Manifold Alignment using Procrustes Analysis

Chang Wang and Sridhar Mahadevan
Computer Science Department
Motivation:

- In real life, we always need to transfer knowledge between two domains.
Motivation:

- In real life, we always need to transfer knowledge between two domains.
Sometimes, the transfer is between two seemingly disparate data sets. For example, an English collection and an Arabic collection.
Sample Applications (can be solved by manifold alignment):

- Automatic machine translation,
- Cross-lingual information retrieval,
- Representation and control transfer between different Markov decision processes.
Outline

- Background
- Our Algorithm
- Justification
- Experiments
- Conclusions
A **manifold** is an abstract mathematical space in which every point has a neighborhood which resembles Euclidean space, but in which the global structure may be more complicated. **Examples:** the earth surface; a state space in Markov Decision Process; a collection of documents; a set of images collected by a robot.

- Manifolds are usually embedded in a high-dimensional Euclidean space, but their intrinsic dimensionality is much lower due to fewer degrees of freedom.

- We can collect data points from manifolds.
Manifold Alignment:

- **Manifold alignment approaches** align the manifolds using a set of correspondences between data points.

- For example:
  - **Semi-supervised manifold alignment** maps the points of the two data sets to the same space by solving a constrained embedding problem, where the embeddings of the corresponding points from different sets are constrained to be identical. [Ham et al, 05]
Outline

- Background
- Our Algorithm
  - Justification
- Experiments
- Conclusions
The Framework of the Algorithm

- Step 1: Map the $S_1$, $S_2$ data sets (coming from manifolds) to low dimensional spaces reflecting their intrinsic geometries.

- Step 2: Apply **Procrustes analysis** to align the two low dimensional embeddings based on a number of landmark points ($S_1 \leftrightarrow S_2'$).
  - Procrustes Analysis removes the **translational, rotational** and **scaling components** from one set so that the optimal alignment between the two sets can be achieved.

Data set 1: $S_1 = S_1^u + S_1' = \{x_1, x_2, \ldots, x_m\}$;

Data set 2: $S_2 = S_2^u + S_2' = \{y_1, y_2, \ldots, y_n\}$;

$S_1' \leftrightarrow S_2' : in ~ correspondence.$
Step 1 (Dimensionality Reduction):

LLE, ISOMAP, Laplacian eigenmap, PCA, LPP… can be used.

**Laplacian Eigenmap:** $m_i \rightarrow n_i$, where $n_i$ is the embedding of $m_i$

**Locality Preserving Projection** (Linear approximation of Laplacian Eigenmap): $m_i \rightarrow f \rightarrow n_i$, where $f$ is a linear mapping.

- Note: when Locality Preserving Projection or PCA is used, the mapping $f$ is defined everywhere rather than just on the training data points, so they can handle the new test points.

1. **Learning low dimensional embeddings of the data sets (with Laplacian Eigenmap):**

   - Construct the weight matrices $W_1$ for $S_1$ and $W_2$ for $S_2$ using $K_i$, where $W_1(i, j) = K_1(x_i, x_j)$ and $W_2(i, j) = K_2(y_i, y_j)$.
   - Compute Laplacian matrices $L_1 = I - D_1^{-0.5}W_1D_1^{-0.5}$ and $L_2 = I - D_2^{-0.5}W_2D_2^{-0.5}$, where $D_k$ is a diagonal matrix ($D_k(i, i) = \sum_j W_k(i, j)$) and $I$ is the identity matrix.
   - Compute selected eigenvectors of $L_1$ and $L_2$ as the low dimensional embeddings of the data sets $S_1$ and $S_2$. Let $X$, $X_U$ be the $d$ dimensional embeddings of $S_1^t$ and $S_2^u$, $Y$, $Y_U$ be the $d$ dimensional embeddings of $S_2^t$ and $S_2^u$, where $S_1^t$, $S_2^t$ are in pairwise alignment and $|S_1^t| = |S_2^t|$.自主学习实验室
Step 2 (Procrustes Analysis):

Remove the *translational, rotational* and *scaling components* from one set to achieve the optimal alignment between the two sets.

- **X**: lower dimensional embedding of $S_1$
- **Y**: lower dimensional embedding of $S_2$

- a. Translate the configures of $X$ and $Y$ so that their centroids are at the origin.
- b. Do singular value decomposition: $U \Sigma V^T = \text{svd}(Y^TX)$
- c. Optimal Rotation: $Q = UV^T$
- Optimal re-scale factor: $k = \text{trace}(\Sigma)/\text{trace}(Y^TY)$.

$Y^* = kYQ$ is the optimal mapping result that minimizes $\|X-Y^*\|_F$, where $\|\cdot\|_F$ is Frobenius norm. $\|A\|_F = \Sigma \Sigma A_{ij}^2$. 

---

UMass Amherst

Autonomous Learning Laboratory
Step 1: $A$ and $B$ are low dimensional representations of the amino acid set $A$ and $B$.

Step 2: Manifold Alignment. Computing the optimal rescale factor $k$, and optimal rotation $Q$ such that $B$ can be rescaled and rotated to align well with $A$.

$$k = 4.2971,$$

$$Q = \begin{pmatrix}
0.56151 & -0.53218 & 0.63363 \\
0.65793 & 0.75154 & 0.048172 \\
-0.50183 & 0.38983 & 0.77214
\end{pmatrix}.$$
Outline

- Background
- Our Algorithm
  - Justification
- Experiments
- Conclusions
Procrustes Analysis achieves optimal alignment of two structures regarding Frobenius Norm.

Theorem 1: Let $X$ and $Y$ be low dimensional embeddings of the points with known correspondences in data set $S_1$, $S_2$, and $X_i$ matches $Y_i$ for each $i$. If Singular Value Decomposition (SVD) of $Y^T X$ is $U \Sigma V^T$, then $Q = UV^T$ and $k = \text{trace}(\Sigma) / \text{trace}(Y^T Y)$ minimize $\|X - kYQ\|_F$. 
Under what conditions are the manifolds similar?

- Many dimensionality reduction approaches compute a relationship matrix, and use the “top” eigenvectors of the matrix as the low dimensional embeddings.
  
  For example, **Laplacian eigenmap**: the “smallest” eigenvectors of the graph Laplacian.

- For eigenvector-based approaches: given the difference between two relationship matrices, theorem 2 shows under what conditions the manifolds can be aligned well.

  **Theorem 2**: If the absolute value of each element in $E$ is bounded by $\tau$, and $\tau \leq 2\varepsilon d_1/(N(\pi + 2\varepsilon))$, then the difference between the two embedding subspaces $\|Q - P\|$ is at most $\varepsilon$.

- It is in fact Perturbation analysis. Eigengap is important!
Outline

- Background
- Our Algorithm
- Justification
- Experiments
- Conclusions
Protein 3D reconstruction

- Already shown.

Figure 2. (A): Manifold $A$; (B): Manifold $B$; (C): Comparison of Manifold $A$ (red) and $B$ (blue) before alignment; (D): Comparison of Manifold $A$ (red) and $B$ (blue) after alignment.
Cross-lingual Information Retrieval (1)

- **The Problem:** finding the exact correspondences between documents in different languages.

- We treat each collection as a manifold.

- **The Data:** we are given two document collections (English, Arabic), each of which has 2,119 documents (manually translated).

  Correspondences between 25% of them were given and used to learn the alignment. The remaining 75% were used for testing.
The document vectors in each language are represented by language models (multinomial term distributions) estimated using the document text.

- Similarity of documents in the same language: Distributional affinity is used to detect topical relatedness.
- Similarity of documents in different languages: Not available to us.
- Correspondence between a small number (25% for this case) of documents from different collections is given.

- The data/approach used here are from (Diaz et al., IJCAI 2007).
In our first step, two dimensionality reduction approaches (Laplacian eigenmap, LPP) are tested (100 eigenvectors are used).

We also tested:

- **Semi-supervised alignment:**
  - 100 eigenvectors are used.

- **Baseline:** assume that we have $m$ correspondences in the training set, then document $x$ is represented by a vector $V$ with length $m$, where $V(i)$ is the similarity of $x$ and the $i^{th}$ document in the training correspondences.
For each given Arabic document:

- **Laplacian Eigenmap + Procrustes analysis**
  1. If we retrieve 3 most relevant English documents, then the true match has a 60% probability of being among the 3.
  2. If we retrieve 10 most relevant English documents, then we have about 80% probability of getting the true match.
  3. 2 times faster than semi-supervised manifold alignment.

- **LPP + Procrustes analysis**
  1. If we retrieve 10 most relevant English documents, then we have a 60% probability of getting the true match.
  2. Faster than Laplacian eigenmap.

Figure 3. Cross-lingual information retrieval test.
Transfer **Proto-Value Functions** (Mahadevan, ICML05)

- Value function $V(x) = k_1 \langle f_1, x \rangle + \ldots + k_n \langle f_n, x \rangle$, where $f_i$ is a proto value function (eigenvector of the normalized graph Laplacian), $k_i$ is a coefficient.

- We want to transfer $f_i$ from the old domain to the new domain.

- In this paper, we did a perturbation analysis. If the state space is only slightly different, the old proto-value functions can be directly transferred to the new space.
If two domains are quite different:

- $\mathbf{Y}$: PVFs of the current MDP; //column vectors; also the low dimensional embedding of the old state space manifold. //row vectors;
- $\mathbf{X}$: low dimensional embedding of the new state space manifold. //row vectors;

Align $\mathbf{X}$ and $\mathbf{Y}$, and get $\mathbf{k}$ and $\mathbf{Q}$. The new PVFs will be $\mathbf{YQ}$.

Future work:
How to compute $\mathbf{X}$? (low dimensional embedding)
How to find correspondence between two domains?
Outline

- Background
- Our Algorithm
- Justification
- Experiments
- Conclusions
Summary & Future work:

- **Method:** (under the manifold framework)
  - Map the data sets to lower dimensional spaces reflecting their intrinsic geometries.
  - Align the low dimensional embeddings based on a number of landmark points.
  - Using rescale factor and rotation factor to transfer knowledge.

- **Applications:** Cross-lingual IR, Representation Transfer in MDPs

- **Future work:**
  - Transfer learning in MDPs.