Learning Multi-View Neighborhood Preserving Projections

Novi Quadrianto1 | Christoph H. Lampert2

1: SML-NICTA & Australian National University | 2: IST Austria (Institute of Science and Technology Austria)

Abstract

- We address the problem of projecting data in different representations into a shared space, such that the Euclidean distance in this space provides a meaningful within-view as well as between-view similarity.
- We formulate an objective function that expresses the intuitive concept that matching samples are mapped closely together in the output space, whereas non-matching samples are pushed apart.
- We show that the resulting objective function can be efficiently optimized using the convex-cone procedure (CCCP).
- Our proposed approach has a direct application for cross-media and content-based retrieval tasks.

Motivating Example

A Cross-Media and Content-Based Retrieval

Goal:
- Building an object cross-retrieval system that allows query objects and objects in the database to have different representations.
- Building a content-based object retrieval system where several representations can be used to describe a content, such as, for image objects: SIFT, Color, GIST, SURF, HOG, pHOG, Text, ...

A Multi-View Neighborhood Preserving Projection

Problem Setting

What we have:
- Two sets of m observed data points, \( \{x_1, \ldots, x_m\} \subset X \) and \( \{y_1, \ldots, y_m\} \subset Y \) describing the same objects;
- A cross-neighborhood set \( S_x \) for each \( x_i \in X \) that corresponds to a set of data points from \( Y \) that are deemed similar to \( x_i \).

What we want:
- Projection functions, \( g_1 : X \rightarrow \mathbb{R}^p \) and \( g_2 : Y \rightarrow \mathbb{R}^p \), that respect the neighborhood relationship \( S_x \)

Assumption:
- A linear parameterization of the functions \( g^*_1(x) = \langle w_1, d(x) \rangle \) for \( H_1 \) basis functions \( \{d_0(x), \ldots, d_{|Y|}\} \) and \( w_1 \in \mathbb{R}^{2|Y|} \) and likewise for \( g_2 \) with the weight parameter \( w_2 \in \mathbb{R}^{2|Y|} \).

Regularized Risk Functionals

What so special about the wisdom loss:
- The wisdom loss function \( L^W(x, y) \) consists of the friends term \( L_f^W \) and the enemies term \( L_e^W \):

\[
L_f^W(x, y) = \|f_1(x_1)-f_2(\gamma_1^*)\|_2^p + \|f_1(x_2)-f_2(\gamma_2^*)\|_2^p \\
L_e^W(x, y) = \|f_1(x_1)-f_2(\gamma_1^*)\|_2^p + \|f_1(x_2)-f_2(\gamma_2^*)\|_2^p
\]

where \( f_1(x_1) = \langle w_1, d(x) \rangle \) and \( f_2(\gamma_1^*) = \langle w_2, \gamma_1 \rangle \), \( f_1(x_2) = \langle w_1, d(x) \rangle \) and \( f_2(\gamma_2^*) = \langle w_2, \gamma_2 \rangle \).

Optimization

What so special about the wisdom loss:
- The wisdom loss function is non-convex; the friends term is convex, however the enemies term is non-convex;
- Though the enemies term is non-convex, it has a decomposition form as a difference of two convex functions:

\[
L_e^W(x, y) = L_1(x,y) - L_2(x,y)
\]

\[
L_1(x,y) = \|f_1(x_1)-f_2(\gamma_1^*)\|_2^p + \|f_1(x_2)-f_2(\gamma_2^*)\|_2^p
\]

\[
L_2(x,y) = \|f_1(x_1)-f_2(\gamma_1)\|_2^p + \|f_1(x_2)-f_2(\gamma_2)\|_2^p
\]

Experiments

Dataset Statistics:
- 1000 images with 11 categories from the Israeli-Images dataset (http://www.cs.technion.ac.il/~romb/)
- We use global color descriptors as one view and local SIFT descriptors as another.
- Performance metric: k-Nearest Neighbor classification metric.

Algorithm A vs. Baselines (PCA and CCA) for a Cross-Retrieval Task (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>5S.N</th>
<th>16N.N</th>
<th>30S.N</th>
<th>50S.N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>9.6±1.7</td>
<td>9.7±1.9</td>
<td>10.0±2.1</td>
<td>9.9±1.7</td>
</tr>
<tr>
<td>CCA</td>
<td>9.6±1.9</td>
<td>9.8±2.0</td>
<td>9.9±2.0</td>
<td>9.8±1.8</td>
</tr>
<tr>
<td>Ours</td>
<td>9.6±1.9</td>
<td>9.8±2.0</td>
<td>9.9±1.8</td>
<td>9.8±1.8</td>
</tr>
</tbody>
</table>

Algorithm A–Multi-View Neighborhood Preserving Projection:

Input: Data sources \( X = \{x_1, \ldots, x_m\} \) and \( Y = \{y_1, \ldots, y_m\} \), an inter-view neighborhood relationship \( S_x \), number of alternations \( N \)

Output: \( w_1 \) and \( w_2 \)

Initialize \( w_1 \) and \( w_2 \)

for \( t = 1 \) to \( N \) do

Solve the convex optimization problem w.r.t. \( w_1 \) and obtain \( w_1^{t+1} \)

Solve the convex optimization problem w.r.t. \( w_2 \) and obtain \( w_2^{t+1} \)

end for

Algorithm B–Hybrid-PCA and Multi-NNP:

Input: Data sources \( X = \{x_1, \ldots, x_m\} \) and \( Y = \{y_1, \ldots, y_m\} \) and an inter-view neighborhood relationship \( S_x \)

Output: \( w_1^{PCA} \) and \( w_2 \)

Initialize \( w_2 \)

Solve the optimization problem w.r.t. \( w_2 \) while fixing \( w_1 = w_1^{PCA} \) and obtain \( w_2^* \)

Extensions

Kernels:
- By the Representer Theorem, the projection matrices admits \( w_1 = \sum_{i=1}^{|X|} \langle w_1, \phi(x_i) \rangle \) and \( w_2 = \sum_{j=1}^{|Y|} \langle w_2, \psi(y_j) \rangle \) for a positive-definite kernel \( \phi \) and a kernel \( \psi \) on \( Y \).

Beyond 2-View:
- For the case with more than two data sources we build an analogous objective function by summing up the terms of all pairwise objectives.

Cross-Retrieval Results with Algorithm B (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>5S.N</th>
<th>16N.N</th>
<th>30S.N</th>
<th>50S.N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>9.6±1.9</td>
<td>9.9±2.0</td>
<td>10.1±2.1</td>
<td>9.9±1.7</td>
</tr>
<tr>
<td>CCA</td>
<td>9.6±1.9</td>
<td>9.8±2.0</td>
<td>9.9±2.0</td>
<td>9.8±1.8</td>
</tr>
<tr>
<td>Ours</td>
<td>9.6±1.9</td>
<td>9.8±2.0</td>
<td>9.9±1.8</td>
<td>9.8±1.8</td>
</tr>
</tbody>
</table>

Color Query – SIFT Database

SIFT Query – Color Database

For more experimental results, please refer to the paper.