Functional Data Analysis and Cluster Analysis: a Marriage with some Constraints

Simone VANTINI

joint with L.M. SANGALLI, P. SECCHI, V. VITELLI

MOX – Dept. of Mathematics, Politecnico di Milano
A marriage…

Cluster Analysis (Identification of groups) + Functional Data Analysis (Analysis of functions) → Functional Data Clustering (Identification of groups of functions)

They lived happily ever after…
There are indeed **some serious issues** going on in this marriage:

- **No model at hand**
  (probability density function does not exist, basically impossible to assess the validity of the model)
- **Choice of the Smoothing**
  (functions and their derivatives need to be estimated from point-wise noisy evaluations)
- **Choice of the Metric**
  (huge variety of distances wrt to the multivariate framework)
- **Choice of the Group of Warping Functions**
  (data should generally be horizontally aligned)

**An Example:**

*K-mean Clustering of Misaligned Data Using a Derivative-based Metric*
Choice of the Smoothing

Same row data (smoothed using B-splines with different number of knots)

Weak Smoothing
One Cluster

Strong Smoothing
Two Clusters

Choice of the Metric

- $\xi_1(s), \ldots, \xi_7(s) : 1$D i.i.d. random variable
- $\{e_k, k \geq 1\} :$ Fourier basis

$$\chi_s^{(1)} = \sum_{k=1}^{7} \xi_k(s)e_k$$

- From an $L^2$ perspective the two datasets shows the same variability across curves.

- From an $H^1$ perspective the two datasets shows a different variability across curves (lower the former, larger the latter).


simone.vantini@polimi.it
Registration of a set of functions

Find suitable warping functions \( h_1, \ldots, h_n \) such that \( c_1 \circ h_1, \ldots, c_n \circ h_n \) are the most similar.

**Landmark Approach** (similar means that functions are warped along the x-axis such that each (known) landmark occurs at the same point along the x-axis)

**Continuous Approach** (similar means that functions are warped along the x-axis such that for each point along the x-axis functions present close values along the y-axis)
Choice of the Group of Warping Functions

200 periodic curves (spike-trains of neuronal activity)

Not Aligned Curves
Four Clusters

Aligned Curves
Two Clusters

Clustering of Misaligned Functional Data

- The Algorithm
- A Toy Example
- The Theory
- The Case Study
The Algorithm
Functional Clustering

Alignment / Registration

K-mean Clustering

K-mean Alignment

Continuous Alignment
Goals

Goal of Continuous Alignment: Decoupling Phase and Amplitude Variability

Goal of $K$-mean Clustering: Decoupling Within and Between-cluster (Amplitude) Variability

Goal of $K$-mean Alignment: Decoupling Phase Variability, Within-cluster Amplitude Variability, and Between-cluster Amplitude Variability
Continuous Alignment
(e.g. Sangalli, Secchi, Vantini, Veneziani 2009)

Estimation

Estimate the template curve $c_0$

Maximization

For each curve, find $h_i$ that maximizes similarity between template curve $c_0$ and patient warped curve $c_i \circ h_i$

$n$ registered curves and $n$ warping functions

$n$ unregistered curves
**K-mean Clustering**
(e.g. Tarpey and Kinateder 2003)

**Estimation**

Estimate the \( K \) centroids curves \( \{c_0^k\}_{k=1,2,...,K} \)

**Assignment**

Assign \( c_i \) to the \( k \)-th cluster if the similarity between \( c_0^k \) and \( c_i \) is maximal over \( k = 1, 2, \ldots, K \)

\( n \) unregistered curves

\( K \) clusters
**K-mean Clustering vs Continuous Alignment**

### Continuous Alignment
- **Estimation**
  - Estimate the template curve \( c_0 \)

- **Maximization**
  - For each curve, find \( h_i \) that maximizes similarity between template curve \( c_0 \) and patient warped curve \( c_i \circ h_i \)

- **Output**
  - \( n \) registered curves and \( n \) warping functions

### K-mean Clustering
- **Estimation**
  - Estimate the \( K \) centroids curves \( \{c_0^k\}_{k=1,2,...,K} \)

- **Assignment**
  - Assign \( c_i \) to the \( k \)-th cluster if the similarity between \( c_0^k \) and \( c_i \) is maximal over \( k = 1, 2, \ldots, K \)

- **Output**
  - \( K \) clusters
K-mean Alignment

Estimation

Estimate the $K$ template/centroid curves $\{c^k_0\}_{k=1,2,...,K}$

Maximization

For each curve, find $h^k_i$ that maximizes similarity between each template curve $c^k_0$ and the candidate warped curve $c_i \circ h^k_i$

Assignment

Assign $c_i$ to the $k$-th cluster if the similarity between $c^k_0$ and $c_i \circ h^k_i$ is maximal over $k = 1, 2, \ldots, K$ and then warp $c_i$ along $h_i = h^k_i$

$n$ unregistered curves

$n$ registered curves, $n$ warping functions and $K$ clusters

$K = 1$ Continuous Registration Algorithm (e.g. Sangalli et al. 2009)

$W = \{1\}$ Functional $K$-mean Clustering (e.g. Tarpey and Kinateder 2003)
Functional Clustering  \hspace{5cm} \text{Alignment / Registration}

- $K$-mean Clustering
- $K$-mean Alignment
- Continuous Alignment

It is a $K$-mean Clustering Algorithm where warping is allowed

It is an Alignment Algorithm with $K$ templates
A Toy Example
A Simulated Toy Example: Simulation Details

2 Amplitude Clusters (2 template curves) with further clustering in the phase

Variability in both Amplitude and Phase
A Simulated Toy Example: Algorithm Results

Aligned and clustered curves

Warping functions

K = 1

K = 2

K = 3
A Simulated Toy Example: Algorithm Results

Boxplots of the similarity indices between curves and templates

Means of the similarity indices between curves and templates
The Theory
Abstract Visualization of Alignment

Original Functions

Aligned Functions

Equivalence Classes generated by the Action of the Group of Warping Functions
Meta-Equivalence Result

Analysis of a registered functional data set with respect to the metric $d$

(e.g., K-mean Alignment in the Functional Space)

Analysis of a set of equivalence classes (induced by the application of $W$ to the original functions) with respect to a new metric $d_F$

(jointly defined by $d$ and $W$).

(e.g., K-mean Clustering in the Quotient Space)
Meta-Equivalence Theorem

(a) $F$ is a metric space according to a distance $d : F \times F \to \mathbb{R}_0^+$ whose elements are functions: $\Omega \subseteq \mathcal{R}^p \to \Psi \subseteq \mathcal{R}^q$.

(b) $W$ is a compact (with respect to a metric $d_G$) subgroup (with respect to ordinary composition $\circ$) of the group $G$ of the continuous automorphisms: $\Omega \subseteq \mathcal{R}^p \to \Omega \subseteq \mathcal{R}^p$.

(c) $\forall f \in F$ the map $f \circ : h \in W \longmapsto (f \circ)(h) = (f \circ h) \in F$ is continuous;

(d) Given any couple of elements $f_1, f_2 \in F$ and an element $h \in W$, the distance between $f_1$ and $f_2$ is invariant under the composition of $f_1$ and $f_2$ with $h$, i.e.:

$$d(f_1, f_2) = d(f_1 \circ h, f_2 \circ h);$$

we will refer to this property as $W$-invariance of $d$.

$$d_{\mathcal{F}}([f_1], [f_2]) = \min_{h_1, h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2)$$

is a metric on the quotient set $\mathcal{F}$. 
The introduction of a metric/semi-metric \( d \) and of a group \( W \) of warping functions, with respect to which the metric/semi-metric is invariant, enables a not ambiguous definition of phase and amplitude variability.

**Total Variability**
(variability between elements of \( F \))

**Amplitude Variability**
(variability *between* equivalence classes)

**Phase Variability**
(variability *within* equivalence classes)
In many situations, $d$ is not a metric but a semi-metric, i.e.: $d(f_1, f_2) = 0 \not\Rightarrow f_1 = f_2$. The presented theory still holds if $F$ is replaced with $\overline{F} = F / \circ$.

$$f_1 \circ f_2 \iff d(f_1, f_2) = 0$$

**Hierarchy of Quotient Spaces**

Functions belonging to $F$ are grouped in equivalence classes belonging to $\overline{F}$ that are grouped in equivalence classes belonging to $\overline{F}$.
Some examples of $W$-invariant semi-metrics

<table>
<thead>
<tr>
<th>Metric / Semi-metric</th>
<th>Maximal $W$ (Phase Variability)</th>
<th>Ancillary Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$|f_1 - f_2|_{L^2}$</td>
<td>H-translations</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$|f_1' - f_2'|_{L^2}$</td>
<td>H-translations</td>
<td>V-translations</td>
</tr>
<tr>
<td>$|\left((f_1 - f_1) - (f_2 - f_2)\right)|_{L^2}$</td>
<td>H-translations</td>
<td>V-translations</td>
</tr>
<tr>
<td>$|\left((f_1' - f_1') - (f_2' - f_2')\right)|_{L^2}$</td>
<td>H-translations</td>
<td>V-translations V-linear trends</td>
</tr>
<tr>
<td>$|\frac{f_1}{</td>
<td>f_1</td>
<td>_{L^2}} - \frac{f_2}{</td>
</tr>
<tr>
<td>$|\frac{f_1'}{</td>
<td>f_1'</td>
<td>_{L^2}} - \frac{f_2'}{</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$|\text{sign}(f_1')\sqrt{</td>
<td>f_1</td>
<td>} - \text{sign}(f_2')\sqrt{</td>
</tr>
</tbody>
</table>
The Case Study
Cerebral aneurysms: malformations of cerebral arteries, in particular of arteries belonging to or connected to the Circle of Willis.

**EPIDEMIOLOGICAL STATISTICS**

- Incidence rate of cerebral aneurysms: 1/20 people
- Incidence rate of ruptured cerebral aneurysms per year: 1/10000 people per year
- Mortality due to a ruptured aneurysm:
  - > 50%: Out of 9 patients with a ruptured aneurysm:
    - 3 are expected to die before arriving at the hospital
      - 2 to die after having arrived at the hospital
      - 2 to survive with permanent cerebral damages
    - 2 to survive without permanent cerebral damages
Pathological Classification

Upper
High Aneurysm
33

Lower
Low Aneurysm
25

Healthy
No Aneurysm
7
From X-rays to Centerlines

1. Contrast Fluid Injections
2. X-rays (one direction)
3. 3d-array (one slice)
4. Gradient 3d-array (one slice)

Surface Points

Voronoi Diagram

Eikonal Equation

Centerline
Discrete Data

Observational Study conducted at Ospedale Ca’ Granda Niguarda – Milano relative to 65 patients hospitalized from September 2002 to October 2005.
The sample of 65 ICA: each patient is represented by the centerline of their ICA.
ICA centerline First Derivatives
Theoretical Choices

**Similarity Index between Curves**

\[ \rho(c_i, c_j) = \frac{1}{3} \left[ \rho(x_i, x_j) + \rho(y_i, y_j) + \rho(z_i, z_j) \right] \]

**Group of Warping Functions**

\[ W = \{ h : h(s) = ms + p \text{ with } m \in \mathbb{R}^+, p \in \mathbb{R} \} \]

**Properties**

\[ |\rho(c_i, c_j)| \leq 1 \]

\[ \rho(c_i, c_j) = 1 \iff \exists A \in (\mathbb{R}^+)^3, B \in \mathbb{R}^3 : \begin{cases} x_i = A_x x_j + B_x \\ y_i = A_y y_j + B_y \\ z_i = A_z z_j + B_z \end{cases} \]

\[ \rho(c_i, c_j) = \rho(c_i \circ h, c_j \circ h) \quad \forall h \in W \]

\[ \rho(c_i \circ h, c_j) = \rho(c_i, c_j \circ h^{-1}) \quad \forall h \in W \]

\[ \sup_{h \in W} \rho(c_i \circ h, c_j) = \sup_{h \in W} \rho(c_i, c_j \circ h) \]

**Minimal Joint Properties**

\[ \text{find } \varphi = \{ \varphi_1, \ldots, \varphi_k \} \subset \mathcal{C} \text{ and } \mathbf{h} = \{ h_1, \ldots, h_N \} \subset W \text{ such that} \]

\[ \frac{1}{N} \sum_{i=1}^{N} \rho(\varphi_{\lambda(\varphi, c_i)}, c_i \circ h_i) \geq \frac{1}{N} \sum_{i=1}^{N} \rho(\psi_{\lambda(\varphi, c_i)}, c_i \circ g_i) \]
K-mean Alignment Performances

Original

Similarity indexes

Mean similarity indexes

with alignment

without alignment
One-mean Alignment

Original

\[ X' \]

\[ Y' \]

\[ Z' \]

\[ x' \]

\[ y' \]

\[ z' \]

\[ k = 1 \]
One-mean Alignment: aneurysm location on registered ICA radius and curvature profiles

Unregistered Radius and Curvature Profiles

Registered Radius and Curvature Profiles

Unregistered Profiles

Registered Profiles

Radius Profiles

Curvature Profiles

$R(s)$

$\hat{R}(s)$

$C(s)$

$\hat{C}(s)$
One-mean Alignment
Two-mean Alignment

Original

\[ x', y', z' \]

\[ -100 \quad -80 \quad -60 \quad -40 \quad -20 \quad 0 \]

\[ 0 \quad 1 \]

\[ -10 \quad 0 \quad 10 \]

\[ -10 \quad 0 \quad 10 \]

\[ -10 \quad 0 \quad 10 \]

\[ -100 \quad -80 \quad -60 \quad -40 \quad -20 \quad 0 \]

\[ 0 \quad 1 \]

\[ 0 \quad 1 \]

\[ 0 \quad 1 \]

\[ k = 2 \]

\[ x', y', z' \]

\[ -100 \quad -80 \quad -60 \quad -40 \quad -20 \quad 0 \]

\[ 0 \quad 1 \]

\[ -10 \quad 0 \quad 10 \]

\[ -10 \quad 0 \quad 10 \]

\[ -10 \quad 0 \quad 10 \]
Two-mean Alignment vs Two-mean Clustering

Mean similarity indexes

- with alignment
- without alignment

Two-mean Alignment

Two-mean Clustering
Clustering

Two-mean Alignment

Clusters that are morphologically different

30 S-shaped ICAs vs 35 Ω-shaped ICAs

Krayenbuehl et al. (1982)

<table>
<thead>
<tr>
<th></th>
<th>No Aneurysm</th>
<th>Aneurysm along ICA</th>
<th>Aneurysm downstream ICA</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-shaped ICAs</td>
<td>100%</td>
<td>52%</td>
<td>30%</td>
</tr>
<tr>
<td>Ω-shaped ICAs</td>
<td>0%</td>
<td>48%</td>
<td>70%</td>
</tr>
</tbody>
</table>

\[ P \text{-value of Pearson's Chi-squared test for independence equal to 0.0013} \]

Fluid-dynamical interpretation of the onset of cerebral aneurysms


• AneuRisk65 data are freely downloadable at http://mox.polimi.it/it/progetti/aneurisk/
  http://ecm2.mathcs.emory.edu/aneuriskweb/a65