Neurocomputing 74 (2011) 2018–2027

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

PPoSOM: A new variant of PolSOM by using probabilistic assignment for multidimensional data visualization

Yang Xu, Lu Xu, Tommy W.S. Chow*

Department of Electronic Engineering, City University of Hong Kong, Hong Kong

ARTICLE INFO

Available online 4 March 2011

Keywords:
SOM
PolSOM
Probabilistic Polar SOM (PPoSOM)
Data visualization

ABSTRACT

A new Self-Organizing Map algorithm, called the probabilistic polar self-organizing map (PPoSOM), is proposed. PPoSOM is a new variant of PolSOM, which is constructed on 2-D polar coordinates. Two variables: radius and angle are used to reflect the data characteristics. PPoSOM, developed to enhance the visualization performance, provides more data characteristics compared with the traditional methods that use Euclidian distance as the only variable. The weight-updating rule of PPoSOM is associated with a cost function. Instead of using the hard assignment, PPoSOM employs the soft assignment that the assignment of data to neuron is based on a probabilistic function. The obtained results are compared with the conventional SOM and ViSOM. The presented results show that the proposed PPoSOM is an effective method for multidimensional data visualization. In addition, the quality measurement of mapping, synthetical cluster density (SCD) is applied and it shows PPoSOM exhibits an improved result compared with PolSOM.

1. Introduction

Data visualization is a graphical presentation of multidimensional data. It has been widely used in many applications such as signal compression, pattern recognition, image processing, etc. Data visualization aims at producing a simplified graphical display, usually in 2 dimensions, that makes the perceiving of high dimensional data possible and easy. Compared with the data appear in their original high dimensional space, we may be more capable of finding the possible relationship among the data points in the low dimensional space, i.e., 2-D. Principal component analysis (PCA) [1] and multidimensional scaling (MDS) [2] are two classical methods for data reduction and visualization. PCA is one of the most widely used methods because it is effective and robust for performing linear projection [3,4]. Linear projection means the projection of data is conducted by multiplying each component of the original vector with a scalar. Thus, PCA is not the most suitable approach when one is dealing with highly nonlinear data [5,6]. MDS is another classical projection but its final visualization map is difficult to perceive, when one is handling high-dimensional and highly unsymmetrical data set. Sammon's mapping [7] is the earliest approach in nonlinear projecting data into low dimensional space for visualizing multivariate data. Sammon's mapping tries to minimize the distances between input data in the original high dimensional space and the output data in the projected space. It is capable of preserving the topological structure of the original data and their corresponding inter-point distances, but Sammon's mapping is computationally demanding especially when one is handling huge numbers of data points. In addition, it requires re-computation when new data points are added [8].

Self-organizing map (SOM) [9–11] is an unsupervised learning neural network to visualize high-dimensional data in a low-dimensional map. SOM forms a 2 dimensional map by mapping the input space onto the output space. The SOM map is capable of displaying the data topology by assigning each datum to a neuron with the highest similarity. SOM is computationally effective because its updating rule is based on a simple winner-take-all algorithm. All these characteristics have made SOM applicable to many physical problems. Despite all these advantages, conventional SOM is usually defined under a uniform grid map making the output map relatively rigid for preserving the data relationship among clusters or within one cluster. In addition, the requirement of pre-defining the map size is another shortcoming of SOM. As SOM does not preserve the inter-neuron distances on the map, some coloring schemes such as U-matrix [12,13] and interpolation [14], were proposed to imprint the inter-neuron distances. The clusters and boundaries of U-matrix can be marked as a result. In U-matrix, the data structure is visualized by distance matrix, in which large distances stand for borders and small distances represent clusters. The coloring scheme enhances the visualization effect, but the data structure and distribution often appear in distorted forms.

Since Kohonen derived the classical SOM, many modified SOM [15–23] have been proposed. Visualization-induced SOM (ViSOM) [15–18], a relatively new variant of SOM, is aimed at improving

* Corresponding author. Tel.: +852 2788 7756; fax: +852 2788 7791.
E-mail address: eetchow@cityu.edu.hk (T.W.S. Chow).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.028

the visualization of SOM. ViSOM regularizes the Euclidian distance of any pair of neurons, which is proportional to the distance of corresponding data in the input space. It uses a parameter, derived from the inter-neuron distance, to define and control the resolution of the output map. Thus, ViSOM is able to preserve the data topology as well as the inter-neuron distances. It gives an outstanding multidimensional data visualization. But ViSOM is based on heuristic and no cost function is assigned to it. Probabilistic Regularized SOM (PRSOM) [19] is another recently developed SOM. Instead of hard assignment by ViSOM, PRSOM employs a soft assignment approach and extends the weight-updating rule from ViSOM. PRSOM hybridizes SOM and MDS in order to reduce the computational burden and preserve the inter-neuron distances. When one compares ViSOM with PRSOM, ViSOM can be considered a special case of PRSOM. Simulation results presented in [19] demonstrate that PRSOM is an effective algorithm for data visualization. But the disadvantage of both ViSOM and PRSOM lies in the fact that certain data points may be projected onto one neuron, which results in blurring the relationship among the data represented by the same neuron.

In this paper, a new algorithm, namely Probabilistic Polar SOM (PPoSOM), is proposed for data visualization. It is derived from the concept of a new visualization map PolSOM [20] which is designed to visualize data in a 2-D polar coordinates map. PPoSOM is generated with the introduction of probabilistic assignment to form a cost function which results in the principled weight-updating rule. Instead of Cartesian coordinates, PPoSOM visualizes data in a 2-D polar coordinates map with two variables: radius and angle. These two variables represent data weight and feature, respectively. The neurons learn data feature by a probabilistic data assignment method, and the projected data points are updated to approach their winning neurons. As a result, the data topology as well as the inter-data distances is preserved. Simulation results and the comparisons with SOM and ViSOM show that PPoSOM exhibits remarkable performance on data visualization. In addition, it is worth noting that the introduction of the probabilistic function made the PPoSOM visualization performance, measured by synthetical cluster density (SCD) [20], markedly increase compared with those without the probabilistic mechanism.

This paper is organized as follows. Section 2 presents an overview of PolSOM, ViSOM and a mapping quality measurement named synthetical cluster density (SCD). Section 3 details the principle of PPoSOM. In Section 4, six data sets are used to demonstrate the abilities of PPoSOM and other visualization algorithms. The detailed comparisons and analyses are given. The conclusion is presented in Section 5.

2. Background

2.1. Polar self-organizing map (PolSOM)

The traditional algorithms such as SOM and ViSOM, are only capable of clustering, but fall short in exhibiting the data characteristics and the differences among clusters. Polar SOM (PolSOM) [20] is designed to provide a new kind of visualization and overcome the above problem. Instead of using Cartesian coordinate, PolSOM is constructed on 2-D polar coordinates. The projected data points on the map are expressed by two variables: angle and radius, representing the data feature and weight, respectively.

The PolSOM map is evenly divided into n angles and p tori, where n is the dimensionality of data and p is the number of parts from center to maximum data weight. Each angle stands for an attribute of the data feature and different torus represents different data weight value. The $n \cdot p$ neurons are set on the map as benchmarks of data characteristics. Their weight initializations are determined by their positions in a way that the significant attribute of a neuron feature is represented by its angle and the weight value is determined by its radius.

Define $N_r(i)$ and $N_\alpha(i)$ as neuron i's radius and angle, respectively, and r_x and α_x as an output datum x's radius and angle, respectively. During each training process, a datum $x_j=(x_{j1},x_{j2},\ldots,x_{jn})^T$ is randomly chosen from the input space. The winning neuron c is found according to $c=\text{argmin}_i\|x-w_i\|$, $i\in\{1,\ldots,n\cdot p\}$. Update the weights of neuron c and its neighborhood set N_c by

$$w_i(t+1)=\begin{cases} w_i(t)+\eta_1(t)(x(t)-w_i(t)) & \text{for} \quad i=c, \\ w_i(t)+\eta_2(t)(x(t)-w_i(t)) & \text{for} \quad i\in N_c. \end{cases} \tag{1}$$

The polar coordinate of the corresponding selected datum is updated according to

$$\begin{cases} r_x(t+1)=r_x(t)+\beta_1(t)(N_r(c)-r_x(t)), \\ a_x(t+1)=a_x(t)+\beta_2(t)(N_a(c)-\alpha_x(t)), \end{cases} \tag{2}$$

where η_1, η_2, β_1 and β_2 are the learning rates that monotonically decrease with time.

Upon the completion of training process, the visualization map is created so that each input datum is represented by its radius and angle on the polar map. Data with similar features are grouped together, and their characteristics are also reflected by their positions. Because the data are not projected onto neurons, the disadvantage of SOM and ViSOM mentioned above is overcome so that the data relationships among different clusters and within a single cluster are well preserved. Since there are the benchmark neurons representing data characteristics, PolSOM does not require re-computation when new instances are added. This characteristic facilitates the incremental training process in most practical applications in which one usually needs to update the map with new data.

In the following, we demonstrate the property of topology preservation of PolSOM. We use the 2009 Times Higher Education World's University Rankings [34] because it is easy for us to illustrate the topology preservation property through comparing the characteristics of different well-known universities. More detailed analyses are described in Section 4. The visualization result is shown in Fig. 1. Note that the top 10 universities, i.e., Harvard University, University of Cambridge, Yale University and Imperial College London are located with large radii and around the angular 0°. The similar length of radii implies that these universities carry the similar overall high scores or the similar magnitudes of the 6-dimensional vectors. They all located around the angular 0° that

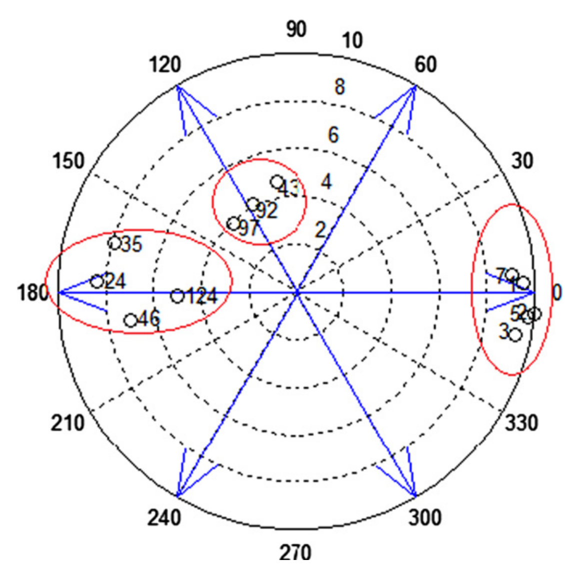

Fig. 1. The topology preservation property of PolSOM.

Table 1
Part of the 2009 Times High Education World's University Rankings data set.

Rank	Name	Peer review	Employer review	Student/faculty	International faculty	International student	Citation/faculty
1	Harvard	100	100	98	85	78	100
2	CAM	100	100	100	98	96	89
3	Yale	100	99	100	85	77	94
5	Oxford	100	100	100	96	97	80
7	Chicago	100	99	97	77	83	88
24	HKU	96	89	87	100	95	56
35	HKUST	89	86	84	100	99	54
43	Osaka	92	73	90	24	33	68
46	CUHK	87	77	77	97	79	55
92	Nagoya	67	77	89	28	34	61
97	Tohoku	67	58	98	36	32	58
124	CityU	66	58	63	100	60	48

means they all have the same significant attribute (peer review). In the PolSOM output map, universities with relatively low overall scores are clustered with relatively small radii. In Table 1, it is clear that the four universities of Hong Kong (Rank 24, 35, 46, 124) and the three universities of Japan (Rank 43, 92, 97) have relative high scores in the attribute of International Faculty and Student/Faculty, respectively. Their different strengths are displayed in different angular coordinates of the output map as shown in Fig. 1, which reflects that the universities from the same country/city may have similar characteristics. The visualization result depicted in Fig. 1 demonstrates that PolSOM is able to preserve the topological nature of the input data. It not only groups the similar data together, it also reveals the particular strength of the cluster in a way of arranging them in a particular angular coordinate.

2.2. Visualization-induced SOM (ViSOM)

ViSOM [15–18] is proposed to preserve the inter-neuron distances as well as the data topology. ViSOM uses the similar map structure as SOM, but the training of winning neuron's neighbors is different. The weight updating formula of ViSOM is defined by

$$w_i(t+1)=w_i(t)+\varepsilon(t)h_{ic}(t)\left(\left(x(t)-w_c(t)+(w_c(t)-w_i(t))\frac{d_{ci}-\Delta_{ci}\lambda}{\Delta_{ci}\lambda}\right)\right),\quad i\in N_c, \tag{3}$$

where d_{ci} and Δ_{ci} are the distances between nodes c and i in the input space and output space, respectively. $\varepsilon(t)$ is the learning rate that decrease monotonically with time; $h_{ic}(t)$ is the neighborhood function of winning neuron c. λ is a positive pre-specified resolution parameter.

ViSOM decomposes the updating force $F_{ix}=x(t)-w_i(t)$ into two forces: $F_{ix}=(x(t)-w_c(t))+(w_c(t)-w_i(t))=F_{cx}+F_{ci}$. F_{cx} is the updating force from the winning neuron c to the input data x; F_{ci} is a lateral contraction force bringing neighboring neuron i to the winner c. The second force regularizes the inter-neuron distance in the output space to resemble that in the input space. ViSOM delivers an excellent visualization result. However, ViSOM has a drawback that certain input data points are mapped on the same neuron, making the relationship of these data difficult or even impossible to be preserved.

2.3. Quality measurement of mapping

In order to compare the visualization performance of PolSOM and the proposed method PPoSOM, Synthetical Cluster Density (SCD) proposed in [20] is employed. SCD demonstrates the combination of the intra cluster density which shows the data compactness within one cluster, and the inter cluster density which represents the separation between different clusters. Assuming a data set has c clusters, in which ith cluster is denoted as $X_i=\{x_1,x_2,\ldots,x_{p_i}\}$, where p_i is the number of data in the ith cluster.

The intra cluster density is high when the cluster is compact. The intra density of ith cluster is defined as

$$Intra\_density(i)=\frac{1}{p_i}\sum_{k=1}^{p_i} f(x_k)/std_r^i,\quad x_k\in X_i, \tag{4}$$

where $f(x_k)=\begin{cases}1 & \text{if } \|x_k-\overline{X_i}\|\le std_d^i\\ 0 & \text{otherwise}\end{cases}$; std_d^i and std_r^i are the standard deviations of data and radii in the ith cluster, respectively, and $\overline{X_i}$ is the mean weight of the ith cluster. The inter cluster density is low when two clusters are well separated. Its definition between ith and jth cluster is given by

$$Inter\_density(i,j)=\frac{1+\sum_{k=1}^{p_i+p_j} g(x_k)}{(h(r_i,r_j)+1)(p_i+p_j)},\quad x_k\in\{X_i,X_j\}, \tag{5}$$

where $g(x_k)=\begin{cases}1 & \text{if } \left\|x_k-\frac{\overline{X_i}+\overline{X_j}}{2}\right\|\le\frac{std_d^i+std_d^j}{2},\\ 0 & \text{otherwise,}\end{cases}$

$$h(r_i,r_j)=\begin{cases}|\overline{r_i}-\overline{r_j}| & \text{if } \left[\overline{r_i}-std_r^i,\overline{r_i}+std_r^i\right]\cap\left[\overline{r_j}-std_r^j,\overline{r_j}+std_r^j\right]=\emptyset\\ 0 & \text{otherwise}\end{cases}$$

and $\overline{r_i}$ is the average radius of the ith cluster.

Overall, the SCD of a data set with c clusters is defined by

$$SCD=\sum_{i=1}^{c} Intra\_density(i)\Big/\sum_{a=1}^{c}\sum_{\substack{b=1\\ b\ne a}}^{c} Inter\_density(a,b). \tag{6}$$

A larger SCD value implies a better clustering performance for the given data set.

3. Probabilistic polar self-organizing map (PPoSOM)

In this section, a probabilistic PolSOM (PPoSOM) algorithm is introduced. The PPoSOM extends the PolSOM algorithm by using a soft assignment, which assigns an input datum to a neuron with a certain probability. A subsequent weight updating rule is derived by optimizing the cost function. It is noted that the nature of soft assignment on SOM has been widely studied. These include the investigation on convergence property [24]. More work and theoretical analyses on using soft assignment can be found in [19,25,26,29–31]. The soft topographic vector quantization (STVQ) [25,26] was derived by applying a deterministic annealing algorithm for optimizing the cost function of TVQ [27,28]. This process is taken as a method to avoid local minima of the cost function for clustering. Instead of using hard assignment, the PPoSOM proposed in this paper uses the soft assignment to obtain an effective visualization.

PPoSOM employs the soft assignment so that it is associated with a cost function. By minimizing the cost function, a principled rule for updating weight is obtained, which results in a good mapping effect. First, the noised probabilistic assignment $p_i(x(t))$ of neuron i is introduced as follows, and the term "noised" means that $p_i(x(t))$ is affected by probabilistic assignments of neighboring neurons.

$$p_i(x(t)) = \sum_{j=1}^{N} h_{ij} P_j(x(t)), \tag{7}$$

where N is the number of neurons, $P_j(x(t))$ is the probabilistic assignment of neuron j for input $x(t)$, and h_{ij} is a neighborhood constant satisfying $\sum_{i=1}^{N} h_{ij} = 1$. They can be taken as

$$P_j(x(t)) = \frac{1}{C}\left(\frac{1}{\left\| \sum_{k=1}^{N} h_{jk}(x(t)-w_k)\right\|^2}\right) \tag{8}$$

$$h_{ij} = \frac{\exp\left(-\frac{\|Pos_i - Pos_j\|^2}{2\sigma^2}\right)}{\sum_{k=1}^{N}\exp\left(-\frac{\|Pos_i - Pos_k\|^2}{2\sigma^2}\right)}, \tag{9}$$

where C is a normalization constant and the neighborhood radius σ is a constant, Pos_i is the coordinates of neuron i. $P_j(x(t))$ is reversed to the Euclidian distance between input data and the neuron, that $P_j(x(t))$ is highest if neuron w_j is the nearest to $x(t)$. The cost function of PPoSOM computes the sum of square errors between the input data and all the neurons with consideration of the probabilistic assignment. The cost function is defined as

$$E = \frac{1}{2}\sum_{t=1}^{M}\left\| \sum_{i=1}^{N} p_i(x(t))(x(t)-w_i)\right\|^2 = \sum_{t=1}^{M} E(t), \tag{10}$$

where $E(t) = \frac{1}{2}\left\| \sum_{i=1}^{N} p_i(x(t))(x(t)-w_i)\right\|^2$, M is the number of input data.

Because $E(t)$ is positive, minimizing E can be simplified to minimizing every $E(t)$ which gives

$$\frac{\partial E(t)}{\partial w_i} = -p_i(x(t))\sum_{j=1}^{N} p_j(x(t))(x(t)-w_j) \tag{11}$$

As a result, the principled weight updating rule is

$$w_i(t+1) = w_i(t) - \varepsilon(t)\frac{\partial E(t)}{\partial w_i} = w_i(t) + \varepsilon(t) p_i(x(t))\sum_{j=1}^{N} p_j(x(t))(x(t)-w_j(t)). \tag{12}$$

In order to avoid or to minimize the influence of sequential training order, the above updating-weight rule is rewritten for a batch algorithm of PPoSOM. It gives

$$w_i(k+1) = w_i(k) + \frac{\varepsilon(t)}{M}\sum_{t=1}^{M} p_i(x(t))\sum_{j=1}^{N} p_j(x(t))(x(t)-w_j(k)). \tag{13}$$

The definitions of $N_r(i)$, $N_\alpha(i)$, r_x and α_x are the same as PolSOM. The executing steps of PPoSOM are as follows:

Step 1. Initialize each neuron according to its position. Normalize the input data, and initialize the polar coordinates of the data by setting the radii proportional to their weights and the angles with random values.

Step 2. Compute the noised assignment probability of all the neurons for all input data according to Eq. (7).

Step 3. Update the weights of all neurons by batch algorithm according to Eq. (13).

Step 4. Update the polar coordinates of this datum according to Eq. (2).

Step 5. Repeat Steps 2–4 until the map converges.

Upon the completion of training, the visualization map is generated such that all advantages of PolSOM are preserved. The most important property of PPoSOM is that it is associated with a cost function, which gives a principled rule for weight updating and obtains a good visualization effect. A probabilistic data assignment instead of a hard assignment makes PPoSOM more flexible in converging to the final map in accordance with different data characteristics. This effect will be investigated in later section of this paper.

4. Simulation results

Six data sets, including two synthetic data sets, Iris data set [32], Wine data set [33], Wisconsin breast cancer data set [33] and the Times Higher Education 2009 World University Rankings [34] are used to illustrate the characteristics of PPoSOM. The visualization results are compared with SOM and ViSOM. The map size of SOM is 20×20, the number of maximum iterations is 1000 and the learning rate monotonically decreases from 1 to 0.018 with time. The neighborhood range also monotonically decreases from 14.78 to 2. In ViSOM, the map size is the same as that of SOM, and λ is set to 0.1. In PPoSOM, η_1 and η_2 are set to 0.05 and 0.1, respectively.

5. Three-dimensional synthetic data sets

In order to demonstrate the characteristics of PPoSOM, two types of 3-D synthetic data sets are used in this section. Each of them consists of two clusters named Cluster "1" and Cluster "2", and each cluster is formed by 100 data points.

In the first data set, the mean weight vectors of two clusters are $[0.45\ \ 0.54\ \ 0.54]^T$ and $[2.48\ \ 2.48\ \ 2.52]^T$, respectively. The data weights in Cluster "1" are smaller than those in Cluster "2". The simulation results of PPoSOM, SOM and ViSOM are presented in Fig. 2. As shown in Fig. 2, Cluster "1" and Cluster "2" are well separated from each other in PPoSOM, SOM and ViSOM. In PPoSOM map Fig. 2(a), all the radii of the data from Cluster "1" are smaller than those from Cluster "2". This is in the agreement with the fact that the average data weight of Cluster "1" is smaller than that of Cluster "2". Besides, the evenly distributed data angles indicate that the attributes of every data are similar.

Based on Eq. (4), the intra cluster densities of Cluster "1" and Cluster "2" are 1.1189 and 1.0191, respectively, which indicate that these two have almost the same compactness. And according to Eq. (5), the inter cluster density between Cluster "1" and Cluster "2" is 0.00069. This small value shows that the two clusters are well separated. From Eq. (6), the clustering criterion SCD of the first synthetic data set is 3098.55. The SCD is used to compare PPoSOM and PolSOM. In Table 2, the SCD value of PolSOM is less than that of PPoSOM, which means PPoSOM provides a better visualization effect for the studied data set.

In the second data set, the second attributes in Cluster "1" and the third attributes in Cluster "2" are larger than others. The mean vectors are $[0.48\ 2.49\ 0.54]^T$ and $[0.44\ 0.50\ 2.49]^T$, respectively. The visualizations of PPoSOM, SOM and ViSOM are shown in Fig. 3.

Fig. 3 shows that the two clusters are well separated in all the three algorithms. In PPoSOM, data from Cluster "1" and data from Cluster "2" are located in the neighborhood of the angular 120° and 240°, respectively. It is worth noting that the second and the third attributes of Cluster "1" and Cluster "2" in input space are more significant. These important characteristics cannot be exhibited in SOM or ViSOM.

The intra cluster densities of Cluster "1" and Cluster "2" are 0.511 and 0.5071, respectively. It shows that the compactnesses of these two are similar. The inter cluster density between Cluster

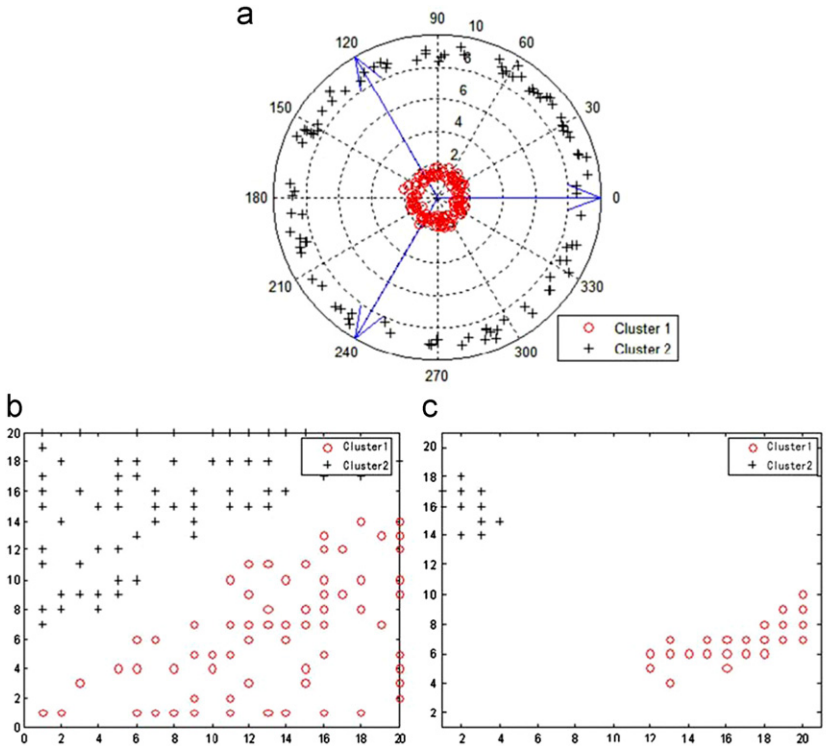

Fig. 2. Visualization of the first 3-D synthetic data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

Table 2
Comparison of mapping by using SCD.

SCD	First 3D data	Second 3D data	Iris data	Wine data	Wisconsin
PPoSOM	3098.55	40.724	13.98	8.097	0.7
PolSOM	3015.66	32.60	11.41	7.97	0.6792

"1" and Cluster "2" is 0.025. This indicates the two clusters are well separated. The SCD value is 40.724 which is in agreement with the visualization of PPoSOM. As shown in Table 2, the relatively larger SCD value obtained by PPoSOM shows it exhibits an improved mapping ability.

5.1. *Iris data set*

Iris data set [32] contains three clusters; each cluster has 50 four-dimensional instances. The mean vectors of these three clusters after normalization are $[0.20\ 0.59\ 0.08\ 0.06]^T$, $[0.45\ 0.32\ 0.55\ 0.51]^T$ and $[0.64\ 0.41\ 0.77\ 0.80]^T$, respectively. The visualization results of PPoSOM, SOM and ViSOM are shown in Fig. 4. The characteristics of Iris data are clearly shown in Fig. 4(a). Cluster "1" is clearly separated from Cluster "2" and Cluster "3". The latter two are overlapped in some extent and not linearly separable from each other. The average radius of the Cluster "1" is the smallest, and that of the Cluster "3" is the largest. In addition, Fig. 4(a) illustrates different significant attributes in the three clusters: the second attribute in Cluster "1", the third attribute in Cluster "2" and the fourth attribute in Cluster "3". The visualization is in agreement with the Iris data characteristics. It is worth noting that the above characteristics cannot be shown in SOM or ViSOM.

The intra cluster densities of three clusters are 0.577, 0.4069 and 0.4172. The inter cluster density between Cluster "1" and Cluster "2", Cluster "1" and Cluster "3", and Cluster "2" and Cluster "3" are 0.0023, 0.0013 and 0.0966, respectively. This result is identical with the PPoSOM visualization. In Table 2, the SCD value of PPoSOM is 13.98, while the SCD value of PolSOM is 11.41. It indicates the PPoSOM is able to deliver improved results.

5.2. *Wine data set*

The Wine data set [33] consists of 178 13-D data points which are divided into three clusters. The number of data points in each cluster is 59, 71 and 48. It is noted that these three clusters are not well separated.

The visualization results of PPoSOM, SOM and ViSOM are presented in Fig. 5. The characteristics of Wine data set are clearly shown in Fig. 5(a). The average radius in Cluster "1" is the largest, and the average radii in Cluster "2" and Cluster "3" are similar. This is in agreement with the fact that the data weight of Cluster "1" is larger than the other two clusters. PPoSOM demonstrates different significant attributes of each cluster along different corresponding angular coordinates. It is worth pointing out that, SOM and ViSOM are incapable of exhibiting these data characteristics as shown in Fig. 5(b) and (c).

The intra cluster densities of three clusters are 0.7806, 0.4933 and 0.5706. The inter cluster density between Cluster "1" and Cluster "2", Cluster "1" and Cluster "3", and Cluster "2" and Cluster "3" are 0.0857, 0.0073 and 0.1348, respectively. The SCD value is 8.097. This result is also in agreement with the PPoSOM visualization. As shown in Table 2, PPoSOM also has a larger SCD value compared to PolSOM.

5.3. *Wisconsin breast cancer data set*

The Wisconsin breast cancer data set [33] consists of 569 instances with 32 attributes (30 real-valued input features). The data set is divided into two clusters: benign and malignant. The numbers of benign and malignant are 357 and 212, respectively. There is no clear separation between the two clusters.

The visualization results of PPoSOM, SOM and ViSOM are shown in Fig. 6. The characteristics of breast cancer data set are

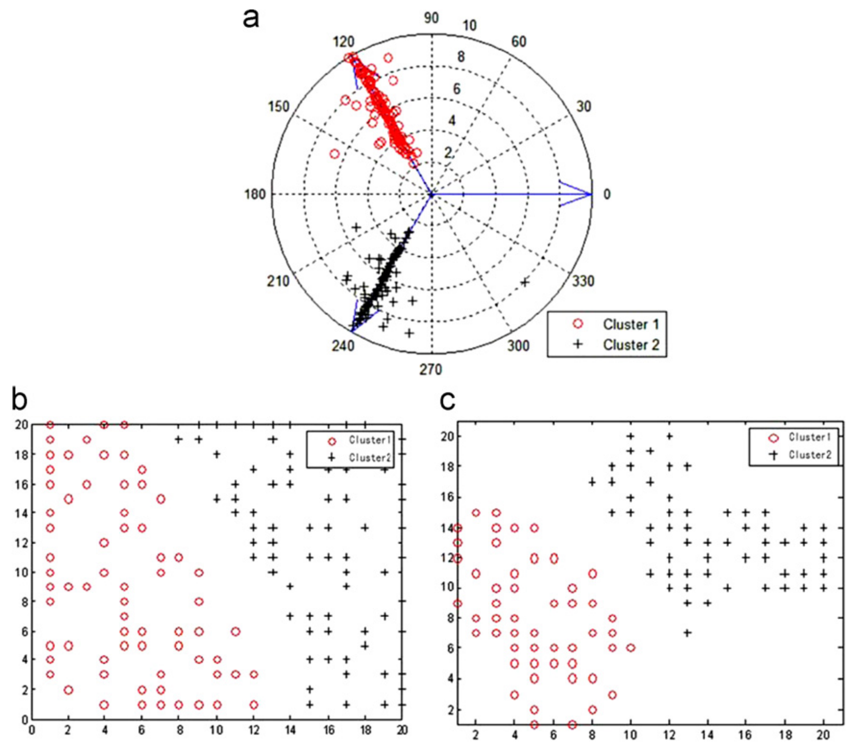

Fig. 3. Visualization of the second 3-D synthetic data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

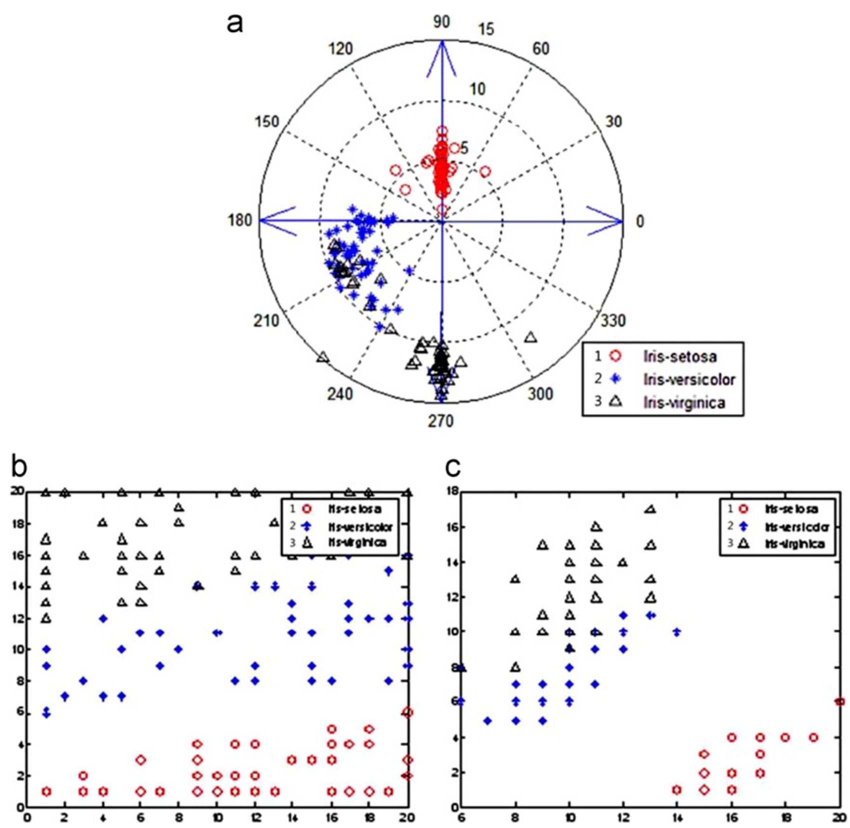

Fig. 4. Visualization of the Iris data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

clearly shown in Fig. 6(a). The average radius of benign is smaller than that of malignant. This is in agreement with the fact that the data weight of benign is smaller than that of malignant. By comparing the data along different corresponding angular coordinates, PPoSOM demonstrates these two clusters have different significant attributes. It is worth noting that SOM and ViSOM are not capable of exhibiting these data characteristics as shown in Fig. 6(b) and (c).

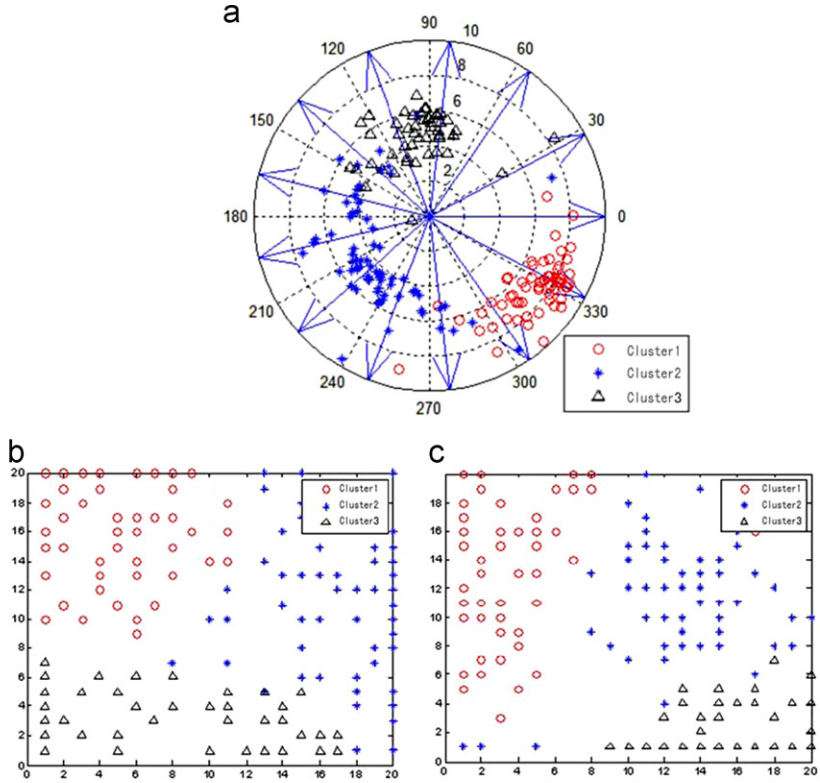

Fig. 5. Visualization of the Wine data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

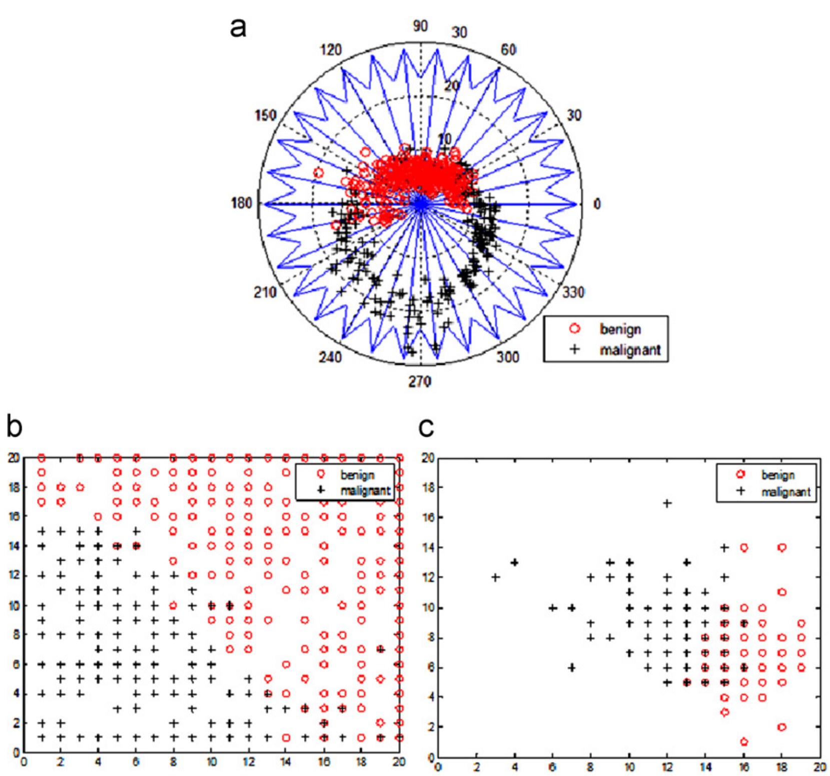

Fig. 6. Visualization of the Wisconsin breast cancer data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

The intra cluster densities of benign and malignant are 0.1824 and 0.1239, respectively. It shows that the compactness of benign is higher than that of malignant. The inter cluster density between benign and malignant is 0.4376. The SCD value is 0.7. From Table 2, the SCD value indicates that PPoSOM delivers the slightly better mapping result than PolSOM.

5.4. 2009 World University Rankings data set

The 2009 World University Rankings [34] used in this study is the one reported by the Times Higher Education. The top two hundreds institutions from 20 countries are evaluated by six attributes in the ranking list. This ranking is to present the Times views on the overall strengths of the world's universities, including academic peer review 40%, employer review 10%, faculty student ratio 20%, proportion of international faculty 5%, proportion of international students 5% and citations per faculty 20%. The weight vector of a university is formed by these six attributes, which are the scores of the six aspects. Their ranking calculation can be considered as a kind of simple dimension reduction by putting different weights on the 6 attributes. Under this simple calculation, the final ranking can be significantly varied by changing the weighting scheme, i.e., increase the international faculty from 5% to 20%; there are a thousand of Hamlets in a thousand people's eyes. Apparently, this simple ranking approach cannot reflect the strength and specialty of certain universities clearly and accurately. In this section, we use this data set to help illustrating the characteristics of the proposed algorithm, because readers have a clear idea of the topological nature of the data (universities). The visualizations by PPoSOM, SOM and ViSOM are illustrated in Figs. 7 and 8. Note that Fig. 8(a)(d), (b)(e) and (c)(f) are samples of Fig. 7(a), (b) and (c), respectively.

In Fig. 7(a)–(c), the dots represent the top 100 universities, and the crosses represent universities ranking from 101 to 200. In Fig. 7(a), it is clear that the higher ranking universities are located with larger radii. This is in agreement with the fact that higher ranking universities have higher scores in most aspects. It is worth noting that this characteristic is not available in the SOM and ViSOM; it is clear that PPoSOM provides more data characteristics. Take certain universities as example, the 20th data ($[97\ 99\ 84\ 93\ 86\ 65]^T$) and the 21st data ($[97\ 80\ 55\ 100\ 94\ 99]^T$) represent University of Edinburgh and ETH Zurich, respectively. Both of them are ranked 20 according to the Times Higher Education, but their features are significantly different in a way that the former one has a good response from employer, whilst the latter one owns more international faculty. In their corresponding output map shown in Fig. 8(a), it shows different significant attributes in the 20th data (the second attribute) and the 21st data (the fourth attribute). The same observations can also be perceived in the remaining data, like the 177th data (University of Antwerp $[47\ 36\ 99\ 59\ 56\ 37]^T$) and the 178th data (University of Athens $[46\ 44\ 65\ 0\ 91\ 76]^T$). They are both at the same 177th ranking, but exhibit different specialties. It is also worth noting that one can easily identify the universities with similar features via visualizing the data located with similar angles. In Fig. 8(d), the 12th data (University of Pennsylvania $[96\ 99\ 85\ 82\ 60\ 98]^T$), the 91st data (University of Nottingham $[70\ 99\ 61\ 84\ 86\ 48]^T$) and the 131st data (Ohio State University $[69\ 77\ 40\ 69\ 46\ 64]^T$) are all located around the same angle, which indicates they all have relatively high scores in the attribute of employer review. In Fig. 8(e) and (f), it is clear that SOM and ViSOM are unable to reflect these characteristics compared with Fig. 8(d).

The results show that the PPoSOM not only can group similar data, it can also make use of the data positions to reflect the data characteristics. In other words, PPoSOM is capable of preserving data topology and exhibiting data characteristics. Compared with SOM and ViSOM, which map data on Cartesian coordinates by using Euclidian distance as the only variable, PPoSOM can manifest more data characteristics. In addition, by applying the probabilistic mechanism, PPoSOM delivers improved performance compared with PolSOM according to the SCD analysis.

6. Conclusions

In this paper, a new self-organizing map called Probabilistic Polar SOM (PPoSOM) is developed for providing a new type of visualization. PPoSOM introduces a soft assignment to obtain a cost function, which gives the principled rule for weight-updating

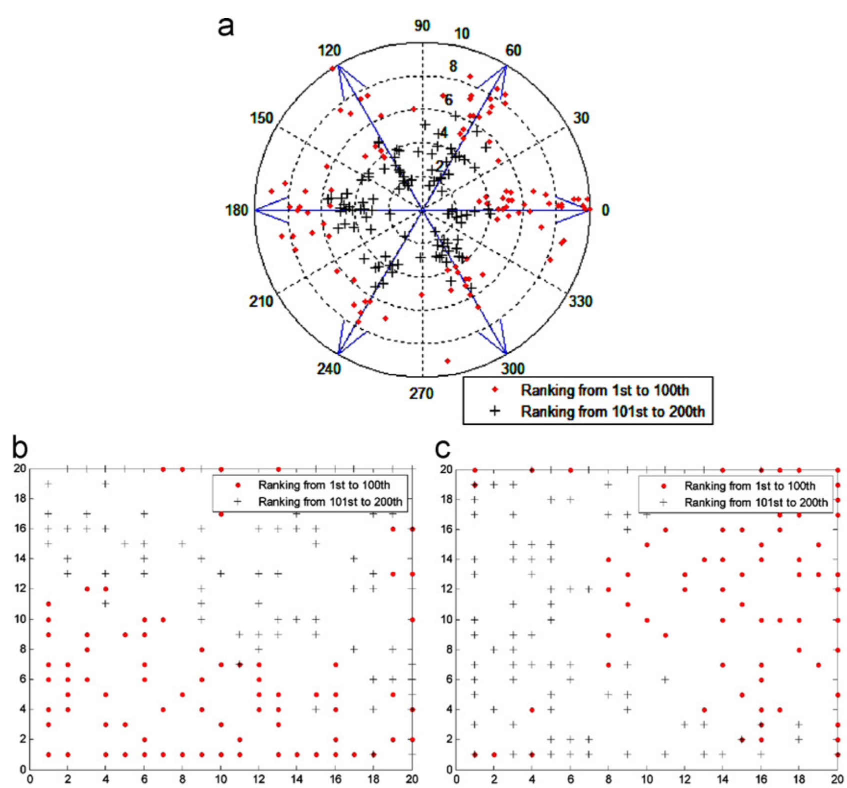

Fig. 7. Visualization of the 2009 Times High Education World's University Rankings data set: (a) PPoSOM, (b) SOM, and (c) ViSOM.

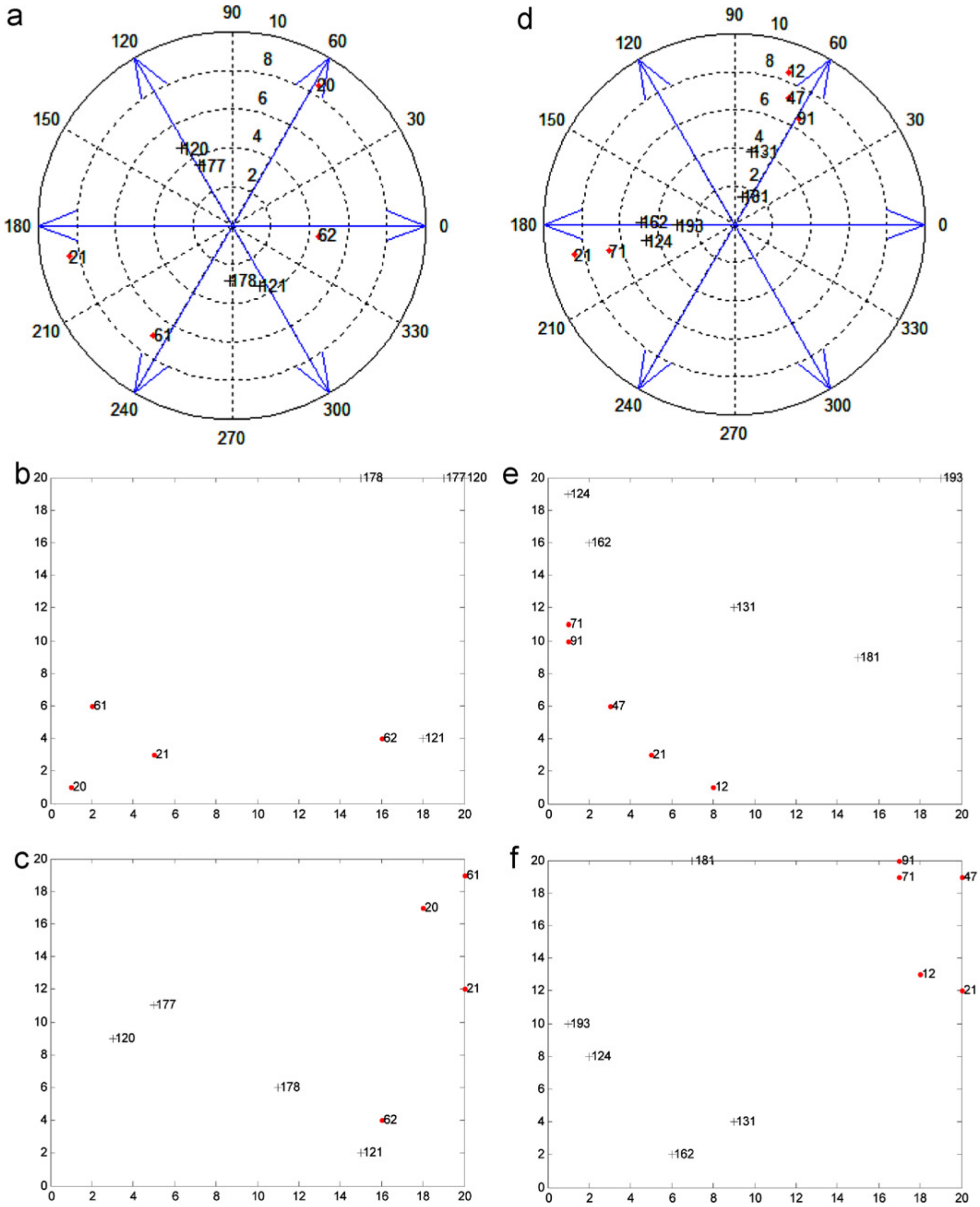

Fig. 8. Visualization of the 2009 Times High Education World's University Rankings data set: (a) PPoSOM (Sample 1), (b) SOM (Sample 1), (c) ViSOM (Sample 1), (d) PPoSOM (Sample 2), (e) SOM (Sample 2), and (f) ViSOM (Sample 2).

to the recently developed Polar SOM. The probabilistic function of a neuron for an input datum in the soft assignment is reversely proportional to the difference between the input datum and the neuron. This enables a more flexible convergence of the output map in accordance with the characteristic of the input data. Combining with the excellent characteristic of PolSOM, the proposed PPoSOM is able to exhibit improved visualization performance. This is demonstrated in the presented results and the synthetical cluster density (SCD). This study shows that the PPoSOM is effective for multi-dimensional data visualization.

References

[1] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1992.

[2] R.N. Shepard, J.D. Carroll, Parametric representation of nonlinear data structures, in: P.R. Krishnaiah (Ed.), Proceedings of the International Symposium on Multivariate Analysis, Academic, New York, 1965, pp. 561–592.

[3] A.C. Rencher, Methods of Multivariate Analysis, Wiley, New York, 1995.

[4] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, J. Roy. Statist. Soc., Ser. B: Statist. Methodol. 61 (1999) 611–622.

[5] J. Rubner, P. Tavan, A self-organizing network for principal component analysis, Europhysics Letters 10 (1989) 693–698.

[6] M. LeBlanc, R.J. Tibshirani, Adaptive principal surfaces, J. Am. Statist. Assoc. 89 (1994) 53–64.

[7] J.W. Sammon, A nonlinear mapping for data structure analysis, IEEE Trans. Comput. C18 (1969) 401–409.

[8] N.R. Pal, V.K. Eluri, Two efficient connectionist schemes for structure preserving dimensionality reduction, IEEE Trans. Neural Networks 9 (1998) 1142–1154.

[9] T. Kohonen, Self-Organizing Maps, Springer, Berlin (1997).

[10] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybernet. 43 (1982) 59–69.

[11] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Networks 11 (2000) 586–600.

[12] A. Ultsch, Self-organizing neural networks for visualization and classification, Inform. Classification (1993) 864–867.
[13] M.A. Kraaijveld, J. Mao, A.K. Jain, A nonlinear projection method based on Kohonen's topology preserving maps, IEEE Trans. Neural Networks 6 (1995) 548–559.
[14] H. Yin, Interpolating self-organizing map (iSOM), Electron. Lett. 35 (1999) 1649–1650.
[15] H. Yin, A. ViSOM, novel method for multivariate data projection and structure visualization, IEEE Trans. Neural Networks 13 (2002) 237–243.
[16] H. Yin, Data visualization and manifold mapping using the ViSOM, IEEE Trans. Neural Networks 15 (2002) 1005–1016.
[17] H. Yin, Self-organizing maps: statistical analysis, treatment and applications, Ph.D. Thesis, Department of Electronics, University of York, UK, 1996.
[18] H. Yin, On multidimensional scaling and the embedding of self-organizing maps, IEEE Trans. Neural Networks 21 (2008) 160–169.
[19] S. Wu, T.W.S. Chow, PRSOM: a new visualization method by hybridizing multidimensional scaling and self-organizing map, IEEE Trans. Neural Networks 16 (2005) 1362–1380.
[20] L. Xu, Y. Xu, T.W.S. Chow, PolSOM: A new method for multidimensional data visualization, Pattern Recognition 43 (2010) 1668–1675.
[21] D. Alahakoon, S.K. Halgamuge, B. Srinivasan, Dynamic self-organizing maps with controlled growth for knowledge discovery, IEEE Trans. Neural Networks 11 (2000) 601–614.
[22] T.W.S. Chow, S. Wu, An online cellular probabilistic self-organizing map for static and dynamic data sets, IEEE Trans. Circuits Systems I: Regular Papers 51 (2004) 732–747.
[23] D. Merkl, A. Rauber, Alternative ways for cluster visualization in self-organizing maps, in: Proceedings of the Workshop on Self-Organizing Maps (WSOM97), Espoo, Finland, Helsinki University of Technology, 1997, pp. 106–111.
[24] E. Erwin, K. Obermayer, K. Schulten, Self-organizing maps: ordering, convergence properties and energy functions, Biol. Cybernet. 67 (1992) 47–55.
[25] T. Graepel, M. Burger, K. Obermayer, Phase transitions in stochastic self-organization maps, Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. 56 (1997) 3876–3890.
[26] T. Graepel, M. Burger, K. Obermayer, Self-organizing maps: generalizations and new optimization techniques, Neurocomputing 21 (1998) 173–190.
[27] S.P. Luttrell, Self-organization: a derivation from fist principles of a class of learning algorithm, in Proceedings of the IEEE International Joint Conference on Neural Network, 1989, pp. 495–498.
[28] S.P. Luttrell, Code vector density in topographic mappings: scalar case, IEEE Trans. Neural Network 2 (1991) 427–436.
[29] J. Sum, C.S. Leung, L.W. Chan, L. Xu, Yet another algorithm which can generate topography map, IEEE Trans. Neural Network 8 (1997) 1204–1207.
[30] F. Anouar, F. Badran, S. Thiria, Probabilistic self-organizing map and radial basis function networks, Neurocomputing 20 (1998) 93–96.
[31] T. Heskes, Energy functions for self-organizing maps, in: E. Oja, S. Kaski (Eds.), Kohonen Maps, Elsevier, Amsterdam, The Netherlands, , 1999, pp. 303–315.
[32] R.A. Fisher, The use of multiple measure in taxonomic problems, Ann. Eugenics 7 (1936) 179–188 Part II.
[33] D.J. Newman, S. Hettich, C.L. Blake, C.J. Merz, Uci Repository of Machine Learning Databases, Department of Information and Computer Sciences, University of California, Irvine, 1998.
[34] 〈http://www.usnews.com/articles/education/worlds-best-universities/2009/10/20/worlds-best-universities-top-200.html?PageNr=1〉.

Yang Xu is now pursuing M.Phil. degree in the Department of Electronic Engineering of City University of Hong Kong, Hong Kong, China. She obtained B.E. degree in the Department of Electrical Engineering and Information of Sichuan University, China in 2008. Her research interest areas are neural networks, pattern recognition, and their applications.

Lu Xu is working towards the Ph.D. degree in Department of Electronic Engineering of City University of Hong Kong. She obtained the B.Eng. and M.Sc. in electronic engineering from Beijing University of Technology in 2004 and 2007, respectively. Her research interests include data clustering, optimization, and network routing.

Tommy W.S. Chow (M'93) received the B.Sc. (First Hons.) and Ph.D. degrees from the University of Sunderland, Sunderland, UK. He joined the City University of Hong Kong, Hong Kong, as a Lecturer in 1988. He is currently a Professor in the Electronic Engineering Department. His research interests include machine fault diagnosis, HOS analysis, system identification, and neural networks learning algorithms and applications.