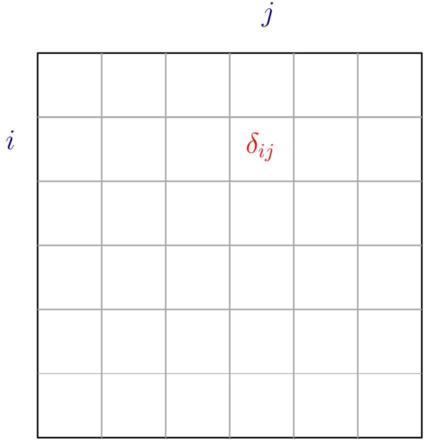
Joint Optimization of Fidelity And Commensurability for Manifold Alignment

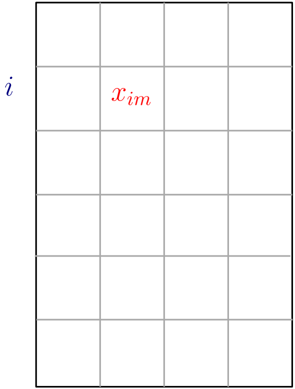
Sancar Adali, Carey E. Priebe

Department of Applied Mathematics & Statistics
Johns Hopkins University

July 8, 2012

Dissimilarity Representation vs Feature Representation





$$\Delta = [\delta_{ij}]$$

$$x_i \in \mathbf{R}^d$$

$$\delta_{ij} >= 0 \ , \ \delta_{ii} = 0 \ , \ \delta_{ij} = \delta_{ji}$$
$$i = \{1, \ldots, n\} \quad j = \{1, \ldots, n\}$$

How do we embed dissimilarities in Euclidean space?

Multidimensional Scaling [BorgGroenen1997]

Given $n \times n$ dissimilarity matrix
$\Delta = [\delta_{st}]; s = 1, \ldots, n;\ t = 1, \ldots, n,$

- Finds an embedding of the dissimilarities in the Euclidean space (with a chosen dimension d) such that the distances between the embeddings are as close as possible (in various senses) to the original dissimilarities in Δ.
- One function to measure closeness is the weighted raw-stress:

$$\sigma_W(X) = \sum_{1 \leq s \leq n; 1 \leq t \leq n} w_{st}(d_{st}(X) - \delta_{st})^2$$

where $d_{st}(X)$ is the Euclidean distance between s^{th} and t^{th} embedded points, and
weight matrix $W = [w_{st}]; s = 1, \ldots, n;\ t = 1, \ldots, n.$

How do we embed dissimilarities in Euclidean space?

Multidimensional Scaling [BorgGroenen1997]

Given $n \times n$ dissimilarity matrix
$\Delta = [\delta_{st}]; s = 1, \ldots, n;\ t = 1, \ldots, n,$

- Finds an embedding of the dissimilarities in the Euclidean space (with a chosen dimension d) such that the distances between the embeddings are as close as possible (in various senses) to the original dissimilarities in Δ.
- One function to measure closeness is the weighted raw-stress:

$$\sigma_W(X) = \sum_{1 \leq s \leq n; 1 \leq t \leq n} w_{st}(d_{st}(X) - \delta_{st})^2$$

where $d_{st}(X)$ is the Euclidean distance between s^{th} and t^{th} embedded points, and
weight matrix $W = [w_{st}]; s = 1, \ldots, n;\ t = 1, \ldots, n$.

Problem Formulation for Dissimilarity Representation

Given K $n \times n$ dissimilarity matrices

$$\mathbf{\Delta}^{(k)} = [\delta_{ij}]^{(k)}; 1 \leq i < j \leq n; k = 1, \cdots, K;$$

- Dissimilarities between n objects are each measured under K different conditions.
- The set of dissimilarities under different conditions for the same object are “matched".

Also given

$\mathbf{D}^{(k)} = \{\delta_i; i = 1, \ldots, n\} \qquad k = 1, \ldots, K$: dissimilarities between K new measurements and previous n objects under the same K conditions.

Question we can answer

Do dissimilarities $\mathbf{D}^{(k)}$ come from K new measurements that represent a single object?

Problem Formulation for Dissimilarity Representation

Given K $n \times n$ dissimilarity matrices

$$\mathbf{\Delta}^{(k)} = [\delta_{ij}]^{(k)}; 1 \leq i < j \leq n; k = 1, \cdots, K;$$

- Dissimilarities between n objects are each measured under K different conditions.
- The set of dissimilarities under different conditions for the same object are “matched".

Also given

$\mathbf{D}^{(k)} = \{\delta_i; i = 1, \ldots, n\} \qquad k = 1, \ldots, K$: dissimilarities between K new measurements and previous n objects under the same K conditions.

Question we can answer

Do dissimilarities $\mathbf{D}^{(k)}$ come from K new measurements that represent a single object?

Problem Formulation for Dissimilarity Representation

Given K $n \times n$ dissimilarity matrices

$$\mathbf{\Delta}^{(k)} = [\delta_{ij}]^{(k)}; 1 \leq i < j \leq n; k = 1, \cdots, K;$$

- Dissimilarities between n objects are each measured under K different conditions.
- The set of dissimilarities under different conditions for the same object are “matched".

Also given

$\mathbf{D}^{(k)} = \{\delta_i; i = 1, \ldots, n\} \qquad k = 1, \ldots, K$: dissimilarities between K new measurements and previous n objects under the same K conditions.

Question we can answer

Do dissimilarities $\mathbf{D}^{(k)}$ come from K new measurements that represent a single object?

Problem Formulation for Dissimilarity Representation

Given K $n \times n$ dissimilarity matrices

$$\mathbf{\Delta}^{(k)} = [\delta_{ij}]^{(k)}; 1 \leq i < j \leq n; k = 1, \cdots, K;$$

- Dissimilarities between n objects are each measured under K different conditions.
- The set of dissimilarities under different conditions for the same object are “matched".

Also given

$\mathbf{D}^{(k)} = \{\delta_i; i = 1, \ldots, n\} \qquad k = 1, \ldots, K$: dissimilarities between K new measurements and previous n objects under the same K conditions.

Question we can answer

Do dissimilarities $\mathbf{D}^{(k)}$ come from K new measurements that represent a single object?

Meaning of “conditions” and “matched”

	Condition 1	Condition 2
A linked collection of wikipedia articles	Textual content of an article	Hyperlink graph of articles
A linked collection of English and French wikipedia articles	The articles in English	The articles in French
Photos of same objects taken under different condition	Indoor photos	Outdoor photos

Hypotheses

For K=2,

$$H_0 : \mathbf{D}^{(1)} \sim \mathbf{D}^{(2)} \text{ (Matched)}$$

versus

$$H_A : \mathbf{D}^{(1)} \nsim \mathbf{D}^{(2)} \text{ (Not matched)}$$

(Hypothesis testing with α=constraint on the probability of missing a true match)

Omnibus Embedding Approach

$$\overset{2n\times 2n}{M} = \begin{bmatrix} \overset{n\times n}{\Delta_1} & \overset{n\times n}{L} \\ L^{T} & \overset{n\times n}{\Delta_2} \end{bmatrix}$$

- Impute L (dissimilarities between different conditions) to obtain an *omnibus dissimilarity matrix* M
- Embed the omnibus matrix M as $2n$ points (one point for each condition of each object) in commensurate space ($\mathbb{R}^d$)

Out-of-sample Embedding

$$\Delta \longrightarrow X$$

$$[D^{(k)}; k = 1, \ldots, K] \xrightarrow[X]{\text{OOS-embed}} \qquad Y$$

Omnibus Embedding Approach (OOS-embedding)

$$\overset{2n\times 2n}{M} = \begin{bmatrix} \overset{n\times n}{\Delta_1} & \overset{n\times n}{\color{red}{L}} \\ \color{red}{L^{T}} & \overset{n\times n}{\Delta_2} \end{bmatrix} \begin{array}{cc} \overset{n\times 1}{D^{(1)}} & \overset{n\times 1}{\vdots} \\ \overset{n\times 1}{\vdots} & \overset{n\times 1}{D^{(2)}} \end{array}$$

$$\begin{array}{ccc} y_1 & {D^{(1)}}^T & \\ y_2 & & {D^{(2)}}^T \end{array}$$

General Approach

1. Map/Embed via MDS in-sample ($\mathbf{\Delta}_k; k = 1, \ldots, K$) dissimilarities into commensurate space.
2. Map/Embed (Out-of-sample) test dissimilarities ($\mathbf{D}^{(k)}; k = 1, \ldots, K$) into commensurate space.
3. Compute test statistic τ (the distance between embeddings of test dissimilarities).
4. Reject null hypothesis if τ is large.

Fidelity and Commensurability Criteria

Fidelity is how well the mapping to commensurate space preserves original dissimilarities. Within-condition *fidelity error* is given by

$$\epsilon_{f_k} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} (d(\widetilde{\boldsymbol{x}}_{ik}, \widetilde{\boldsymbol{x}}_{jk}) - [\delta_{ij}]^{(k)})^2,$$

where $\widetilde{\boldsymbol{x}}_{ik}$ is the embedded configuration of ith object for kth condition.

Fidelity and Commensurability Criteria

Commensurability is how well the mapping to commensurate space preserves matchedness. Between-condition *commensurability error* is given by

$$\epsilon_{c_{k_1 k_2}} = \frac{1}{n} \sum_{1 \le i \le n} (d(\widetilde{\boldsymbol{x}}_{ik_1}, \widetilde{\boldsymbol{x}}_{ik_2}) - [\delta_{ii}]^{(k_1,k_2)})^2$$

for different conditions k_1 and k_2.

Although $[\delta_{ii}]^{(k_1,k_2)}$ is not available (the between-condition dissimilarities), it is not unreasonable in this setting to set $[\delta_{ii}]^{(k_1,k_2)} = 0$ for all i, k_1, k_2. (We want embedded points of the same object to be close.)

Fidelity and Commensurability Criteria

Commensurability is how well the mapping to commensurate space preserves matchedness. Between-condition *commensurability error* is given by

$$\epsilon_{c_{k_1 k_2}} = \frac{1}{n} \sum_{1 \le i \le n} (d(\widetilde{\boldsymbol{x}}_{ik_1}, \widetilde{\boldsymbol{x}}_{ik_2}) - [\delta_{ii}]^{(k_1,k_2)})^2$$

for different conditions k_1 and k_2.

Although $[\delta_{ii}]^{(k_1,k_2)}$ is not available (the between-condition dissimilarities), it is not unreasonable in this setting to set $[\delta_{ii}]^{(k_1,k_2)} = 0$ for all i, k_1, k_2. (We want embedded points of the same object to be close.)

Commensurability and Fidelity Terms in Raw Stress

The weighted raw stress criterion for MDS with different conditions is

$$\sigma_W(X) = \sum_{i \leq j, k_1 \leq k_2} w_{ijk_1k_2}(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2$$

(i and j are the indices for the object, k_1 and k_2 are condition indices).

Ignoring some error terms (We can set $w_{ijk_1k_2} = 0$ for $i \neq j$ and $k_1 \neq k_2$)

$$\sigma_W(X) = \underbrace{\sum_{i=j, k_1 < k_2} w_{ijk_1k_2}(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Commensurability} +$$

$$\underbrace{\sum_{i<j, k_1 = k_2} w_{ijk_1k_2}(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Fidelity}$$

Commensurability and Fidelity Terms in Raw Stress

The weighted raw stress criterion for MDS with different conditions is

$$\sigma_W(X) = \sum_{i \le j, k_1 \le k_2} w_{ijk_1k_2} (d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2$$

(i and j are the indices for the object, k_1 and k_2 are condition indices).
Ignoring some error terms (We can set $w_{ijk_1k_2} = 0$ for $i \neq j$ and $k_1 \neq k_2$)

$$\sigma_W(X) = \underbrace{\sum_{i=j, k_1 < k_2} w_{ijk_1k_2} (d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Commensurability} + \underbrace{\sum_{i<j, k_1 = k_2} w_{ijk_1k_2} (d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Fidelity}$$

Which one is critical for hypothesis testing performance: Fidelity or Commensurability?

Setting $w_{ijk_1k_2}$ to $(1-w)$ for Fidelity terms, and to w for Commensurability terms,

$$\sigma_w(X) = \underbrace{\sum_{i=j,k_1<k_2} w(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Commensurability} + \underbrace{\sum_{i<j,k_1=k_2} (1-w)(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Fidelity}.$$

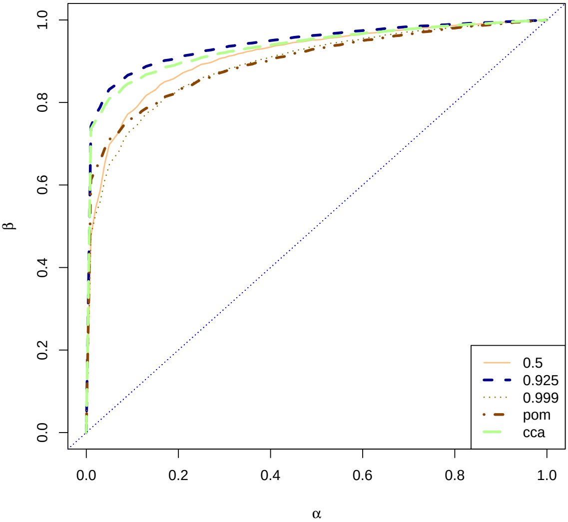
w controls the tradeoff between Fidelity and Commensurability.
Define w^* to be the value of w which maximizes power.

Which one is critical for hypothesis testing performance: Fidelity or Commensurability?

Setting $w_{ijk_1k_2}$ to $(1-w)$ for Fidelity terms, and to w for Commensurability terms,

$$\sigma_w(X) = \underbrace{\sum_{i=j,k_1<k_2} w(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Commensurability} + \underbrace{\sum_{i<j,k_1=k_2} (1-w)(d_{ijk_1k_2}(X) - [\delta_{ij}]^{(k1,k2)})^2}_{Fidelity}.$$

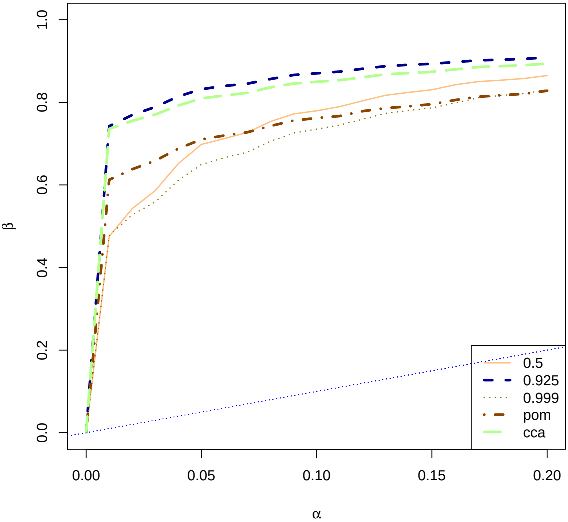
w controls the tradeoff between Fidelity and Commensurability. Define w^* to be the value of w which maximizes power.



1.0
0.8
0.6
0.4
0.2
0.0
β
0.0
0.2
0.4
0.6
0.8
1.0
α
0.5
0.925
0.999
pom
cca

Simulation Results

ROC curves: Power(β) against allowable Type I error(α)



Matching Dissimilarities

If multiple dissimilarities are available $\mathbf{D}_i^{(k)}, i = 1, \ldots, n$ from two different conditions, what is the best matching of dissimilarities from two different conditions?

- Embed the dissimilarities
- Interpret the distances between embeddings as costs in an assignment problem
- Solve the linear assignment problem by the Hungarian algorithm:
  Find the best assignment $f : \{1, \ldots, s\} \to \{1, \ldots, s\}$ that minimizes the cost
  $$\sum_{i=1}^{n} C(i, f(i))$$
  where the cost $C(\cdot, \cdot)$ is the distance between the embeddings of each dissimilarity from the different conditions.

Matching Dissimilarities

If multiple dissimilarities are available $\mathbf{D}_i^{(k)}, i = 1, \ldots, n$ from two different conditions, what is the best matching of dissimilarities from two different conditions?

- Embed the dissimilarities
- Interpret the distances between embeddings as costs in an assignment problem
- Solve the linear assignment problem by the Hungarian algorithm:
  Find the best assignment $f : \{1, \ldots, s\} \to \{1, \ldots, s\}$ that minimizes the cost

$$\sum_{i=1}^{n} C(i, f(i))$$

  where the cost $C(\cdot, \cdot)$ is the distance between the embeddings of each dissimilarity from the different conditions.

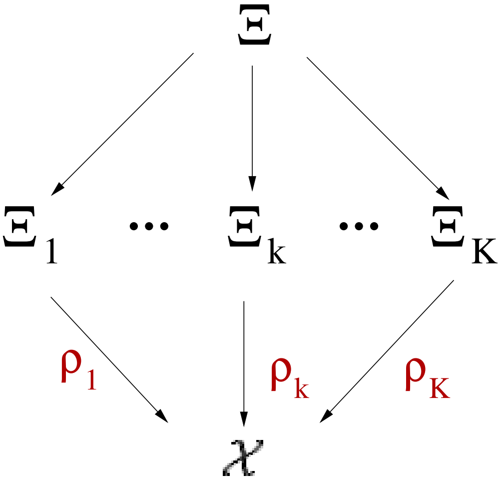
To solve the inexact graph matching problem with hard seeds where the objects we want to match are graph vertices,

- Compute dissimilarities between vertices
- Embed the hard seeds which are known to be matched with JOFC.
- OOS-embed the vertices to be matched.
- Compute dissimilarities between embeddings of vertices between different graphs
- Solve the linear assignment problem with dissimilarities as costs.

Conclusions and Future Work

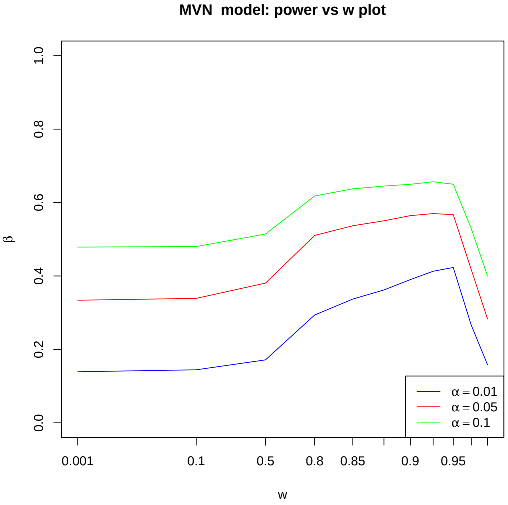
- Our JOFC approach is a general tool that can solve many kinds of related problems.
- The tradeoff between fidelity and commensurability , controlled by the parameter w has a big effect on the performance for the inference task.
- A particular type of graph matching problem can be solved using JOFC followed by the Hungarian algorithm.

Measurement Spaces and Commensurate Space



Simulation Results

Power(β) curves: Power(β) against w at fixed α



I. Borg and P. Groenen.
Modern Multidimensional Scaling. Theory and Applications.

Springer, 1997.

Carey E. Priebe, David J. Marchette, Zhiliang Ma, Sancar Adali
Manifold Matching: Joint Optimization of Fidelity and Commensurability
Brazilian Journal of Probability and Statistics, accepted for publication
February, 2012.