3] F. Murtagh, “A Survey of Recent Advances in Hierarchical Clustering Algorithms,” The Computer Journal, vol. 26, 1983, pp. 354–359.

[4] B. McWilliams and G. Montana, “Subspace Clustering of High-Dimensional Data: A Predictive Approach,” Data Mining and Knowledge Discovery, vol. 28, 2014, pp. 736–772.