Manifold Matching:
Joint Optimization
of
Fidelity & Commensurability

Carey E. Priebe
Department of Applied Mathematics & Statistics
Johns Hopkins University

July 15, 2010
Pittsburgh
Collaborators

David J. Marchette
Zhiliang Ma
Sancar Adali
&c.

Support: AFOSR, NSSEFF, ONR, HLTCOE, ASEE
Problem Formulation

Given $x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK}$, $i = 1, \ldots, n$
Problem Formulation

Given \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK}, \ i = 1, \ldots, n \)

- \( n \) objects are each measured under \( K \) different conditions
- \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK} \)
  
  denotes \( K \) matched feature vectors representing a single object \( O_i \)
- \( x_{ik} \in \Xi_k \)
Problem Formulation

Given \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK}, \ i = 1, \ldots, n \)

- \( n \) objects are each measured under \( K \) different conditions
- \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK} \) denotes \( K \) matched feature vectors representing a single object \( O_i \)
- \( x_{ik} \in \Xi_k \)
- \( K \) new measurements \( \{y_k\}_{k=1}^K, \ y_k \in \Xi_k \)
Problem Formulation

Given \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK}, \ i = 1, \ldots, n \)

- \( n \) objects are each measured under \( K \) different conditions
- \( x_{i1} \sim \cdots \sim x_{ik} \sim \cdots \sim x_{iK} \)
denotes \( K \) matched feature vectors representing a single object \( O_i \)
- \( x_{ik} \in \Xi_k \)
- \( K \) new measurements \( \{y_k\}_{k=1}^K, \ y_k \in \Xi_k \)

Question

Are \( \{y_k\}_{k=1}^K \) matched feature vectors representing a single object measured under \( K \) conditions?
Hypotheses

<table>
<thead>
<tr>
<th></th>
<th>$\Xi_1$</th>
<th>$\cdots$</th>
<th>$\Xi_K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object $O_1$</td>
<td>$x_{11}$</td>
<td>$\sim \cdots \sim$</td>
<td>$x_{1K}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>Object $O_n$</td>
<td>$x_{n1}$</td>
<td>$\sim \cdots \sim$</td>
<td>$x_{nK}$</td>
</tr>
</tbody>
</table>
Hypotheses

\[ \Xi_1 \quad \cdots \quad \Xi_K \]

Object \( O_1 \quad x_{11} \sim \cdots \sim x_{1K} \)

\vdots

Object \( O_n \quad x_{n1} \sim \cdots \sim x_{nK} \)

- Each space \( \Xi_k \) comes with a dissimilarity \( \delta_k \), yielding dissimilarity matrices \( \Delta_1, \cdots, \Delta_K \)
Hypotheses

\[ \Xi_1 \quad \cdots \quad \Xi_K \]

Object \( O_1 \quad x_{11} \sim \cdots \sim x_{1K} \)

\vdots

\vdots

Object \( O_n \quad x_{n1} \sim \cdots \sim x_{nK} \)

- Each space \( \Xi_k \) comes with a dissimilarity \( \delta_k \), yielding dissimilarity matrices \( \Delta_1, \cdots, \Delta_K \)
- Given new measurements \( \{y_k\}_{k=1}^K \) we can obtain within-condition dissimilarities

\[ \delta_k(y_k, x_{ik}), \ i = 1, \ldots, n, \ k = 1, \ldots, K \]
Hypotheses

Each space $\Xi_k$ comes with a dissimilarity $\delta_k$, yielding dissimilarity matrices $\Delta_1, \cdots, \Delta_K$.

Given new measurements $\{y_k\}_{k=1}^K$ we can obtain within-condition dissimilarities

$$\delta_k(y_k, x_{ik}), \ i = 1, \ldots, n, \ k = 1, \ldots, K$$

Goal ($K = 2$): determine whether $y_1$ and $y_2$ are a match.
Hypotheses

$$\Xi_1 \quad \cdots \quad \Xi_K$$
Object $$O_1 \quad x_{11} \sim \cdots \sim x_{1K}$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$
Object $$O_n \quad x_{n1} \sim \cdots \sim x_{nK}$$

- Each space $$\Xi_k$$ comes with a dissimilarity $$\delta_k$$, yielding dissimilarity matrices $$\Delta_1, \cdots, \Delta_K$$
- Given new measurements $$\{y_k\}_{k=1}^K$$ we can obtain within-condition dissimilarities

$$\delta_k(y_k, x_{ik}), \ i = 1, \ldots, n, \ k = 1, \ldots, K$$

- Goal ($$K = 2$$): determine whether $$y_1$$ and $$y_2$$ are a match

$$H_0 : y_1 \sim y_2 \quad \text{versus} \quad H_A : y_1 \not\sim y_2$$

(we control the probability of missing a true match)
what are these “conditions” and what does it mean to be “matched”

- let condition be language for a text document, and “matched” mean “on the same topic”
- let condition be modality for an photo, and “matched” mean “of the same person”
  - indoor lighting vs outdoor lighting
  - two cameras of different quality
  - passport photos and airport surveillance photos
- let condition 1 be wiki text document and condition 2 be wiki hyperlink structure
- let condition 1 be text document and condition 2 be photo
- ... or just a single space with multiple dissimilarities
The English is clear enough to lorry drivers — but the Welsh reads “I am not in the office at the moment. Send any work to be translated.”

<http://news.bbc.co.uk/2/hi/uk_news/wales/7702913.stm>
Conditional distributions are induced by maps $\pi_k$ from “object space” $\Xi$

\[ \Xi \quad \xrightarrow{\pi_1} \quad \ldots \quad \xrightarrow{\pi_K} \quad \Xi_1 \quad \Xi_K \]

Conditional spaces $\Xi_k$ are *not* commensurate
Conditional distributions are induced by maps $\pi_k$ from “object space” $\Xi$

\[ \Xi \xrightarrow{\pi_1} \cdots \xrightarrow{\pi_K} \Xi_1 \xrightarrow{\phi?} \Xi_K \]

Conditional spaces $\Xi_k$ are *not* commensurate
Dirichlet Setting

Let $S^p$ be the standard $p$-simplex in $\mathbb{R}^{p+1}$

Let $\Xi_1 = S^p$ and $\Xi_2 = S^p$
(but the fact that the two spaces are the same
is unknown to the algorithms ...)

Let $\alpha_i \sim iid \text{ Dirichlet}(1)$ represent $n$ “objects” or “topics”
Let $X_{ik} \sim iid \text{ Dirichlet}(r\alpha_i + 1)$ represent $K$ languages (WCHs)
Dirichlet Setting

Let $S^p$ be the standard $p$-simplex in $\mathbb{R}^{p+1}$

Let $\Xi_1 = S^p$ and $\Xi_2 = S^p$
(but the fact that the two spaces are the same is unknown to the algorithms ...)

Let $\alpha_i \sim^\text{iid} \text{Dirichlet}(1)$ represent $n$ “objects” or “topics”
Let $X_{ik} \sim^\text{iid} \text{Dirichlet}(r\alpha_i + 1)$ represent $K$ languages (WCHs)

• $r$ controls “what it means to be matched”
  (document variability & translation quality analogy)
Matched points are used to define maps $\rho_k$ to the same space $\mathcal{X}$ (with distance $d$)

\[
\begin{array}{c}
\Xi \\
\pi_1 \quad \ldots \quad \pi_K \\
\Xi_1 & \quad \ldots & \quad \Xi_K \\
\rho_1 & \quad \ldots & \quad \rho_K \\
\mathcal{X}
\end{array}
\]

Reject for $d(\tilde{y}_1, \tilde{y}_2)$ “large”
Matched points are used to define maps $\rho_k$ to the same space $\mathcal{X}$ (with distance $d$)

\[ \Xi \xrightarrow{\pi_1} \cdots \xrightarrow{\pi_K} \Xi_1 \xrightarrow{\rho_1} \cdots \xrightarrow{\rho_K} \Xi_K \]

\[ \mathcal{X} = \mathbb{R}^d \]

Reject for $d(\tilde{y}_1, \tilde{y}_2)$ "large"
multidimensional scaling yields high-dimensional embeddings: $\Delta_1 \mapsto X'_1$ and $\Delta_2 \mapsto X'_2$

- Canonical correlation finds $U_1 : X'_1 \mapsto X_1$ and $U_2 : X'_2 \mapsto X_2$ to maximize correlation

- Out-of-sample embedding: $y_1 \mapsto y'_1$, $y_2 \mapsto y'_2$

- Both $\tilde{y}_1 = U_1^T y'_1$ and $\tilde{y}_2 = U_2^T y'_2$ are in $\mathbb{R}^d$ with same coordinate system (i.e., they are commensurate)

- Reject for $d(\tilde{y}_1, \tilde{y}_2)$ “large”
procrustes $\circ$ mds

- Multidimensional scaling yields low-dimensional embeddings: $\Delta_1 \mapsto X_1$ and $\Delta_2 \mapsto X_2$
- Procrustes$(X_1, X_2)$ yields
  \[ Q^* = \arg \min_{Q^TQ=I} \|X_1 - X_2Q\|_F \]
- Out-of-sample embedding: $\mathbf{y}_1 \mapsto \tilde{\mathbf{y}}_1$, $\mathbf{y}_2 \mapsto \tilde{\mathbf{y}}'_2$
- Both $\tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2 = Q^*\tilde{\mathbf{y}}'_2$ are in $\mathbb{R}^d$ with same coordinate system (i.e., they are commensurate)
- Reject for $d(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2)$ “large”
fidelity & commensurability

Fidelity is how well the mapping preserves original dissimilarities; our within-condition \textit{fidelity error} is given by

$$
\epsilon_{f_k} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} (d(\tilde{x}_{ik}, \tilde{x}_{jk}) - \delta_k(x_{ik}, x_{jk}))^2.
$$

Commensurability is how well the mapping preserves matchedness; our between-condition \textit{commensurability error} is given by

$$
\epsilon_{c_{k_1k_2}} = \frac{1}{n} \sum_{1 \leq i \leq n} (d(\tilde{x}_{ik_1}, \tilde{x}_{ik_2}) - \delta_{k_1k_2}(x_{ik_1}, x_{ik_2}))^2.
$$

Alas, $\delta_{k_1k_2}$ does not exist; however, our story seems to suggest that it might be reasonable to let $\delta_{k_1k_2}(x_{ik_1}, x_{ik_2}) = 0$ for all $i, k_1, k_2$.

NB: There is also between-condition \textit{separability error} given by

$$
\epsilon_{s_{k_1k_2}} = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} (d(\tilde{x}_{ik_1}, \tilde{x}_{jk_2}) - \delta_{k_1k_2}(x_{ik_1}, x_{jk_2}))^2.
$$
Methodological Comparison

- **canonical correlation** optimizes commensurability
  *without regard for fidelity*

- **procrustes ○ mds** optimizes fidelity
  *without regard for commensurability*
Methodological Comparison

- **canonical correlation** optimizes commensurability
  *without regard for fidelity*

- **procrustes ○ mds** optimizes fidelity
  *without regard for commensurability*

- compare: *joint optimization of fidelity & commensurability* ...
Omnibus Embedding Approach

\[
\begin{bmatrix}
\Delta_1 & W \\
W^T & \Delta_2
\end{bmatrix}
\]

Under “matched” assumption, impute dissimilarities \(\delta_{12}(x_{i11}, x_{i22})\) to obtain an omnibus dissimilarity matrix \(M\).

- Embed \(M\) as \(2n\) points in \(\mathbb{R}^d\).
- Let \(u_{i1} = \delta_1(y_1, x_{i1})\) and \(v_{i2} = \delta_2(y_2, x_{i2})\).
- Under \(H_0: y_1 \sim y_2\), impute \(v_{i1} = \delta_{12}(y_1, x_{i2})\) and \(u_{i2} = \delta_{12}(y_2, x_{i1})\).
- Out-of-sample embedding of \((u_1^T, v_1^T)^T\) and \((u_2^T, v_2^T)^T\) yields \(\tilde{y}_1\) and \(\tilde{y}_2\).
Simulation results indicate that joint optimization of fidelity & commensurability via omnibus embedding approach is (for this case) superior to **canonical correlation** and **procrustes•mds**.
Spurious Correlation Phenomenon

Let $\Xi_k = S^{p+q} = S^p \times S^q$;

$S^p$ encodes “signal” and $S^q$ encodes “noise”

On $S^p$, let $\alpha_i \sim iid \text{ Dirichlet}(1)$ and $X_{ik}^1 \sim iid \text{ Dirichlet}(r\alpha_i + 1)$ (signal, as before)

On $S^q$, let $X_{ik}^2 \sim iid \text{ Dirichlet}(1)$ (pure noise)

For $c \in [0, 1]$, let $X_{ik} = [(1 - c)X_{ik}^1, cX_{ik}^2]$
Incommensurability Phenomenon I

\[ \mathcal{F}_1 \xleftarrow{\mathcal{F}_2} \xrightarrow{\mathcal{F}_3} \xrightarrow{\mathcal{F}_4} \]
Incommensurability Phenomenon II

Dirichlet

Scale & Polarity

Procrustes

\[ \|M_1 - M_2P_{S,\ell}\| \]

\[ \|M_1 - M_2Q\| \]

\( \text{Sqrt of Commensurability Error} \)
Experimental Data

Wikipedia Documents

- Wikipedia is a free, multilingual encyclopedia project
- 13 million articles (2.9 million in the English Wikipedia) have been written collaboratively by volunteers around the world
- A Wikipedia document has information regarding
  - textual content of the document
  - links in the document to other documents
- Consider a subset of English and French Wikipedias that are 1-1 correspondent
- We take the (directed) 2-neighborhood of the document “Algebraic Geometry” in the English Wikipedia, with the associated documents in the French Wikipedia ($n = 1382$)
Experimental results indicate that joint optimization of fidelity & commensurability via omnibus embedding approach is (for this case) superior to canonical correlation and procrustes−mds.
Exploitation Task: Classification

$$(X, Y, Z) \sim F_{XYZ}$$

$$Y: \mathcal{J} \cup \overline{\mathcal{J}}$$

$$Z: \mathcal{J} \rightarrow \{0, 1\}$$

$$X|Z = 2: \mathcal{J} \rightarrow \sim 2$$

- Available training data:
  $$\mathcal{T}_0 = \{ (x_i, y_i, \xi, Z_i = 0) \}$$
  $$\rightarrow 0 = \{ (x_i, y_i, \xi, Z_i = 0) \}$$
  $$\rightarrow 0 = \{ (x_i, y_i, \xi, Z_i = 1) \}$$

- Data to be classified:
  $$(X, Y, \xi, Z = 1)$$

- No training data $\mathcal{T}$ available of this type ($Z = 1$) for these classes ($Y \in \overline{\mathcal{J}}$)

1. $$\tilde{g}_1 = \hat{g}_0 \cdot \omega \Rightarrow \tilde{g}_0 \left( \hat{w}(x), \tilde{g}_0 \right)$$

2. LDA approach (explicitly) identifies portions $P(x, Z)$ for which $\tilde{g}_0$ identity is a satisfactory estimate

3. $$\tilde{g}_2 = \tilde{g}_0 \Rightarrow \tilde{g}_2$$

   "Known" estimate

   Unknown & needed
Integrated Sensing and Processing

\[ \Delta(\theta) \]

\[ \Xi_1 \rightarrow \Xi_2 \rightarrow \cdots \rightarrow \Xi_{K-1} \rightarrow \Xi_K \]

ISP

\[ \rho_1 \rightarrow \rho_2 \rightarrow \cdots \rightarrow \rho_{K-1} \rightarrow \rho_K \]

MM

\[ W \]

Fidelity

Commensurability

Separability
“The wealth of your practical experience with sane and interesting problems will give to mathematics a new direction and a new impetus.”

– Leopold Kronecker to Hermann von Helmholtz –