A Survey on Dimension Reduction Algorithms in Big Data Visualization

Zheng Sun[1], Weiqing Xing[2], Wenjun Guo[1], Seungwook Kim[1], Hongze Li[1], Wenye Li[1], Jianru Wu[2], Yiwen Zhang[3], Bin Cheng[2(✉)], and Shenghui Cheng[1(✉)]

[1] Shenzhen Research Institute of Big Data and the Chinese University of Hong Kong, Shenzhen, China
chengshenghui@cuhk.edu.cn
[2] Shenzhen Institute of Pharmacovigilance and Risk Management, Shenzhen, China
[3] Anhui University, Hefei, China

Abstract. In practical applications, the data set we deal with is typically high dimensional, which not only affects training speed but also makes it difficult for people to analyze and understand. It is known as "the curse of dimensionality". Therefore, dimensionality reduction plays a key role in the multidimensional data analysis. It can improve the performance of the model and assist people in understanding the structure of data. These methods are widely used in financial field, medical field e.g. adverse drug reactions and so on. In this paper, we present a number of dimension reduction algorithms and compare their strengths and shortcomings. For more details about these algorithms, please visit our Dagoo platform via www.dagoovis.com.

Keywords: High dimension · Dimension reduction · Radar map · Data visualization

1 Introduction

With the rapid development of society, the large amount of data is produced in various field daily and the big data sets with a few attributes are really hard for people to recognize its structure and important information, difficult to analyze and visualize as well, so the requirement of advanced technology in big data analysis is urgent since dimension reduction problem is a big concern [1]. The complexity and vastness of big data lead to the need for various dimensionality reduction algorithms [2] in the analysis process, and these methods are used to improve the performance of the operators of data analytics process [3], so choosing an appropriate dimensionality reduction method is crucial to analysis the big data set. As years' accumulation of data and analysis technologies, big data analysis and applications have entered into a new period. To visualize and analyze the big data set, reduce the dimension of data set and meanwhile preserve characteristics of original data as much as possible, a variety of algorithms is used to reduce dimensions

Z. Sun and W. Xing—Co-first author.

X. Zhang et al. (Eds.): CloudComp 2019/SmartGift 2019, LNICST 322, pp. 375–395, 2020.
https://doi.org/10.1007/978-3-030-48513-9_31

of big data set [4]. After applying the dimensionality reduction methods, the results are always much easier to analyze and get the relevant information.

Nowadays, no matter in which area, the data we deal with are always complex, which have a large volume of information with multiple attributes, and no longer simple to understand. That is why dimension reduction algorithms are so important since these algorithms are exactly used to reduce features when facing high dimensional data sets. So far, many dimension reduction methods have been proposed, like principle component analysis [5], local linear embedding and so on. In this paper, we are going to introduce several dimension reduction algorithms and apply a bunch of dimension reduction methods including Principle Component Analysis this kind of very common algorithm, and visualize the results of different methods with radar maps, which makes results easier to understand. We will analyze the performance of these algorithms, and then make a comparison of the performances of different algorithms applied, which is shown as in the Fig. 1.

Radar map, which is usually used in the financial field, it is also known as the spider map or the network map. We combine the visualization of different algorithms with the radar map, which is used as user portrait, it is a graphical method to display multivariable data in the form of a 2D map, which corresponds to parallel coordinate diagrams, helps us get to know the general information of any points after dimension reduction. The relative positions and angles of the axes in the maps are usually with no information.

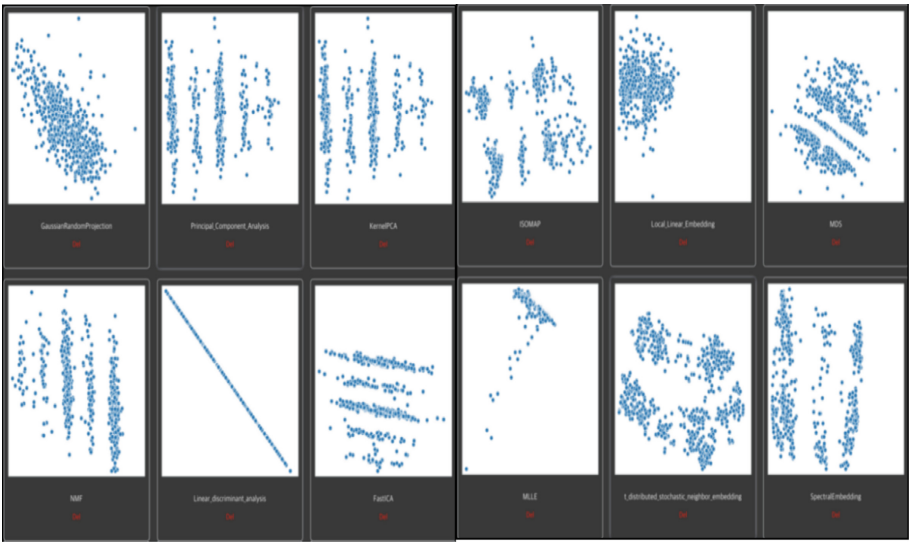

Fig. 1. Comparison of different dimension reduction algorithms

1.1 Problem Statement

The typical high dimensional data is represented as the data matrix X, with m data points and n dimensions.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,n-1} & x_n \\ x_{21} & x_{22} & \dots & x_{2,n-1} & x_{2n} \\ \dots & \dots & \dots & \dots & x_{3n} \\ x_{m-1,1} & x_{m-1,2} & \dots & \dots & x_{m-1,n} \\ x_{m,1} & x_{m,2} & \dots & \dots & x_{mn} \end{bmatrix},$$

where x_{ij} is the value of i-th data point on j-th dimension.

$$X_i = [x_{i1}, x_{i2}, \dots, x_{in}] \text{ is a data point.}$$

The purpose of dimension reduction is to seek the lower dimensional data matrix $X'_{\{m \times p\}}$, where

$$\left\| X_i - X_j \right\| \approx \left\| X'_i - X'_j \right\|$$

Here, $p < n$ ensures the dimensions reduction. Typical p = 2.

1.2 Data Encoding

The most important function of the computer is to process information. The first problem needs to be solved is how to express, store and transfer information in the computer. Obviously, information in the computer can only be expressed, stored and transferred after being digitally encoded. There are various types of data, like numerical data, categorical data, image data, text data etc. These types of data are quite common, especially in medical field.

Taking Adverse Drug Reaction (ADR) as an example. ADR is caused by taking medication which is bad for people's health. ADRs may occur by a single or combined drugs. It is a typically "side effect" for our body, and the effects can be beneficial. It uses text data to show the levels like "good", "bad" or some feelings after medicine. The data they typically got are text data. Some text data can also be categorical data. Dimensional reduction can recover the correlations and co-existence among the drugs. During the dimensional reduction, the data should be numbers. For the text, if data is the serial text, it is easy to be encoded as a number. For example, if the text are "low", "middle", "high", we can encode them as 1, 2, 3 or some other numbers by the domain experts. For the texts which are not serial, we can use word2tec, Latent Semantic Analysis methods to encode them as numbers. After that, the text can be processed as numbers.

2 Typical Dimension Reduction Algorithms

First of all, a brief introduction of dimensionality reduction algorithms we use is given necessarily as the following: Random Projection is a powerful dimensionality method

known for its simplicity and less erroneous output, including Gaussian Random Projection and Sparse Random Projection. Linear Discriminant Analysis (LDA) is similar to Principle Component Analysis (PCA), the difference is that the result of LDA is to project the data to different classifications, while PCA is to project the data to the highest similarity group. PCA, which synthesizes multiple indexes into a few independent comprehensive indexes, simplifies the problem to obtain more effective data information. PCA can maintain the maximum variance of the sample and greatly preserve the information of the original data set. Local Linear Embedding is the process of the dimension reduction of manifolds from high dimension to low dimension, consisting of Hessian LLE and Modified LLE. The LLE algorithm keeps the local linear embedding structure. Local Tangent Space Alignment (LTSA) is a kind of new manifold learning algorithm, which can effectively learn low-dimensional embedding coordinates from high-dimensional sampling data. Multidimensional Scaling (MDS) is a mean of the intuitive spatial map using points in multidimensional space to represent the perceptual and psychometric relationship. MDS method keeps the distances between sample points. Isometric Feature Mapping (ISOMAP) is similar to MDS, while the difference lies in the calculation of the distance matrix in the original space. It is a popular dimensionality reduction method for nonlinear data. ISOMAP method can keep the same geodesic distance approximately. Non-negative Matrix Factorization (NMF) makes all components of the decomposed matrix non-negative and realizing the non-linear dimensionality reduction. The NMF method preserves the information under non-negative constraints. Fast Independent Component Analysis is to extract original independent signals from mixed data. T-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear method to reduce the dimension of high-dimensional data and can be visualized through points. Spectral Clustering uses the eigenvalues of the data similarity matrix to reduce dimensions. Figure 1 is the comparison of different dimension reduction algorithms which we are going to introduce in this paper, we can vividly see the differences between various methods from visualization graphs.

2.1 Gaussian Random Projection Algorithm

Gaussian random projection is a powerful method for dimensionality reduction. The general idea of Gaussian random projection is, for high dimension data X, Y, with dimension n, we can first produce a random matrix R, and then use these data to multiply the random matrix R to get a lower-dimensional data. The data from a mixture of k Gaussians can be projected into log k dimensions while maintaining the approximate level of separation between clusters. Controlling the dimension and distribution of the random projection matrix so as to retain the paired distance between any two samples of the data set, therefore, random projection is a suitable approximation technique for solving the distance-based problems, and also random projection makes original clusters more spherical although the original clusters are highly eccentric.

As shown in Fig. 2, it is the result of visualization of Gaussian random projection algorithm, the data set is from a poverty alleviation project of a county. We can vividly see that only one large cluster exists, so it is hard to figure out the differences between points in the graph, each point represents a person in the poverty alleviation project, except for a bunch of points, which we cannot get any details. Then we combine the

visualization of Gaussian random projection with radar map, as we choose a small part from the visualization graph, the corresponding radar map will change as well to show a general situation of each attribute, for example, the points we choose in the graph on the right in Fig. 2 has a much higher level of education than the points we pick in the graph on the left.

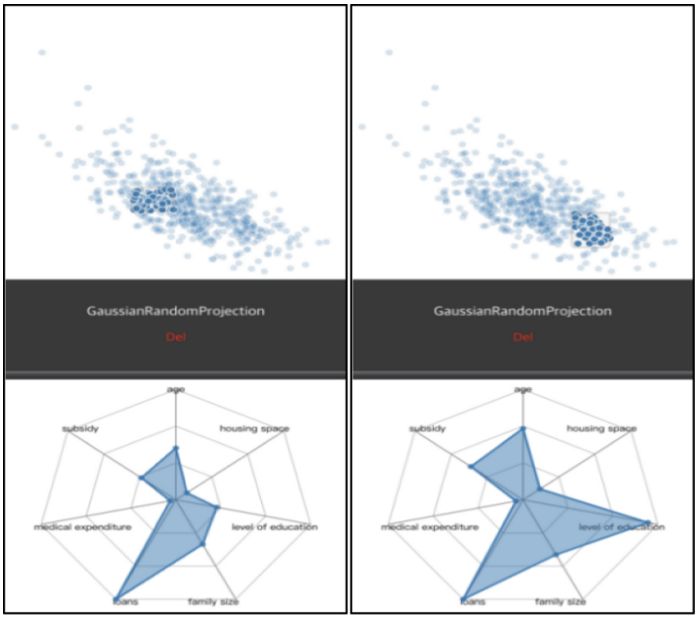

Fig. 2. Gaussian random projection (selected parts are in dark blue) (Color figure online)

The random projection can also be used in conjunction with the expectation-maximization (EM) algorithm easily, which is tested with experiments on synthetic data from a variety of Gaussian mixtures [6], EM algorithm with random projection can save a lot of time in dimension reduction.

2.2 Linear Dimension Reduction Algorithms

According to the relationship between high-dimensional space and low-dimensional space, dimension reduction algorithms can be divided into linear and nonlinear methods. The PCA algorithm and the LDA algorithm are the representative methods for linear dimension reduction algorithms.

Linear Discriminant Analysis Algorithm. LDA belongs to supervised learning, that is, for each given n-dimensional sample S, with a corresponding expected value or category label y. The idea of LDA is pretty simple, which projects or transforms the labeled data into lower dimensional space. In the lower dimensional space, homogeneous samples are as close as possible, while heterogeneous samples are as far as possible,

which means that the points projected are grouped by category, clusters by clusters, points in the same cluster will be closer in the projected space. For LDA bi-classification problem, since only two categories exist, all samples will eventually be projected to the one-dimensional space. While for LDA multi-classification problem, for example, original data has C categories in n dimension, then problem comes that how to figure out the optimal method to project the original data to C-1 dimensional space.

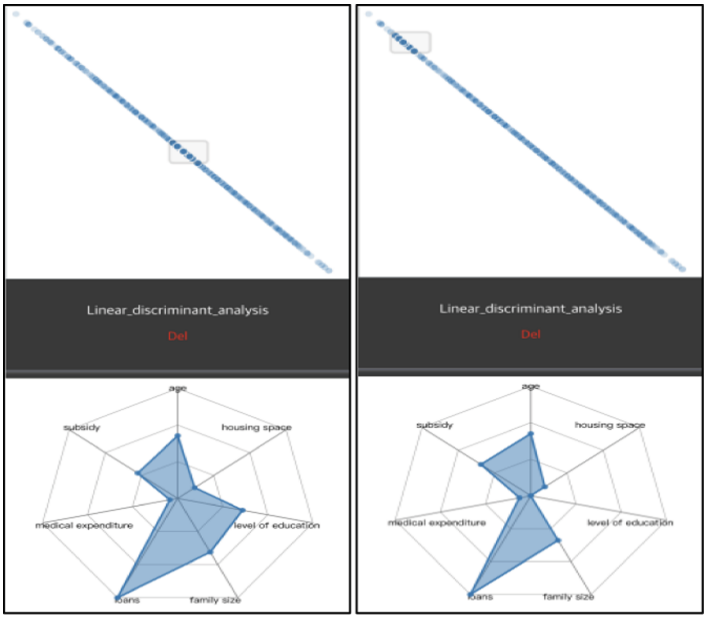

Fig. 3. LDA (selected parts are in dark blue) (Color figure online)

From Fig. 3, we can vividly see that all the points are projected on a straight line, while the selected points change along the line in the visualization graph, the corresponding radar map keep changing as well, then we can figure out the general tendency of different attributes along the line. LDA is well-known for feature extraction and dimension reduction, it is widely used in many areas like image classification [7] and face recognition [8]. Actually, in practical applications, the performance of the LDA algorithm is not that good compared with other algorithms in general, so it often used in conjunction with other algorithms, such as PCA+LDA.

Principle Component Analysis Algorithm. PCA belongs to unsupervised learning, which is probably most widely used dimension reduction algorithm [9] and the most popular multivariable statistical technique [10]. PCA is always used when the data is unlabeled, reduces the dimension of features and meanwhile maintains the features which contribute the most to the variance. Keeping lower-order principle components and ignoring high-order principle components, since lower-order components tend to retain the most important aspects of the data. PCA relies on the original data, the accuracy

of the data has a great impact on the analysis results. The main idea of PCA is to get the principal components and their weights by the eigen decomposition of the covariance matrix. The pursuit of PCA is to preserve the internal information of data as much as possible after dimension reduction. The loss of information is minimized after the dimensions are reduced, since PCA does not classify information, while classification becomes more difficult.

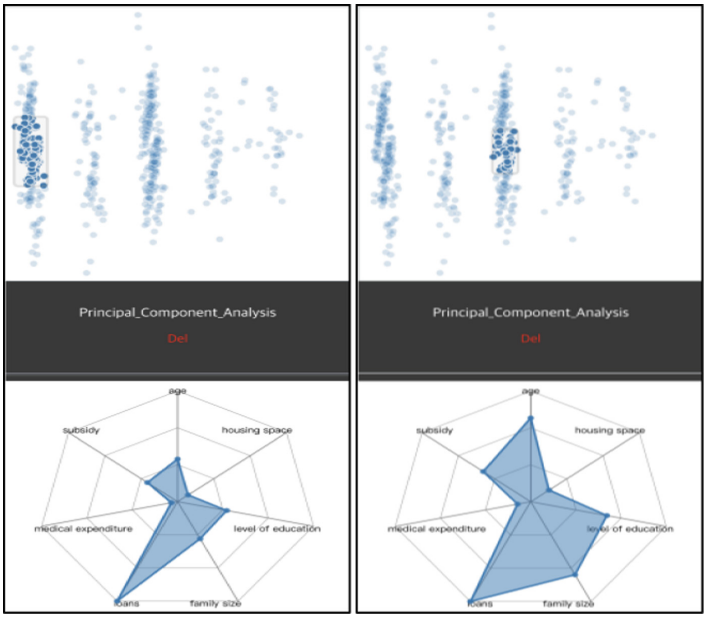

Fig. 4. PCA (selected parts are in dark blue) (Color figure online)

In Fig. 4, clusters can be easily distinguished from the visualization result of PCA, and according to the radar map, we can vividly see the differences between different clusters, and each clusters represent different groups of people. The third group of people are elder than the first group of people with higher subsidy and larger family size, and small housing space, low medical expenditure and high loans are kinds of common characteristics of the poor in this poverty alleviation project.

The LDA algorithm and the PCA algorithm are both linear dimension reduction methods, PCA is an algorithm closely related to LDA. Comparing these two methods, LDA is a supervised learning method, which considers classified information of data set, then the data can be classified in the low dimension space, which cut a lot of computation, while PCA is an unsupervised learning method, and does not take classified information into consideration. LDA pursues to make it easier to distinguish the data points after dimension reduction, and PCA aims to retain the information of data set. The number of dimensions after the LDA method is related to the number of categories, if the original data has n dimensions and c categories, then the dimensions after LDA are 1, 2,… and c−1. While the number of dimensions after the PCA algorithm is related to the data

dimensions, if the original data has n dimensions, the dimensions after PCA are 1, 2,… and n dimension. The PCA projections on a coordinate system are orthogonal, while the LDA method does not guarantee that its projections are orthogonal on the coordinate system.

2.3 Nonlinear Dimension Reduction Algorithms

The representative methods of nonlinear dimension reduction methods are mainly kernel-based nonlinear dimension reduction algorithms like kernel PCA and manifold learning algorithms like ISOMAP and LLE.

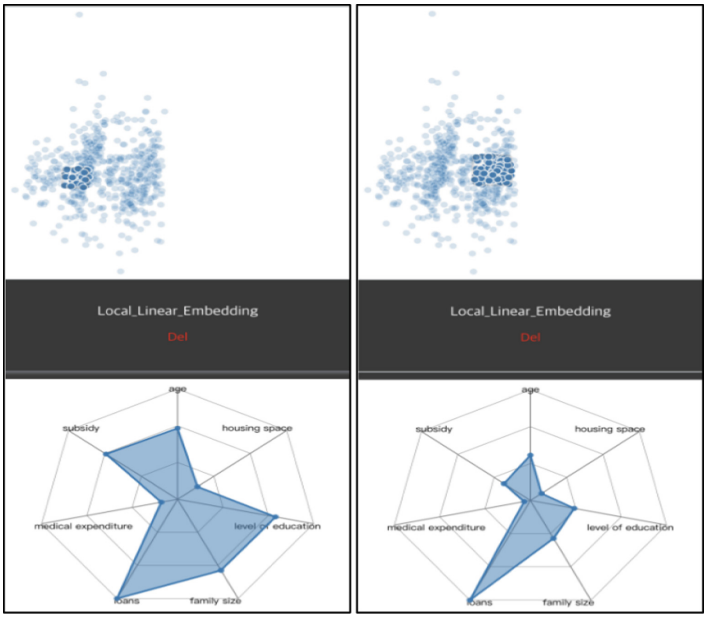

Fig. 5. LLE (selected parts are in dark blue) (Color figure online)

Local Linear Embedding Algorithm. LLE method is one of the Manifold learning algorithms, manifold learning is a kind of framework based on manifolds. Manifolds are abstract in the mathematical sense, but we can regard LLE manifold as an open surface. Dimension reduction algorithm based on manifolds is the process of reducing the dimension of manifolds from high dimension to low dimension, during this process, we hope that some characteristics in the high dimension of manifolds can be preserved. LLE [11] is a very important nonlinear dimensionality reduction method. Compared with traditional dimensionality methods like LDA and PCA, LLE focuses on maintaining local linear features of samples when reducing dimensions, since it, it is widely used in image recognition, data visualization, and other fields. The LLE algorithm is mainly

divided into three steps. The first step is the process of finding the nearest neighbor of K, which uses the same method of finding the nearest neighbor as the K-Nearest Neighbor algorithm. The second step is to find the linear relation of K nearest neighbors for each sample and obtain the weight coefficient W of the linear relation. The third step is to use the weight coefficient to reconstruct the sample data in low dimension space.

According to Fig. 5, two clusters are very close to each other, but when we select the dense points of both clusters, the differences in radar maps are easily to tell, the age, subsidy and family size of the left selected points are much higher than them of right selected points.

The LLE algorithm is efficient, but it has some problems. For example, if the number of neighbors k is great than the dimension of input data, the weight matrix is not full rank. To solve such problems, modified locally linear embedding (MLLE) and Hessian-based locally linear embedding (HLLE) come into existence. For HLLE, it maintains the quadratic relationship of the local Hessian matrix instead of the local linear relationship [12]. For MLLE, it considers the distribution of the weight of neighbor while looking for k nearest neighbor, it hopes to find the neighbor weight distribution in the sample in all directions, rather than focusing on one side. LLE is simple to implement with small computational complexity, it can learn local linear low-dimensional manifolds of any dimension. While it has strict requirements in manifold distribution characteristics of data for example, it can not be a closed manifold, sparse data sets, non-uniformly distributed data sets and so on, which limit its application. The algorithm is sensitive to the selection of the nearest neighbor sample number, and different nearest neighbor numbers have a great impact on the final dimension reduction results.

Local Tangent Space Alignment Algorithm. LTSA is a manifold learning algorithm for nonlinear dimension reduction and mainly considers the local tangent space at each point to represent the geometric characteristics of the point [13, 14]. LTSA can be regarded as a kind of variation of LLE, which hopes to maintain local geometric relations of the data set, and meanwhile uses the technique of transition from local geometry to overall properties. The implementation steps for LTSA are as the following, the first step is neighborhood selection, for a sample point, k nearest neighbor points including itself are selected as neighbors. The second step is local linear fitting and the third step is local coordinates integration, the global coordinates are obtained by step by step calculation according to n local projections. Since the LTSA is linear and not implicit, it has a good effect on nonlinear manifolds dimension reduction.

The visualization result of LTSA method is shown in Fig. 6, almost all the points are centered on the bottom left, while only one point is on the bottom right. We select the only one point on the bottom right, and find its housing space is much larger than the average of other people's housing space.

Based on LSTA, linear LTSA(LLTSA), improved LTSA(ILTSA) and generalized ILTSA(GILTSA) are proposed in recent years to deal with the problem of increasing data set, also the new LSTA methods increase the overall accuracy compare to traditional LTSA algorithm.

LLE and LTSA are nonlinear methods that retain local properties, while MDS, ISOMAP, and kernel PCA are the nonlinear algorithms preserving global properties.

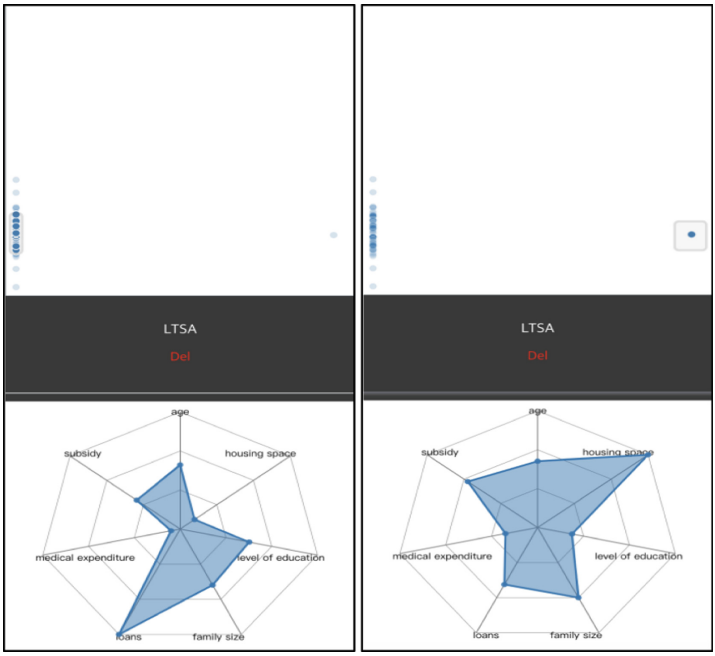

Fig. 6. LTSA (selected parts are in dark blue) (Color figure online)

Multidimensional Scaling Algorithm. MDS is a very traditional method of dimensionality reduction, which takes distance as the standard to project points in high dimensional coordinates to low dimensional coordinates, keeping the relative distance between each other with the minimum change [15–17]. Assuming that the distance matrix of the original high-dimensional data sample is Dh, and the distance matrix of the low-dimensional data sample is Dl, we can select the initial point by using the optimization algorithm, and use the gradient descent method to obtain the best approximation, so that ||Dh-Dl|| is the minimum. Meanwhile, we can also use the inner product to obtain the low-dimensional mapping. The former is prone to fall into local optimum when the sample size is large and the latter is stable, but the latter is worse than the former when the sample size is small. According to daily observation or data collection, many data features are not needed, while learning tasks may be limited to a low-dimensional distribution, which is a low-dimensional "embedding" in high-dimensional space. MDS algorithm is an effective low-dimensional embedding algorithm, which means to reduce the dimension of high-dimensional data on the premise that the distance between the original space and low-dimensional space samples is consistent. When applying MDS to data dimension reduction, the basic idea is to ensure that the distance of all data point pairs in low dimensional space is equal to the distance in high dimensional space. MDS uses the similarity between paired samples, aiming to use this information to construct an appropriate low-dimensional space, that is, the distance between samples in this space and the similarity between samples in high-dimensional space are consistent as far as possible. The MDS algorithm calculates the distance matrix for all n k-dimensional data

through the distance function, which measures the distance in the original feature space, mainly Euclidean distance [18].

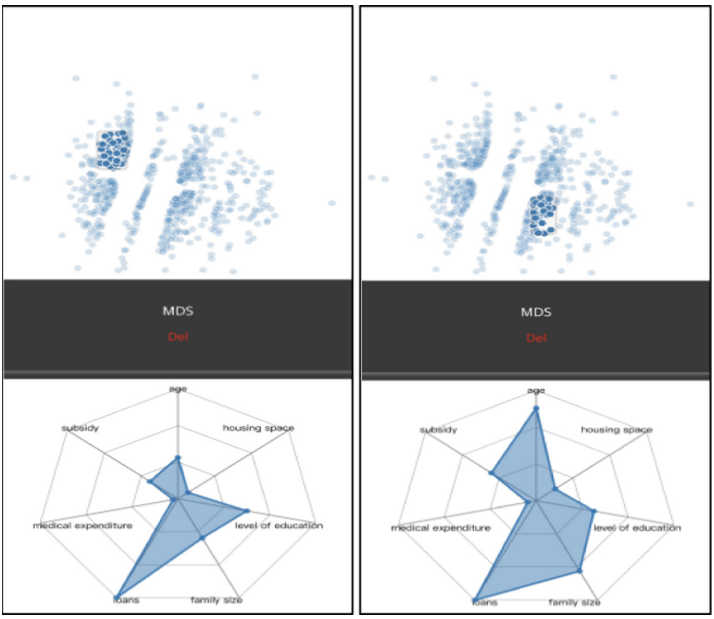

Fig. 7. MDS (selected parts are in dark blue) (Color figure online)

In Fig. 7, we select the top left cluster and the bottom right cluster to see the differences, from the radar map, we can see the housing space, medical expenditure and loan are in common, while the average age of the people in the bottom right cluster is much higher then it in the top left cluster, with relative higher subsidy and family size.

Isometric Feature Mapping Algorithm. ISOMAP is a manifold learning method, and it is used in nonlinear data dimensionality reduction. It is an unsupervised algorithm and its core algorithm is consistent with MDS, but the difference lies in the calculation of the distance matrix in the original space. Many data are nonlinear structure, which is not suitable for direct use of PCA. In a nonlinear data structure, two data points that are far apart on a manifold (geodesic distance) may be very close in a high-dimensional space (Euclidean distance). Only geodesic distance reflects the low-dimensional geometry of manifolds. ISOMAP is based on MDS [19] and retains the essential geometry of nonlinear data, which are geodesic distances between arbitrary point pairs [20]. The process of implementing the ISOMAP method is as the following. First, determining neighborhood for each data point, there are two methods, one is taking the nearest k points as neighbors, and another one is taking all the points in the circle with a selected radius as neighbors. Then we connect these points with edges manifold G build into a weighted flow diagram showing the relationship between adjacent G. Second, calculating the shortest path between all pairs to the geodesic distance matrix on the manifold,

Dijkstra algorithm can be used to calculate the shortest path. Finally, the distance matrix determined according to the shortest path is taken as the input of the MDS algorithm to obtain the data representation that best preserves the essential structure of manifolds in low-dimensional space. When calculating nearest neighbors, if the neighborhood range is specified to be large, then the distant points may be considered as nearest neighbors, causing a "short circuit" problem. If the neighborhood range is specified to be small, then some areas in the figure may not be connected with other areas, resulting in the "circuit breaking" problem. A short circuit or circuit break can mislead later calculations of the shortest path.

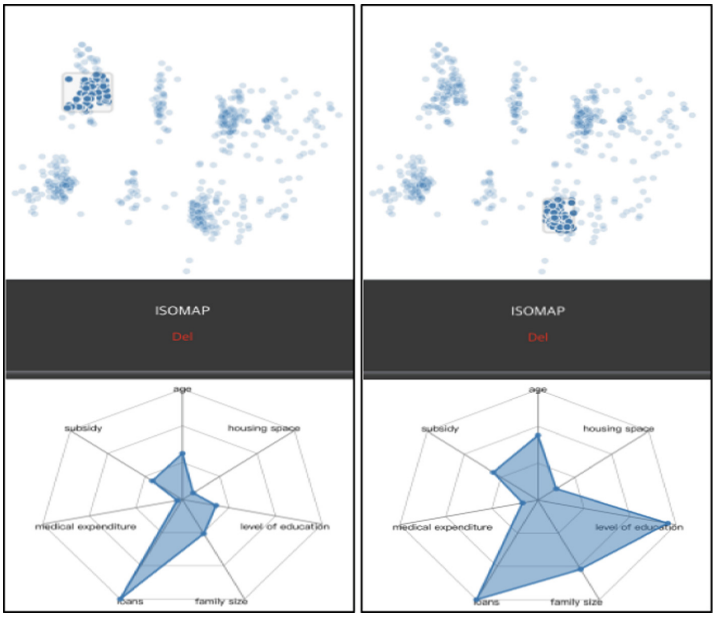

Fig. 8. ISOMAP (selected parts are in dark blue) (Color figure online)

From Fig. 8, the distances between clusters are much larger than the distances between clusters of MDS method. ISOMAP method seems to be more suitable for analyzing this data set than the methods above, since the points in the same cluster are close and the distances between clusters are large, the visualization result performs well.

Comparing ISOMAP with MDS, MDS reduces dimensions and keeps the distance between the samples at the same time, ISOMAP creates a graph by connecting each sample to its nearest neighbors and then tries to preserve the geodesic distance between the samples while lowering the dimension.

Kernel PCA Algorithm. People are able to know the kernel technique if getting to learn support vector machine before, which is to achieve nonlinear classification by using mathematical methods to add features similar to functions. A similar technique can be applied to PCA, allowing complex nonlinear projection dimensionality reduction,

which is known as kernel PCA, a method for performing a nonlinear form of PCA [21]. This algorithm is good at maintaining the after-projection of clusters after clustering, and sometimes the expanded data is close to the distorted manifold. By using the kernel technique, we can calculate the similarity of vectors in two high-dimensional eigenspaces in the original eigenspace.

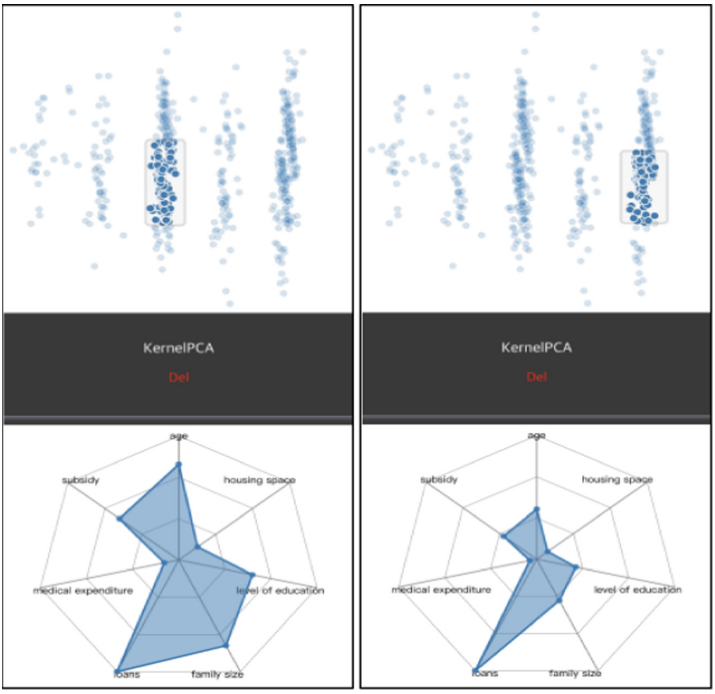

Fig. 9. Kernel PCA (selected parts are in dark blue) (Color figure online)

Kernel PCA is an improved version of PCA, which converts nonlinear, separable data to a new low-dimensional subspace suitable for linear classification of alignment, A kernel PCA can convert data to a high-dimensional space through nonlinear mapping, map it to another low-dimensional space in a high-dimensional space using PCA, and divide samples by a linear classifier. The disadvantage of this algorithm is the high computation cost. The kernel function is an important item in the kernel PCA algorithm, which measures the similarity between vectors by the dot product of two vectors. Common functions include the Gaussian kernel function, polynomial kernel, and hyperbolic tangent kernel.

Kernel PCA is realized in three steps: First, in order to calculate the similarity matrix, the value between any two samples needs to be calculated. Then the kernel matrix is aggregated to make the similarity matrix more clustered. After that, the eigenvalues of the aggregated similarity matrix are arranged in descending order, and the eigenvectors corresponding to the first k eigenvalues are selected, The vectors here are not principal component axes, but the samples are mapped to these axes.

The visualization result of kernel PCA which is shown in Fig. 4, is pretty similar to the result of PCA as shown in Fig. 9, the clusters are easily distinguished, while for the left two clusters, the distances between points in the same cluster are not that close to each other.

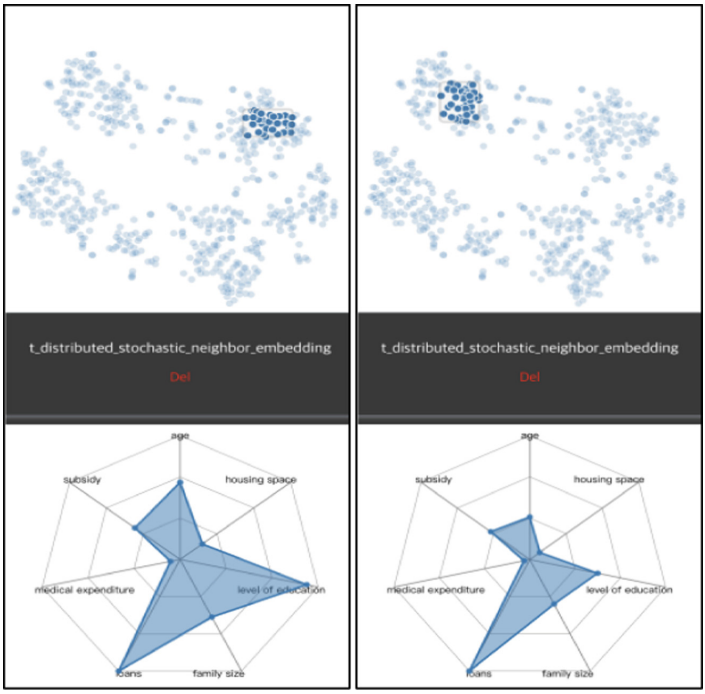

Fig. 10. t-SNE (selected parts are in dark blue) (Color figure online)

t-Distributed Stochastic Neighbor Embedding Algorithm. t-SNE algorithm was proposed by Laurens van der Maaten and Geoffrey Hinton in the year 2008 [22], it can effectively map high-dimensional data to low-dimensional space and maintain the local structure of data in high-dimension space. t-SNE algorithm is an extension of the SNE algorithm, the SNE method utilizes the distribution of adjacent data points of each data point for dimension reduction. When we talk about t-SNE, we should first know the idea of SNE, which is to express the similarity between points through conditional probability with high-dimensional data. t-SNE is an unsupervised dimension reduction method.

The reason for speaking of t-SNE is that in the deep learning test, visualization is needed to analyze the characteristics of the data, so as to know whether the interval between similar categories is small enough and the interval between different categories is large enough in the classification task [23]. t-SNE adopts the method of symmetric SNE for the distribution in high dimensions, while the distribution in low dimensions

adopts the more general T distribution, which is also symmetric. The reason for using T distribution is that T distribution has a little bit taller and longer tail than the normal distribution, which helps the data to be distributed more evenly in two dimensions [24–26]. Although t-SNE greatly improves SNE, they have a very common problem, great time and energy consumption [27–30]. It is difficult to build the network and the gradient descent is too slow when the samples are large.

The visualization result of t-SNE algorithm is shown as Fig. 10, which is similar to the result of ISOMAP, shown in Fig. 8, however, the distance between clusters are smaller than them in the visualization result ISOMAP, and the distances between points in the same cluster of t-SNE method are larger than the distances between points in the same cluster of ISOMAP.

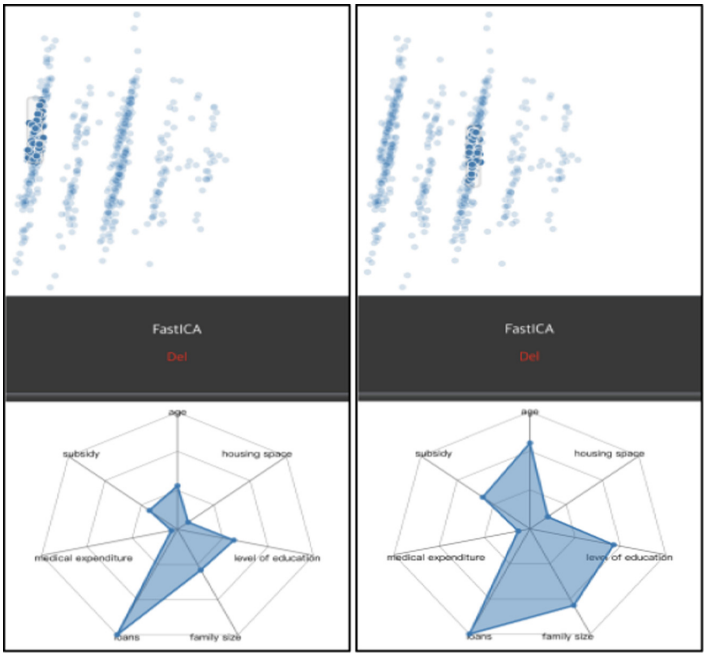

Fig. 11. Fast ICA (selected parts are in dark blue) (Color figure online)

Fast Independent Component Analysis Algorithm. Independent component analysis (ICA) is a linear transformation to calculate using statistical principle. ICA is divided into iterative algorithms based on information theory criteria and algebraic methods based on statistics, such as Fast ICA algorithm, Infomax algorithm, and maximum likelihood estimation algorithm and so on. Fast ICA algorithm, which is also known as a fixed-point algorithm [31], is a fast optimization iterative algorithm, which adopts batch processing, and each iteration involves a large number of sample data. Fast ICA is based on kurtosis, maximum likelihood, maximum negative entropy and so on.

From Fig. 11, the visualization result of Fast ICA is almost the same as the results of PCA and kernel PCA. We can see that the distances between points in the same cluster are much smaller than those in the results of PCA or kernel PCA, since the points in FastICA result only take about two-thirds of canvas, while the points in PCA and kernel PCA results almost spread all over the canvas.

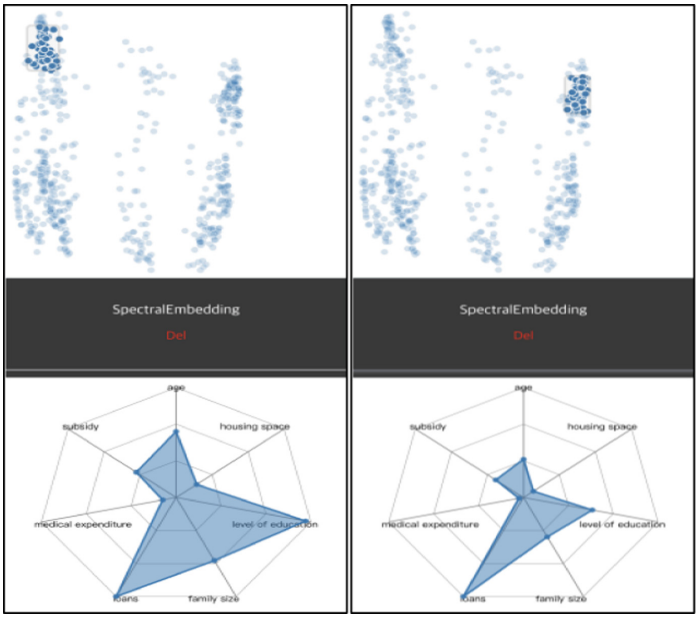

Fig. 12. Spectral Clustering (selected parts are in dark blue) (Color figure online)

Spectral Clustering Algorithm. In spectral clustering, the Laplacian matrix is used for dimension reduction. The general idea of Spectral clustering is that assuming the original matrix is A and infinite relation is 0, the resulting relation matrix is B, then we get diagonal matrix D, the diagonal is the sum of each row of B, and all other elements are 0. Then the Laplacian matrix can be obtained according to the formula, we reduce the dimension of the Laplacian matrix, which is the same as that of PCA, while this time, we are not going to get the eigenvector corresponding to the maximum eigenvalue, the minimum eigenvalue instead. Spectral clustering is simple to implement, but it is not obvious to figure out why it works and what it really does. To solve the problem, different versions of spectral clustering algorithm have been proposed [32, 33]. As one of the most popular modern clustering algorithms, spectral clustering has many applications in machine learning computer vision.

In Fig. 12, the visualization of spectral clustering is similar to the results of MDS, ISOMAP, while the distances between points in the same cluster are too large, so the performance of spectral clustering for analyzing this poverty alleviation data set is not that good as the performance of ISOMAP.

3 Display with the Original Data Set

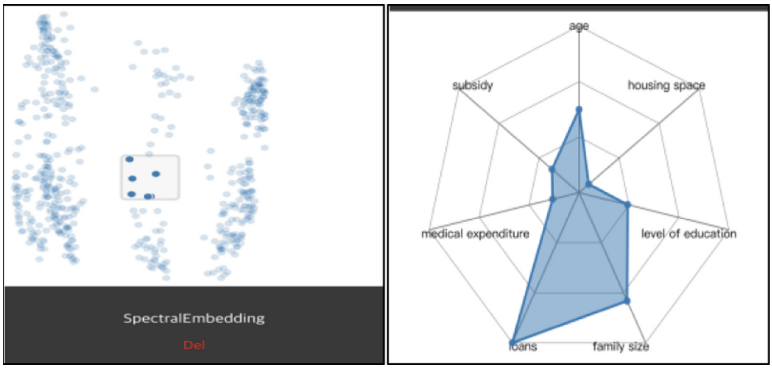

Fig. 13. Points selected

name	age	subsidy	medical expenditure	loans	family size	level of education	housing space
105.0	41	30472.02	10296.87	40000	5	1	75
178.0	55	52569.24	8596.67	40000	10	1	80
195.0	39	8897.77	2990.12	40000	3	1	80
220.0	46	26645.55	89614.8	40000	5	1	83
278.0	59	31037.16	45598.29	40000	10	1	90
296.0	48	10705	9072.87	40000	3	1	90

Fig. 14. Corresponding data of the selected points

We can not only get the general information from radar map, but also can vividly see the original data as we pick any parts of the visualization graph of different algorithms, as shown in Fig. 13 and Fig. 14. The original data set we use has multiple features, such as subsidy, loans, and family size, it is poverty alleviation data set for a county, and since the radar graph is especially suitable for giving a holistic evaluation for multivariable architecture, then we combine the different dimension reduction algorithms with radar map, once we pick an area from the visualization graph after using dimensionality reduction methods, then the radar map will vividly show the general information, and the original data corresponding to the selected area will be given automatically as well, which means that we can learn not only the general information after reducing dimension but also the characteristics of original data at the same time. Combining information before and after reducing dimensions, which makes these high-dimensional data set much easier to understand.

4 Conclusion

In this paper, we introduce and visualize some typical dimension reduction algorithms and make comparisons. From these figures, we can vividly see the differences of performance results of different dimension reduction algorithms using the same data set. From the visualization, the results of using PCA, kernel PCA and Fast ICA are pretty similar. For Gaussian random projection and local linear embedding, unlike PCA and t-distribution neighbor embedding having several clusters, just one cluster appears in each of the results of Gaussian random projection and local linear embedding algorithms. Comparing the results of ISOMAP and t-distribution stochastic neighbor embedding, the distances of nodes in the same cluster with the ISOMAP method are smaller than them with the t-distribution method. Also, we can clearly see the distances between clusters using the PCA algorithm are greater than the distances between clusters using Fast ICA.

We also study the distribution of points. For Fig. 15, we select a part from the visualization graph of PCA, the positions of the same points we pick in PCA will be displayed in other visualization graphs, and we can get the distribution situations of these points in the graph of other algorithms. We pick a small part from the same cluster in the graph of PCA, and we can see that these points are almost in the same cluster in the graph of other algorithms, for LLE, there are not many clear clusters, but we can also see these points distribute closely.

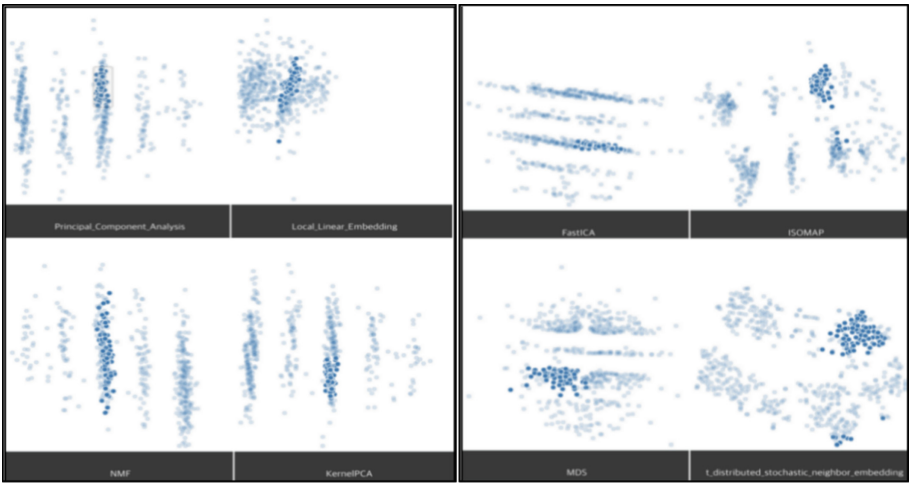

Fig. 15. Distribution of the same points with different algorithms (selected parts are in dark blue) (Color figure online)

In general, we cover 11 dimension reduction algorithms, which are introduced in detail above, while in fact, we actually study more than these, we do not give detailed introduce in this paper, like Modified LLE in Fig. 1. Just comparing the introduced algorithms, we can learn the differences between these methods. For example, although the random projection method sacrifices some accuracy, it greatly reduces the computation.

Table 1. The comparison table of different algorithms

Algorithm	Advantages	Disadvantages	Attributes
Random projection	Greatly reduce computation.	Sacrifice some accuracy	
LDA	Compared with the ambiguity of PCA, its purpose is more clear, and it can reflect the differences between samples better	Limited by types of samples; The maximum number of dimensions of the projection space is C-1 (the original data has C categories)	Supervised learning; Linear
PCA	The concept is simple; Minimum error; The main information is extracted	Calculate co-variance matrix causes large calculation; When gross corruption exists, PCA cannot grasp the real subspace structure of data	Unsupervised learning; Linear
LLE	Can process and analyze nonlinear signals; The selection of parameters is few, so the characteristic parameters can be optimized better	Require dense sampling; Embedding dimension and noise in the signal will affect the performance of dimension reduction in high-dimension space	Unsupervised learning; Nonlinear
LTSA	Reflect the local geometric features of manifolds	Not suitable for processing data sources with high curvature distribution and sparse distribution	Manifold learning; Nonlinear
MDS	Retain the relative relation of original data well	Cannot deal with high-dimensional nonlinear data	Nonlinear; Unsupervised
ISOMAP	Suitable for the internal flat low-dimensional manifold	Not suitable for the manifolds with large intrinsic curvature	Manifold learning; Unsupervised
NMF	Wide range of Applications	The choice of K is ambiguous; NMF training takes time	Nonlinear; Unsupervised
Fast ICA	Fast convergence; Simple and effective	Kurtosis value can only be estimated from the measured sample	Nonlinear; Unsupervised
t-SNE	Very useful for reducing high-dimensional data to 2D or 3D	High computational complexity; The global structure is not explicitly retained	Manifold learning; Unsupervised
Spectral clustering	Effective for sparse data	Rely on similarity matrix, different matrices cause various results	Nonlinear; Unsupervised

t-SNE has high computational complexity, it is very useful for reducing high-dimensional data to 2D or 3D. The following comparison Table 1 is a summary of the advantages and shortages of various typical dimension reduction algorithms. For more details, please visit our Dagoo platform [34]. Dimensional reduction is a typical method that aims to find the features and patterns in the high dimensional space. These methods have wide applications in financial field, medical field e.g. ADR.

Funding Acknowledgement. This paper is support by the Program for Guangdong Introducing Innovative and Enterpreneurial Teams (Grant No.: 2017ZT07X183), the Pearl River Talent Recruitment Program Innovative and Entrepreneurial Teams in 2017 (Grant No.: 2017-ZT07X152), the Shenzhen Fundamental Research Fund (Grants No.: JCYJ20170306141038939, KQJSCX20170728162302784, KQTD2015033114415450 and ZDSYS201707251409055), and Department of Science and Technology of Guangdong Province Fund(2018B030338001), and Shenzhen Science and Technology Innovation Committe, (Basic Research (Free Exploration No.: CYJ20170818104824165).

References

1. Song, L., Ma, H., Wu, M., Zhou, Z., Fu, M.: A brief survey of dimension reduction. In: Peng, Y., Yu, K., Lu, J., Jiang, X. (eds.) IScIDE 2018. LNCS, vol. 11266, pp. 189–200. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02698-1_17
2. Fodor, I.K.: A Survey of Dimension Reduction Techniques. No. UCRL-ID-148494. Lawrence Livermore National Lab., CA, US (2002)
3. Tsai, C.-W., et al.: Big data analytics: a survey. J. Big Data **2**(1), 21 (2015)
4. Engel, D., Hüttenberger, L., Hamann, B.: A survey of dimension reduction methods for high-dimensional data analysis and visualization. In: Visualization of Large and Unstructured Data Sets: Applications in Geospatial Planning, Modeling and Engineering-Proceedings of IRTG 1131 Workshop 2011. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2012)
5. Zhang, T., Yang, B.: Big data dimension reduction using PCA. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud). IEEE (2016)
6. Brigham, E.: Random projection in dimension reduction: applications to image and text data. In: ACM SIGKDD ICKDDM (2001)
7. Ye, F., Shi, Z., Shi, Z.: A comparative study of PCA, LDA and kernel LDA for image classification. In: 2009 International Symposium on Ubiquitous Virtual Reality, Gwangju, pp. 51–54 (2009)
8. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using LDA-based algorithms. IEEE Trans. Neural Networks **14**(1), 195–200 (2003)
9. Jolliffe, I.: Principal Component Analysis. Springer, Berlin Heidelberg (2011). https://doi.org/10.1007/b98835
10. Abdi, H., Williams, L.J.: Principal component analysis. Wiley Interdisc. Rev. Comput. Stat. **2**(4), 433–459 (2010)
11. Roweis, S.T., Lawrence, K.S.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
12. Dou, J., Qin, Q., Tu, Z.: Robust edit propagation based on Hessian local linear embedding. In: 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, pp. 3336–3339 (2017)
13. Zhang, Z., Zha, H.: Nonlinear dimension reduction via local tangent space alignment. In: Liu, J., Cheung, Y.-m., Yin, H. (eds.) IDEAL 2003. LNCS, vol. 2690, pp. 477–481. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45080-1_66

14. Yang, G., Xu, X., Zhang, J.: Manifold alignment via local tangent space alignment. In: 2008 International Conference on Computer Science and Software Engineering, Hubei, pp. 928–931 (2008)
15. Cox, T.F., Cox, M.A.A.: Multidimensional Scaling. Chapman and hall/CRC, Boca Raton (2000)
16. Borg, I., Groenen, P.: Modern multidimensional scaling: theory and applications. J. Educ. Meas. **40**(3), 277–280 (2003)
17. Kruskal, J.B., Wish, M.: Multidimensional Scaling, vol. 11. Sage, London (1978)
18. Li, Y.: Locally multidimensional scaling for nonlinear dimensionality reduction. In: 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, pp. 202–205 (2006)
19. Bengio, Y., et al.: Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering. In: Advances in Neural Information Processing Systems (2004)
20. Fan, M., et al.: Isometric multi-manifold learning for feature extraction. In: 2012 IEEE 12th International Conference on Data Mining. IEEE (2012)
21. Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D. (eds.) ICANN 1997. LNCS, vol. 1327, pp. 583–588. Springer, Heidelberg (1997). https://doi.org/10.1007/BFb0020217
22. Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
23. Mounce, S.: Visualizing smart water meter dataset clustering with parametric t-distribution stochastic neighbor embedding. In: 2017 13th International Conference on Natural Computation, Fuzzy System and Knowledge Discovery (ICNC FSKD), Guilin, pp. 1940–1945 (2017)
24. Cheng, S., Zhong, W., Isaacs, K.E., Mueller, K.: Visualizing the topology and data traffic of multi-dimensional torus interconnect networks. IEEE Access **6**, 57191–57204 (2018)
25. Cheng, S., Xu, W., Mueller, K.: RadViz Deluxe: a component-aware display for multivariate chemical data. Processes **5**(4), 75 (2017)
26. Cheng, S., Mueller, K.: The data context map: fusing data and attributes into a unified display. IEEE Trans. Visual Comput. Graphics **22**(1), 121–130 (2016)
27. Cheng, S., Mueller, K., Xu, W.: A framework to visualize temporal behavioral relationships in streaming multivariate data. In: New York Scientific Data Summit, pp. 1–10, New York, August 2016
28. Cheng, S., Mueller, K.: Improving the fidelity of contextual data layouts using a generalized Barycentric coordinates framework. In: 2015 IEEE Pacific Visualization Symposium (PacificVis), pp. 295–302 (2015)
29. Cheng, S., De, P., Jiang, S.H., Mueller, K.: TorusVisˆND: unraveling high-dimensional torus networks for network traffic visualizations. In: First Workshop on Visual Performance Analysis, pp. 9–16 (2014)
30. Cheng, S., Xu, W., Mueller, K.: ColorMapND: a data-driven approach and tool for mapping multivariate data to color. IEEE Trans. Visual Comput. Graphics **25**(2), 1361–1377 (2019)
31. Spurek, P., Jacek, T., Śmieja, M.: Fast independent component analysis algorithm with a simple closed-form solution. Knowl.-Based Syst. **161**, 26–34 (2018)
32. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems (2002)
33. Von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
34. Cheng, S., et al.: Dagoo – A platform for big data visualization (2018). http://www.dagoovis.com