Multi-task Regularization of Generative
Similarity Models

International Workshop on Similarity-based Pattern Analysis and
Recognition – SIMBAD ‘11

September 28-30, 2011

Dr. Luca Cazzanti
Applied Physics Lab
Univ. Washington
Seattle, USA

Prof. Maya Gupta
Dept. EE
Univ. Washington
Seattle, USA

Mr. Sergey Feldman
Dept. EE
Univ. Washington
Seattle, USA

Dr. Michael Gabbay
Applied Physics Lab
Univ. Washington
Seattle, USA

Work supported by U.S Office of Naval Research – PM Dr. Ivy Estabrooke
Outline

1. Review local similarity discriminant analysis (local SDA)

2. Need for regularization

\[
\{v_{gh}^*\}_{g,h=1}^G = \arg \min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2 + \eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, \hat{v}_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.
\]

3. Multi-task regularization for local SDA

4. Computer experiments and discussion
Euclidean Features

Class 1: fruit

Class 2: OS logos

Goal: classify as either fruit of OS logo

This is the conventional statistical learning set-up

\[ P(x_i | Y = g) \]

\( x_i \) described by four numbers, \( x_i \in \mathbb{R}^4 \):
RGB color triplet image texture
Similarities

Class 1: fruit

Class 2: OS logos

Information about the relationship between samples:

If you like
then you like

Human judgments of similarity

Given taxonomy of objects

www.indiana.edu/~hlw/meaning/AppleTaxonomy2.gif
Similarities

Information about the relationship between samples:

If you like then you like

Human judgments of similarity

Given taxonomy of objects

Matrix of pairwise similarities

\[
P(s(x_i, x_j) | Y = g)
\]
Local Similarity Discriminant Analysis (local SDA)
(Cazzanti ’07, Cazzanti & Gupta ’07)

Classify $x$

Class 1: fruit
$N_1(x)$

$T_1(x) = \{s(x, z)\}, z \in N_1(x)$

Similarities of test sample $x$ to all its $k$-most similar training samples from class 1.

Class 2: OS logos
$N_2(x)$
Local Similarity Discriminant Analysis (local SDA)  
(Cazzanti '07, Cazzanti & Gupta '07)

Classify $x$

Class 1: fruit  
$\mathcal{N}_1(x)$

Class 2: OS logos  
$\mathcal{N}_2(x)$

$T_1(x) = \{s(x, z)\}, z \in \mathcal{N}_1(x)$

Similarities of test sample $x$ to all its $k$-most similar training samples from class 1.

$P_1(T_1(x)|Y = 1)$

Class $x$ is compared to

Assumed class for $x$
Local Similarity Discriminant Analysis (local SDA)
(Cazzanti ’07, Cazzanti & Gupta ’07)

Classify $x$:
- Class 1: fruit $\mathcal{N}_1(x)$
- Class 2: OS logos $\mathcal{N}_2(x)$

$T_1(x) = \{s(x, z)\}, \ z \in \mathcal{N}_1(x)$

Similarities of test sample $x$ to all its $k$-most similar training samples from class 1.

$P_1(T_1(x)|Y = 1)$

Class $x$ is compared to

$P_1(T_1(x)|Y = 2)$

Assumed class for $x$
Local Similarity Discriminant Analysis (local SDA)  
(Cazzanti '07, Cazzanti & Gupta '07)

Classify $x$

Class 1: fruit  
$\mathcal{N}_1(x)$

Class 2: OS logos  
$\mathcal{N}_2(x)$

$T_1(x) = \{s(x, z)\}, z \in \mathcal{N}_1(x)$

$P_1(T_1(x)|Y = 1)$

$P_1(T_1(x)|Y = 2)$

$T_2(x) = \{s(x, z)\}, z \in \mathcal{N}_2(x)$

$P_2(T_2(x)|Y = 1)$

$P_2(T_2(x)|Y = 2)$
Local Similarity Discriminant Analysis (local SDA)
(Cazzanti ’07, Cazzanti & Gupta ’07)

Classify $x$

Class 1: fruit $\mathcal{N}_1(x)$

Class 2: OS logos $\mathcal{N}_2(x)$

$T_1(x) = \{s(x, z)\}, z \in \mathcal{N}_1(x)$

$P_1(T_1(x)|Y = 1)$

$P_1(T_1(x)|Y = 2)$

$T_2(x) = \{s(x, z)\}, z \in \mathcal{N}_2(x)$

$P_2(T_2(x)|Y = 1)$

$P_2(T_2(x)|Y = 2)$

$y = \arg \max_g \prod_{h=1}^{G} P_h(T_h(x)|Y = g)P(Y = g)$
Local Similarity Discriminant Analysis (local SDA)

\[
P_h(T_h(x)|Y = g) \triangleq \frac{1}{k_h} \sum_{z \in (N)_h(x)} \hat{P}_h(s(x, z)|Y = g)
\]

\[
= \frac{1}{k_h} \sum_{z \in (N)_h(x)} \gamma_{gh} e^{\lambda_{gh} s(x, z)}
\]

\[
\hat{P}_h(s(x, z)|Y = g)
\]

+ \ldots +

\[
P_h(T_h(x)|Y = g)
\]
Estimating the Local SDA Parameters

\[ E_{P_h(\mathcal{T}_h(x) \mid Y=g)}[s(X, z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h} \]

Class 1: fruit \( \mathcal{N}_1(x) \)

Class 2: OS logos \( \mathcal{N}_2(x) \)
Estimating the Local SDA Parameters

\[
E_{\mathcal{P}_h}(T_h(x) | Y=g) [s(X, z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h}.
\]

Class 1: fruit
\[\mathcal{N}_1(x)\]

Class 2: OS logos
\[\mathcal{N}_2(x)\]
Estimating the Local SDA Parameters

\[ E_{P_h(T_h(x)|Y=g)}[s(X, z)] = \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} \frac{s(z_a, z_b)}{k_g k_h}. \]

Class 1: fruit \( \mathcal{N}_1(x) \)

Class 2: OS logos \( \mathcal{N}_2(x) \)

\[ P_1(s(x, z)|Y = 1) \]
Estimating the Local SDA Parameters

\[ E_{P_h}(\mathcal{T}_h(x)|Y=g)[s(X, z)] = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h} \]

Class 1: fruit \( \mathcal{N}_1(x) \)

Class 2: OS logos \( \mathcal{N}_2(x) \)

normalizing constant

solution to mean-similarity constraint

\[ \gamma_{gh} e^{\lambda_{gh} s(x, z)} \]
Need for Regularization

If mean similarity = maximum (or minimum) similarity
→ No exponential solution exists!

Approach: mean class-conditional similarities regularize each other

\[ v_{11} \leftrightarrow v_{12} \implies v_{11}^*, v_{12}^* \]
Need for Regularization

If mean similarity = maximum (or minimum) similarity
\[ \rightarrow \text{No exponential solution exists!} \]

Approach: mean class-conditional similarities regularize each other

\[ \mathcal{V}_{11} \leftrightarrow \mathcal{V}_{12} \implies \mathcal{V}^*_{11}, \mathcal{V}^*_{12} \]
Need for Regularization

If mean similarity = maximum (or minimum) similarity → No exponential solution exists!

Approach: mean class-conditional similarities regularize each other

\[ u_{11} \leftrightarrow u_{12} \implies u^*_1, u^*_2 \]
Multi-task Regularization

Single task: estimate mean similarity $v_{gh}$

Multi-task:

$$\{v^*_{gh}\}_{g,h=1}^G = \arg \min_{\{\hat{v}_{gh}\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2) + \eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.$$
Multi-task Regularization

Single task: estimate mean similarity $v_{gh}$

Multi-task:

$$\left\{ v^*_g \right\}_{g,h=1}^G = \arg \min_{\left\{ \hat{v}_{gh} \right\}_{g,h=1}^G} \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2) + \eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.$$
Multi-task Regularization – Closed Form Solution

For $A$ symmetric and invertible:

$$v^* = (I - \tilde{A})^{-1} \tilde{v},$$

$$\tilde{v}_{gh} = \frac{\sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} s(z_a, z_b)}{k_g k_h + \eta \sum_{l,m \neq g,h} A(v_{gh}, v_{lm})}$$

and

$$\tilde{A}(v_{gh}, v_{lm}) = \left\{ \begin{array}{ll} \frac{\eta A(v_{gh}, v_{lm})}{k_g k_h + \eta \sum_{g,h \neq l,m} A(v_{gh}, v_{lm})} & \text{for } \{g, h\} \neq \{j, k\} \\ 0 & \text{for } \{g, h\} = \{j, k\} \end{array} \right.$$}

Then solve the $G^2$ regularized mean-similarity constraints:

$$E_{P_h}(\mathcal{T}_h(x)|Y=g)[s(X, z)] = v_{gh}^*$$

$$\rightarrow \lambda_{gh}^*$$
Choice of Task Relatedness Matrix A

Symmetric and invertible \( \Rightarrow A(v_{jk}, v_{lm}) = e^{-\frac{(v_{jk} - v_{lm})^2}{\sigma}} \)

Emphasizes mean similarities close to each other
De-emphasizes distant mean similarities

Can use any problem-relevant task relatedness.
Side information easily incorporated into problem.
Benchmark Datasets

AMAZON (fiction & nonfiction): similarities between books based on user statistics from amazon.com

SONAR (target & clutter): similarities between sonar signals rated by human subjects.

PATROL (8 patrol units): membership in patrol unit reported by other patrol members.

VOTING (2 political parties): value difference metric on congressional votes.

FACE RECOGNITION (139 faces): cosine similarity between features from 3D face data.
# Benchmark Datasets

<table>
<thead>
<tr>
<th></th>
<th>Amazon 2 classes</th>
<th>Sonar 2 classes</th>
<th>Patrol 8 classes</th>
<th>Protein 4 classes</th>
<th>Voting 2 classes</th>
<th>FaceRec 139 classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-task Local SDA</td>
<td>8.95</td>
<td>14.50</td>
<td>11.56</td>
<td>9.77</td>
<td>5.52</td>
<td>3.44</td>
</tr>
<tr>
<td>Local SDA</td>
<td>11.32</td>
<td>15.25</td>
<td>11.56</td>
<td>10.00</td>
<td>6.15</td>
<td>4.23</td>
</tr>
<tr>
<td>Similarity $k$-NN</td>
<td>12.11</td>
<td>15.75</td>
<td>19.48</td>
<td>30.00</td>
<td>5.69</td>
<td>4.29</td>
</tr>
<tr>
<td>SVM-KNN (sims-as-features)</td>
<td>13.68</td>
<td><strong>13.00</strong></td>
<td>14.58</td>
<td>29.65</td>
<td><strong>5.40</strong></td>
<td>4.23</td>
</tr>
</tbody>
</table>

Percent test error averaged over 20 random train/test splits.
RBF task relatedness for multi-task local SDA

Multi-task local SDA at least as good as local SDA.
Benchmark Datasets

<table>
<thead>
<tr>
<th></th>
<th>Amazon 2 classes</th>
<th>Sonar 2 classes</th>
<th>Patrol 8 classes</th>
<th>Protein 4 classes</th>
<th>Voting 2 classes</th>
<th>FaceRec 139 classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-task Local SDA</td>
<td>8.95</td>
<td>14.50</td>
<td>11.56</td>
<td>9.77</td>
<td>5.52</td>
<td>3.44</td>
</tr>
<tr>
<td>Local SDA</td>
<td>11.32</td>
<td>15.25</td>
<td>11.56</td>
<td>10.00</td>
<td>6.15</td>
<td>4.23</td>
</tr>
<tr>
<td>Similarity $k$-NN</td>
<td>12.11</td>
<td>15.75</td>
<td>19.48</td>
<td>30.00</td>
<td>5.69</td>
<td>4.29</td>
</tr>
<tr>
<td>SVM-KNN (sims-as-features)</td>
<td>13.68</td>
<td>13.00</td>
<td>14.58</td>
<td>29.65</td>
<td>5.40</td>
<td>4.23</td>
</tr>
</tbody>
</table>

Percent error averaged over 20 random train/test splits.
RBF task relatedness for multi-task local SDA

Multi-task local SDA at least as good as local SDA.
Multi-task local SDA competitive with other similarity-based classifiers.
Insurgent Rhetoric Experiment

1924 documents (press releases)

Which of 8 Iraqi insurgent groups authored the document?

Document similarity: KL divergence of pmfs over 173 keywords

\[
A(v_{jk}, v_{lm}) = e^{-\left(Q_{jk} - Q_{lm}\right)^2 / \sigma}
\]

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-task Local SDA (w/ joint statements task relatedness)</td>
<td>52.34</td>
</tr>
<tr>
<td>Multi-task Local SDA (w/ Gaussian kernel task relatedness)</td>
<td>52.75</td>
</tr>
<tr>
<td>Local SDA</td>
<td>54.52</td>
</tr>
<tr>
<td>Similarity k-NN</td>
<td>53.53</td>
</tr>
<tr>
<td>Guessing Using Class Priors</td>
<td>77.91</td>
</tr>
</tbody>
</table>

Leave-one-out cross validation error
Summary

Reviewed local SDA

Need for regularization

Multi-task regularization

\[
\{v_{gh}^*\}_{g,h=1}^G = \arg \min \left\{ \sum_{g,h=1}^G \sum_{z_a \in \mathcal{N}_g(x)} \sum_{z_b \in \mathcal{N}_h(x)} (s(z_a, z_b) - \hat{v}_{gh})^2 \right\} + 
\eta \sum_{j,k=1}^G \sum_{l,m=1}^G A(v_{jk}, v_{lm})(\hat{v}_{jk} - \hat{v}_{lm})^2.
\]

Illustrated different choices for task relatedness A with benchmark and real datasets.


Software and data available: http://staff.washington.edu/lucagc
To Learn More


