A comment on "Alternative c-means clustering algorithms"

Dao-Qiang Zhang, Song-Can Chen*

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, People's Republic of China

In their paper [1], Wu and Yang proposed two alternative (hard) c-means and fuzzy c-means clustering algorithms, by replacing the original Euclidean distance in the c-means and fuzzy c-means algorithm with a new Gaussian function based distance. Although the new distance function is more robust than the Euclidean distance, we will show in this comment that the distance Wu and Yang used is not a metric.

Suppose $\mathbf{x}, \mathbf{y} \in \boldsymbol{X}$, and $\boldsymbol{X}$ is a compact subset of $\mathbf{R}^n$. Let $d(\boldsymbol{x}, \boldsymbol{y})$ denote the distance function between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, then the Euclidean distance is represented by $d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x}-\boldsymbol{y}||$, where $||\bullet||$ denote the two norm. Wu and Yang [1] defined a new distance as

$$d(x, y) = 1 - \exp(-\beta \, || \, x - y \, ||^2) . \quad (1)$$

To prove a distance function $d(\boldsymbol{x}, \boldsymbol{y})$ is a metric, the necessary and sufficient condition is that $d(\boldsymbol{x}, \boldsymbol{y})$ satisfies the following three conditions [2]

(i) $d(x, y) > 0, \forall x \neq y, d(x, x) = 0$,

(ii) $d(x, y) = d(y, x)$,

(iii) $d(x, y) \leq d(x, z) + d(z, y), \forall z$.

Wu and Yang [1] claimed that the distance function in equation (1) is a metric, i.e, the above three conditions were satisfied. However, we will show in Theorem 1 that it is not true.

* Corresponding author. Tel.:+86-25-489-2805.
E-mail address: daoqz@mail.com (D. Zhang), s.chen@nuaa.edu.cn (S. Chen).

Theorem 1. The distance function in equation (1) is not a metric.

Proof. To prove the distance is not a metric, we just need to find a counterexample for which one of the above conditions (i)-(iii) don't satisfied any more. Let $\boldsymbol{x}$= [0.2 0.2], $\boldsymbol{y}$= [0.4 0.4], $\boldsymbol{z}$= [0.3 0.3] and β=1, then according to equation (1), $d(\boldsymbol{x}, \boldsymbol{y}) = 0.0769$, $d(\boldsymbol{x}, \boldsymbol{z}) = 0.0198$, $d(\boldsymbol{z}, \boldsymbol{y}) = 0.0198$, so $d(\boldsymbol{x}, \boldsymbol{y}) = 0.0769 > d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}) = 0.0396$, i.e., the condition (iii) is not satisfied any more, thus the distance function in equation (1) is not a metric. □

The reason why the distance function in equation (1) is not a metric is that it is a type of squared distance, and a squared distance is not necessarily a metric. For example, the Euclidean distance is a metric, while the squared Euclidean distance is not necessarily a metric. So in order to remedy the mistake, we propose a new distance function as follows

$$d(x, y) \overset{\Delta}{=} \sqrt{1 - \exp(-\beta \| x - y \|^2)}\ . \tag{2}$$

Theorem 2. The distance function in equation (2) is a metric.

Proof. We prove Theorem 2 from another viewpoint, i.e., kernel-based learning method in machine learning community [3].

Define Gaussian kernel function $K(\boldsymbol{x}, \boldsymbol{y}) = exp(-\beta\|\boldsymbol{x}-\boldsymbol{y}\|^2)$, and according to the Mercer theorem [3], there exists some nonlinear map $\boldsymbol{\Phi}: \boldsymbol{X} \rightarrow \boldsymbol{F}$ (here $\boldsymbol{F}$ is a compact subset of Hilbert space), satisfying $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Phi}(\boldsymbol{x})^{\boldsymbol{T}} \boldsymbol{\Phi}(y)$, where $(.)^{\mathrm{T}}$ is vector transpose. From the definition of K, we have

$K(\mathbf{x}, \mathbf{x}) = K(\mathbf{y}, \mathbf{y}) = 1$, Thus

$$\begin{aligned} \| \Phi(x) - \Phi(y) \|^2 &= \Phi(x)^T \Phi(x) + \Phi(y)^T \Phi(y) - 2\Phi(x)^T \Phi(y) \\ &= K(x,x) + K(y,y) - 2K(x,y) = 2(1 - K(x,y)) \end{aligned} . \quad (3)$$

According to (2) and (3), and the definition of Gaussian kernel function, the following can be derived

$$d(x,y) = \sqrt{\frac{1}{2} \| \Phi(x) - \Phi(y) \|^2} = \frac{1}{\sqrt{2}} \| \Phi(x) - \Phi(y) \| . \quad (4)$$

According to (2), $\forall x \neq y$, $d(\mathbf{x}, \mathbf{y})=d(\mathbf{y}, \mathbf{x})>0$, and $d(\mathbf{x}, \mathbf{y})=0$, so condition (i) and (ii) are satisfied. And from (4)

$$\begin{aligned} d(x,y) &= \frac{1}{\sqrt{2}} \| \Phi(x) - \Phi(y) \| \\ &\leq \frac{1}{\sqrt{2}} \left(\| \Phi(x) - \Phi(z) \| + \| \Phi(z) - \Phi(y) \| \right) = d(x,z) + d(z,y) \end{aligned} \quad (5)$$

Thus condition (iii) is satisfied due to the properties of the norm. So the distance function in equation (2) is a metric. □

Acknowledgements

We thank National Science Foundations of China and of Jiangsu under Grant Nos. 60271017 and BK2002092 for partial supports respectively.

Reference

1. K.L. Wu, M.S. Yang, Alternative c-means clustering algorithms, Pattern Recognition 35 (2002) 2267-2278.
2. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill Book Company, New York, 1976
3. N. Cristianini, J.S. Taylor, An Introduction to SVMs and Other Kernel-based Learning Methods, Cambridge University Press, 2000