Conference Proceedings Paper

A Comparison between Wasserstein Distance and a Distance Induced by Fisher-Rao Metric in Complex Shapes Clustering

Angela De Sanctis [1,*,†] **and Stefano A. Gattone** [2]

[1] Department of Management and Business Administration, University "G. d'Annunzio" of Chieti-Pescara, 66100 Chieti, Italy; a.desanctis@unich.it

[2] Department of Philosophical, Pedagogical and Quantitative Economic Sciences, University "G. d'Annunzio" of Chieti-Pescara, 66100 Chieti, Italy; gattone@unich.it

[*] Correspondence: a.desanctis@unich.it

[†] These authors contributed equally to this work.

Abstract: Shape Analysis studies geometrical objects, as for example a flat fish in the plane or a human head in the space. The applications range from structural biology, computer vision, medical imaging to archaeology We focus on the selection of an appropriate measurement of distance among observations with the aim of obtaining an unsupervised classification of shapes. Data from a shape are often realized as a set of representative points, called landmarks. For planar shapes, we assume that each landmark is modeled via a bivariate Gaussian, where the means capture uncertainties that arise in landmarks placement and the variances the natural variability across the population of shapes. At first we consider the Fisher-Rao metric as a Riemannian metric on the Statistical Manifold of the Gaussian distributions. The induced geodesic-distance is related with the minimization of information in the Fisher sense and we can use it to discriminate shapes. Another suitable distance is the Wasserstein distance, which is induced by a Riemannian metric and is related with the minimal transportation cost. In this work, a simulation study is conducted in order to make a comparison between Wasserstein and Fisher-Rao metrics when used in shapes clustering.

Keywords: information geometry; Fisher-Rao metric; Wasserstein metric; landmarks; Shape Analysis; Cluster Analysis

1. Introduction

Shapes clustering is of interest in various fields such as geometric morphometrics, computer vision and medical imaging. In the clustering of shapes is crucial to find an appropriate measurement of distance among observations. In particular we are interested to classify shapes which derive from complex systems as expression of self-organization phenomenon. We consider objects whose shapes are based on landmarks ([1–3]). These objects can be obtained by medical imaging procedures, curves defined by manually or automatically assigned feature points or by a discrete sampling of the object contours.

Since the shape space is invariant under similarity transformations, that is translations, rotations and scaling, an Euclidean distance function on such a space is not really meaningful. In Shape Analysis [4], in order to apply standard clustering algorithms to planar shapes, the Euclidean metric has to be replaced by the metric of the shape space. Examples are provided in [5,6] where the Procrustes distance was integrated in standard clustering algorithms such as the *k*-means. Similarly, [7] applied standard hierarchical or *k*-means clustering using dissimilarity measures based on the inter-landmark distances. In a model-based clustering framework [8,9] developed a mixture model of offset-normal shape distributions.

Considering for simplicity a planar shape of a population, we assume that each landmark is modeled via a bivariate Gaussian, where the means are the landmark geometric coordinates and capture uncertainties that arise in landmark placement while the variances derived from the natural variability across the population of shapes. According to Information Geometry, we consider the space of bivariate Gaussian densities as a Statistical Manifold ([10,11]) with the local coordinates defined by the model parameters. Next, we define distances between objects related with different Riemannian metrics. These distances are induced by the geodesics of the two metrics (geodesic distances). Applications of geodesics to shape clustering techniques are provided, in a landmark-free context, by [12,13].

At first we consider the Fisher-Rao metric as a Riemannian metric on the Statistical Manifold of the Gaussian densities. The induced geodesic-distance is related with the minimization of information in the Fisher sense and we can use it to define a shape distance.

Another suitable distance is the Wasserstein distance, which is induced by a Riemannian metric and is related with the minimal transportation cost.

The geodesic distances induced by Wasserstein and Fisher-Rao metrics can be used to discriminate shapes. The discriminative power of these shapes distances will be evaluated, in the setting of shapes clustering on simulated data.

2. The Method

Suppose we are given a planar shape configuration, C, consisting of a fixed number K of labeled landmarks

$$C = \{\mu_1, \mu_2, \ldots, \mu_K\}$$

with generic element $\mu_k = \{\mu_{k1}, \mu_{k2},\}$ for $k = 1, \ldots, K$.

Following [14], the k-th landmark may be represented by a bivariate Gaussian density as follows:

$$f(\mathbf{x}; \mu_k, \Sigma_k) = (2\pi)^{-1} |\Sigma_k|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_k)'\Sigma_k^{-1}(\mathbf{x} - \mu_k)\right\} \tag{1}$$

with $\mathbf{x}$ being a generic 2-dimensional vector and Σ_k given by

$$\Sigma_k = \sigma_k^2 \mathbf{I}_2 = \text{diag}(\sigma_{k1}^2, \sigma_{k2}^2) \tag{2}$$

where $\{\sigma_{k1}^2, \sigma_{k2}^2\}$ is the vector of the variances in the horizontal and vertical directions of the k-th landmark coordinates, for $k = 1, \ldots, K$. We remark that the means capture uncertainties that arise in landmark placement and the variances the natural variability across a population of shapes.

From Equation (1) we can assign to the k-th landmark a new set of coordinates given by $\theta_k = (\mu_k, \sigma_k)$ on the 4-dimensional manifold which is the product of two upper half planes. So two planar shapes S and S' can be parametrized as follows: $S = (\theta_1, \ldots, \theta_K)$ and $S' = (\theta'_1, \ldots, \theta'_K)$. For every k, let $\gamma_k(t)$ with $t \in [0, 1]$ be a path of the manifold such that $\gamma_k(0) = \theta_k$ and $\gamma_k(1) = \theta'_k$. From differential geometry we know that a given Riemannian metric g induces an inner product $< .,. >_g$ on the tangent space of the manifold such that the length of $\gamma_k(t)$ is defined as follows

$$l(\gamma_k) = \int_0^1 \|\dot{\gamma}_k(t)\|_{\mathbf{g}}^2 dt \tag{3}$$

The distance between the k-th landmarks of the two shapes is given by the minimum length of the trajectories $\gamma_k(t)$ (geodesic distance)

$$d_g(\theta_k, \theta'_k) = \inf_{\gamma_k}\{\sqrt{l(\gamma_k)} : \gamma_k(0) = \theta_k, \gamma_k(1) = \theta'_k\}. \tag{4}$$

Finally, the sum of the distances between each pair of landmarks is used to define a distance between two shapes S and S' [15].

In the statistical manifold of bivariate Gaussian densities, we will consider two different Riemannian metrics which in turn induce two types of geodesic distances.

One is the Fisher-Rao metric g_f. For this metric, the closed form of the geodesic distance between two densities with diagonal covariance matrices is available and given by [16]:

$$d_{g_f}((\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}), (\mu_{21}, \sigma_{21}, \mu_{22}, \sigma_{22})) =$$
$$\sqrt{2\sum_{i=1}^{2}\left(\ln \frac{|(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}) - (\frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i})| + |(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}) - (\frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i})|}{|(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}) - (\frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i})| - |(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}) - (\frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i})|}\right)^2}. \tag{5}$$

For Gaussian densities with Σ being any symmetric positive definite covariance matrix, a closed form for the associated distance is not available.

The other Riemannian metric we consider is g_w, which induces the Wasserstein distance ([17]). For Gaussian densities the explicit expression of the Wasserstein distance is the following:

$$d_{g_w}(\theta, \theta') = \|\mu - \mu'\| + tr(\Sigma) + tr(\Sigma') - 2tr(\sqrt{\Sigma^{\frac{1}{2}}\Sigma'\Sigma^{\frac{1}{2}}}) \tag{6}$$

where $\|.\|$ is the euclidean norm and $\Sigma^{\frac{1}{2}}$ is defined for a symmetric positive definite matrix Σ so that $\Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} = \Sigma$.

[18] proved that, with respect to the Riemannian metric which induces the Wasserstein distance, the manifold of Gaussian densities has non-negative sectional curvature. We deduce that the Wasserstein metric is different from the Fisher-Rao metric. For example in the univariate case, it is well known that the statistical manifold of Gaussian densities with the Fisher-Rao metric can be regarded as the upper half plane with the hyperbolic metric, which has negative curvature.

3. A Simulation Experiment

In this section we report the results of a simple simulation experiment. In order to test the discriminative power of the proposed shape distances we first simulate shapes from two different mean shapes. The Fisher-Rao distance and the Wasserstein distance are evaluated between each pair of shapes and stored in two different pairwise distance matrices. Then we run a hierarchical cluster algorithm which takes as input the pairwise distance matrices computed with the two shapes distances. The quality of the clusters identified with the two shapes distances is measured by means of the Adjusted Rand index ([19]).

The shapes are simulated from the following Gaussian perturbation model

$$X_{ih} = (\mu_h + E_i)\Gamma_i + \mathbf{1}_K \gamma_i^T \tag{7}$$

where

- E_i are zero mean $K \times 2$ random error matrices simulated from the multivariate Normal distribution with covariance structure Σ_E
- μ_h is the mean shape for cluster h
- Γ_i is an orthogonal rotation matrix with an angle θ uniformly produced in the range $[0, 2\pi]$
- γ_i^T is a 1×2 uniform translation vector in the range $[-2, 2]$

  Two different covariance structures are considered:

- Isotropic with $\Sigma_E = \sigma \mathbf{I}_K \otimes \sigma \mathbf{I}_2$ with independent spherical variation around each mean landmark
- Heteroscedastic with $\Sigma_E = \text{diag}[\sigma_1, \sigma_2 \ldots, \sigma_K] \otimes \sigma \mathbf{I}_2$ with a heteroscedastic variation around each mean landmark

The number of clusters was set to $h = 2$ and the number of configurations to 30. The mean shapes in the two clusters were taken from the mean skull of 21 rats collected at ages of 7 and 14 days (the rat calvarial data set, [1]). In the isotropic case, two values of σ were used: $\sigma = 10$ (small error) and $\sigma = 13$ (high error). The heteroscedastic case was simulated by multiplying the value of $\sigma = 10$ of 3 landmarks (out of $K = 8$ landmarks) by a factor of 1.69 (small error) and 3 (high error). Results from 500 random simulations are reported in Table 1.

Table 1. Adjusted-Rand index.

Model	Fisher-Rao	Wasserstein
Isotropic-small error	0.8848	0.9553
Isotropic-high error	0.6265	0.7457
Heteroscedastic-small error	0.8854	0.9606
Heteroscedastic-high error	0.8367	0.6616

Results show that the Wasserstein distance has a good performance especially in the Isotropic case. As the variability around the landmarks becomes heteroscedastic then the Fisher-Rao distance performs better (Heteroscedastic-high error case).

A more thorough study, however, is needed to analyze advantages and limits of these metrics as a tool for evaluating the differences between shapes. This is the aim of our future work.

References

1. Bookstein, F.L. *Morphometric Tools for Landmark Data: Geometry and Biology*; Cambridge University Press: Cambridge, UK, 1991.
2. Cootes, S.; Taylor, C.; Cooper, D.; Graham, J. Active shape models-their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59.
3. Kendall, D.G. Shape manifolds, Procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.* **1984**, *16*, 81–121.
4. Dryden, I.L.; Mardia, K.V. *Statistical Shape Analysis*; John Wiley & Sons: London, UK, 1998.
5. Amaral, G.; Dore, L.; Lessa, R.; Stosic, B. *k*-Means Algorithm in Statistical Shape Analysis. *Commun. Stat. Simul. Comput.* **2010**, *39*, 1016–1026.
6. Stoyan, D.; Stoyan, H. A further application of D.G. Kendall's Procrustes Analysis. *Biom. J.* **1990**, *32*, 293–301.
7. Lele, S.; Richtsmeier, J. *An Invariant Approach to Statistical Analysis of Shapes*; Chapman & Hall/CRC: New York, NY, USA, 2001.
8. Huang, C.; Styner, M.; Zhu, H. Clustering High-Dimensional Landmark-based Two-dimensional Shape Data. *J. Am. Stat. Assoc.* **2016**, *19*, 702–723.
9. Kume, A.; Welling, M. Maximum likelihood estimation for the offset-normal shape distributions using EM. *J. Comput. Graph. Stat.* **2010**, *19*, 702–723.
10. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; Translations of Mathematical Monographs; AMS & Oxford University Press: Providence, RI, USA, 2000; Volume 191.
11. Murray, M.K.; Rice, J.W. *Differential Geometry and Statistics*; Chapman & Hall: Boca Raton, FL, USA, 1984.
12. Srivastava, A.; Joshi, S.H.; Mio, W.; Liu, X. Statistical Shape analysis: Clustering, learning, and testing. *IEEE Trans. PAMI* **2005**, *27*, 590–602.
13. Mio, W.; Srivastava, A.; Joshi, S.H. On Shape of Plane Elastic Curves. *Int. J. Comput. Vis.* **2007**, *73*, 307–324.
14. De Sanctis, A.; Gattone, S.A. Methods of Information Geometry to model complex shapes. *Eur. Phys. J. Spec. Top.* **2016**, *225*, 1271–1279.
15. Gattone, S.A.; DeSanctis, A. A shape distance based on the Fisher-Rao metric and its application for shapes clustering. *Phys. A* **2017**, *487*, 93–102.
16. Costa, S.; Santos, S.; Strapasson, J. Fisher information distance: A geometrical reading. *Discret. Appl. Math.* **2015**, *197*, 59–69.
17. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.

18. Otto, F. The geometry of dissipative evolution equations: the porous medium equation. *Commun. Part. Differ. Equ.* **2001**, *26*, 101–174.
19. Hubert, L.; Arabie, P. Comparing Partitions. *J. Classif.* **1985**, 2, 193–218.