REGULAR PAPER

Bo Wu · James S. Smith · Bogdan M. Wilamowski · R. M. Nelms

DCMDS-RV: density-concentrated multi-dimensional scaling for relation visualization

Received: 16 February 2018 / Revised: 21 September 2018 / Accepted: 25 October 2018 / Published online: 22 November 2018
© The Author(s) 2018

Abstract This paper proposes a novel unsupervised multi-dimensional scaling (MDS) method to visualize high-dimensional data and their relations in a low-dimensional (e.g., 2D) space. Different from traditional MDS approaches where the main purpose is to embed high-dimensional data into a low-dimensional space, this study aims to both embed data into a low-dimensional space and reveal data relations, thus providing better visualization as graph. By taking into account the density relationships inherent in data, this paper proposes a new density-concentrated multi-dimensional scaling algorithm DCMDS-RV to perform visualization of high-dimensional data and their relations. One benefit of the proposed DCMDS-RV algorithm is the ability to embed data more accurately than traditional MDS techniques by using second-order gradient optimization instead of first-order gradient. A key advantage of the presented DCMDS-RV algorithm is the capability to show relations as categorical information. In the resulting embedding, data are compact in clusters. The results demonstrate that the proposed DCMDS-RV algorithm outperforms conventional MDS methods regarding Kruskal stress factor and ACC value. The relations between data as graph are clearly viewed as well.

Keywords Density · Clustering · Multi-dimensional scaling (MDS) · Optimization · Relation · Visualization

1 Introduction

Modern data can be overwhelming to interpret due to their size and dimensionality. As a result, the demand for understanding these data is growing rapidly. For analyzing high-dimensional data (Yuan et al. 2013) such as human facial expression images (Zhang et al. 2003), visualization techniques can provide instinctive knowledge of data. A normally used method to achieve an accurate visualization of high-dimensional data is learning a low-dimensional embedding of the high-dimensional data (Fujiwara et al. 2011). The low-dimensional representation of data should reveal corresponding relationships in higher dimensions.

B. Wu (✉) · J. S. Smith · B. M. Wilamowski · R. M. Nelms
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA
E-mail: bowu1004@gmail.com; bowu@auburn.edu

J. S. Smith
E-mail: jss0036@tigermail.auburn.edu

B. M. Wilamowski
E-mail: wilam@ieee.org

R. M. Nelms
E-mail: nelmsrm@auburn.edu

Specifically, data in close proximity represent similarity and data separated by long distances represent dissimilarity.

Conventional visualization methods derive from dealing with the problem of dimensionality reduction. Different methods are proposed in dimensionality reduction techniques, such as principal components analysis (PCA) (Jolliffe 1986) and nonnegative matrix factorization (NMF) (Lee and Seung 2001). Matrix transformations are taken to obtain the principal components in a smaller matrix fulfilling dimensionality reduction. In general, matrix operations are easy to be realized and can provide results quickly. However, these dimensionality reduction approaches are not capable of preserving the dimensionality information except the principal components. One can expect a general dimensionality reduction result but at the expense of meticulous embedding.

Multi-dimensional scaling (MDS) techniques (Cox and Cox 2001) are commonly used as a means of visualizing data while preserving dimensionality information in the form of distances. A set of MDS techniques can be found in the literature, such as Isomap (Tenenbaum et al. 2000), locally linear embedding (LLE) (Donoho and Grimes 2003), Sammon mapping (Sammon 1969), and LAMP (Joia et al. 2011). In general, an MDS algorithm is designed to place data iteratively in low-dimensional space such that the distances between data are preserved as well as possible. The majority of these techniques attempt to simulate the short pairwise distances between data, which are considered to be dependable in high-dimensional space. For example, LLE considers preserving only the local, small distances at the expense of not including remaining distances. Furthermore, there is much uncertainty as to what defines the “local” range. Even though the famous Sammon mapping method optimizes all the mutual distances (i.e., not just local small, local distances), it suffers from many overlaps between categories and is not guaranteed to converge. LAMP is one of the MDS techniques based on landmarks. Different landmarks give different MDS results. In summary, traditional MDS methods still have the following shortcomings:

1. The optimization on distance preservation uses only the first-order gradient method, which is easily trapped in local minima.
2. Narrow margins among clusters create overlapping in the mapped results.
3. Most existing methods only consider distances between data in embedding, whereas additional factors could provide desirable information.
4. Connected relations are not shown in the embedded results.

In this paper, we revisited MDS techniques and asked the question: *Can we use an unsupervised learning approach to conduct MDS purpose for both data visualization and clustering while discovering data relations?* In order to improve the stated shortcomings of traditional MDS methods, optimization-based MDS with unsupervised clustering may provide both a promising and effective solution. Optimization methods using second-order gradient descent have the ability to produce more accurate results than first-order methods due to their stronger ability to escape from local minima in our case. However, second-order gradients often bring heavy computation load. In unsupervised learning, clustering is one of the most advanced techniques that can provide data category information. Regarding what the users want to get from the MDS results, cluster automatic formation is thought to be one of the highest demanded expectations. Other expectations include larger cluster margins and individual data relationships. Therefore, in order to utilize the second-order gradient approach to fulfill MDS purpose with cluster automatic formation as well as possible, this paper proposes a new density-concentrated multi-dimensional scaling algorithm for relation visualization, called DCMDS-RV. The key idea behind the proposed DCMDS-RV algorithm is to use density-based clustering and incorporate it with Levenberg–Marquardt (LM) optimization-based MDS. This algorithm presents an alternative MDS approach using LM optimization and density concentration, yielding improved MDS performance. The main contributions of this paper are summarized as follows:

1. We propose a new unsupervised algorithm for general MDS purpose based on LM optimization method. As LM method can automatically switch between first-order gradient and second-order gradient optimization methods, the MDS technique using LM optimization shows great improvement in mapping data because of its ability to escape from local minima.
2. To obtain a better visualization on data category information, density-based clustering is integrated in the MDS process as an auxiliary methodology. In the mapping results, data move based on their mutual distances as well as their density relationships. Because of this, better cluster gathering and larger cluster margins are achieved during mapping without knowing any category knowledge.
3. Our proposed algorithm is evaluated and compared with other MDS approaches, including Sammon mapping, Isomap, LLE, and LAMP. When evaluated on experiments involving mapping several real-

life data on a 2D plane, DCMDS-RV algorithm outperforms traditional MDS approaches. Moreover, it has the ability to provide mapping results in any desired dimensional space.

4. In order to reveal the relations between data, nearest neighbor map (NNM) is generated along the algorithm and no extra step is needed. The relations between data are shown in the embedded results as connected lines, providing a more vivid and intuitive graph view of data.

The rest of this paper is organized as follows: Section 2 briefly reviews the related concepts. Section 3 details our proposed DCMDS-RV algorithm for data relation visualization. Section 4 demonstrates the experimental results on several high-dimensional datasets to evaluate the effectiveness of the proposed algorithm. Finally, Sect. 5 draws the conclusions.

2 Related works

2.1 Overview of MDS approaches

MDS is a technique used for embedding high-dimensional data into a low-dimensional space where the distances in the low dimension well represent the distances in the original, high-dimensional space. As data with short distances are similar to each other, MDS can also be applied to analyze the similarity, dissimilarity, and relationships between data. The general cost function of MDS onto a 2D plane is defined in (1), where n is the data number. The general goal of MDS approaches is to minimize the cost function.

$$\text{Err} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} - d_{ij} \right)^2 \quad (1)$$

where x_i and y_i are the x and y coordinates of mapped points i on 2D plane, respectively. d_{ij} is the distance calculated in the original high-dimensional space.

In general, there exist two types of MDS algorithms: metric and non-metric.

1. In metric MDS approaches, the actual values of the dissimilarities are used. The distances between data points are then set to be as close as possible to the similarity or dissimilarity data. Sammon mapping and LAMP techniques are two of the standard metric MDS approaches.
2. In non-metric MDS approaches, the order of the distances is preserved, and instead a monotonic relationship between the similarities/dissimilarities and the distances in the embedded space is found. In most non-metric MDS techniques, such as Isomap and LLE, criterion choices become undefined when two data points are at the same location (“colocated”), which means there are same distances between some data. When “colocated” happens, it will suspend non-metric MDS processing without an MDS solution.

2.2 Density-based clustering

Given the assumption that cluster centers are surrounded by neighbors with lower local densities and are at a relatively large distance from any data with a higher local density, density-based clustering methods (Bo and Wilamowski 2017; Rodriguez and Laio 2014) are capable of fulfilling unsupervised clustering task with satisfying results. There are three steps according to density-based clustering approaches (Rodriguez and Laio 2014).

Step 1: Calculate the data local density ρ_i . Regarding density computation, a Gaussian kernel-based density contribution is given as (2).

$$\rho_i = \sum_{j \neq i} \exp \left(- \left(\frac{d_{ij}}{d_c} \right) \right)^2 \quad (2)$$

here ρ_i is the local density of data i . d_{ij} is the distance between data i and j . d_c is the cutoff distance.

Step 2: Calculate the minimum distance δ_i between data i and any other data with a higher local density. The calculation of δ_i is given in (3).

$$\delta_i = \begin{cases} \min(d_{ij}), & \rho_j > \rho_i \\ \max(d_{ij}), & \rho_j < \rho_i \end{cases} \quad (3)$$

here $j = 1, 2, \dots, np$ and $j \neq i$.

Step 3: Generate the decision graph. Aim to choose the data with large local density ρ_i as well as large minimum distance δ_i as the cluster centers. The decision graph generated with x -coordinate is ρ_i and y -coordinate is δ_i . After the selection of cluster centers in the decision graph, data will be assigned into different clusters based on the minimum distances δ_i .

3 Proposed density-concentrated multi-dimensional scaling for relation visualization (DCMDS-RV)

The proposed algorithm focuses on projecting high-dimensional data into a low-dimensional (2D/3D) space with data relations revealed. Overall description of the proposed DCMDS-RV is illustrated in Fig. 2. The benefits of the proposed DCMDS-RV algorithm include applying first-order and second-order gradient descent methods to optimize data locations and using density relationships between data to concentrate clusters and enlarge margins between them. It simultaneously learns MDS embedding and clustering. Moreover, data relations in density field are found during the process, which can provide us a better understanding of the original data and their relations.

3.1 MDS process in the proposed DCMDS-RV

The cost function of MDS methods in (1) leads to an alternating nonlinear least squares optimization process, where we alternate between re-computing different data, and each step and iteration is guaranteed to lower the value of the cost function. In most cases, the optimization process is fulfilled using first-order instead of second-order gradient approaches, allowing the process to be trapped in local minima. Levenberg–Marquardt (LM) method, which is developed to solve nonlinear least squares problems iteratively, finds the best solutions by switching between first-order and second-order gradient approaches via a damping parameter. Unlike the second-order gradient methods, which are of heavy computation, LM method approximates the second-order gradient with the first-order gradient. In this paper, LM method is adopted as the MDS technique to determine data positions. In the proposed DCMDS-RV algorithm, MDS embedding process has two phases: (1) data position initialization using matrix eigen-decomposition and (2) data position optimization using LM method.

3.1.1 Getting initial data positions for LM method via matrix eigen-decomposition

Instead of randomly generating initial positions, the matrix eigen-decomposition technique is used to provide initial positions very fast on a 2D plane or 3D space. It can accelerate the nonlinear least squares optimization process.

Suppose $D\text{Mat}$ is the distance matrix that contains all the between-data distances and $D\text{Mat}(i, j)$ returns the distance between data i and j . Then, the initial data positions on a 2D plane are given as (10) via matrix eigen-decomposition.

$$SD = D\text{Mat} * D\text{Mat} \quad (4)$$

$$ti = \text{sum}(SD(i, :)) \quad (5)$$

$$tj = \text{sum}(SD(:, j)) \quad (6)$$

$$ta = \text{sum}(\text{sum}(SD)) \quad (7)$$

$$M(i, j) = \frac{1}{2}(ti + tj - ta - SD(i, j)) \quad (8)$$

$$[V, D] = \text{eigs}(M) \quad (9)$$

More details on eigen operations can be found in Abdi (2007). Diagonal matrix D contains the eigenvalues on the main diagonal, and the columns of matrix V are the corresponding eigenvectors. Initial data positions in a 3D space are shown as (11). Note that the first eigenvalue $D(1, 1)$ is zero, which is neglected here.

$$\text{2D mapping: } P^{(0)} = \sqrt{D(2:3, 2:3)} * V(:, 2:3)' \quad (10)$$

$$\text{3D mapping: } P^{(0)} = \sqrt{D(2:4, 2:4)} * V(:, 2:4)' \quad (11)$$

Next, with the initial positions, embedded data are optimized iteratively using LM method.

3.1.2 LM method

LM method (Marquardt 1963) is for minimizing the cost function (1) in DCMDS-RV algorithm. In order to return better gradient-based optimization results, the second-order derivatives of the total error function are considered. However, the calculation of Hessian matrix H , which contains the second-order derivatives of cost function, is often complicated. In order to simplify the computing process (Wilamowski and Yu 2010), the Jacobian matrix J is introduced to approximate Hessian matrix H . J is the matrix of all first-order derivatives of the cost function with respect to data's coordinates. For the cost function (1) of 2D mapping, the m 'th row of Jacobian matrix is $J_m = \begin{bmatrix} \frac{\partial \text{Err}}{\partial x_m} & \frac{\partial \text{Err}}{\partial y_m} \end{bmatrix}$.

$$H \approx J^T J \quad (12)$$

In order to make sure that the approximated Hessian matrix $J^T J$ is invertible, LM algorithm introduces another approximation to Hessian matrix:

$$H \approx J^T J + \mu I \quad (13)$$

where μ is combination coefficient with positive value; I is the identity matrix with the size of 2×2 for 2D MDS. From Eq. (13), one may notice that the elements on the main diagonal of the approximated Hessian matrix will be larger than zero. Therefore, with this approximation (13), it can be sure that matrix H is always invertible.

Now, the update rule of LM algorithm for 2D MDS case can be presented as (14)–(16):

$$\Delta = (J^T J + \mu I)^{-1} * J * \text{Err} \quad (14)$$

$$x = x + \Delta(1) \quad (15)$$

$$y = y + \Delta(2) \quad (16)$$

As the combination of the steepest descent and second-order gradient algorithms, LM algorithm switches between the two algorithms during the least squares minimization process. When the combination coefficient μ is very small (nearly zero), (14) approaches the second-order algorithm; when the combination coefficient μ is very large, (14) approaches the steepest descent method.

3.2 Density concentration and relations generation process in the proposed DCMDS-RV

Density-based clustering methods assume that cluster centers are surrounded by neighbors with lower local densities. In other words, data with small local densities should move close to data with large local densities when embedding. The general new idea behind this is the density concentration process, where each datum will be mapped closer to its nearest neighbor data in the density field.

Step 1: Calculate the data local density ρ_i using (2).

Step 2: Generate nearest neighbor map (NNM) in the meantime of calculating the minimum distance δ (3). Nearest neighbor information will be stored in NNM, where data connect to its nearest neighbor of larger local density, with their distance equal to δ . For instance, if the minimum distance δ for data point m is δ_m from data point m to data point n , then data point n is the nearest neighbor of data point m in the density field. Notice that data's nearest neighbor in NNM always has a larger local density, with their distance as δ . The datum with the largest local density connects to itself in the NNM. This also generates the relations between data in the density field, which are shown connected in the resulting embedding. An NNM example of a two-dimensional dataset is shown in Fig. 1.

With the involvement of NNM, it is possible to present data structures as graphs. In a variety of research and application areas, data graph modeling and analysis (Kennedy et al. 2017) are an important task to be accomplished, which makes graph visualization (Cheng et al. 2018) necessary when developing visualization techniques.

Fig. 1 An illustration example of NNM for two-dimensional dataset-FLAME (Fu and Medico 2007). The colors of data stand for different density values. In NNM, data's nearest neighbor always has a larger local density, with their distance as δ

$$\text{Moving rate: } \lambda = 1 - \frac{\delta(i)}{\max(\delta)} * \frac{\rho(i)}{\max(\rho)} \quad (17)$$

If λ is larger, the data will move more closer to its nearest neighbor in the density field but might not in the distance field. The embedded data $P^{(t)}$ in the t -th iteration after density concentration process will be as follows.

$$P^{(t)} = P^{(t-1)} + \lambda * (P_{\text{NN}} - P^{(t-1)}) \quad (18)$$

3.3 DCMDS-RV algorithm

Although it is illustrated in our paper that LM method is capable of providing a better visualization than traditional MDS techniques, LM alone is not enough to generate impressive MDS results. Fortunately, by combining LM with the benefit of density concentration and NNM, the proposed DCMDS-RV algorithm can generate inspiring MDS results with data topology revealed. Detailed DCMDS-RV algorithm is presented in Fig. 2.

3.4 Summary of DCMDS-RV algorithm

In summary, the proposed DCMDS-RV algorithm has several important merits as follows:

1. *General MDS purpose* DCMDS-RV algorithm applies LM method to find the embedded locations for data based on their mutual distances and therefore is a general-purpose MDS approach. Thus, it is suitable to conduct dimensionality reduction, visualization, and other purposes that general MDS approaches are used for. Besides, it is capable of escaping “colocated” problems that appear in traditional MDS techniques.
2. *Micro-clusters forming* Density-based clustering methodology (an unsupervised technique) is used to concentrate clusters so that the margins between micro-clusters are enlarged. As clusters are formed and separations between clusters expand, a better visualization of micro-clusters is expected.
3. *Absence of parameter setting/integrated* Simple and efficient, DCMDS-RV algorithm integrates cluster concentration based on local densities with MDS approach based on LM optimization by linear combination (19). It is easy to be interpreted and fulfilled. Most importantly, it is parameter-setting-free and unsupervised.

4. *Revealing data topology* NNMs are generated along the algorithm, which can provide data nearest neighbor information in density field. Unlike the traditional MDS approaches with only one purpose of embedding data, the proposed DCMDS-RV algorithm is capable of showing how the data are related/connected to the others. With the connections between related data shown, more information can be revealed in the embedded results.

4 Experimental results and analysis

In this section, the proposed DCMDS-RV algorithm is evaluated using experiments on four real-world data.

Algorithm: DCMDS-RV
Purpose: MDS for Relation Visualization Input: Unlabeled raw high-dimensional data Output: Embedded data P in a desired dimension (e.g. 2D) with relations connected Algorithm: compute d ; % between-data distances compute ρ ; % each datum's local density (2); generate NNM; shrink data's distance to its near neighbors with larger densities by a factor of $(\rho_{data}/\rho_{neighbor})^2$; % make neighbor distances smaller initial solution $P^{(0)}$ from matrix eigen-decomposition; % (10) $\alpha = 0.5$; $\beta = 0.5$; % $\alpha + \beta = 1$ for $t = 1$ to T do % T is the number of iterations compute $P^{(t)}$ using LM based on the previous solution $P^{(t-1)}$; compute moving rate λ ; % (17) set $P^{(t)} = \alpha * P^{(t)} + \beta * (P^{(t-1)} + \lambda * (P_{NN} - P^{(t-1)}))$; (19) % P_{NN} is data's nearest neighbor found in NNM. end connect each datum to its nearest neighbor in NNM.

Fig. 2 DCMDS-RV algorithm

Fig. 3 Sammon mapping on HAR

4.1 Human activity recognition (HAR) using smartphone dataset

HAR (Anguita et al. 2013) consists of 7352 data with 561 dimensions/attributes, built from the recordings of volunteers performing activities of daily living. These data belong to one of the following six classes: C1—walking, C2—walking upstairs, C3—walking downstairs, C4—sitting, C5—standing, and C6—laying.

In Fig. 3, Sammon mapping shows two distinct clusters with overlapping classes in each cluster. In Fig. 4, Isomap shows two distinct, dense clusters with overlapping classes of even less distinction compared to Fig. 3. LLE in Fig. 5 shows nearly no distinction among classes outside of being in the left or right loosely collected clusters. Similar to Sammon mapping, LAMP in Fig. 6 shows two distinct clusters with little distinction between classes within each cluster.

In contrast, DCMDS-RV in Fig. 7 shows distinct clusters with better class separation within each cluster. Moreover, the connections between data are clearly shown as well. Whereas some of the competing methods show sharp class distinction for no more than a few classes, DCMDS-RV shows distinct separation in almost

Fig. 4 Isomap on HAR

Fig. 5 LLE on HAR

Fig. 6 LAMP on HAR

Fig. 7 DCMDS-RV on HAR

every part of the map, save the two overlapping classes designated in pink and yellow. The optimal performance of DCMDS-RV for this dataset is verified in a later performance evaluation.

4.2 MNIST handwritten digits dataset

MNIST handwritten digit recognitions have been considered as one of the most complex and difficult problems to be solved. It consists of 60,000 photographs with the size of 28×28 pixels (784 dimensions) (LeCun et al. 2017).

Figures 8, 9, 10, and 11 contain experiment results on first 6000 photographs of the MNIST dataset using Sammon mapping, Isomap, LLE, and LAMP, respectively. In Fig. 8, Sammon mapping shows that all data are contained in a large circle with many overlapping classes due to being trapped in local minima, demonstrating poor visualization; this technique is only able to distinguish a single class. Figure 9 shows that Isomap is able to distinguish the same class as Sammon mapping in a more compact, distinctive

Fig. 8 Sammon mapping on MNIST

Fig. 9 Isomap on MNIST

manner, yet struggles to show good visualization on the remaining overlapping classes. LLE visualization in Fig. 10 demonstrates better visualization than the former methods with a tight, distinct class clusters on the edges, but still has little distinction in the center. LAMP visualization also has small distinction in the edges, with most of the data barely distinguishable in the center. Unlike the other methods, the proposed DCMDs-RV in Fig. 12 shows distinct classes with significantly less overlap with visual separation between clusters. Additionally, the relations between data can be favorably visualized. Data relations given by NNM are critical to be analyzed in some cases, e.g., social media relations analysis (Cao et al. 2015). Similar to the HAR dataset, the proposed DCMDs-RV's superiority is later verified in a performance evaluation.

4.3 Olivetti Face dataset

Face images (Du et al. 2018) are widely analyzed and applied to different applications. Here, Olivetti Face images are tested through our approach for visualization. Olivetti Face data (Samaria and Harter 1994) is a set of 112×92 (or 10,304 dimensions) images of different persons with different face angels, expressions,

Fig. 10 LLE on MNIST

Fig. 11 LAMP on MNIST

and even with or without glasses wearing. In Fig. 13, Sammon mapping shows relatively strong visualization results, with some overlapping classes seen throughout the map. Isomap, LLE, and LAMP in Figs. 14, 15, and 16 all fail to provide distinct class separation. In contrast, DCMDS-RV results in Fig. 17 provide significant class distinction with each datum collected to another, far outperforming all other methods. Only a single datum is incorrectly related. The Olivetti Face dataset clearly shows great performance of DCMDS-RV algorithm compared to the competing methods. Raw face images are eventually categorized into classes using the presented approach, as shown in Fig. 18.

4.4 Wikipedia 2014 words dataset

Analysis on text such as text relation extraction reveals useful information. Here, Wikipedia corpuses are visualized using the proposed approach. Wikipedia corpuses (<https://corpus.byu.edu/wiki/>, accessed 24 January 2018) contain the full text of Wikipedia, and they contain 1.9 billion words in more than 4.4 million articles. Wikipedia 2014 words dataset is the Wikipedia corpuses of year 2014, with a vocabulary of the top

Fig. 12 DCMDS-RV on MNIST

Fig. 13 Sammon mapping on Olivetti Faces

200,000 most frequent words. In the result of Fig. 19, the proposed approach provides an impressive visualization on text. The graph topology is successfully shown between texts. Four zoomed-in regions in the resulting graph are clearly illustrated in Fig. 20.

4.5 Evaluation metrics

4.5.1 Kruskal stress

The concept of using a loss function to evaluate the performance of MDS came with Kruskal (1964) and gave us the concept of minimizing a loss function called *stress*. The disparity in *stress* is a measure of how well the Euclidean distance in low-dimensional space matches the dissimilarity, which is usually the Euclidean distance in high-dimensional space.

It's proven in the original paper that the order of the original dissimilarities is preserved by the disparities. A loss function L , which is really *stress*, is defined as follows:

Fig. 14 Isomap on Olivetti Faces

Fig. 15 LLE on Olivetti Faces

$$L = \text{stress} = \sqrt{\frac{\sum_{r < s} (d_{rs} - \hat{d}_{rs})^2}{\sum_{r < s} d_{rs}^2}} \quad (20)$$

where d_{rs} is the Euclidean distance in low-dimensional space between point r and s . $\hat{d}_{rs}$ is the disparity corresponding to d_{rs} . The order that indicates $\{r < s\}$ is determined by the Euclidean distance relationships in high-dimensional space. When the MDS map perfectly reproduces the input data, $d_{rs} - \hat{d}_{rs}$ is zero for all r and s , so stress is zero. Thus, generally speaking, the smaller the stress, the better the representation.

4.5.2 ACC

The standard unsupervised evaluation metric and protocols for evaluations and comparisons to other algorithms are used (Yang et al. 2010). Intuitively, this metric considers a cluster assignment from an unsupervised algorithm and a ground-truth assignment and then finds the best matching between them. The best mapping can be efficiently computed using the Hungarian algorithm (Kuhn 1955). For all the

Fig. 16 LAMP on Olivetti Faces

Fig. 17 DCMDS-RV on Olivetti Faces

approaches, the number of clusters is set to be the number of ground-truth categories. Clustering performance is evaluated with *unsupervised clustering accuracy (ACC)*:

$$ACC = \max_m \frac{\sum_{i=1}^n 1\{l_i = m(c_i)\}}{n} \quad (21)$$

where l_i is the ground-truth label, c_i is the cluster assignment produced by the algorithm, and m ranges over all possible one-to-one mappings between clusters and labels.

4.6 Evaluation results and discussions

Kruskal stress factors using different MDS approaches for different datasets are shown in Table 1. Although other methods may have positive outcomes on some of the datasets, our method achieves better or same-level results each time.

Experimental results (Table 2) using *K*-means (MacQueen 1967) to cluster embedded data clearly demonstrate a consistent superior performance of DCMDS-RV approach compared to other MDS methods.

Fig. 18 Corresponding face images of Fig. 17

Fig. 19 DCMDS-RV on Wikipedia 2014 words (956 samples)

As the iterative methods are not advanced in processing time, e.g., Sammon mapping and DCMDS-RV, our method has one limitation, which makes DCMDS-RV not favorable in fast mapping techniques. However, the visualization on data relations, in combination with the ability to produce category distinctions, favors DCMDS-RV approach as a highly successful data relation visualization method.

5 Conclusion

This paper proposes a new MDS algorithm called DCMDS-RV for data relation visualization in a low-dimensional space. The major innovation of the new algorithm is the incorporation of density concentration into our new MDS with LM optimization. DCMDS-RV algorithm can reveal both the distance relationships

Fig. 20 Zoomed-in visualization results in the corresponding marked region in Fig. 19

Table 1 Comparison of Kruskal stress (s) and processing time (t in seconds): (s , t)

	HAR ($N_p=2000$)	MNIST ($N_p=6000$)	Olivetti Face ($N_p=100$)	Wiki words ($N_p=956$)
Sammon mapping	(0.1551, 88.77)	(0.4320, 11,362.07)	(0.3180, 1.04)	(0.3849, 138.42)
Isomap	(0.2354, 34.13)	(0.3930, 163.25)	(0.3223, 0.13)	(0.4588, 28.43)
LLE	(0.7150, 1.39)	(0.7896, 29.65)	(0.6187, 0.09)	(0.6320, 0.97)
LAMP	(0.1566, 9.73)	(0.4051, 67.92)	(0.3234, 0.12)	(0.4692, 1.67)
DCMDS-RV	(0.1680, 27.21)	(0.3408, 1283.85)	(0.3257, 1.09)	(0.3826, 20.72)

The proposed algorithm achieves better or same-level performance regarding Kruskal stresses compared to other conventional MDS approaches in bold

Table 2 Comparison of ACC on four embedded datasets using K-means

	HAR ($N_p=2000$) (%)	MNIST ($N_p=6000$) (%)	Olivetti Face ($N_p=100$) (%)	Wiki words ($N_p=956$)
Sammon mapping	53	38	73	N/A
Isomap	55	36	62	N/A
LLE	27	39	69	N/A
LAMP	54	31	49	N/A
DCMDS-RV	71	68	98	N/A

N_p : test data number

The proposed algorithm has state-of-the-art performance on the preservation of clusters compared to other conventional MDS approaches in bold

and the density relationships between data, and therefore successfully provide us a better visualization on all sorts of high-dimensional data as graph topology. Relations between data are revealed as NNM, showing how data are connected to the others. Experimental figures also give us vivid visualization results with relations shown. Compared with other state-of-the-art MDS methods, DCMDS-RV achieves better performance. Although it may suffer in embedding time, the proposed DCMDS-RV algorithm is a useful and heuristic technique for combining unsupervised clustering methods with traditional MDS techniques in further study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abdi H (2007) Eigen-decomposition: eigenvalues and eigenvectors. In: Salkind NJ (ed) Encyclopedia of measurement and statistics. Sage, Thousand Oaks, pp 304–308
- Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In: 21th European symposium on artificial neural networks, computational intelligence and machine learning, 24–26 April 2013
- Bo W, Wilamowski BM (2017) A fast density and grid based clustering method for data with arbitrary shapes and noise. *IEEE Trans Industr Inf* 13(4):1620–1628
- Cao N, Lu L, Lin Y-R, Wang F, Wen Z (2015) SocialHelix: visual analysis of sentiment divergence in social media. *J Vis* 18(2):221–235
- Cheng S, Zhong W, Isaacs KE, Mueller K (2018) Visualizing the topology and data traffic of multi-dimensional torus interconnect networks. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2018.2872344>
- Cox TF, Cox MAA (2001) Multidimensional scaling, 2nd edn. Chapman and Hall, Boca Raton
- Donoho DL, Grimes C (2003) Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci* 100(10):5591–5596
- Du J, Song D, Tang Y et al (2018) ‘Edutainment 2017’ a visual and semantic representation of 3D face model for reshaping face in images. *J Vis*. <https://doi.org/10.1007/s12650-018-0476-4>
- Fu L, Medico E (2007) FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinform* 8(1):3
- Fujiwara T, Iwamaru M, Tange M, Someya S, Okamoto K (2011) A fractal-based 2D expansion method for multi-scale volume data visualization. *J Vis* 14(2):171–190
- Joia P, Paulovich F, Coimbra D, Cuminato JA, Nonato LG (2011) Local affine multidimensional projection. *IEEE Trans Visual Comput Graphics* 17(12):2563–2571
- Jolliffe IT (1986) Principal component analysis. Springer, Berlin, p 487
- Kennedy A, Klein K, Nguyen A, Wang FY (2017) The graph landscape: using visual analytics for graph set analysis. *J Vis* 20(3):417–432
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika* 29:1–28
- Kuhn HW (1955) The Hungarian method for the assignment problem. *Naval Res Logist Q* 2(1–2):83–97
- LeCun Y, Cortes C, Burges CJC (1998) THE MNIST DATABASE of handwritten digits. <http://yann.lecun.com/exdb/mnist/>. Accessed 31 Oct 2018
- Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems 13: proceedings of the 2000 conference, pp 556–562
- MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley symposium on mathematical statistics and probability, University of California Press, pp 281–297
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11(2):431–441
- Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. *Science* 344(6191):1492–1496
- Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: Proceedings of 1994 IEEE workshop on applications of computer vision, pp 138–142
- Sammon JW Jr (1969) A nonlinear mapping for data structure analysis. *IEEE Trans Comput* 18(5):401–409
- Tenenbaum J, Silva V, Langford J (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
- Wilamowski BM, Yu H (2010) Improved computation for Levenberg Marquardt training. *IEEE Trans Neural Networks* 21(6):930–937
- Yang Y, Xu D, Nie F, Yan S, Zhuang Y (2010) Image clustering using local discriminant models and global integration. *IEEE Trans Image Process* 19(10):2761–2773
- Yuan X, Ren D, Wang Z, Guo C (2013) Dimension projection matrix/tree: interactive subspace visual exploration and analysis of high dimensional data. *IEEE Trans Visual Comput Graphics* 19:2625–2633
- Zhang Yu, Prakash EC, Sung E (2003) Hierarchical facial data modeling for visual expression synthesis. *J Vis* 6(3):313–320