6C-5

Hierarchical Map to Transform Distances among Data

Ryo SAEGUSA† Kazuyuki WATANABE†† Shuji HASHIMOTO†
†School of Science and Engineering, Waseda University
††Graduate School of Science and Engineering, Waseda University

1 Introduction

In the analysis of multi-dimensional data, it is important to operate distances among the data, since the distances characterize the data structure. For instance, dimension reduction of the data with preserving the distances is effective in terms of data compression. Data mapping to transform interpoint distances provides us a new interpretation for the data in the meaning of the scaling given by the distances.

Conventionally, some approaches for distance operation have been studied. Multi-dimensional scaling [1] determines the coordinates of the data from the given distances. Local linear embedding [2] provides the low-dimensional coordinates of high-dimensional data with preserving the distances among the neighbor vectors. These approaches are not effective to obtain the coordinate of the new given vector. A neural-network based approach to transform the distances is proposed in [3], which overcomes this problems, however, has uncertainties over control of the degree of freedom such as the number of the hidden units, and selection of the mapping-space dimension.

In this research, we propose a new approach to construct the map to transform distances among data. As the element of the map, we use the Piecewise Linear Map (PLM) which consists of several linear maps. The PLM has advantages in easy optimization of their parameters due to linearity of the map components, and the degree of freedom which can be controlled intuitively by selecting the number of piecewise regions. Consequently, we construct the distance transformation map with PLMs to obtain efficient low-dimensional coordinates, which satisfies given distances. The constructed hierarchical map is applicable to data retrieval based on subjective similarity, and dimension reduction with preserving the data structure in terms of distances.

2 Proposed Approach

2.1 Piecewise Linear Map

The PLM is defined as the nonlinear map which divides input space into some piecewise regions and performs linear mapping in the each region. For the input vector $\vec{x} \in R^n$ and the output vector $\vec{y} \in R^m$, the linear map L_i corresponding to the region $D_i (i = 1, ..., d)$ and the determinant function S_i of the corresponding region are defined as follows,

$$L_i(\vec{x}) = W_i\vec{x} + \vec{w}_{i0}, \tag{1}$$

$$S_i(\vec{x}) = \vec{v}_i \cdot \vec{x} + v_{i0}, \tag{2}$$

$$v_{i0} = -|\vec{v}_i|^2/2. \tag{3}$$

W_i, $\vec{w}_{i0}$ and v_{i0} are the coefficients. Hereafter, we represent $\vec{v}_i$ as the *reference* of the D_i. The PLM: Φ integrating all regions is defined as the following equations.

$$\Phi(\vec{x}) = L_r(\vec{x}), \tag{4}$$

$$r = \mathrm{ArgMax}_i \ S_i(\vec{x}), \tag{5}$$

where D_r represents the region to which the input vector $\vec{x}$ belongs. Under the constraint of Eq.(3), $\vec{v}_r$ is the nearest reference of the $\vec{x}$ (The proof is skipped here)

The Φ is evaluated by the following mean-squared error function regarding the dataset $\{\vec{x_p}, \vec{y_p}\}_{p=1,...,N}$:

$$E = \frac{1}{N}\sum_p E(\vec{x^p}),\ E(\vec{x^p}) = |\vec{y}^p - \Phi(\vec{x}^p)|^2. \tag{6}$$

The parameters of the PLM are trained with the conventional gradient method in the following manner.

$$\Delta w^r_{ij} = -\eta_e \frac{\partial E}{\partial w^r_{ij}}(\vec{x}^p),\quad \Delta v^r_j = +\eta_s \frac{\partial S}{\partial v^r_j}(\vec{x}^p), \tag{7}$$

$$\Delta w^k_{ij} = 0,\quad \Delta v^k_i = 0 \quad (k \neq r), \tag{8}$$

where η_e and η_s are set to small positives.

This optimization procedure lets $\vec{x}$ approach to the nearest reference $\vec{v}_r$ and moreover lets the output $\Phi(\vec{x}^p)$ of the corresponding linear map L_r approach to the expected output $\vec{y}^p$. This procedure generates the references distribution according to the density of the input vectors.

2.2 Distance Transformation Map

Here, we formulate the distance transformation map: Ψ which provides the new coordinates of the input vectors with arbitrarily transforming the distances among them. Let us define a pair of input vectors as $(\vec{x}^p,\ \vec{x}^q)$ and the corresponding pair of output vectors as $(\vec{y}^p,\ \vec{y}^q)$. The mean squared error function regarding the expected distance: δ^{pq} in output space are described as

$$E = \sum_p\sum_q E^{pq},\quad E^{pq} = |\delta^{pq} - d^{pq}|^2/2, \tag{9}$$

$$d^{pq} = \sqrt{|\vec{y}^p - \vec{y}^q|^2} \ = \sqrt{|\Psi(\vec{x}^p) - \Psi(\vec{x}^q)|^2}. \tag{10}$$

We construct the distance transformation map Ψ to minimize the above-mentioned error. The quasi-optimal parameters of Ψ are iteratively determined

† Ryo SAEGUSA (ryos@ieee.org)
†† Kazuyuki WATANABE (wata@shalab.phys.waseda.ac.jp)
† Shuji HASHIMOTO (shuji@waseda.jp)

by the conventional gradient method as follows,

$$\Delta w = -\eta \frac{\partial E_r}{\partial w} = \eta \left(\frac{\delta^{pq}}{d^{pq}} - 1 \right) \cdot \sum_k (y_k^p - y_k^q)\left(\frac{\partial y_k^p}{\partial w} - \frac{\partial y_k^q}{\partial w}\right), \tag{11}$$

where the η is set to a small positive.

Until now, we have not assumed a concrete function form of the distance transformation map. Here, let us formulate the distance transformation map Ψ based on the above-mentioned PLM. For a pair of input vectors $(\vec{x}^p, \vec{x}^q)$, the respective regions of the PLM are described as

$$r^p = \mathrm{ArgMax}_i\ S_i(\vec{x}^p),\ r^q = \mathrm{ArgMax}_i\ S_i(\vec{x}^q). \tag{12}$$

In case of $r^p = r^q (= r)$, the parameter optimization of Eq.(11) results in

$$\Delta w_{ij}^r = -\eta_e \left\{ \frac{\partial E_r}{\partial w_{ij}^r}(\vec{x}^p) + \frac{\partial E_r}{\partial w_{ij}^r}(\vec{x}^q) \right\}, \tag{13}$$

$$\Delta v_i^r = +\eta_s \left\{ \frac{\partial S_r}{\partial v_i^r}(\vec{x}^p) + \frac{\partial S_r}{\partial v_i^r}(\vec{x}^q) \right\}, \tag{14}$$

$$\Delta w_{ij}^k = 0,\quad \Delta v_i^k = 0 \quad (k \neq r). \tag{15}$$

Otherwise $(r^p \neq r^q)$, it results in

$$\Delta w_{ij}^{r^p} = -\eta_e \frac{\partial E_{r^p}}{\partial w_{ij}^{r^p}}(\vec{x}^p),\ \Delta w_{ij}^{r^q} = -\eta_e \frac{\partial E_{r^q}}{\partial w_{ij}^{r^q}}(\vec{x}^q) \tag{16}$$

$$\Delta v_i^{r^p} = +\eta_s \frac{\partial S_{r^p}}{\partial v_i^{r^p}}(\vec{x}^p),\ \Delta v_i^{r^q} = +\eta_s \frac{\partial S_{r^q}}{\partial v_i^{r^q}}(\vec{x}^q) \tag{17}$$

$$\Delta w_{ij}^k = 0,\quad \Delta v_i^k = 0 \quad (k \neq r^p \text{and } k \neq r^q). \tag{18}$$

3 Experimental Results

To evaluate the proposed method, we assumed a problem to map a pair of the input vectors: $\vec{x}^p = (x_1^p, x_2^p)$ and $\vec{x}^q = (x_1^q, x_2^q)$ onto the pair of the output vectors: $\vec{y}^p = (y_1^p, y_2^p)$ and $\vec{y}^q = (y_1^q, y_2^q)$ with satisfying the following distance among the output vectors:

$$d^{pq} = \begin{cases} \sqrt{\left(\frac{x_1^p x_2^p}{x_1^p+x_2^p} - \frac{x_1^q x_2^q}{x_1^q+x_2^q}\right)^2 + \left(\frac{(x_2^p)^2}{x_1^p+x_2^p} - \frac{(x_2^q)^2}{x_1^q+x_2^q}\right)^2} & (\text{if } x_1 < x_2) \\ \sqrt{\left(\frac{(x_1^p)^2}{x_1^p+x_2^p} - \frac{(x_1^q)^2}{x_1^q+x_2^q}\right)^2 + \left(\frac{x_1^p x_2^p}{x_1^p+x_2^p} - \frac{x_1^q x_2^q}{x_1^q+x_2^q}\right)^2} & (\text{if } x_1 \geq x_2). \end{cases} \tag{19}$$

The output vector: $\vec{y}$ provided by the following equation is one of the theoretical solutions to satisfy the distance constraint of the Eq.(19).

$$(y_1, y_2) = \begin{cases} \left(\frac{x_1 x_2}{x_1+x_2}, \frac{x_2^2}{x_1+x_2}\right) & (\text{if } x_1 < x_2) \\ \left(\frac{x_1^2}{x_1+x_2}, \frac{x_1 x_2}{x_1+x_2}\right) & (\text{if } x_1 \geq x_2) \end{cases} \tag{20}$$

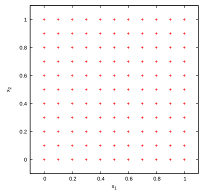

Figure 1: Input

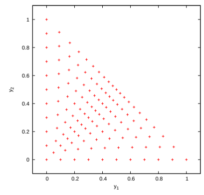

Figure 2: Theoretical

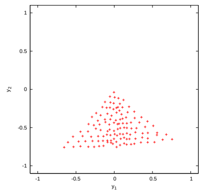

Figure 3: Output

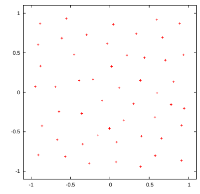

Figure 4: Reference

We solved this problem with the proposed distance transformation map. We used the randomly generated samples inside the region of $(0,1) \times (0,1)$ as input vectors when training, and the grid points shown in Fig.1 as ones when testing. We useed the following parameters: $d = 50, \eta_e = 0.8, \eta_s = 0.2$ in the experiment and performed 10,000 iterations in training.

The $\vec{x}^p$ and the theoretical solution $\vec{y}^p$ are shown in Fig.1 and Fig.2. The result is shown in Fig.3. As show in the figure, the grid points are mapped onto the triangle-like region which has a similar shape of the theoretical solusion. The differences in the poison and the direction of the triangle region are caused by the redundancy of the distance transformation map in translation and rotation. The Fig.4 suggests that the references reflecting the distribution of the input data are achieved by training.

4 Conclusion

We proposed a novel approach of the distance transformation map based on the piecewise linear map. The constructed hierarchical map is effective to transform the high-dimensional data into the low-dimensional coordinates with satisfying the given distances. The piecewise linear map has some advantages that the parameters are easily optimized due to linearity of the component maps, and degree of freedom can be controlled intuitively by selecting the number of piecewise regions.

The future works encompass the introduction of hierarchy into the proposed approach to obtain less-dimensional output space. Applications of the proposed approach will be considered, e.g. data retrieval based on subjective similarity, dimension reduction with preserving the distances and etc.

Acknowledgements
This work was supported in part by the Japanese Ministry of Education, Culture, Sports, Science and Technology (17700247) and Gifu Prefecture, Japan for the WABOT-HOUSE Project.

References

[1] Kruskal, J. B., Wish, M., "Multidimensional scaling," Sage Publications (1978).

[2] Roweis, S., Saul,L., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science vol.290, no.5500, pp.2323–2326 (2000).

[3] Suzuki, K., Hashimoto, S., "On the Evaluation of Relevance Learning by a Multi-layer Perceptron," Proc. IJCNN, pp.3204-3209 (2005).