ARCHETYPAL CONTOUR SHAPES

Aleix Alcacer[1], Irene Epifanio[1], M.Victoria Ibáñez[1] and Amelia Simó[1]

[1] Department of Mathematics, Jaume I University, (e-mail: aalacacer@uji.es, epifanio@uji.es, mibanez@uji.es, simo@uji.es)

ABSTRACT: Shapes are represented by contour functions from planar object outlines. Functional archetypal analysis is proposed to describe closed contour shapes. Each contour function is approximated by a convex combination of functional contour archetypes, which are a mixture of cases in the data set. Archetypes represent extreme shape patterns and improve the interpretability of highly complex distributions. The archetypal contours of feet from an anthropometric database of the adult Spanish population are extracted, which is useful for improving the fit in footwear.

KEYWORDS: shape analysis, archetype analysis, functional data analysis, footwear.

1 Introduction

Archetype Analysis (AA) (Cutler & Breiman, 1994) is an unsupervised technique that describes cases of a sample as a mixture of archetypes, which in turn, are mixtures of the cases in the sample. This multivariate technique was extended to functional data (Epifanio, 2016; Vinué & Epifanio, 2017).

Shape is all the geometrical information that remains after location, scale and rotational effects are removed from an object. Shapes can be analyzed from three approaches (Stoyan & Stoyan, 1994): objects can be treated as subsets of $\mathbb{R}^2$, they can be described by landmarks, or by using functions that represent their contours. Epifanio *et al.*, 2018 propose archetypal shapes based on landmarks. Here we propose archetypal shapes based on contour functions. In particular, we consider the natural parametrization of the contour, i.e. when the contour is parametrized by its arc length. This can be applied to any contour (other contour functions have limitations (Kindratenko, 2003)).

In Sect. 2 the methodology is introduced and it is applied on a foot shape data set in Sect. 3. The work ends with some conclusions in Sect. 4.

2 Methodology

Let $\mathbf{X}$ be an $n \times m$ matrix with n observations and m variables. AA seeks to find k archetypes, i.e a $k \times m$ matrix $\mathbf{Z}$, in such a way that $\mathbf{x}_i$ is approximated by a mixture of $\mathbf{z}_j$'s (archetypes): $\sum_{j=1}^{k} \alpha_{ij}\mathbf{z}_j$, with the mixture coefficients contained in the $n \times k$ matrix α. Additionally, $\mathbf{z}_j$'s is expressed as a mixture of the data through the mixture coefficients found in the $k \times n$ matrix β: $\mathbf{z}_j = \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l$. To obtain the archetypes, AA computes two matrices α and β that minimize the following residual sum of squares (RSS): $\sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij}\mathbf{z}_j\|^2$ $= \sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l\|^2$, under the constraints 1) $\sum_{j=1}^{k} \alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ for $i = 1, \dots, n$ and 2) $\sum_{l=1}^{n} \beta_{jl} = 1$ with $\beta_{jl} \geq 0$ for $j = 1, \dots, k$.

2.1 Functional Archetype Analysis (FAA)

In the functional context, the values of the m variables in the standard multivariate context are replaced by function values with a continuous index t. Similarly, summations are replaced by integration to define the inner product. See Epifanio, 2016 for details about extension of AA to functional data.

In our problem, two functions characterize each contour, so FAA for bivariate functions must be considered. Let $f_i(t) = (x_i(t), y_i(t))$ be a bivariate function. Its squared norm is $\|f_i\|^2 = \int_a^b x_i(t)^2 dt + \int_a^b y_i(t)^2 dt$. Let $\mathbf{b}^{x_i}$ and $\mathbf{b}^{y_i}$ be the vectors of length m of the coefficients for x_i and y_i respectively for the basis functions B_h. Therefore, FAA is defined by $RSS = \sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} z_j\|^2 =$ $\sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} f_l\|^2 = \sum_{i=1}^{n} \|x_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} x_l\|^2 + \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} y_l\|^2 = \sum_{i=1}^{n} \mathbf{a}^{x_i\prime}\mathbf{W}\mathbf{a}^{x_i} + \sum_{i=1}^{n} \mathbf{a}^{y_i\prime}\mathbf{W}\mathbf{a}^{y_i}$, where $\mathbf{a}^{x_i\prime} = \mathbf{b}^{x_i\prime}$ - $\sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b}^{x_l\prime}$ and $\mathbf{a}^{y_i\prime} = \mathbf{b}^{y_i\prime} - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b}^{y_l\prime}$, with the corresponding constraints for α and β; and where $\mathbf{W}$ is the order m symmetric matrix with elements $w_{m_1,m_2} = \int_a^b B_{m_1} B_{m_2} dt$. In the case of an orthonormal basis, $\mathbf{W}$ is the order m identity matrix, and FAA is reduced to AA of the basis coefficients. But, in other cases, we may have to resort to numerical integration to evaluate $\mathbf{W}$, but once $\mathbf{W}$ is computed, no more numerical integrations are necessary.

3 Application

Knowledge of foot shape has a great relevance for the appropriate design of footwear. It is a main issue for manufacturing shoes, since a proper fit is a key factor in the buying decision, besides improper footwear can cause foot pain and deformity, especially in women. Therefore, the objective is to identify the shapes that represent the fitting problems of the population by means of archetypal shapes, which are extreme patterns. Then the shoe designer may adapt the design to the measurements of the extremes of a size.

Footprints have been extracted from an database of 775 3D right foot scans representing Spanish adult population. The anthropometric study was carried out by the Instituto de Biomecánica de Valencia. Data was collected in different regions across Spain using an INFOOT laser scanner. The binary images have been centered and scaled to remove the effects of translations and changes of scale as explained by Epifanio & Ventura-Campos, 2011.

In order to obtain the contour functions, the tracing begins counterclockwise in the most eastern outline point in the same row as the centroid, using *bwtraceboundary* of the image toolbox of MatLab. We normalize these functions in such a way that the perimeter length is eliminated, and the functions are defined on [0,1]. We approximate each curve by a linear combination of 51 Fourier basis (note that this basis system is periodic with period 1). All this work has been done by means of *fda* library (Ramsay & Silverman, 2005). We have therefore two pairs of functions (representing coordinates) $\{X(t), Y(t)\}$ for each foot, with $t \in [0,1]$.

3.1 Results

FAA is applied to the database. The screeplot is represented in Fig. 1, with the number of archetypes versus the respective RSS, and an elbow is found at $k = 3$. Fig. 1 also shows the contour of the 3 archetypes and the ternary plot (black circles and red triangles indicate women and men, respectively), where α values are displayed. The feet distribute more densely between archetype 2 and 3. The first archetype correspond with the solid black contour, the second one with the dashed red contour, while the third one is the dotted green contour.

4 Conclusions

AA for contour functions has been proposed. We have applied it to a novel data set of foot images. Knowing the extreme shapes can help shoe designers adjust their designs to a larger number of the population and be aware of the characteristics of the users that will not be comfortable to use them, whether to consider a line of special sizes or modify any shoe feature to cover more customers. As future work, we can extend AA to surface functions in order to analyze 3D foot shapes.

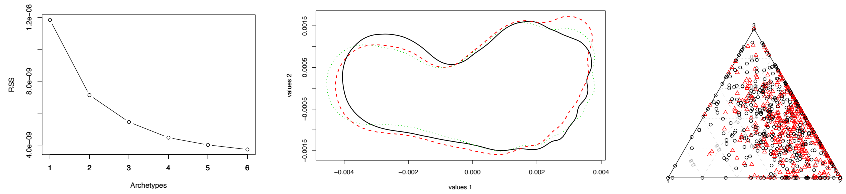

Figure 1. *Screeplot (left-handed). Archetypes (central panel) and ternary plot (right-handed). See text for details.*

5 Acknowledgments

This work is supported by the grant DPI2017-87333-R from the Spanish Ministry of Science, Innovation and Universities (AEI/FEDER, EU) and UJIB2017-13 from Universitat Jaume I.

References

CUTLER, A., & BREIMAN, L. 1994. Archetypal Analysis. *Technometrics*, **36**(4), 338–347.

EPIFANIO, I. 2016. Functional archetype and archetypoid analysis. *Computational Statistics & Data Analysis*, **104**, 24 – 34.

EPIFANIO, I., & VENTURA-CAMPOS, N. 2011. Functional data analysis in shape analysis. *Computational Statistics & Data Analysis*, **55**(9), 2758–2773.

EPIFANIO, I., IBÁÑEZ, M. V., & SIMÓ, A. 2018. Archetypal shapes based on landmarks and extension to handle missing data. *Advances in Data Analysis and Classification*, **12**(3), 705–735.

KINDRATENKO, V. V. 2003. On Using Functions to Describe the Shape. *Journal of Mathematical Imaging and Vision*, **18**, 225–245.

RAMSAY, J. O., & SILVERMAN, B. W. 2005. *Functional Data Analysis*. 2nd edn. Springer.

STOYAN, D., & STOYAN, H. 1994. *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics*. Wiley.

VINUÉ, G., & EPIFANIO, I. 2017. Archetypoid Analysis for Sports Analytics. *Data Mining and Knowledge Discovery*, **31**(6), 1643–1677.