Local Multidimensional Scaling

For Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis

Lisha Chen (Statistics, Yale)
Joint work with Andreas Buja (Upenn)

Feb 6th, 2007
Facial images vary from moment to moment.
Machine Learning Facial Images

- Facial images vary from moment to moment.
- The mystery of perception:
  How does brain perceive constancy in flux?
- Goal: Construct machines to capture the variability of facial images.
Characterizing Image Variability

High-dimensional data . . .

- An image is regarded as a collection of numbers.
- Each number is the light intensity at an image pixel.
- An Example: 64 by 64 images $\Rightarrow$ 4096-D data.
Characterizing Image Variability

High-dimensional data . . .

- An image is regarded as a collection of numbers.
- Each number is the light intensity at an image pixel.
- An Example: 64 by 64 images ⇒ 4096-D data.

Low-dimensional manifold . . .

Degree of freedom
- Left-right pose
- Up-down pose
- Lighting-direction
- Expression change (not shown)
Local Multidimensional Scaling
For Nonlinear Dimension Reduction, Graph Layout
and Proximity Analysis
L. Chen

Motivation

Some Existing Methods

New proposal: LMDS

The LC Meta-criterion for Measuring Local Continuity

Examples

Graph-layout and Energy functions

Generalized LMDS

Conclusion and Discussion

Classical Method 1: PCA

- Principal Components: Fit the best ellipsoid to multivariate data $x_1, \ldots, x_N \in \mathbb{R}^p$.
- Linear method: Provide a sequence of best linear approximation to the data.
- Uses: Plot the wide dimensions; interpret their directions.
Multidimensional Scaling (MDS) for proximity data...

- Make a map from given pairwise dissimilarities.
- A Toy Example:

\[ D = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 5 \\ 4 & 5 & 0 \end{bmatrix} \]

MDS for dimension reduction...

- Calculate dissimilarities \((D_{ij})\) in high-dimensional space.
- Match \(D_{ij}\) in low-dimensional space.
Nonlinear Structure in High Dimensional Data

- Classical methods fail to find nonlinear structure.

Nonlinear dimension reduction methods:
- Isomap (Tenenbaum et. al., 2000)
- LLE (Roweis & Saul, 2000)
- Hessian Eigenmaps (Donoho & Carrie, 2003)
- Laplacian Eigenmaps (Belkin & Niyogi, 2003)
- Diffusion Maps (Coifman et. al., 2005)
Isometric Feature Mapping (Isomap)

The idea . . .

- Characterize the intrinsic geometry by geodesic distances (shortest path).
- Match geodesic distances by Euclidean distances in low-dimensional space.

The algorithm . . .

1. Construct neighborhood graph.
2. Compute shortest path.
3. Construct embedding by MDS.
Local Linear Embedding (LLE)

The idea . . .

- Nearby points lie on a locally linear patch.
- Characterize the local geometry by linear coefficients that reconstruct each data point from its neighbors.

The algorithm . . .

1: Select neighbors.
2: Construct weights.
   \[ \varepsilon(W) = \sum_i |X_i - \sum_{j \in N(i)} W_{ij} X_j|^2 \]
3: Embedding.
   \[ \Phi(Y) = \sum_i |Y_i - \sum_{j \in N(i)} W_{ij} Y_j|^2 \]
Commonality of The Proposals: Localization

- Discard the information of large distances.
- Use local information:
  - Fixed-radius metric neighborhoods.
  - K nearest neighborhoods (KNN).
- Find a way to reassemble local structure and flatten the manifold.
- Our goal: Apply localization to MDS.
Revisit MDS

- **Data:**
  
  *Objects* $i = 1, 2, ..., N$ (subjects, stimuli, graph nodes,...) 
  *Dissimilarities* $D_{ij}$ for all pairs of objects $i$ and $j$

- **Goal:** Visualization of objects
  
  Find *configuration* $\{x_i \in \mathbb{R}^k | i = 1..N\}$ such that:

  $$\|x_i - x_j\| \approx D_{ij}$$

  $k$: embedding dimension

- **Stress function:** Goodness of fit criterion
  
  The simplest version:

  $$\text{MDS}_D(x_1, ..., x_N) = \sum_{i,j=1..N} (D_{ij} - \|x_i - x_j\|)^2$$
Localization Problem in MDS

A simple proposal . . .

- Localization: Drop the large dissimilarities.

- Stress function becomes:

\[
MDS_D(x_1, \ldots, x_N) = \sum_{(i,j) \in E} (D_{ij} - ||x_i - x_j||)^2
\]

\(E = \) edge set of local links
Localization Problem in MDS

A simple proposal . . .

- Localization: Drop the large dissimilarities.
- Stress function becomes:

\[ MDS_D(x_1, \ldots, x_N) = \sum_{(i,j) \in E} (D_{ij} - ||x_i - x_j||)^2 \]

\( E = \) edge set of local links

Unstable system: Graef and Spence (1979)
New Proposal: Local MDS (LMDS)

The idea . . .

- Approximate the small dissimilarities.
- Give a proper amount of repulsive force among the points to avoid instability.
New Proposal: Local MDS (LMDS)

- **Step 1:** Localization: \( E \) = edge set of local links; \( E^C \) = edge set of non-local links.
- **Step 2:** Replace large distances \((E^C)\) with \(\infty\), but with infinitesimal weight.

\[
S_D(x_1, \ldots, x_N) = \sum_{(ij) \in E} (D_{ij} - \|x_i - x_j\|)^2 \\
+ w \cdot \sum_{(ij) \in E^C} (D_{\infty} - \|x_i - x_j\|)^2
\]

- **Step 3:** Take limit \( w \to 0 \) and \( D_{\infty} \to \infty \) subject to \( 2w D_{\infty} = t \).

\[
\text{LMDS}_D(x_1, \ldots, x_N) = \sum_{(ij) \in E} (D_{ij} - \|x_i - x_j\|)^2 \\
- t \cdot \sum_{(ij) \in E^C} \|x_i - x_j\|
\]
New Proposal: Local MDS (LMDS)

\[
\text{LMDS}_D(x_1, \ldots, x_N) = \sum_{(ij) \in E} (D_{ij} - ||x_i - x_j||)^2
- t \cdot \sum_{(ij) \in E^C} ||x_i - x_j||
\]

- The first term: Local stress to match small dissimilarities.
- The second term: Repulsive force for spreading out.
- The choice of \( t \): \( t = \tau \cdot (K/N) \cdot \text{median}_{E}(D_{ij}) \)
  - \( \tau \) repulsion tuning parameter, usually 0.001 \( \sim \) 1.
  - \( K \): the number of neighbors for each point.
The LC Meta-criterion: Motivation and Definition

- **Motivation**
- **Notation:**
  - $N^D_K(i)$: Data point $i$’s $K$ nearest neighbors with regard to the target dissimilarities $D_{ij}$.
  - $N^X_K(i)$: Data point $i$’s $K$ nearest neighbors with regard to the configuration distances $\|x_i - x_j\|$.
  - $|S|$: Cardinality of a set $S$.
- The local version of the LC meta-criterion:
  \[ N_K(i) = |N^D_K(i) \cap N^X_K(i)| \]
- The global version of the LC meta-criterion:
  \[ N_K = \frac{1}{N} \sum_{i=1}^{N} N_K(i) \]
The LC meta-criterion: Normalization and Adjustment for Baselines

- Compare LC meta-criteria for different values of $K$.
- Normalization:
  \[ M_K = \frac{1}{K} \cdot N_K \]
- The problem of meta-criterion for large $K$: $M_{N-1} = 1$.
- Adjusted LC Meta-criterion:
  \[ M_K^{adj} = M_K - \frac{K}{N-1} \]
Example 1: Sculpture Face

3D LMDS

Left-right Pose

Up-down Pose

Lighting Direction

Left-right Pose
Example 1: Sculpture Face

Adjust the repulsion tuning parameter $\tau$ in LMDS:

![Graph](image)
Example 1: Sculpture Face

<table>
<thead>
<tr>
<th>Methods</th>
<th>PCA</th>
<th>MDS</th>
<th>Isomap</th>
<th>LLE</th>
<th>LMDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_6$</td>
<td>2.6</td>
<td>3.1</td>
<td>4.5</td>
<td>2.8</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Motivation

Some Existing Methods

New proposal: LMDS

The LC Meta-criterion for Measuring Local Continuity

Examples

Graph-layout and Energy functions

Generalized LMDS

Conclusion and Discussion
Example 1: Sculpture Face

3D PCA ($M_6 = 2.6$)

3D ISOMAP ($M_6 = 4.5$)

3D LLE ($M_6 = 2.8$)

3D LMDS ($M_6 = 5.2$)
Example 1: Sculpture Face

![Graph showing example](image-url)
Local Multidimensional Scaling
For Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis

L. Chen

Motivation
Some Existing Methods
New proposal: LMDS
The LC Meta-criterion for Measuring Local Continuity

Examples
Graph-layout and Energy functions
Generalized LMDS

Conclusion and Discussion

Example 1: Sculpture Face

![Graph showing the variations of Mk and Adjusted Mk with respect to K for different values of K0: K0=4, K0=6, K0=8, K0=10, K0=25, K0=125, and K0=650. The graphs display the trends as K increases, illustrating how the values change for each K0.]
Example 2: Frey Face

3D PCA

3D ISOMAP

3D MDS

3D LLE

Local Multidimensional Scaling
For Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis

L. Chen

Motivation
Some Existing Methods
New proposal: LMDS
The LC Meta-criterion for Measuring Local Continuity
Examples
Graph-layout and Energy functions
Generalized LMDS
Conclusion and Discussion
Example 2: Frey Face

3D LMDS $K_0=12$

3D LMDS $K_0=4$
Local Multidimensional Scaling
For Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis

L. Chen

Motivation

Some Existing Methods

New proposal: LMDS

The LC Meta-criterion for Measuring Local Continuity

Examples

Graph-layout and Energy functions

Generalized LMDS

Conclusion and Discussion

Example 2: Frey Face
Example 2: Frey Face

3D PCA \( (M_{12}=3.6) \)

Left−right Pose
Serious−Smiling

3D MDS \( (M_{12}=4.8) \)

Left−right Pose
Serious−Smiling

3D ISOMAP \( (M_{12}=4.2) \)

Left−right Pose
Serious−Smiling

3D LLE \( (M_{12}=3.2) \)

Left−right Pose
Serious−Smiling

3D LMDS \( K_0=12 \) \( (M_{12}=4.6) \)

Left−right Pose
Serious−Smiling

3D LMDS \( K_0=4 \) \( (M_{12}=5.1) \)

Left−right Pose
Serious−Smiling
Example 1: Frey Face

![Graph showing Frey Face example with line graphs representing different values of Ko ranging from 4 to 1964. The x-axis represents K, and the y-axis represents M.

- The left graph shows the relationship between K and M for different Ko values, with solid lines indicating linear relationships.
- The right graph illustrates the adjusted M with MDS, highlighting the comparison between different Ko values with markers for each Ko.

The graphs demonstrate how M changes with respect to K for various Ko values, providing insights into the behavior of the Frey Face dataset under different scaling conditions.
Outline for the Rest of the Talk

- Graphs and energy-based models.
- Stress in LMDS and energy in graph layout.
- Clustering postulate.
- Generalized LMDS: A Box-Cox family of energy functions.
Graph-layout

▶ A graph $G = (V, E)$:
  ▶ $V \leftrightarrow$ Entities: Telephone numbers, carbon atoms . . .
  ▶ $E \leftrightarrow$ Relations: Phone calls, chemical bonds . . .

▶ Goal: Two-dimensional layout $\leftrightarrow$ Visualize the relational information.

▶ Algorithm: Force-directed Placement (Eades, 1987)
  ▶ A physical system:
    Vertices $\rightarrow$ Steel rings;
    Edges $\rightarrow$ Springs.
  ▶ Best layout is achieved at the minimal energy state.
Energy-based Models

- **Notation:**
  - $E =$ edge set; $E^C =$ non-edge set; $V^2 = E \cup E^C$.
  - Denote $d_{ij} = \|x_i - x_j\| =$ drawn distances.
  - $U =$ Energy of the system.
  - $f(\cdot)$: energy corresponding to attraction
    $g(\cdot)$: energy corresponding to repulsion

- Common in graph layout: attraction-repulsion forces

  $$U = \sum_{E} f(d_{ij}) - \sum_{V^2} g(d_{ij})$$

- **LMDS and energy-based models.**

  $$\text{LMDS}_D = \sum_{E} (D_{ij} - d_{ij})^2 - t \cdot \sum_{E^C} d_{ij}$$
Energy-based Models

Some energy functions . . .

- Eades (1984)
  \[ U = \sum_{E} d_{ij} \cdot \left( \log(d_{ij}) - 1 - d_{ij}^{-2} \right) + \sum_{V^2} d_{ij}^{-1} \]

- Fruchterman & Reingold (1991)
  \[ U = \sum_{E} d_{ij}^{3/3} - \sum_{V^2} \log(d_{ij}) \]

- Davidson & Harel (1996)
  \[ U = \sum_{E} d_{ij}^{2} + \sum_{V^2} d_{ij}^{-2} \]

- Noack (2003), LinLog model
  \[ U = \sum_{E} d_{ij} - \sum_{V^2} \log(d_{ij}) \]
Clustering Postulate

- A graph with two clusters:

\[
\begin{array}{c}
\bullet \\
\text{n}_1 \\
\text{n}_e \\
\text{n}_2 \\
\bullet
\end{array}
\]

- Problem: What should be the inter-cluster distance \(d\)?
- Energy:

\[
U(d) = n_efa(d) - n_1n_2g(d)
\]

- Clustering Postulate (Noack 2004):
  - \(U(d)\) should be minimized at \(d = n_1n_2/n_e = 1/C\).
  - \(C = n_e/n_1n_2\) is called coupling strength, \(C < 1\).
- Stationary condition yields:

\[
f'(d) = d \cdot g'(d)
\]

- Weak Clustering Postulate: More strongly coupled clusters should be placed more closely to each other.
Distance-weighted Graphs

- Assume target distances $D_{ij}$ for $(i, j) \in E$.

- **Extended clustering condition**: In the 2-cluster situation, require the minimum at
  
  $$d = D \cdot \left( \frac{1}{C} \right)^\lambda$$

  - $D$: the target distance
  - $\left( \frac{1}{C} \right)^\lambda$: a generalized clustering factor.
  - $\lambda > 0$: clustering power; $C < 1$: coupling strength.

- **New stationarity condition**:

  $$f'(d) = \left( \frac{d}{D} \right)^{\frac{1}{\lambda}} g'(d)$$

- $f(d)$ and $g(d)$ satisfy the extended coupling condition:

  $$f(d) = \frac{d^{\mu + \frac{1}{\lambda}} - 1}{\mu + \frac{1}{\lambda}}, \quad g(d) = \frac{d^{\mu} - 1}{\mu} D^{\frac{1}{\lambda}}$$

  with logarithmic fill-in for zero exponents.
A Box-Cox Family of Energy Functions

- **Energy:**

\[ U \sim \sum_{E} \left( \frac{d_{ij}^{\mu + \frac{1}{\lambda}} - 1}{\mu + \frac{1}{\lambda}} - D_{ij}^{\frac{1}{\lambda}} \frac{d_{ij}^{\mu} - 1}{\mu} \right) - t \sum_{E^c} \frac{d_{ij}^{\mu} - 1}{\mu} \]

- **Special cases:**
  - LMDS: \( \lambda = \mu = 1 \)
  - Energy for unweighted graph (\( D_{ij} = 1 \)):
    - Noack (LinLog): \( \lambda = 1, \mu = 0 \)
    - Fruchterman & Reingold: \( \lambda = \frac{1}{3}, \mu = 0 \)
    - Davison & Harel: \( \lambda = \frac{1}{4}, \mu = -2 \)

**Advantages . . .**

- Flexibility for achieving desirable configurations.
- Clustering power \( \lambda \): visualize the clusters.
Example 3: Olivetti Faces
Example 3: Olivetti Faces

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\lambda = 0.25$</th>
<th>$\lambda = 0.5$</th>
<th>$\lambda = 1.0$</th>
<th>$\lambda = 1.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_4$</td>
<td>0.64</td>
<td>1.01</td>
<td>1.44</td>
<td>1.57</td>
</tr>
</tbody>
</table>
Conclusion and Discussion

- Nonlinear dimension reduction, proximity analysis and graph-layout
  - Contiguity
  - Distances
  - Localization
- Graph Layout: Two Fundamental Approaches
  - Complete the graph (Isomap, Krustal and Seery(1980))
  - Apply repulsive force
- B-C Energy Functions
- The Selection Problem: LC Meta-Criteria
Diagnostic Plot: Frey Face

Diagnostic Plot

Motivation

Some Existing Methods

New proposal: LMDS

The LC Meta-criterion for Measuring Local Continuity

Examples

Graph-layout and Energy functions

Generalized LMDS

Conclusion and Discussion