A Nonlinear Dimensionality Reduction Using Combined Approach to Feature Space Decomposition

Myasnikov E.V.

Samara State Aerospace University,
Image Processing Systems Institute of the Russian Academy of Science
Samara, Russia
mevg@geosamara.ru

Abstract. In this paper we propose a new combined approach to feature space decomposition to improve the efficiency of the nonlinear dimensionality reduction method. The approach performs the decomposition of the original multidimensional space, taking into account the configuration of objects in the target low-dimensional space. The proposed approach is compared to the approach using hierarchical clustering in the original space and to the approach based on the decomposition of the target space using KD-Tree.

Keywords: dimensionality reduction; multidimensional scaling; Sammon's mapping; hierarchical clustering; KD-Tree

1   Introduction

Dimensionality reduction methods, operating on the principle of preserving the pairwise distances between objects can be used as a means to display multidimensional data in scientific research and in production activities in a number of areas: biology, genetics, sociology, economics, etc. In modern information systems, such methods can be used to create navigation systems for multimedia databases, as well as in interface design to virtual directories. In the field of image analysis and processing nonlinear dimensionality reduction techniques have been applied not only for research but also to solve a number of applied problems: creation of automated systems for image segmentation, thematic mapping from satellite imagery and others.

   One of the most well-known dimensionality reduction methods is a method [9], which minimizes the following error of multidimensional data representation

$$\varepsilon = \frac{1}{\sum\limits_{\substack{o_i,o_j \in O \\ i<j}} d(o_i,o_j)} \cdot \sum\limits_{\substack{o_i,o_j \in O \\ i<j}} \frac{\left(d(o_i,o_j) - d^*(o_i,o_j)\right)^2}{d(o_i,o_j)} \qquad (1)$$

where o_i is an object from a set O, $d(o_i,o_j)$ is the distance between objects in the original space, $d^*(o_i,o_j)$ is the distance between objects in the target space.

Iterative procedure that minimizes the error (1) may be based on the following recurrence relations for the coordinates $\mathbf{y}_i$ in the target low-dimensional space (gradient descent):

$$\mathbf{y}_i(t+1) = \mathbf{y}_i(t) + m \cdot \sum_{o_j \in O} \varsigma(o_i,o_j) \tag{2}$$

where (m can be calculated once):

$$m = \frac{2 \cdot \alpha}{\displaystyle\sum_{\substack{o_i,o_j \in O \\ i<j}} d(o_i,o_j)}$$

$$\varsigma(o_i,o_j) = \frac{d(o_i,o_j) - d^*(o_i,o_j)}{d(o_i,o_j) \cdot d^*(o_i,o_j)} \cdot \left(\mathbf{y}_i - \mathbf{y}_j\right) \tag{3}$$

It is worth noting that the performance of the nonlinear mapping has been improved by extending multidimensional data representation error (1) with Bregman divergences [11, 12].

Unfortunately, a significant drawback of the nonlinear dimensionality reduction techniques, operating according to the iteration scheme (similar to that shown above) is the high computational complexity ($\mathrm{O}(K^2)$ per iteration, where $K = |O|$ is the number of objects). For this reason a number of techniques with a reduced computational complexity have been proposed [3, 4, 5, 6, 7].

One of the most effective ways to reduce the computational complexity is the hierarchical decomposition of space. After performing such decomposition objects can be considered not individually, but in groups, that allows to speed up the iterative optimization process. So, if all the objects of the original set O are divided into groups $s_j \in S$, then (2) takes the form

$$\mathbf{y}_i(t+1) = \mathbf{y}_i(t) + m \cdot \sum_{s_j \in S} \widetilde{\varsigma}(o_i,s_j) \tag{4}$$

$$\widetilde{\varsigma}(o_i,s_j) = \left|s_j\right| \cdot \frac{d(o_i,s_j) - d^*(o_i,s_j)}{d(o_i,s_j) \cdot d^*(o_i,s_j)} \cdot \left(\mathbf{y}_i - \mathbf{y}_{s_j}\right) \tag{5}$$

where $d(o_i,s_j)$ is the distance from the object to the center of the group s_j in the original space, $d^*(o_i,c_j)$ is the distance from the object to the center of the group in the target space, $\mathbf{y}_{s_j}$ is the coordinates of the center of the group in the target space.

Such decomposition allows to reduce the computational complexity to O(LK) per iteration, where $L = |S|$ is the number of groups.

It is obvious that the expression (4) allows to approximate (2) with a certain accuracy that depends on how well objects are divided into groups. For this reason, there is a question about how to better carry out such decomposition. It is understood that the decomposition can be performed in both the source and target space. Approach based on decomposition of the original space, is implemented in a nonlinear method for dimensionality reduction of data using reference nodes [5] proposed by the author of this article. Approach based on decomposition of the target space has been implemented, for example, in [1] (regular decomposition) and [8] (hierarchical decomposition) to solve the problem of drawing graphs to approximate the forces acting on the vertices of the graph.

In this paper we propose a new method that uses the combined approach to feature space decomposition to improve the efficiency of the nonlinear dimensionality reduction method. The proposed method is based on the nonlinear method for dimensionality reduction of data using reference nodes [5] that performs the decomposition of the original space, and complemented with the additional control on the configuration of objects in the target space. The proposed method is compared to the method based on the decomposition of the original space, and to the method based on the decomposition of the target space, both in terms of the quality of the mapping, as well as in terms of the operating time.

The paper is organized as follows. The next section is devoted to the description of the known approaches used in the present study, and consists of two subsections. The third section is devoted to the description of the proposed method. The fourth section presents the results of numerical experiments. At the end of the paper the conclusion is given.

2   Brief Description of the Methods Used in the Study

2.1   Nonlinear Method for Dimensionality Reduction of Data Using Reference Nodes

Approach based on the decomposition of the original space, is implemented in a nonlinear method for dimensionality reduction of data using reference nodes [5]. This method consists of four steps, briefly described below.

The inputs to the method are the feature vectors describing the objects in the original multidimensional space. The outputs of the method are the feature vectors describing the objects in the target low-dimensional space.

Step 1: Construction of a hierarchy of clusters. At the first stage of the method we make a hierarchical partition of the initial set of objects into clusters (subsets of objects with similar characteristics in the original feature space). When we refer to the hierarchy of clusters we mean tree-like data structure, the root of which is a cluster of

top-level, and each vertex-cluster contains either subclusters or the objects of the original set.

Step 2: Initialization of the coordinates of objects in the target space. Initialization of the coordinates of objects in the target space can be performed in various ways, e.g., using random values, the results of the PCA method, etc.

Step 3: Construction of the lists of reference nodes. At the third stage, for each object of the original set the lists of reference nodes are built. The reference node of the object o refers to a certain object $o_i \neq o$ or group of objects $\{o_i\}_{i=1..N}$ that possesses close characteristics in the original multidimensional space and considered as a single object.

Let o be an object for which we need to form a list S of the reference nodes and let C be an arbitrary cluster of the hierarchy. By using the hierarchy of clusters obtained at the first stage the construction of the reference nodes list may be implemented using the following recursive algorithm.

1. If the cluster C satisfies the decomposition criterion with respect to the considered object and subclusters $C_i \in C, i = 1..N$ are contained in cluster C, then one must apply this recursive algorithm to each of the subclusters $C_1, C_2, ..., C_N$.
2. If the cluster C satisfies the decomposition criterion with respect to the considered object and objects $o_i \in C, i = 1..N$ are contained in cluster C, then one must:
   (a) add all objects that close to the considered object to the list of reference nodes S;
   (b) distinguish the set V of objects situated at a distance from the considered object as an incomplete cluster;
   (c) in the case when this set V is nonempty, add it to the list of reference nodes.
3. If the cluster C doesn't satisfy the decomposition criterion with respect to the considered object then add it to the list of reference nodes S.

Decomposition criterion used in this paper is slightly different from the criterion used in [5] and restricts (by the value δ) the estimation of the angle at which the hypersphere of the cluster is observed from the object o in the original space.

Figure 1 illustrates this process. The cluster, which is observed from an object o at an angle α will be divided into three clusters if $\alpha \geq \delta$. The cluster, which is observed at an angle β will be considered as a whole if $\beta < \delta$.
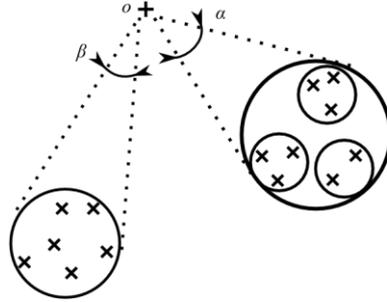
Fig. 1. Illustration to the construction of a list of reference nodes

Step 4: Iterative optimization procedure. At the final stage of the method iterative procedure is performed, which allows to clarify the position of objects in the target low-dimensional space. The work of the iterative procedure is based on (4), (5).

2.2 Nonlinear Dimensionality Reduction Method Using KD-Tree for the target space decomposition

In contrast to the approach discussed above decomposition of the target space inevitably leads to the need for periodic renewal or adjustment of the structure by which the decomposition is performed. The reason for this consists in changing the coordinates of objects in the target space as you perform the optimization process.

In this paper, KD-trees are used for the decomposition of the target space. Tree construction is performed at each iteration of the optimization process using the following recursive procedure.

1. Calculation of the characteristics of the current tree node (average coordinates in the source and target space, boundaries of the node in the target space).
2. In the case when the node contains only one object, then exit.
3. Splitting the node into the two child nodes using the perpendicular to the most elongated boundary so that the number of objects in the child nodes is approximately the same.
4. Perform this procedure for the newly formed child nodes.

Constructed tree is used to perform the next iteration of the optimization process. For each object, the sum (4) is calculated by traversing the tree as follows.

1. If the decomposition criterion is not satisfied and the objects contained in the current node of the tree can be considered as a group, then the next term in (4) is calculated using (5).
2. Otherwise decomposition criterion is satisfied, and both subtrees (child nodes) traversed similarly.

A decomposition criterion used in the tree traversal is similar to the above, and restricts (by the value δ) the estimation of the angle at which the minimum bounding rectangle of the node is observed from the position of the object o in the target space.

Figure 2 illustrates this process. Different decomposition levels are shown using different lines: the first level is shown by the solid line, the second level is shown by the dash-dot line, the third level is shown by the dashed line. A group of four objects (minimum bounding rectangle is shown by dots) will be devided into two groups if the angle β at which the minimum bounding rectangle is observed from the position of the object o is greater than the decomposition angle δ.
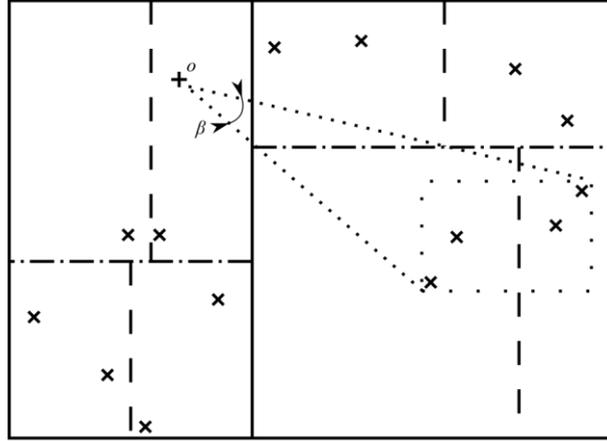
Fig. 2. Illustration of decomposition in the target space using KD-Tree

It is worth noting that this algorithm is executed at each iteration for each object, so to improve the efficiency the constructed tree is traversed not recursively, but iteratively using special pointers.

Note that some additional experimental results on comparison of methods described in this section can be found in [6].

3   Proposed Method

As noted above the expression (4) approximates (2) with some accuracy. The error of approximation for some object o and the set s of individual objects can be represented using (5) as follows:

$$\rho(o,s) = \left\| \sum_{o_j \in s} \frac{d(o,o_j) - d^*(o,o_j)}{d(o,o_j) \cdot d^*(o,o_j)} \cdot (\mathbf{y} - \mathbf{y}_j) - \widetilde{\varsigma}(o,s) \right\|. \tag{6}$$

Unfortunately, the use of the described above approaches can cause significant (in some cases unbounded) approximation errors as in the first case the decomposition criterion takes into account only the characteristics of nodes in the original space and in the second case criterion takes into account only the characteristics of nodes in the target space. The way out of this situation may be a combination of the original space decomposition with the additional control on the configuration of objects in the target space.

In this paper, for this purpose a nonlinear method for dimensionality reduction of data using reference nodes is complemented by the preservation and analysis of the boundaries of nodes in the target space. The proposed method consists of the same steps as the base method [5]:

1. Construction of a hierarchy of clusters
2. Initialization of objects coordinates in the target space
3. Construction of reference nodes lists
4. Iterative optimization procedure

At the first step the difference between the proposed method and the base one is in the structure of nodes in the cluster tree. In the proposed method each node contains information about its boundaries in the target low-dimensional space. The second and the third steps are the same as in the base method. At the fourth step iterative optimization procedure is different in the proposed method.

At the each step of the optimization process for each object $o_i \in O$ with the corresponding reference nodes list S_i it recalculates the coordinates $\mathbf{y}_i$ in the target low-dimensional space. To do this according to (4) it iterates through the list of reference nodes S_i and calculates the approximated term for a given node $s_j \in S_i$ using the following recursive algorithm.

1. If the given node s_j is the object $o_j \in O$ of the initial set of objects, then the exact value of the corresponding term is calculated using (3).
2. If the decomposition criterion is not satisfied for given node s_j in the target low-dimensional space and the objects contained in the given node can be considered as a group, then the approximated term is calculated using (5).
3. Otherwise the decomposition criterion is satisfied for the given node s_j in the target low-dimensional space, and all subtrees of the given node in the cluster tree traversed similarly using this algorithm.

4   Experimental Study

In this paper all the studied methods have been implemented in C++ language in the integrated development environment Borland Turbo C++ 2006 Explorer. The PC

based on Intel Core i5-3470 CPU 3.2 GHz has been used in the experiments. Corel Image Features Data Set (http://archive.ics.uci.edu/ml/databases/CorelFeatures/CorelFeatures.data.html) has been used in the experiments as an initial dataset. This dataset contains a set of features, calculated from the digital images from the Corel image collection (http://corel.digitalriver.com/). In particular, the following feature sets have been used.

Feature set 1 contains color moments [10]. Three features were calculated for each color component: mean, standard deviation, and skewness. The dimensionality of the feature space is 9.

Feature set 2 contains color histograms [13] constructed in the HSV color space. Color space was divided into 8 ranges of H and 4 ranges of S. The dimensionality of the feature space is 32.

Feature set 3 contains texture features based on co-occurrence matrices [2]. Four co-occurrence features (second angular moment, contrast, inverse difference moment, and entropy) were computed in four directions (horizontal, vertical, and two diagonal). The dimensionality of the feature space is 16.

Fragments containing features information for the required number of images has been selected from these feature sets. Then for sets 1 and 3, the identity component has been added to the feature information and normalization has been performed.

To evaluate effectiveness of the methods the value of multidimensional data representation error (1) has been calculated and the average per iteration execution time of the optimization procedure has been measured.

Work of the methods stopped when the rate of decreasing the error slowed down (i.e. the relative decrease in error for ten iterations does not exceed 0.05). In all cases, the dimension of the target space has been set equal to two (two-dimensional data mapping). Initialization is performed using the PCA method.

Some results for the feature set 1 are shown in Fig. 3-6.

Fig. 3 and 4 show the dependence of the qualitative and temporal characteristics on the decomposition angle δ at which the algorithm moves to the child nodes of the corresponding hierarchical structure (cluster tree or KD-tree).

As can be seen from these results, the large values of the angle δ leads to the expected deterioration of the mapping quality due to a coarser approximation, which is reflected in the higher values of the multidimensional data representation error ε (see Fig. 3). The time it takes to perform a single iteration, decreases with increasing angle δ (see Fig. 4), due to the smaller number of processed nodes of the corresponding hierarchical structure during the construction of approximations.

It should be noted that the approach using the original space decomposition provides less value of multidimensional data representation error ε than the approach using the decomposition of the target space.

(a)   (b)
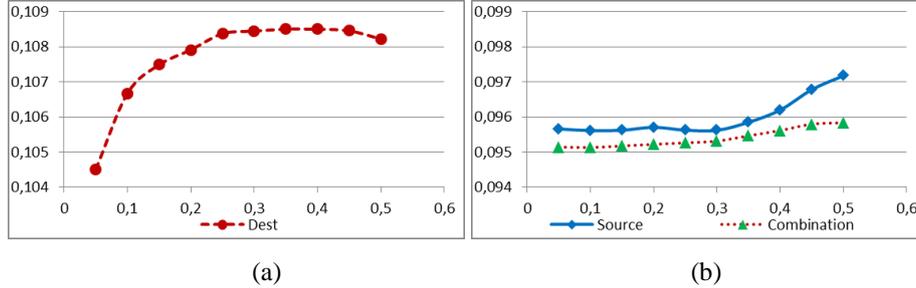
Fig. 3. Dependence of the multidimensional data representation error ε
on the decomposition angle δ (in fractions of π radian)
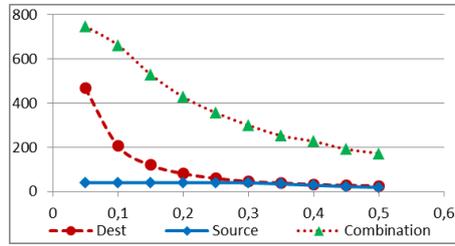
Fig. 4. Dependence of the average execution time per iteration (in ms)
on the decomposition angle δ (in fractions of π radian)

At the same time the average execution time per iteration is much smaller when using the approach with the decomposition of the original space in the range $\delta \in [0.05\ \pi,\ 0.2\ \pi]$.

The use of the combined approach allows slightly reduce the multidimensional data representation error in comparison with the approach based on the decomposition of the original space. At the same time the average time per iteration using a combined approach is significantly larger than when using other considered approaches.

Dependencies of quality and temporal characteristics on the number of objects shown in the figures 5 and 6 confirm said above. The approach using the original space decomposition provides less value of multidimensional data representation error ε than the approach using the decomposition of the target space (fig. 5a). The quality of mapping, measured by the multidimensional data representation error ε, formed using decomposition in the original feature space is only slightly inferior to the combined approach (fig. 5b). The quality of mapping formed using combined approach is virtually identical to the quality obtained with the base method without decomposition.

At the same time approach, using the decomposition of the original space, allows us to construct mapping much faster than the other considered approaches (Fig. 6).

93

Figure 5a also shows the multidimensional data representation error obtained after initialization of objects coordinates in the target low-dimensional space using the PCA.

(a)   (b)
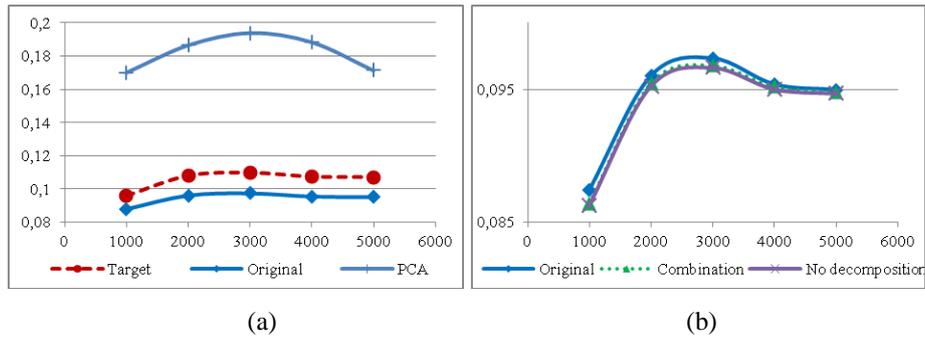
Fig. 5. Dependence of the multidimensional data representation error ε
on the number of objects

Note that the experiments performed on other datasets described above, show similar results.
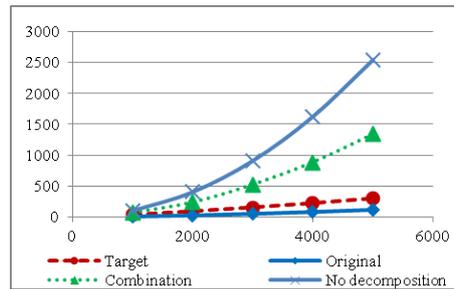
Fig. 6. Dependence of the average execution time per iteration (in ms)
on the number of objects

5   Conclusion

In this paper, we conducted a study of different approaches to the decomposition of space in terms of quality and speed of the nonlinear dimensionality reduction method. The study showed that for the presented data sets decomposition of the original space is more preferable than the decomposition of target space in terms of both quality and time. Combining the decomposition of the original space with the decomposition of target space has allowed slightly improve quality.

94

6　References

1. Fruchterman, T., Reingold, E.: Graph Drawing by Force-directed Placement. Software – Practice and Experience, vol. 21, no. 11, pp. 1129-1164 (1991)
2. Haralick, R.M., Shanmugam, K., Dinstein, I.: Texture Features for Image Classification. IEEE Trans. on Sys. Man. and Cyb. SMC-3(6) (1973)
3. Lee, R.C.T., Slagle, J.R., Blum, H. A.: Triangulation Method for the Sequential Mapping of Points from N-Space to Two-Space. IEEE Transactions on Computers, vol. 26, no. 3, pp. 288-292 (1977)
4. Morrison, A., Ross, G., Chalmers, M.: Fast Multidimensional Scaling through Sampling, Springs and Interpolation. Information Visualization, vol. 2, pp.68-77 (2003)
5. Myasnikov, E.V.: A Nonlinear Method for Dimensionality Reduction of Data Using Reference Nodes. Pattern Recognition and Image Analysis, vol. 22, no. 2, pp. 337–345 (2012)
6. Myasnikov, E.V.: The Choice of a Method for Feature Space Decomposition for Non-Linear Dimensionality Reduction. Computer optics, vol. 38, no. 4, pp. 790-797 (2014)
7. Pekalska, E., de Ridder, D., Duin, R.P.W., Kraaijveld, M.A.: A New Method of Generalizing Sammon Mapping with Application to Algorithm Speed-up. Proc. ASCI'99, 5th Annual Conf. of the Advanced School for Computing and Imaging, pp. 221-228. Heijen, Netherlands (1999)
8. Quigley, A., Eades, P.: FADE: Graph Drawing, Clustering, and Visual Abstraction. Proceedings of the 8th International Symposium on Graph Drawing, pp. 197–210 (2001)
9. Sammon, J.W. Jr.: A Nonlinear Mapping for Data Structure Analysis. IEEE Transactions on Computers, vol. C-18, no. 5, pp. 401-409 (1969)
10. Stricker, M., Orengo, M.: Similarity of color images. In Proc. SPIE Conf. on Vis. Commun. and Image Proc (1995)
11. Sun, J., Crowe, M., Fyfe, C.: Extending metric multidimensional scaling with Bregman divergences. Pattern Recognition, 44 (5), pp. 1137–1154 (2011)
12. Sun, J., Fyfe, C., Crowe, M.: Extending Sammon mapping with Bregman divergences. Information Sciences (2011)
13. Swain, M., Ballard, D.: Color indexing. International Journal of Computer Vision, 7(1) (1991)