Analysis of 2D Foot morphology by functional archetypal analysis

A. Alcacer[1], *I. Epifanio*[1], *M.V. Ibáñez*[1], *A. Simó*[1]
aalcacer@uji.es, epifanio@uji.es, mibanez@uji.es, simo@uji.es
[1]Departament de Matemàtiques, Universitat Jaume I, Spain

Abstract

Improving the fit in footwear is an important issue both for manufacturers and customers. For that reason, an anthropometric database of the adult Spanish population is analyzed. Shapes of feet are represented by 2D images. Archetype analysis is the appropriate statistical tool to describe the extreme patterns. We have extended Archetype analysis to functions with two arguments. We have applied this new methodology to the images of feet of women and men.

Keywords: Archetype analysis, shape analysis, functional data analysis.

1. Introduction

Knowledge of foot shape has a great relevance for the appropriate design of footwear. It is a very important 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 target 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. Archetype Analysis (AA) [1] is an unsupervised data mining technique that describes instances of a sample as a convex combination of certain number of elements called archetypes, which in turn, are convex combinations of the observations in the sample. This multivariate technique was extended to functional data with one argument ([3, 5]).

A 3D anthropometric study of the feet of Spanish adult population was carried out by the Instituto de Biomecánica de Valencia. We consider their footprints, i.e. 2D images, which can be seen as functions with two arguments and functional data techniques ([4]) are used. The purpose of this work is to extend functional archetype analysis (FAA) to 2D binary images, i.e. to functions with two arguments, and to apply it to the novel data set. Furthermore, FAA will help an image data set easier to understand, displaying and describing their features.

Section 2 describes our data. In Section 3 the methodology is introduced, and results are analyzed in Section 4. Finally, some conclusions are given in Section 5.

2. Data

Footprints have been extracted from an anthropometric database of 775 3D right foot scans representing Spanish adult female and male population, 382 correspond to women and 393 to men. Data was collected in different regions across Spain at shoe shops and workplaces using an INFOOT laser scanner. The binary images have been centered and scaled in order to remove the effects of translations and changes of scale and to consider only the shape, as explained by [2].

3. Methodology

3.1 AA for (standard) multivariate data

Let $\mathbf{X}$ be an $n \times m$ matrix with n cases and m variables. The objective of AA is 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):

$$\sum_{j=1}^{k} \alpha_{ij} \mathbf{z}_j, \tag{1}$$

with the mixture coefficients contained in the $n \times p$ 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:

$$\mathbf{z}_j = \sum_{l=1}^{n} \beta_{jl} \mathbf{x}_l. \tag{2}$$

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) $\displaystyle\sum_{j=1}^{k} \alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ for $i = 1, \ldots, n$ and 2) $\displaystyle\sum_{l=1}^{n} \beta_{jl} = 1$ with $\beta_{jl} \geq 0$ for $j = 1, \ldots, k$.

3.2 AA for functional data

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 [3] for details about extension of AA to functional data. Here, we extend the work by [3] to functional data with two arguments.

Let $f_i(s,t)$ be a function defined in $[a,b] \times [c,d]$. Its squared norm is $\|f_i\|^2 = \int_c^d \int_a^b f_i(s,t)^2 ds dt$. In the basis approach, each function f_i is expressed as a linear combination of known basis functions B_h with $h = 1, ..., m$: $f_i(s,t) = \sum_{h=1}^{m} b_i^h B_h(s,t) = \mathbf{b}_i' \mathbf{B}$, where $'$ stands for transpose and $\mathbf{b_i}$ indicates the vector of length m of the coefficients and $\mathbf{B}$ the functional vector whose elements are the basis functions. In FAA with two arguments the objective functions is: RSS = $\displaystyle\sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} z_j\|^2$ $= \sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} f_l\|^2 = \sum_{i=1}^{n} \|\mathbf{b'}_i \mathbf{B} - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b'}_l \mathbf{B}\|^{\mathbf{2}} = \sum_{i=1}^{n} \|(\mathbf{b'}_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b'}_l) \mathbf{B}\|^{\mathbf{2}} = \sum_{i=1}^{n} \|\mathbf{a'}_i \mathbf{B}\|^2 = \sum_{i=1}^{n} < \mathbf{a'}_i \mathbf{B}, \mathbf{a'}_i \mathbf{B} > = \sum_{i=1}^{n} \mathbf{a'}_i \mathbf{W} \mathbf{a}_i$, with the corresponding constraints for α and β; and where $\mathbf{a'}_i = \mathbf{b'}_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b'}_l$ and $\mathbf{W}$ is the order m symmetric matrix with elements $w_{m_1,m_2} = \int_c^d \int_a^b B_{m_1} B_{m_2} ds dt$, i.e. the matrix containing the inner products of the pairs of basis functions. 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.

4. Results

Images have been expressed in the 2D Discrete Cosine Transform, which is an orthonormal base. We have only considered the 0.62% of the first coefficients, since details of feet are kept with these number of coefficients. We have applied FAA for the groups of women and men, separately. In both cases, we have represented the screeplot, with the number of archetypes versus the respective RSS, and we have found an elbow at $k = 2$, in both cases. Figure 1 shows the archetypes (A_1 and A_2) for women and men. The intersection between both archetypes is displayed in white, while the set difference A_1 A_2 is displayed in red, and the set difference A_2 A_1 is displayed in blue.

Results for women and men are very similar, so they are commented both together. In both cases, the difference in shapes of A_1 and A_2 are in the lateral zones in front of the top and bottom zones of the foot image. We have also displayed the α values of A_1 versus the foot length for women and men, and points are uniformly distributed. Therefore, the different shapes are found in all the sizes.

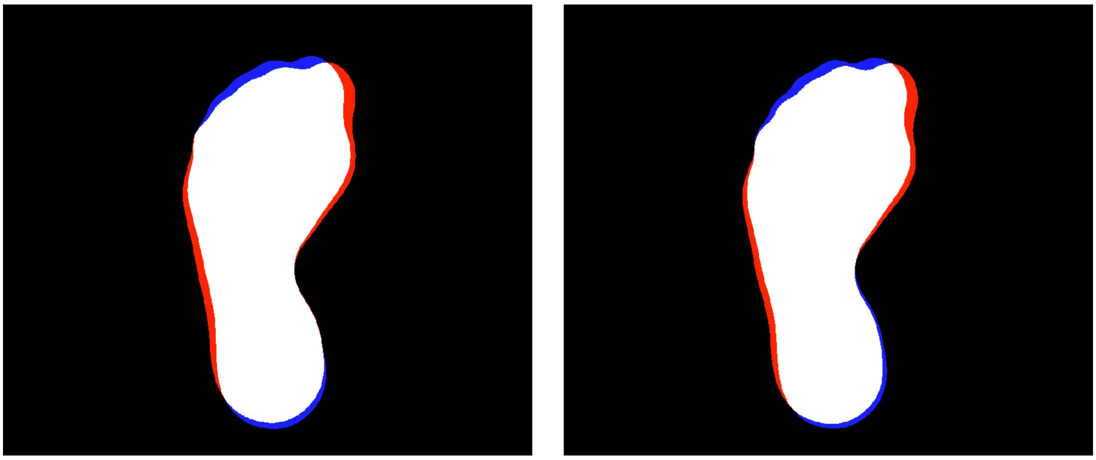

Figure 1: Archetypes for women (left-handed) and men (right-handed). See text for details.

5. Conclusions

AA has been extended to functions with two arguments, and 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 functions with three arguments in order to analyze 3D foot shapes.

Acknowledgments

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

Bibliography

[1] Cutler, A., Breiman, L.: Archetypal Analysis. Technometrics **36**(4), 338–347 (1994)

[2] Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis: With Applications in R, 2nd edn. Wiley (2016)

[3] Epifanio, I.: Functional archetype and archetypoid analysis. Computational Statistics & Data Analysis **104**, 24 – 34 (2016)

[4] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer (2005)

[5] Vinué, G., Epifanio, I.: Archetypoid analysis for sports analytics. Data Mining and Knowledge Discovery **31**(6), 1643–1677 (2017)

[6] Vinué, G., Epifanio, I., Alemany, S.: Archetypoids: A new approach to define representative archetypal data. Computational Statistics & Data Analysis **87**, 102 – 115 (2015)