Mobile Networks and Applications (2020) 25:1376–1391
https://doi.org/10.1007/s11036-019-01455-9

Double-Arc Parallel Coordinates and its Axes re-Ordering Methods

Liangfu Lu[1] · Wenbo Wang[2] · Zhiyuan Tan[3]

Published online: 8 January 2020

Abstract
The Parallel Coordinates Plot (PCP) is a popular technique for the exploration of high-dimensional data. In many cases, researchers apply it as an effective method to analyze and mine data. However, when today's data volume is getting larger, visual clutter and data clarity become two of the main challenges in parallel coordinates plot. Although Arc Coordinates Plot (ACP) is a popular approach to address these challenges, few optimization and improvement have been made on it. In this paper, we do three main contributions on the state-of-the-art PCP methods. One approach is the improvement of visual method itself. The other two approaches are mainly on the improvement of perceptual scalability when the scale or the dimensions of the data turn to be large in some mobile and wireless practical applications. 1) We present an improved visualization method based on ACP, termed as double arc coordinates plot (DACP). It not only reduces the visual clutter in ACP, but use a dimension-based bundling method with further optimization to deals with the issues of the conventional parallel coordinates plot (PCP). 2)To reduce the clutter caused by the order of the axes and reveal patterns that hidden in the data sets, we propose our first dimensional reordering method, a contribution-based method in DACP, which is based on the singular value decomposition (SVD) algorithm. The approach computes the importance score of attributes (dimensions) of the data using SVD and visualize the dimensions from left to right in DACP according the score in SVD. 3) Moreover, a similarity-based method, which is based on the combination of nonlinear correlation coefficient and SVD algorithm, is proposed as well in the paper. To measure the correlation between two dimensions and explains how the two dimensions interact with each other, we propose a reordering method based on non-linear correlation information measurements. We mainly use mutual information to calculate the partial similarity of dimensions in high-dimensional data visualization, and SVD is used to measure global data. Lastly, we use five case scenarios to evaluate the effectiveness of DACP, and the results show that our approaches not only do well in visualizing multivariate dataset, but also effectively alleviate the visual clutter in the conventional PCP, which bring users a better visual experience.

Keywords PCP · Arc-based parallel coordinate plot · Double arc coordinate plot · Visualization · Dimension-based bundling layout · SVD · Mutual information · Nonlinear correlation coefficient

1 Introduction

Parallel Coordinates Plot (PCP) is a simple but strong geometric high-dimensional data visualization method [1–3], which represents N-dimensional data in a 2-Dimensional space with mathematical rigorousness. This approach has been extensively adopted for visualizing both high dimensional dataset and multivariate dataset [4, 5]. Some approaches have been proposed to improve the legibility of parallel coordinates plot [6]. As the increasing of the number of axes or scale of data items, clutter would come out in the layout. Then dimensional reordering in parallel coordinates was proposed to reduce the clutter by revealing patterns that hidden in the layout before [13]. Either reducing the clutter produced by the multiplicity of overlapping and crossing lines, or enhancing their patterns. One of these alternatives is to change the shape of the axes. Rather than using the traditional line segments as the coordinate axes, the paper used the segments of curve to replace them. As we all know, in the same coordinates system, such as Cartesian coordinates system, the length of arc is longer than the length of the line segments. So it can visualize much more data items in the same screen space.

✉ Zhiyuan Tan
z.tan@napier.ac.uk

[1] School of Mathematics, Tianjin University, Tianjin, People's Republic of China 300072

[2] SIST, ShanghaiTech University, Shanghai, People's Republic of China 200120

[3] School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, UK

However, the existing PCP method lacks the ability to visualize data distribution and also supports a low quality of displaying effect. Moreover, using PCP becomes challenging as the number of data items grows larger and quicker. The visualizing results always cause line occlusion, line ambiguity and hidden information. Therefore, as the increasing of the number of axes or scale of data items, clutter would come out in the layout. Therefore, in this paper, we do three main contributions on the state-of-the-art PCP methods. One approach is the improvement of visual method itself. The other two approaches are mainly on the improvement of perceptual scalability when the scale or the dimensions of the data turn to be large in some mobile and wireless practical applications.

Based on the ACP method, we propose an improved visualization method, double arc coordinates plot (DACP) firstly. We use a pair of axes composed by two back-to-back arc axes to represent one axis in the conventional PCP. Inside the two arc axes, we use line segments to connect adjacent pair of axes to show observation points. Therefore, the data distribution information on each coordinate axes can be clearly displayed by the internal link of each pair of axes. Moreover, we propose a bundling method to optimize the display effect, which is based on dimensions. Each dimension takes each pair of coordinate axes as the basis and bundles the close lines with similar tendencies. In addition, we use a layout of filling the bundled lines with various transparency to know the number of lines in each bundle. The transparency is computed by the number of bundled lines, and all bundles are filled according to their values. Users can learn the amounts of each bundle of lines from the depth of their colors.

Independent of the orientation, the order of axes affects the visual patterns greatly. Therefore, we propose two rational dimension reordering methods to support data visual analytics in DACP. Firstly, a method to re-order axes (or dimensions) is developed on the basis of the singular values decomposition (SVD). The axes are re-organized and visualized as double arc parallel coordinates from left to right according to their contribution rates, which are calculated by the contribution of each dimension. This helps to find out the optimal order of axes in a short time period. Secondly, a similarity-based reordering method is presented in DACP. This method is inspired by Person‘s Correlation Coefficient (PCC), and is a combination of a Nonlinear Correlation Coefficient (NCC) and SVD algorithm. The method is not only more rational than the current PCC method in theory but also significantly improves the quality of multidimensional visualization in terms of effectiveness and correctness.

This paper is organized as follows: we first present previous works on existing enhancements in PCP and researches on dimension reordering in high-dimensional data visualization (Section 2). Then, we describe the double arc coordinate method theoretically in the novel coordinates system and describe the bundling layout based on dimension in our approach (Section 3). In Section 4, we introduce how the two dimensions reordering approaches are working. The experimental evaluation is explored in Section 5. Finally, in Section6 we draw conclusions and present directions for future work.

2 Related works

2.1 Rationale of PCP

Parallel coordinates plot is proposed firstly by Inselberg [2] in 1985, and later in 1990 Wegman [7] applied it in hyper-dimensional data analysis. Here is the method details: A point P in a Cartesian system can be mapped into the joining $P_1(0, a)$ to $P_2(1, ma + b)$ in the parallel coordinates; two points lying on the line L in the Cartesian coordinate plane given by L:$y = mx + b$ can be mapped into two lines in parallel coordinates, as shown in Fig. 1. And the two lines intersect at a point $M\left(\frac{b}{1-m}, \frac{1}{1-m}\right)$, where $m \neq 1$. Therefore, a 2-dimensional plane with parallel axes connected by linear segments can represent the coordinates of N-dimensional data. According to the theory of duality property [6] between two different coordinates, the parallel coordinates' visualization possesses some pleasant duality properties through the usual representation of Cartesian orthogonal coordinates.

Although the visualization system displays data without losing any features, the PCP also suffers from numerous challenges [8]. We focus on crowed dimensions and Dimension Layout.

2.2 Improvements on PCP

One of the most important technical challenges with Parallel Coordinates Plot is Crowed Dimensions. As the volume of datasets and the number of dimensions are increased, the edges are cluttered and overlapping lines obscure patterns. Several enhancements have been proposed to resolve this problem [9]. The majority of these approaches can be placed into one of three categories: line-based approaches, axes-based approaches and external approaches.

2.2.1 Line - based approaches

Line-based approaches represent changing the attributes of lines to reduce visual clutter, such as changing line colors, densities or shapes. For example, Huh et al. [10] proposed an enhanced PCP which has proportionate spacing between variables. The data points were connected by “near smooth” curves rather than straight lines; Zhou et al. [11] also exploited curve lines to form visual bundles for clusters in parallel coordinates to reduce the visual clutter in clustered visualization.

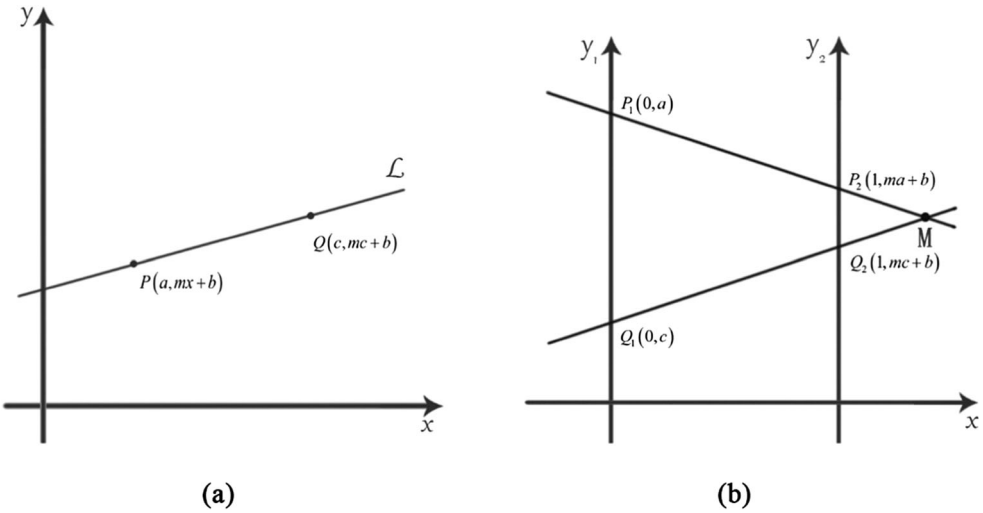


Fig. 1 **a** The Cartesian system (**b**) The parallel coordinates

In our work, we present a bundling layout method based on dimension bundling close lines with similar tendencies to reduce the visual clutter.

2.2.2 Axes – Based approaches

Axes-based approaches extend the axes of parallel coordinates. Claessen et al. [4] developed flexible linked axes to enable users to define and position coordinate axes freely; and Tominski [12] proposed Axes-based techniques with radial arrangements of the axes, termed as TimeWheel and the MultiComb. This method combined some conventional interaction techniques. With the combination between interaction techniques and PCP, Hauser et al. [13] also designed an angular brushing technique to select data subsets which exhibit a data correlation along two axes.These approaches enhanced the quality of visualization to some extent. But the corresponding extensions for the axes in PCP still focus on the line segments between two adjacent axes.

To solve this problem, Huang M L et al. [6] proposed arc-based parallel coordinate plots (ACP) using arc axes rather than line segments as the coordinate axes to display much more items in the same screen space. To strengthen the visualization of high dimensional data, some studies on finding better layouts in PCP have been proposed. For example, Wei Peng et al. [14] defined visual clutter in parallel coordinates as the proportion of outliers against the total number of data points. They tried to use the exhaustive algorithm to find an optimal axes order that can minimize the clutter in a display; Mihael Ankerst et al. [15] also defined similarity measures which determined the partial or global similarity of dimensions. They argued that the reordering based on similarity could reduce visual clutter and do some help in visual clustering; Almir Olivette Artero et al. [16] proposed a method based on similarity to reorder and reduce dimension, called Similarity-Based Attribute Arrangement (SBAA).

The main method of exploring new layouts is dimension reordering. Most of recent dimension reordering methods are established on the basis of Pearson's Correlation Coefficient(PCC). However, from the statistics point of view, PCC is taken as a method for only measuring the linear correlation between two random variables. It is not sufficient to reorder dimensions in similarity if only depends on the calculation of PCC.

Most similar to our method, Aritra Dasgupta et al. [17] developed Pargnostics, a screen-space metrics for parallel coordinates. They calculated for pairs of axes and took into account the resolution of the display as well as potential axis inversions. But the probability and joint probability during the computational process were both denoted as their special axis histograms, which lacked the support of mathematical theories. Moreover, it could be seen from the definition of the mutual information that it did not range in a definite closed interval as the correlation does, which ranges in [−1,1].

2.2.3 External approaches

External approaches represent involving supports from methods other than parallel coordinates plot to uncover clutters in crowed PCP. Such as: user preferences, clustering algorithms and other existed visualization techniques. For example, Dasgupta et al. [9] proposed a model based on screen-space metrics, which is a way of automatically optimizing the results; And Artero et al. [18] developed a frequency and density plots for PCP; While Yuan et al. [19] combined the parallel coordinate's method with scatterplots directly with a seamless transition between them.

Therefore, we make a further improvement and propose a novel axis system in parallel coordinates visualization, termed as double arc coordinate plots (DACP). Comparing with the ACP, our method not only retains the integrity of all advantages the ACP has, but also displays the distribution

information of data items in each pair of coordinate axes. Moreover, two efficient methods for dimensions reordering are proposed. One is contribution-based reordering, based on SVD algorithm, which can not only provide theoretical support for the selection of the first dimension, but also visualize the clear and detailed structure of the dataset with contribution of each dimension; the other is similarity-based reordering method, which is based on combination of NCC and SVD algorithms. Dimensions are reordered in line with the degree of correlations among dimensions. This method is more rational, exact and systemic than the traditional methods. And the combination of this reordering method and DACP makes visualization efficiency better than it works with PCP.

3 Double arc coordinate plot model

3.1 Double arc coordinate system

As mentioned in literature [4], we also involve arcs of circle to replace the original coordinate axes. The purpose is using a longer length segment to replace line segment according to the requirements of displaying more data items, and to remain better geometric structure of some circular datasets.

To further describe the double arc coordinate system, we define the origin as point (0,0) to ensure the generality, which is also the center of the first axis in PCP, marked as point O, displayed in Figure 2. If we consider the length of each axis is 1, a horizontal line in PCP divides all axis into two equal line segments vertically, so the distance of axis $X_1$and axis X_2 is $\frac{3}{2}$; in addition, we argue that the distance between two axis $X_1$and X_2 is $\frac{3}{2}$.

As our PCP system is using arcs to replace lines, as shown in Fig. 2, a vertical axis is replaced by two arcs, a left arc and a right arc. Specifically, the left arc is generated by a circle, which the center point is $O_1\left(-\frac{3}{4},0\right)$, the radius is $\frac{\sqrt{2}}{2}$; While the right arc is generated by a circle with the center point is $O_2\left(\frac{3}{4},0\right)$, and the radius is the same as the radius of the left arc, which is$\frac{\sqrt{2}}{2}$. So we calculate that the position of upper end point of axes X_1 and X_2 are $\left(0,\frac{1}{2}\right)$ and $\left(\frac{3}{2},\frac{1}{2}\right)$ respectively. Therefore, the shortest distance between two arc axes is $(\frac{3}{2}-\sqrt{2}\,)$, and the longest distance is $\frac{1}{2}$.

Based on the calculation formula of the Euclidean distance between two points of the plane with Cartesian coordinates, the equations of the left and right arc of the first pair of axis can be denoted as $\left(x+\frac{3}{4}\right)^2+y^2=\frac{1}{2}$ and $\left(x-\frac{3}{4}\right)^2+y^2=\frac{1}{2}$ respectively. And so on, the equation of the left of i-th pair of arc-arc-axes can be termed as $\left(x+\frac{3}{4}-\frac{3}{2}i\right)^2+y^2=\frac{1}{2}$, where $i=0,\ 1,\ 2\cdots n,\ n\in N$; and the right of i-th pair of arc-axes can be termed as $\left(x-\frac{3}{4}-\frac{3}{2}i\right)^2+y^2=\frac{1}{2}$, where $i=0,\ 1,\ 2\cdots n,\ n\in N$.

To correctly transit information to the arcs, we obtain one to one mapping between the Cartesian coordinates and double arc coordinates. According to the above assumption, we take the first pair of arc-axes as our projection example to explain the mapping of vertices from PCP to DACP. Literature [6] has mentioned that the extension rate of the axis length from PCP to ACP is $\frac{\sqrt{2\pi}}{4}$ when compares PCP and ACP. So because in this paper we use the same radius as literature [6], then the extension rate is $\frac{\sqrt{2\pi}}{4}$ as well.

In Polarimetry, we know that for every two points in space there is a straight line passing through them, and such a line is unique. Therefore, in our visualization projection, when we draw a line segment from point O_1 to point A, defined as line O_1A, there will be only one intersection A_1. And the Point A_1 is also the only one point of the arc axis and line O_1A. Likewise, we can define another intersection A_2. It is worth noting that defining the intersection A_1 as a projection of the vertex A from PCP to DACP in a straightforward approach. However, as there is an extension rate of the axis length from PCP to DACP, which is $\frac{\sqrt{2\pi}}{4}$, so we project A_1 to A_1'in the arc through utilizing the increment of the arc length. And we give

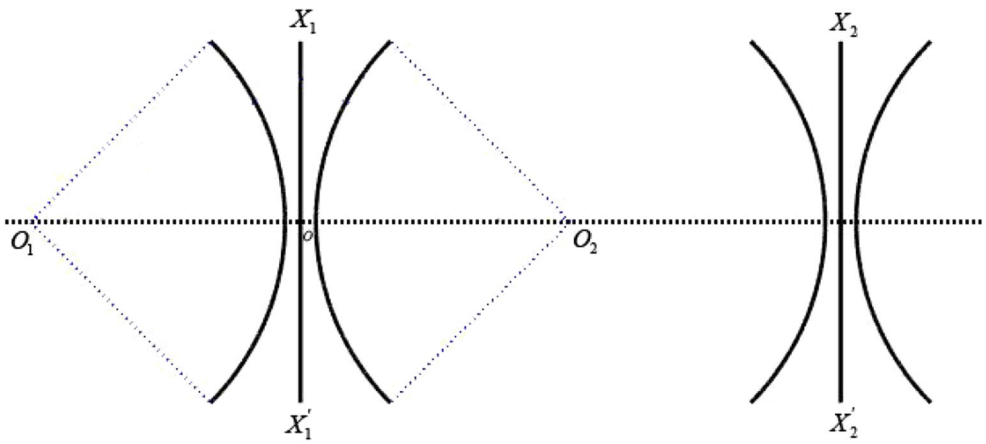


Fig. 2 The double arc coordinates system

details of the computation of increment of the arc length in the following paragraphs.

To simplify the computational complexity, we study the projection of vertices in the positive semi-axis OX_1 of the PCP and arc OM_1 in Fig. 3. In fact, the result of the negative semi-axis is the same as the result of the positive semi-axis. The slope of line O_1X_1 is 2/3, while the angle of OM_1 is exactly half of the right angle, $\pi/4$. The length of the arc is from $\frac{\sqrt{2}}{2}\arctan\frac{2}{3}$ to $\frac{\sqrt{2}\pi}{8}$. Its increment is $\frac{\pi}{4\arctan\frac{2}{3}}$. For all vertices in the positive semi-axis, consider this increment as our extension rate. In addition, due to the symmetry of the axes in PCP and DACP, we can refer to this extension rate as the negative semi-axis. The explanation for this is shown in Fig. 3.

To summarize, there are two steps to build the projection method when we project the point $\left(\frac{3}{2}i, y_0\right)$ in the $(i+1)-th$ PCP axis to the left axis of the DACP.

The main formula is below:

$$\mathrm{F}: \left(\tfrac{3}{2}i, y_0\right) \rightarrow \left(\frac{\cos\theta}{\sqrt{2}} + \tfrac{3}{2}i - \tfrac{3}{4}, \frac{\sin\theta}{\sqrt{2}}\right), \text{where } \theta = \frac{\pi\arctan\left(\frac{4}{3}y_0\right)}{4\arctan\frac{2}{3}}.$$

In the first step, we use the following nonlinear system to obtain the coordinates of the intersection between the line and the arc.

$$\begin{cases} y - y_0 = \dfrac{4}{3}y_0\left(x - \dfrac{3}{2}\mathrm{i}\right) \\ \left(x + \dfrac{3}{4} - \dfrac{3}{2}\mathrm{i}\right)^2 + y^2 = \dfrac{1}{2} \end{cases} \tag{1}$$

The $A_1\left(\frac{3\sqrt{2}}{2\sqrt{16y_0^2+9}} + \frac{3}{2}\mathrm{i} - \frac{3}{4}, \frac{2\sqrt{2}y_0}{\sqrt{16y_0^2+9}}\right)$ coordinates can be obtained from the above system.

In the second step, we use the extension rate $\frac{\pi}{4\arctan\frac{2}{3}}$ as our extension factor, and multiplies the arc length which starts from the point $\left(\frac{3}{2}i, 0\right)$ by the horizontal axis and ends with the intersection coordinates. We can receive the final projection vertices of the original point $\left(\frac{3}{2}i, y_0\right)$. To get the final coordinates, we must associate the arc length with the coordinate system. And we can have the following system:

$$\begin{cases} y_0\cot\theta = x_0 + \dfrac{3}{4} - \dfrac{3}{2}i \\ \left(x + \dfrac{3}{4} - \dfrac{3}{2}i\right)^2 + y^2 = \dfrac{1}{2} \end{cases} \tag{2}$$

Finally, we get the result of projection $\left(\frac{\cos\theta}{\sqrt{2}} + \frac{3}{2}i - \frac{3}{4}, \frac{\sin\theta}{\sqrt{2}}\right)$, where $\theta = \frac{\pi\arctan\left(\frac{4}{3}y_0\right)}{4\arctan\frac{2}{3}}$.

Because the left axis and the right axis of the pair of double-arc coordinates are on X_1X_1' symmetry. So for simplicity, we give the conclusion for the right axis directly as follow:

The following function projects point $\left(\frac{3}{2}i, y_0\right)$ in the $(i+1)-th$ PCP axis to the right axis of the DACP:

$$\mathrm{F}: \left(\tfrac{3}{2}i, y_0\right) \rightarrow \left(-\frac{\cos\theta}{\sqrt{2}} + \tfrac{3}{2}i - \tfrac{3}{4}, \frac{\sin\theta}{\sqrt{2}}\right), \text{where } \theta = \frac{\pi\arctan\left(\frac{4}{3}y_0\right)}{4\arctan\frac{2}{3}}.$$

3.2 Dimension-based bundling layout

3.2.1 Bundling layout

While dealing with large volume datasets, the conventional PCP inevitably generates over-plotting. Overlapping lines between two adjacent axes greatly reduce the visualization effect. We address this problem by using dimension-based bundling layout.

Bundling layout is an effective method for reducing the visual clutter caused by dense edges in parallel coordinates. Specifically, in our paper, we consider two neighboring pair of arc axes X_1 and X_2 and the area between them. We place a virtual binding axis on the right side of X_1 and place a virtual bundling axis on the left side of X_2, denoted as X_1' and X_2', refer to Fig. 1. The distance between a data axis and its binding axis is set to a parameter, and we keep it fixed to 10% of the radius, √2/20, for all screenshots in this paper.

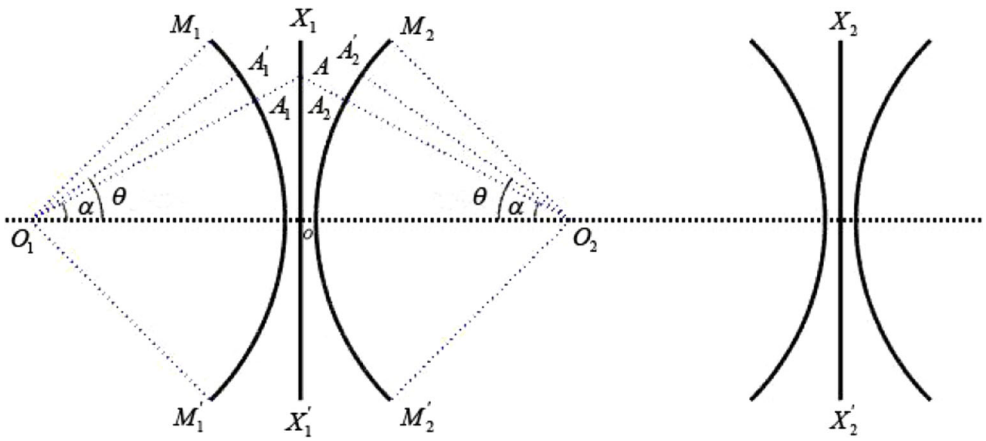


Fig. 3 The rationale of double arc coordinates plane

As we can see from Fig. 1, the area between two axes is segmented into three different parts, which contains $B_1B_1^{'}$, $B_1^{'}C_1^{'}$ and $C_1^{'}C_1$. We also divide arc axis into three sections at equal length, each color represents a part of axis. And there is a cooresponding part on the bundling axis,which is marked by the same color.

To geometrically represent the bundling axis coordinates, we recall the coordinates calculation on DACP axes from the previous section. They are $\left(\frac{\cos\theta}{\sqrt{2}}+\frac{3}{2}i-\frac{3}{4},\frac{\sin\theta}{\sqrt{2}}\right)$ and $\left(-\frac{\cos\theta}{\sqrt{2}}+\frac{3}{2}i+\frac{3}{4},\frac{\sin\theta}{\sqrt{2}}\right)$, where $\theta=\frac{\pi\arctan\left(\frac{4}{3}y_0\right)}{4\arctan\frac{2}{3}}$. Therefore, on the bundling axis, the coordinates can be computed with a different θ'.

Specifically, for the upper part, marked 1, when $\theta\in\left(\frac{\pi}{12},\frac{\pi}{4}\right)$, F: $\theta^{'}=\left(\theta-\frac{\pi}{12}\right)\times\frac{1}{5}+\frac{3\pi}{20}$. For the middle part, marked 2, when $\theta\in\left(-\frac{\pi}{12},\frac{\pi}{12}\right)$, F: $\theta^{'}=\left(\theta+\frac{\pi}{12}\right)\times\frac{1}{5}-\frac{\pi}{60}$. For the third part, marked 3, when $\theta\in\left(-\frac{\pi}{4},\frac{\pi}{12}\right)$, F; $\theta^{'}=\left(\theta+\frac{\pi}{4}\right)\times\frac{1}{5}-\frac{11\pi}{60}$.

And now the observation point (B_1,C_1) in Fig. 4 can be represented by three segments with more details rather than a straight line only.

3.2.2 Further optimization for bundling layout

The bundling layout has reduced the visual clutter by making the close line closer between adjacent pair of axes, but it actually increases the amounts of over-plotting within a bundle. To solve this problem, we optimize the bundling layout by filling the bundles with different color transparency.

For each bundle starts from the same part of the left axes, we count the number of lines and store it in matrix Z. $Z_{i,j}(i,j\in 1,2,3)$ represents the number of lines which runs from the i-th part to the j-th part. The transparency of each bundle is defined as following formula:

$$\alpha_{i,j}=\frac{Z_{i,j}}{\sum_{k=1}^{3}Z_{i,k}}$$

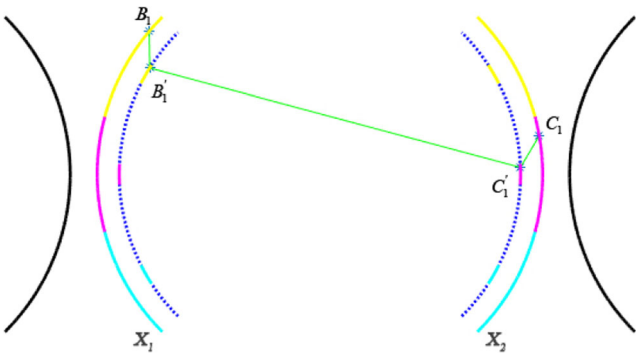


Fig. 4: Dimension-based bundling layout.

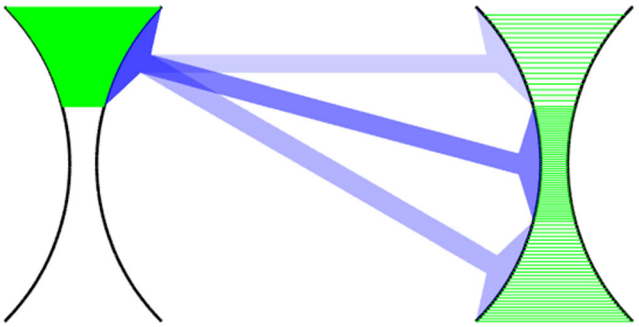


Fig. 5: Transparency-based filling on DACP.

Therefore, if the number of the lines is larger, the transparency level of the bundling is higher, and if the number of lines are smaller, the transparency level of the bundling is lower. For instance, as shown in Fig. 5. There are 100 items start from the upper part of the left pair of axes. There are 20 items end in the first part of the right pair of axes, 50 end in the second and 30 end in the third. And the transparency of the first bundle is 0.2, the second is 0.5, and 0.3 for the third bundle. We can see that there is no over-plotting of lines anymore. The visualization becomes clearer. Meanwhile, we can still know the distribution of data on the axes from the middle part of each pair of axes.

4 Axes re-ordering methods

Due to the clutter problems caused by the order of the axes, we propose axes re-ordering methods in our paper to reveal patterns which are hidden in the datasets. This method is to compute the importance score of attributes (dimensions) of the data using SVD and visualize the dimensions from left to right in DACP according the score in SVD, and we named as contribution-based method in DACP. In addition, to measure the correlation between two dimensions and explains how the two dimensions interact with each other, we propose another reordering method based on non-linear correlation information measurements, which is using mutual information to calculate the partial similarity of dimensions in high-dimensional data visualization.

To be more specific in mathematics, we consider a set of multidimensional data D with n dimensions (variables) and m items for each dimension. Some cases require measuring the statistical characters between the two dimensions X and Y, where$X=(x_1,x_2,\cdots,x_m)^T$, $Y=(y_1,y_2,\cdots,y_m)^T$.

4.1 Contribution-based re-ordering

Singular value decomposition(SVD) is the decomposition of a real matrix. It is a generalization of matrix to an m*n matrix by the extension of the polar decomposition. It has become a popular tool for revealing interesting and attractive algebraic

properties in matrix computation. SVD also plays a prominent role regarding to conveying important geometrical and theoretical insights about transformations. In this paper, we use SVD to measure the contribution of each dimension to the dataset.

The followings are the computation details and properties of SVD. Given a matrix *D with dimension m by n* . The SVD of matrix D, is $U\Sigma V^*$ [15]: where $V^*$is the conjugate transpose of *V*; *U* is a matrix with dimension m by m; V^* is a matrix with dimension n by n; $\Sigma$is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers (singular values of *D*), which aims for decreasing magnitude on the diagonal. There are many properties for SVD matrics, For example, the singular values of a matrix *D* are equal to the square roots of the eigenvalues λ1, λ2, …λm of the matrixD^TD. In this paper, we are impressed by a property that the first r columns of the matrices U and V form the orthonormal basis for the space spanned by columns and rows of *D*. As literature [17] mentioned, characteristic modes can be defined to reconstruct the gene expression patterns by this property. Therefore, we conclude the following property in perspective of the numerical properties for matrix:

Property: The entries of the first column of *V* in the singular value decomposition, which are denoted as v_{1j}, $j = 1, 2, \ldots, n$, show the contributions of columns of *D* to the space spanned by them, *i. e. span*$\{d_1, d_2, \cdots, d_n\}$,$d_i$ is the *ith* column of *D*.

And based on the above property, we propose a contribution-based reordering method. This method uses the entries of the column and uses DACP to visualize the dimensions of the dataset from left to right. On the other hand, considering the representation requirement of data value, this reordering method can provide us effective and clear visualization structure, it also helps us take deeper insight into the dataset. In addition, this method also brings us the idea of the determination of the first dimension with the most contribution. We introduce the details in section 4.2.

4.2 Similarity-based re-ordering

Measuring the correlation between two dimensions (variables/attributes) is a statistical technique. This technique not only represents the magnitude relationship between two dimensions, but also explains how the two dimensions interact with each other. In this section, we propose a reordering method based on non-linear correlation information measurements. Specifically, because mutual information measures how much one variable related with another, and can be thought of as a generalized correlation analogous to linear correlation coefficient. It is always sensitive to any variable relationships, not has an effect on linear correlation only. So we use it to analyze nonlinear correlation in DACP.

Statistically, suppose there is a two-dimensional dataset, x indicates the independent variable, y indicates the dependent variable, then the dataset can be represented as a collection {(xi, yi)|i = 1, 2, 3,. .., n}, where n indicates that there are n pairs data, xi indicates that the ith data of independent variable x, yi indicates the ith data of the dependent variable y, and you can use y = a + b x to represent the linear regression model if x and y are liner related, if x and y are mainly the nonlinear relationship, in this paper, we choose mutual information measures to analyze nonlinear corrlelation. And based on the theory of mutual information [20] and information redundancy [21], Nonlinear correlation coefficient is able to measure nonlinear relationship. In other words, this method can measure any relationships, not only be sensitive to the linear dependence [22]. Some researchers did further studies on its effects of statistical distribution and set it to a closed interval range [0,1], corresponding to the literatures [22, 23]. In this paper, refer to the literature [15], we mainly use NCC to calculate the partial similarity of dimensions in high-dimensional data visualization,

The detailed of NCC is introduced in the following paragraphs.

Mutual information is a critical element in NCC computation, it is denoted as:

$$I(X;Y) = H(X) + H(Y) - H(X;Y) \tag{4}$$

Where H(X)is the information entropy of variable X; H(Y) is the information entropy of variable Y.

$$H(X) = -\sum_{i=1}^{m} p_i \ln p_i$$

$$H(Y) = -\sum_{j=1}^{m} p_j \ln p_j$$

H(X; Y)is the joint entropy of the variables X and Y.

$$H(X;Y) = -\sum_{i=1}^{m}\sum_{j=1}^{m} p_{ij} \ln p_{ij} \ ]$$

Where

p_i is the probability distribution that random variable X takes the value x_i, and p_{ij} is the joint probability distribution $p(X = x_i, Y = y_i)$ of the discrete random variables X and Y.

Then, the revised value of joint entropy of variables X and Y is as formula (5) mentioned.

$$H^r(X;Y) = -\sum_{i=1}^{b}\sum_{j=1}^{b} \frac{n_{ij}}{m} \log_b \frac{n_{ij}}{m} \tag{5}$$

Where: the sample pairs $\{(x_1, y_i)\}_{1 \le i \le m}$ are placed in the $b \times b$ rank grids;

n_{ij} is the number of samples distributed in the ijth rank grid.

In addition, In literature [21], Wang et al. proposed using formula (6) for NCC:

$$NCC(X;Y) = H^r(X) + H^r(Y) - H^r(X;Y)$$

$$= 2 + \sum_{i=1}^{b}\sum_{j=1}^{b} \frac{n_{ij}}{m} \log_b \frac{n_{ij}}{m} \tag{6}$$

Considering the similarity between the problem of dimension reordering and Traveling Salesman problem (TSP), heuristic algorithms had been proposed to overcome exhaustive time. The methods include: genetic algorithms, colony optimization and nearest neighbor heuristic method, etc. [15, 16]. Specifically, In the method Similarity-Based Attribute Arrangement (SBAA), proposed process by A.O.Artero et al. [16]: Once algorithms are applying on a similarity matrix S for searching the largest values of s_{ij}, the two values i and j are considered to be the initial dimension "ij" in the new position of an parallel coordinate arrangement. And then, the algorithm will searches rows and columns of S to compute the similarity and position in the right of it. This method seems reasonable as they reorder the dimensions in line by their similarities. However, there are some dimensions that always get more attentions among the whole visual structures. And their special visual effects cannot be ignored. For example, in DACP system, the first and the last dimensions are more attractive comparing with other axes.

Therefore, differentiate from other proposed methods, we introduce a new dimension reordering algorithm [24] based on NCC and SVD algorithms. As we defined in literature [24], the similarity matrix s is a symmetric matrix which is shown as below:

$$s = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}$$

where

$s_{ij} = s_{ji}\ (i \neq j)$ which we calculate by NCC formula.

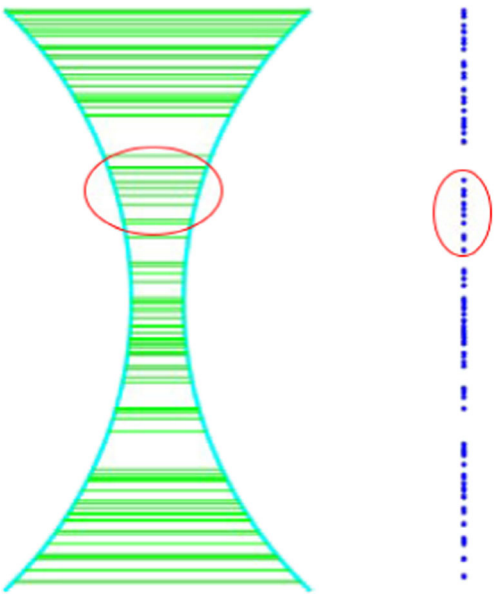
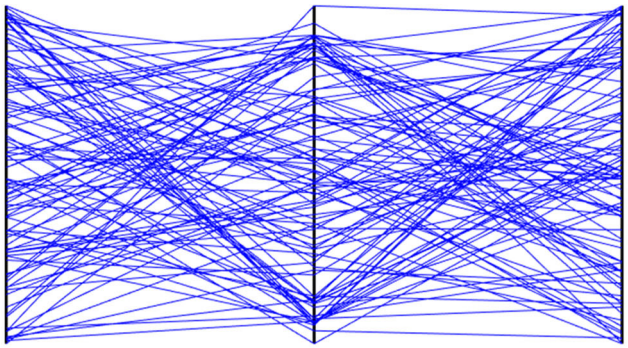
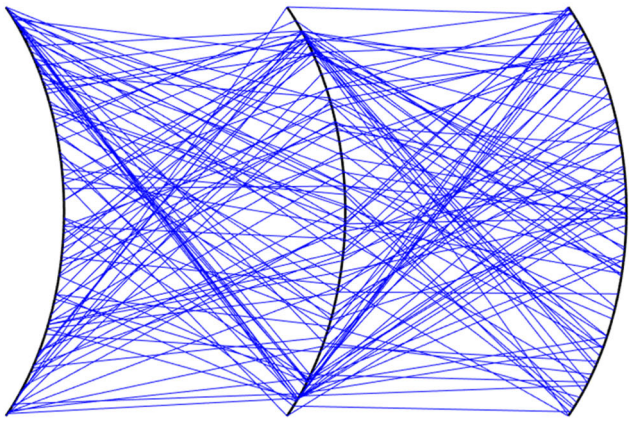


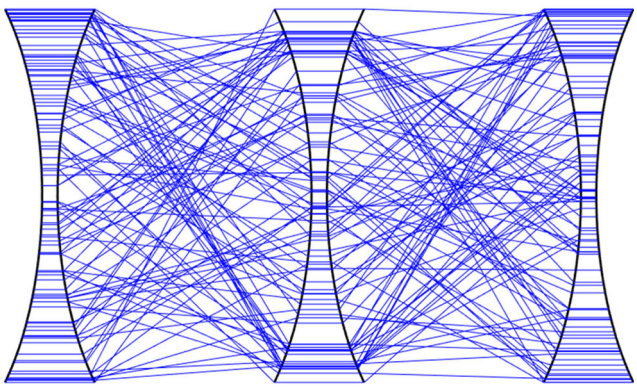
Fig. 6: Random data represented in two different coordinate systems.



(a) Random dataset represented in PCP.



(b) Random dataset represented in ACP.

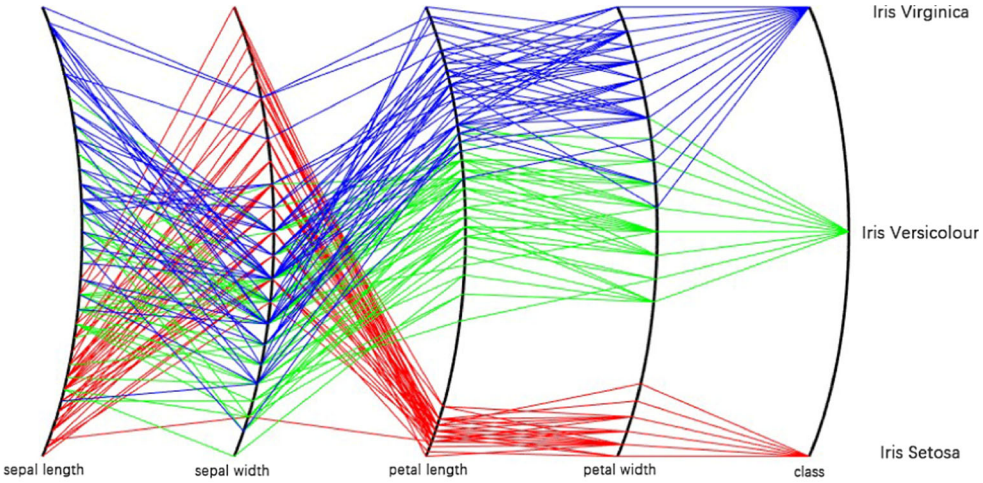


(c) Random dataset represented in DACP.

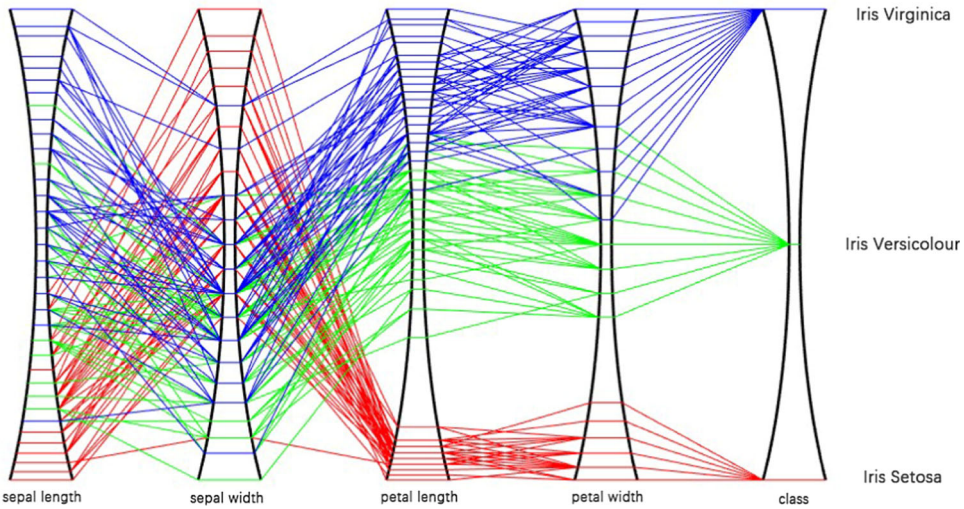
Fig. 7: Random dataset represented in three different visualization systems.

s_{ii} $(i = 1, 2, \cdots, n)$ (we also can denote them as v_{1i}) refers to the contribution value of the i-th dimension towards the whole data values, and we calculate it by SVD algorithm.

According to the similarity matrix s, we reorder dimensions of matrix D and visualize it with different visualization methods. The followings are the steps of Similarity-based Reordering Algorithm, which have been also illustrated in literature [24].



(a) Iris dataset visualized in ACP.



(b) Iris dataset visualized in DACP.

Fig. 8: Iris dataset visualized in ACP and DACP respectively

Step 1. Form the matrix D of the data sets.

Step 2. Calculate the singular value decomposition of matrix D, and get the contribution factors S_{ii}, 1,2, ….. n .

Step 3. Compute the other elements S_{ii} of similarity matrix s, using our nonlinear correlation coefficient method, besides S_{ii}, i n 1,2…..n, which have calculated in step2.

Step 4. Choose the largest value of S_{ii}, i n 1,2…..n, as the extreme left attribute to start display the data sets. We denote this attribute S_{II}, where I belongs to {1,2…..n},

Step 5. Get the largest value S_{II} from $\{S_{II}, I < i\}$`. Therefore, the r_ith attribute is appended to the lth attribute. We get the first two elements of neighbouring sequence NS{I,r_1}.

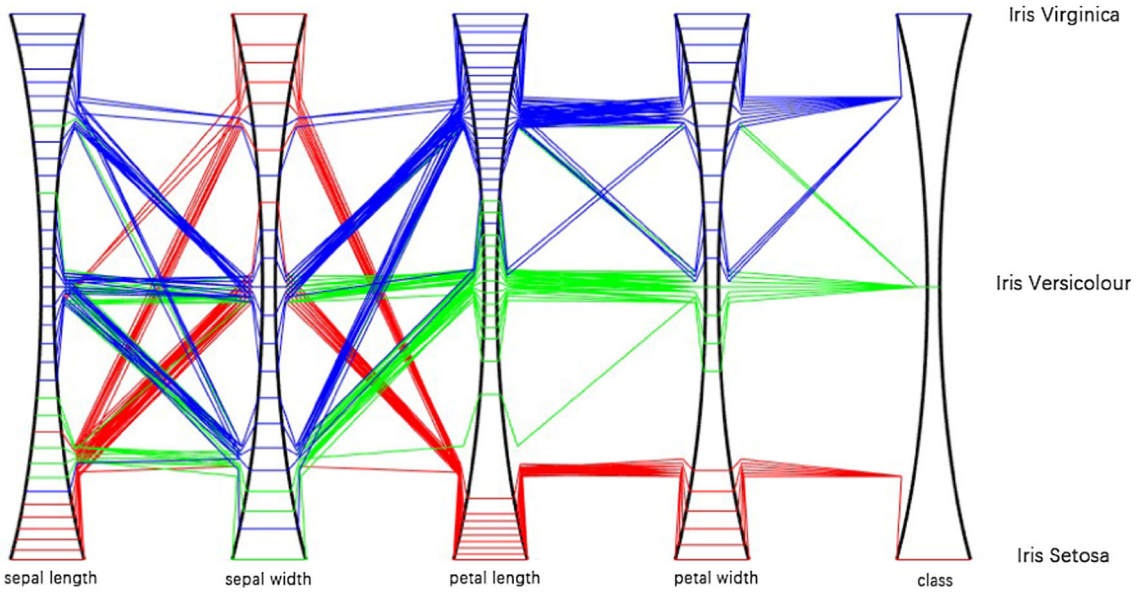
Step 6. Repeat step5 using the r_{1th} attribute as the left neighbouring attribute from $\{_{Sr1i}, r_1 < i\}$, until inserting all attributes into the NS .Our strategy is not only to provide users the dimension similarities between each pair of them, it also expresses some characteristics or patterns of each dimension itself. In the computation process of the NCC, we use b × b rank grids according to the empirical formula, which is mentioned in [25], where:

$$b = 1.87 \times (m-1)^{2/5} \tag{7}$$

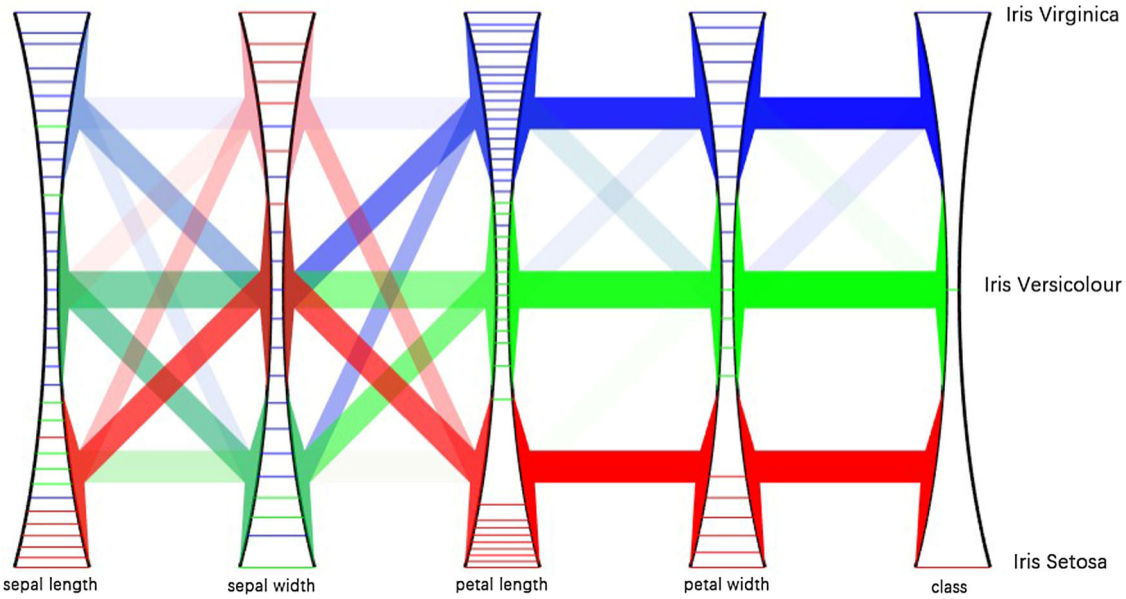
In the experiment section, we will apply this reordering method to our novel visualization method and show that how it works well with our approach and improves the visual readability greatly.

5 Application

We present case scenarios to demonstrate how DACP is effectively used to help experts understand multivariate



(a) Iris dataset visualized with dimension-based bundling layout



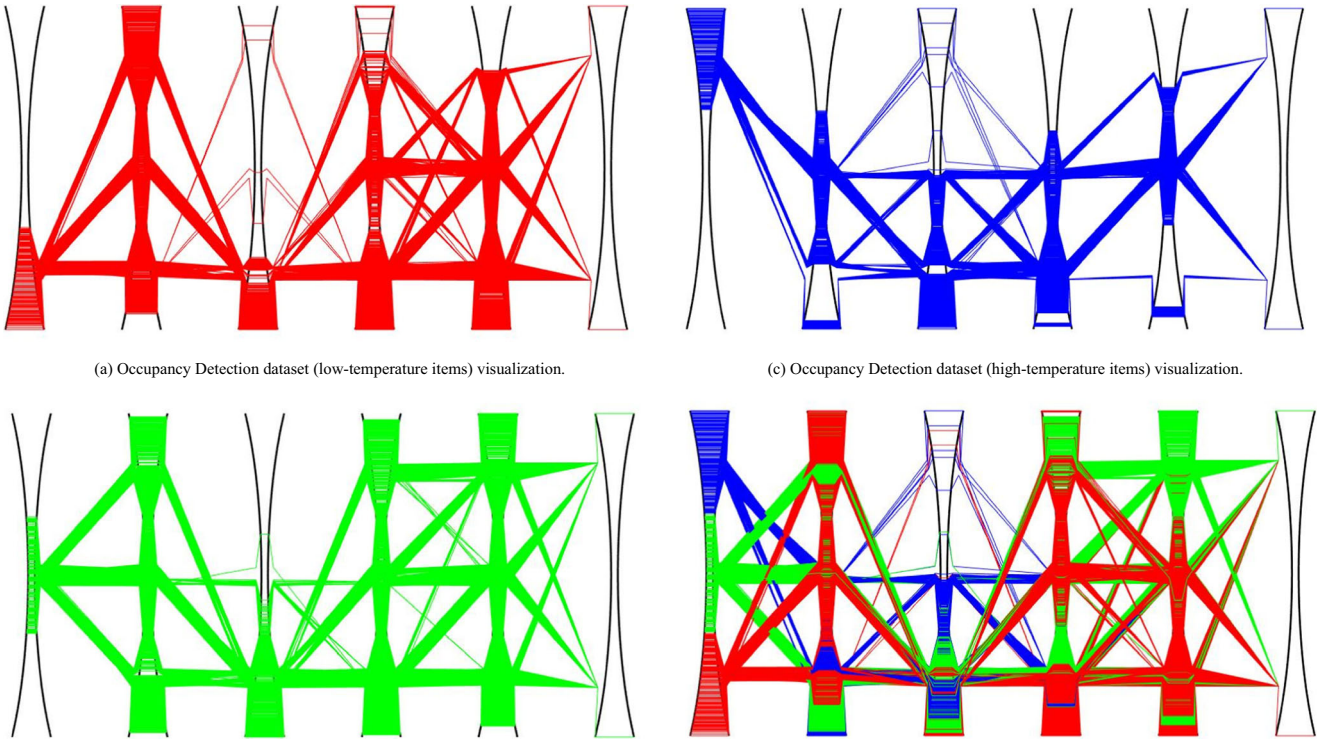
(b) Iris dataset visualized with transparency-based bundling layout.

Fig. 9: Iris dataset visualized with dimension-based bundling layout on DACP.

data and the effectiveness of our new dimension reordering methods. We test several different datasets in this section. Firstly, to illustrate the advantages of DACP comparing with ACP, we use Random dataset; Then to show dimension-based layout in DACP visualization, Iris dataset and Occupancy detection dataset are applied; Lastly, KDD Cup 1999 and Glass Identification dataset are used to test contribution-based and similarity-based reordering methods.

5.1 The comparison between DACP and ACP

We choose random datasets to display the comparisons. This dataset is about 100 data items with one attribute randomly, ranging from −0.5 to 0.5, and we visualize them in parallel coordinate plane. We project these data items by a mapping approach to a pair of axe in the double arc parallel coordinates plan. And it is clear to find that the density of the points in the double arc parallel coordinate plan is different from that is in



(a) Occupancy Detection dataset (low-temperature items) visualization.

(c) Occupancy Detection dataset (high-temperature items) visualization.

(b) Occupancy Detection dataset (middle-temperature items) visualization.

(d) Occupancy Detection dataset (all items) visualization.

Fig. 10: OD dataset visualized with dimension-based bundling layout on DACP.

the traditional parallel coordinate plan. Considering the information readability, the illustration of points by ellipse graph is sparser than the same illustration by PCP, see Fig. 6. In addition, the ellipse graph is also able to display the geometric property of the data.

Then, we use another 100 data items with three attributes randomly, also ranging from −0.5 to 0.5, and we visualize them by using PCP, ACP and our DACP, shown in Fig. 7. (a)Random dataset represented in PCP.

From the comparison, we can see that our double arc axes are not affecting the quality of visualization, they can provide the same visualization quality on datasets as the traditional vertical-line provided. Moreover, our double arc parallel coordinate method enlarges the mean density of points in the geometry and the distribution of items is displayed in the middle of each pair of axes. All these features improve the readability of visualization. (a)Random dataset represented in ACP. (b) Random dataset represented in DACP.

5.2 The dimension-based layout in DACP

In this section, we utilize Iris dataset and Occupancy Detection dataset to demonstrate the effectiveness of our dimension-based bundling layout in DACP in low-density and high-density datasets respectively. Both datasets come from “UCI Machine Learning Repository” [26].

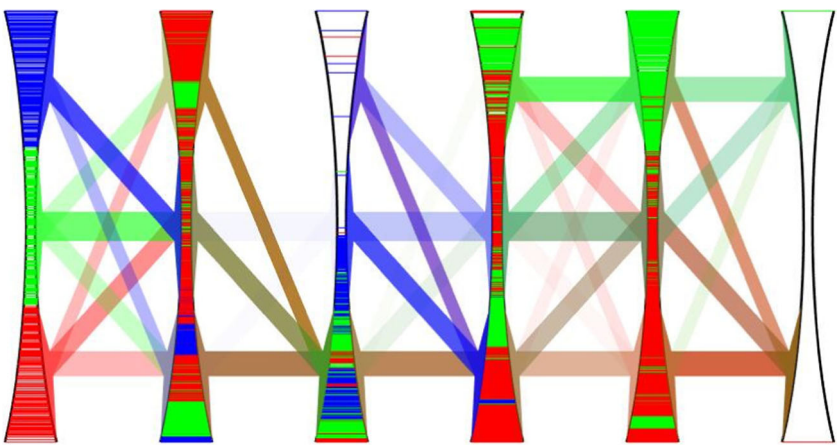


Fig. 11: OD dataset visualized with transparency-based bundling layout on DACP.

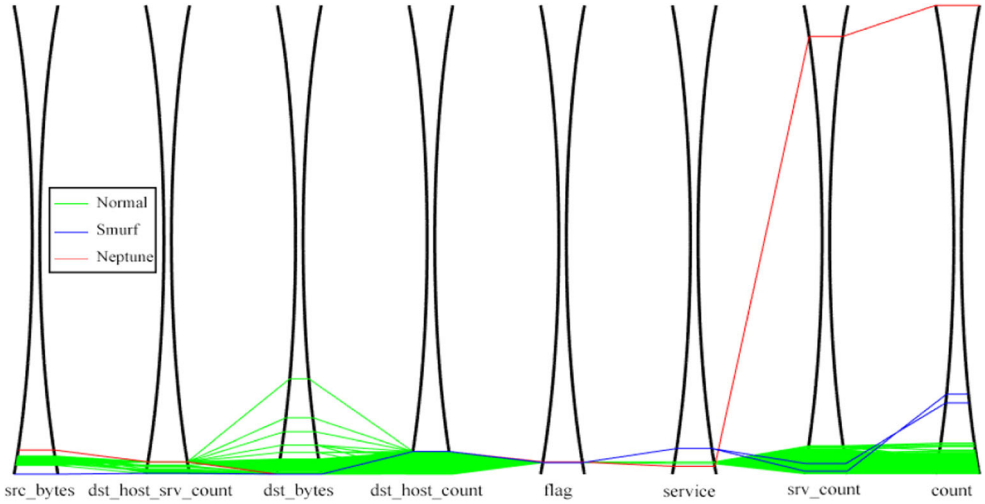


Fig. 12: Contribution-based reordering of KDD 1999 dataset in DACP

5.2.1 Iris dataset

The Iris dataset contains three varieties of Iris, each of them has 50 items, totally 150 items in this dataset. Every item has 4 features, we define the fifth feature as their variety, set 1 for *Iris setosa*, 2 for Iris Versicolour and 3 for *Iris virginica*. The visualizations are shown in Fig. 8 and Fig. 9. Fig. 8 shows the visualization with ACP and DACP respectively. Both of them represent dataset correctly. We can notice that the three clusters in the dataset are clearly represented in the new system as well. So it is to be concluded that our DACP method is able to perform the same as ACP when it comes to this common dataset. (a) Iris dataset visualized in ACP. (b) Iris dataset visualized in DACP.

In Fig. 9, we use our dimension-based bundling layout to visualize the dataset, it is clear in Fig. 9(a) that the lines with similar tendency have been bundled and over-plotting have been alleviated. In Fig. 9(b), the bundled lines are filled with different transparency. The over-plotting does not exist anymore and it is easy for us to observe the tendency in dataset. In addition, we can still know the position of observation points in every pair of axes due to our DACP method. This approach greatly reduces the visual clutter. (a) Iris dataset visualized with dimension-based bundling layout(b) Iris dataset visualized with transparency-based bundling layout.

5.2.2 Occupancy detection dataset

The Occupancy Detection dataset contains some environmental records in a room and also the data about whether the room is occupied or not. There are totally 20,560 items in this dataset, each of them has 7 features. First feature is the date of records, it is useless in our experiment, so we move it out from the dataset. Other features are temperature, relative humidity, light, CO2, humidity ratio and occupancy (0 for not occupied and 1 for occupied status). We use these features to analyze the relationship between environmental records and occupancy of the room.

To better analyze the dataset, we use temperature as classification. We divide temperature records into three parts, low, middle and high. In Fig. 10, we visualize each part of dataset and display them in Fig. 10(a), (b) and (c) respectively, their combination is shown in Fig. 10(d) as well. From these figures, we can see that when temperature is relatively high, relative humidity and CO2 are relatively low, because there are no data on the top of these axes. However, these records increase as temperature declines. Also, there is usually no person in the room when temperature is high while humidity ratio is low, nevertheless, if humidity ratio increases to a middle level, sometimes the room will be occupied. Fig. 11 shows the visualization of dataset by using transparency-based bundling layout. It is obvious that people are more likely to occupy the room when temperature is neither too high nor too low. (a) Occupancy Detection dataset (low-temperature items) visualization. (b) Occupancy Detection dataset (middle-temperature items) visualization. (c) Occupancy Detection dataset (high-temperature items) visualization. (d) Occupancy Detection dataset (all items) visualization.

5.3 Contribution-based reordering visualization

This section we aim to demonstrate the effectiveness of our contribution-based method. We choose two datasets to demonstrate. One dataset is selected from KDD Cup 1999 [27], and the other is Glass Identification dataset.

Firstly, as for the KDD cup 1999 dataset, it contains 1034 data items with 42 attributes, and also includes labels marked as “normal” and “abnormal”. We apply contribution-based reordering method to process the 42 attributes, and use this method as a dimension reduction process for dataset visualization. In this step, to retain as much data characteristics as we can, we set the contribution rate as one of the simplest techniques. And eight attributes are got to retain the 98.6% of the whole datasets. The visualization result is shown in Fig. 12. It

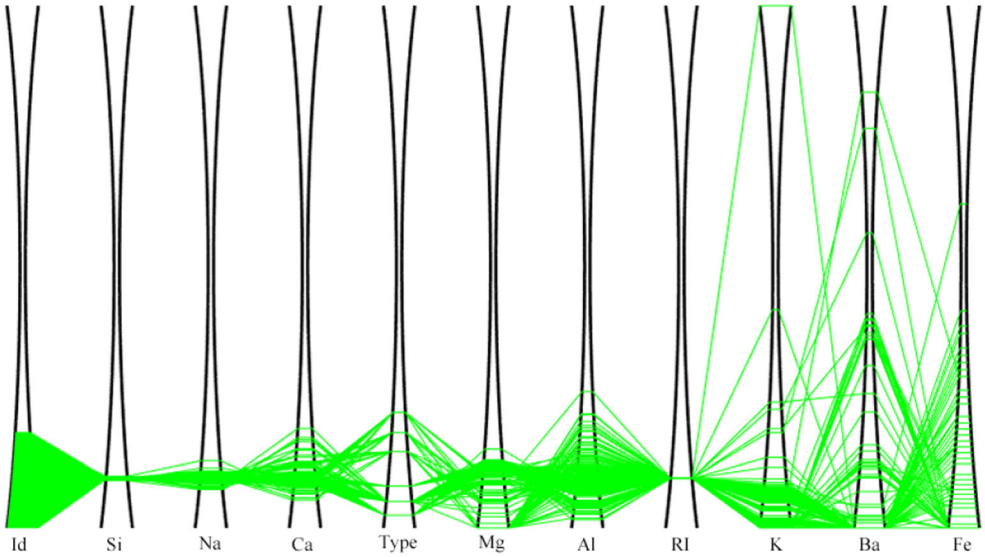


Fig. 13 Contribution-based reordering of Glass Identification dataset in DACP

is easily to discover that two abnormal events are existed in the dataset: one is called "Smurf" represented by red lines; and the other is "Neptune represented by blue lines. It is also to be noticed that some of the polylines among the attributes" srv_count" and "count" are strange. A big fluctuation is existed between the normal and abnormal polylines, and this provides us the pattern of attacks in the datasets.

Secondly, the Glass Identification dataset has 214 values and eleven dimensions. It is used to test our contribution-based reordering method. We compute the contribution of the dimensions by the property mentioned in Section 4. Refer to Fig. 13, the first dimension"Id" has the largest contribution factor 0.8723 (other contribution factors are listed as the diagonal elements in a matrix s of the next section). Dataset in line with their contribution to the values is visualized by DACP. From the data characteristic point of view, this visualization provides a clear description of the contribution order for all dimensions, which is from the highest rate to the lowest rate.

5.4 Similarity-based reordering visualization

This section describes the effectiveness of similarity-based reordering method. We mainly test this method on the Glass identification dataset to arrange dimensions.

Firstly, we visualize the reordered dataset with the conventional PCP and our DACP visualization methods. And then compare their visualization efficiency. As literature [28] mentioned, the relationship between the crossing angle among the polylines and the cognitive load is inversely proportional, but the relationship between the cognitive load and visualization efficiency is proportional. Therefore, to illustrate the benefits of our method from the readability and understandability, we calculate the mean angles among the polylines between two neighboring dimensions. The calculation formula is described below:

$$\text{mean\_angle} = \frac{\text{total\_angle}}{\text{total\_angle crossing}} \tag{8}$$

According to the theory in Section 4, the similarity matrix S of Glass dataset is calculated as below:

$$S = \begin{bmatrix}
0.8723 & 0.0023 & 0.0575 & 0.0709 & 0.0575 & 0.0064 & 0.0229 & 0.0064 & 0.4041 & 0.0926 & 0.5268 \\
0.0023 & 0.0099 & 0.0184 & 0.0021 & 0.0983 & 0.1573 & 0.0575 & 0.2158 & 0.3402 & 0.0935 & 0.1002 \\
0.0575 & 0.0184 & 0.0887 & 0.0074 & 0.0017 & 0.0788 & 0.1925 & 0.0337 & 0.3935 & 0.1098 & 0.1994 \\
0.0709 & 0.0021 & 0.0074 & 0.0150 & 0.0755 & 0.0217 & 0.0117 & 0.0669 & 0.4108 & 0.0906 & 0.2618 \\
0.0575 & 0.0983 & 0.0017 & 0.0755 & 0.0101 & 0.0184 & 0.0041 & 0.0338 & 0.4062 & 0.0883 & 0.1778 \\
0.0064 & 0.1573 & 0.0788 & 0.0217 & 0.0184 & 0.4762 & 0.0124 & 0.0281 & 0.3629 & 0.0920 & 0.0997 \\
0.0229 & 0.0575 & 0.1925 & 0.0117 & 0.0041 & 0.0124 & 0.0033 & 0.1034 & 0.3536 & 0.0881 & 0.1406 \\
0.0064 & 0.2158 & 0.0337 & 0.0669 & 0.0338 & 0.0281 & 0.1034 & 0.0590 & 0.332 & 0.0913 & 0.1359 \\
0.4041 & 0.3402 & 0.3935 & 0.4108 & 0.4062 & 0.3629 & 0.3536 & 0.3329 & 0.0018 & 0.4145 & 0.5575 \\
0.0926 & 0.0935 & 0.1098 & 0.0906 & 0.0883 & 0.0920 & 0.0881 & 0.0913 & 0.4145 & 0.0004 & 0.2160 \\
0.5268 & 0.1002 & 0.1994 & 0.2618 & 0.1778 & 0.0997 & 0.1406 & 0.1359 & 0.5575 & 0.2160 & 0.0232
\end{bmatrix}$$

Table 1 The comparison of mean angles of Glass dataset visualization in PCP and DACP

	Id-Type	Type-Ba	Ba-Fe	Fe-Na	Na-K	K-Ca	Ca-RI	Ri-Si	Si-Mg	Mg-Al	Overall
PCP	–	6.523°	13.249°	6.351°	2.49°	3.192°	0.263°	0.057°2.051°	2.051°	3.641°	2.7731°
DACP	–	9.524°	19.023°	8.911°	3.415°	4.476°	0.363°	0.078°	2.707°	2.707°4.926°	3.8332°

Based on the Similarity-based Reordering Algorithm [24], we first position the first dimension" Id number" as it plays a significant role to the whole dataset. Then we focus on finding out the most similar dimension with this one from the unordered dimensions. The target dimension must hold the largest similarity value to this dimension: $S_{1,11} = 0.5268$. In this case, the 11-th dimensions must be the strongest correlation with the 1st dimension. So we make the 11th dimension to be appended to the 1st one. Repeat the above processes until we put all the dimensions in order, which is $1 \rightarrow 11 \rightarrow 9 \rightarrow 10 \rightarrow 3 \rightarrow 7 \rightarrow$

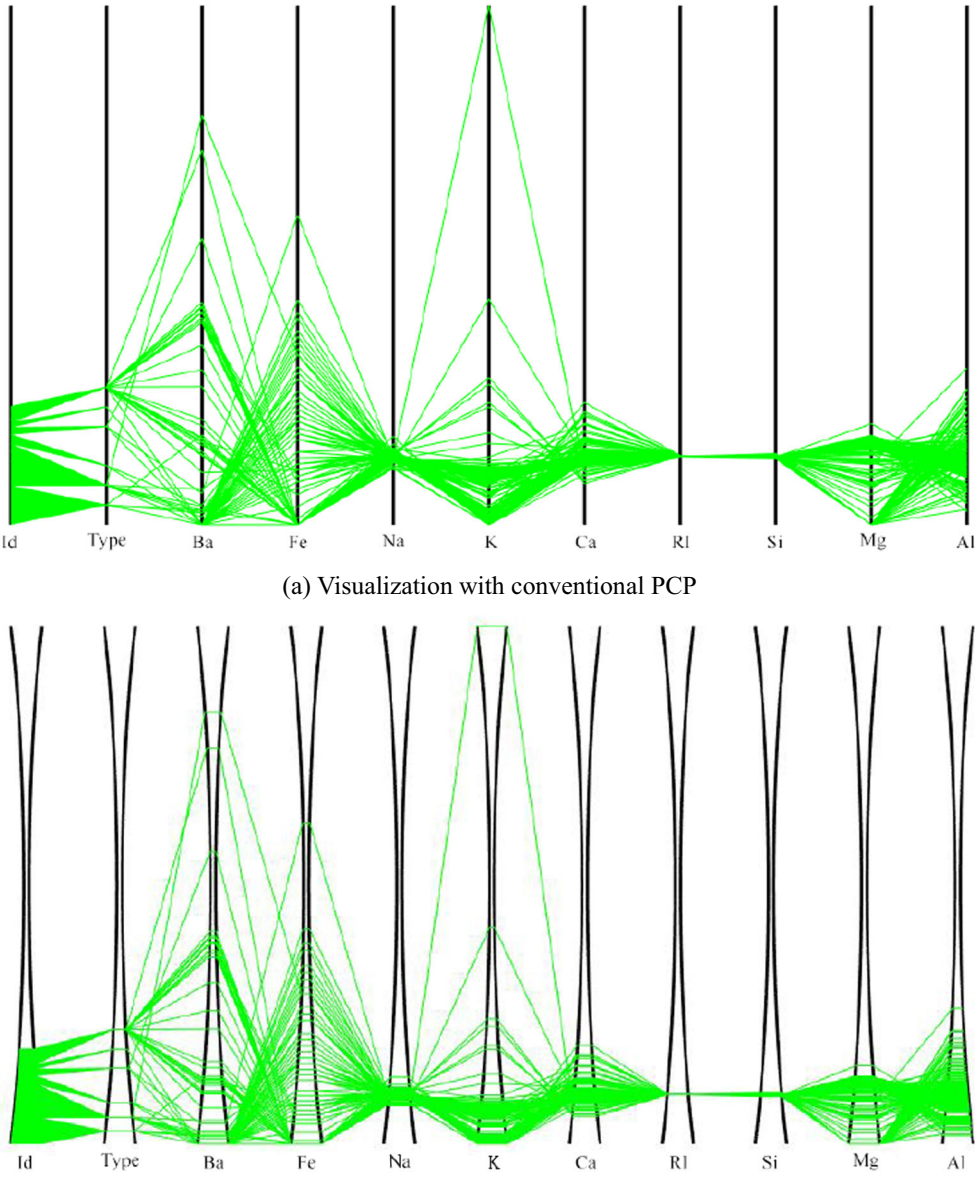


Fig. 14 Dimension reordering visualization of Glass dataset

(a) Visualization with conventional PCP

(b) Visualization with DACP

8 → 2 → 6 → 4 → 5. Related to the original dataset, the reordering dimensions order by our DAPC is: Id number→ Type→Ba→Fe → Na → K → Ca → RI → Si → Mg → Al.

Figure 14 (a) and (b) show the reordering results in conventional PCP and DACP respectively. (a) Visualization with conventional PCP (b) Visualization with DACP

Comparing Fig. 14(b) with Fig. 13, we discover that the visualization structure between the second attribute "Type" and the third attribute" Ba" is much clearer with our similarity-based reordering method.

To evaluate the improvement of visualization efficiency, we calculate mean angles between every two neighboring axes in conventional PCP and DACP, and displayed the results in Table 1. From Table 1, we can find that all the mean angles become larger in DACP than in PCP. For instance, the mean angle between attributes "Ba" and "Fe" gets to 19.023°, which is 5.774° larger than it in PCP. And the mean angle of overall polylines produced in PCP is 2.7731°, while the same mean angle produced by DACP turns to 3.8332°, which is 1.1 times larger than the former.

Therefore, we can conclude that the visual effect of our visualization method is much better than the traditional ones.

6 Conclusion and future work

In this paper, we present a new method for improving the parallel axes in coordinate's plane theoretically. Firstly, we propose DACP, the double arc coordinate plots, which is an arc-based parallel coordinate visualization method. Due to the length of an arc is longer than a line segment, the density of data displayed on each axes can be reduced. Besides this, because there are two arc axes for each pair of axes, the distribution of items can also be displayed in the middle of each pair of axes. So the visualization effect of the parallel coordinate plots is improved. Furthermore, we propose a dimension-based bundling layout to reduce the visual clutter and also fill the bundled lines with different transparency to optimize the bundling method further. Secondly, we propose contribution-based and similarity-based dimension re-ordering methods to find optimal dimension order to display dataset in DACP. Lastly, our evaluation, including five case scenarios, demonstrated the effectiveness and rationale of using our approaches to understand and discover more information from the datasets.

For future work, new ways of strengthening dimension-based bundling layout is our next task. We plan to optimize the classification of data items by using some cluster method rather than artificial approach. Moreover, we consider to apply interaction techniques on our approach for improving this visualization system.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China under No.51877144.

References

1. Hansen CD, Johnson CR (2005) The visualization handbook[M]. Elsevier Butterworth-Heinemann
2. Inselberg A (1985) The plane with parallel coordinates[J]. Vis Comput 1(2):69–91
3. Nguyen H, Rosen P (2018) DSPCP: a data scalable approach for identifying relationships in parallel coordinates. IEEE Trans Vis Comput Graph 24(3):1301–1315
4. Claessen JHT, Van Wijk JJ (2011) Flexible linked axes for multivariate data visualization[J]. IEEE Trans Vis Comput Graph 17(12): 2310–2316
5. Wilkinson L (2018) Visualizing Big Data Outliers through Distributed Aggregation. IEEE Transactions on Visualization & Computer Graphics, (1), pp.1–1
6. Huang ML, Lu LF, Zhang X (2015) Using arced axes in parallel coordinates geometry for high dimensional BigData visual analytics in cloud computing[J]. Computing 97(4):425–437
7. Wegman EJ (1990) Hyperdimensional data analysis using parallel coordinates[J]. J Am Stat Assoc 85(411):664–675
8. Heinrich J, Weiskopf D (2013) State of the art of parallel coordinates[J]. Eurographics 34(1):17–25
9. Dasgupta A, Kosara R (2010) Pargnostics: screen-space metrics for parallel coordinates[J]. IEEE Trans Vis Comput Graph 16(6):1017–1026
10. Huh MH, Park DY (2008) Enhancing parallel coordinate plots[J]. Journal of the Korean Statistical Society 37(2):129–133
11. Zhou H, Yuan X, Qu H et al (2008) Visual clustering in parallel coordinates[C]//computer graphics forum. Blackwell Publishing Ltd 27(3):1047–1054
12. Tominski C, Abello J, Schumann H (2004) Axes-based visualizations with radial layouts[C]// ACM Symposium on Applied Computing. DBLP, 1242–1247
13. Hauser H, Ledermann F, Doleisch H (2002) Angular brushing of extended parallel coordinates. *Proceedings of IEEE Symposium on Information Visualization,* 2002. *INFOVIS* 127–130
14. Peng W, Ward MO, Rundensteiner EA (2004) Clutter reduction in multi-dimensional data visualization using dimension reordering[C]//Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on. IEEE, 89–96
15. Ankerst M, Berchtold S, Keim DA. (1998) Similarity clustering of dimensions for an enhanced visualization of multidimensional data[C]// Information Visualization, 1998. Proceedings. IEEE Symposium on. IEEE, 52–60, 153
16. Artero A O, de Oliveira M C F, Levkowitz H. Enhanced high dimensional data visualization through dimension reduction and attribute arrangement[C]//Information Visualization, 2006.IV 2006. Tenth International Conference on. IEEE, 2006: 707–712
17. Dasgupta A, Kosara R (2010) Pargnostics: screen-space metrics for parallel coordinates[J]. IEEE Trans Vis Comput Graph 16(6):1017–1026

18. Artero AO, de Oliveira MCF, Levkowitz H (2004) Uncovering clusters in crowded parallel coordinates visualizations[C]// Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On. IEEE, 81–88
19. Yuan X, Guo P, Xiao H et al (2009) Scattering points in parallel coordinates[J]. IEEE Trans Vis Comput Graph 15(6):1001–1008
20. Matsuda H (2000) Physical nature of higher-order mutual information: intrinsic correlations and frustration[J]. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 62(3 Pt A):3096–3102
21. Drmota M, Szpankowski W (2004) Precise minimax redundancy and regret[J]. Information Theory IEEE Transactions on 50(11): 2686–2707
22. Shen Z, Wang Q, Shen Y (2011) Effects of statistical distribution on nonlinear correlation coefficient[C]// Instrumentation and Measurement Technology Conference. IEEE, 1–4
23. Wang Q, Shen Y, Zhang JQ (2005) A nonlinear correlation measure for multivariable data set[J]. Physica D Nonlinear Phenomena 200(3–4):287–295
24. Lu LF, Huang ML, Zhang J (2016) Visual pattern mining using parallel coordinates plot. J Vis Lang Comput 33:3–12
25. W.Y.S., Zhuang Chu Qiang, Mathematical Statistics with Applications, South China Science and Technology University Press, Guangzhou, 1992
26. University of California, Irvine. Center for Machine Learning and Intelligent Systems: http://archive.ics.uci.edu/ml/data%20sets.html
27. KDD cup 1999 data [EB/OL], http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html 2015-6-15
28. Huang W, Huang M (2011) Exploring the relative importance of number of edge crossings and size of crossing angles: a quantitative perspective[J]. International Journal of Advanced Intelligence 3(1): 25–42