Learning Multi-View Neighborhood Preserving Projections

Novi Quadrianto
Joint Work with Christoph H. Lampert (IST Austria)

ICML Bellevue, 30th June 2011
Several features can be used to describe content of the images, such as Color, Text, shape, ...
• Query objects and objects in the database have different representations.

DATABASE:

QUERY:
What is the Problem?

- Projecting data in different representations into a shared space, with a special structure;
- In short, it is a multi-view distance metric learning problem;
- Solution: neighborhood structure preservation.
Problem Formulation

What we have:

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

What we want:

- Projection functions, $g_1 : \mathcal{X} \rightarrow \mathbb{R}^D$ and $g_2 : \mathcal{Y} \rightarrow \mathbb{R}^D$, that respect the neighborhood relationship $\{S_{x_i}\}_{i=1}^m$.

Assumption:

- A linear parameterization of the functions $g^w_1(x_i) := \langle w_1, \phi(x_i) \rangle$ for $H_1$ basis functions $\{\phi_h(x_i)\}_{h=1}^{H_1}$ and $w_1 \in \mathbb{R}^{D \times H_1}$ and likewise for $g_2$ with the weight parameter $w_2 \in \mathbb{R}^{D \times H_2}$.
Regularized Risk Functionals

- Folk Wisdom:
  
  Keep your friends (read: matching samples) close and your enemies (read: non-matching samples) closer far far away;

- Turning Wisdom into a Regularized Risk Functional:

  \[
  \sum_{i,j=1}^{m} L^{i,j}(w_1, w_2, x_i, y_j, S_{x_i}) + \eta \Omega(w_1) + \gamma \Omega(w_2)
  \]

  The Wisdom Loss

  The Regularizer
Regularized Risk Functionals

- Folk Wisdom:
  Keep your friends (read: matching samples) close and your enemies (read: non-matching samples) closer far far away;

- Turning Wisdom into a Regularized Risk Functional:

\[
\sum_{i,j=1}^{m} L^{i,j}(w_1, w_2, x_i, y_j, S_{x_i}) + \eta \Omega(w_1) + \gamma \Omega(w_2)
\]

The Wisdom Loss

The Regularizer
The Wisdom Loss Function

\[ L^{i,j}(w_1, w_2, x_i, y_j, S_{x_i}) = \frac{I[y_j \in S_{x_i}]}{2} \times L^{i,j}_1 + \frac{1 - I[y_j \in S_{x_i}]}{2} \times L^{i,j}_2 \]

The Wisdom Loss
The Friends Loss
The Enemies Loss

with

\[ L^{i,j}_1 = \| g_1^{w_1}(x_i) - g_2^{w_2}(y_j) \|_{Fro}^2 \]
\[ L^{i,j}_2(\beta_d) = \max(0, 1 - \| g_1^{w_1}(x_i) - g_2^{w_2}(y_j) \|_{Fro}^2) \]
The Wisdom Loss Function

\[ L_{i,j}(w_1, w_2, x_i, y_j, S_{x_i}) = \frac{I_{[y_j \in S_{x_i}]} \times L_{1,i,j}}{2} + \frac{(1 - I_{[y_j \in S_{x_i}]}) \times L_{2,i,j}}{2} \]

with

\[ L_{1,i,j} = \left\| g_{1}^{w_1}(x_i) - g_{2}^{w_2}(y_j) \right\|_{Fro}^2 \]

\[ L_{2,i,j}(\beta_d) = \max(0, 1 - \left\| g_{1}^{w_1}(x_i) - g_{2}^{w_2}(y_j) \right\|_{Fro}^2) \]

\[ L_{2,i,j}(\beta_d) = \begin{cases} -\frac{1}{2} \beta_d^2 + \frac{a \lambda^2}{2}, & \text{if } 0 \leq |\beta_d| < \lambda \\ \frac{|\beta_d|^2 - 2a \lambda |\beta_d| + a^2 \lambda^2}{2(a-1)}, & \text{if } \lambda \leq |\beta_d| \leq a \lambda \\ 0, & \text{if } |\beta_d| \geq a \lambda, \end{cases} \]

where \( \beta_d = \left\| g_{1}^{w_1}(x_i) - g_{2}^{w_2}(y_j) \right\|_{Fro} \).
What So Special with the Wisdom Loss?

\[ L_{2}^{i,j}(\beta_{d}) = L_{CV}^{1}(\beta_{d}) - L_{CV}^{2}(\beta_{d}) \]
Recall:

- **Objective** = Enemies term + Friends term + Regularizers

\[
\left(1 - I_{y_j \in S_{x_i}}\right) \times \left(\frac{L_1^{1 \times L_2^2}}{2} - \frac{I_{y_j \in S_{x_i}}}{2}\right) \times L_{i,j} + \eta \Omega(w_1) + \gamma \Omega(w_2)
\]

- Concave Function
- Convex Functions
Optimization

The ConCave-Convex Procedure (CCCP):

\[ + \quad = \quad \text{Non - Convex} \]

\[ + \quad = \quad \text{Convex Upper Bound} \]

\[ + \quad = \quad \text{Convex Upper Bound} \]

\[ \ldots \]
Algorithm A—Multi-View Neighborhood Preserving Projection

Assume: $g^w_1(x_i) := \langle w_1, \phi(x_i) \rangle$ and $g^w_2(y_i) := \langle w_2, \psi(y_i) \rangle$

Input: $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_m\}$, $\{S_{x_i}\}_{i=1}^m$, $N$

Output: $w_1^*$ and $w_2^*$

Initialize $w_1$ and $w_2$

for $t = 1$ to $N$ do
  Solve CCCP w.r.t. $w_1$ and obtain $w_1^t$
  Solve CCCP w.r.t. $w_2$ and obtain $w_2^t$
end for
Algorithm

**Algorithm B–Hybrid-\{PCA and Multi-NPP\}**

**Assume:** \( g^w_1(x_i) := \langle w_1, \phi(x_i) \rangle \) and \( g^w_2(y_i) := \langle w_2, \psi(y_i) \rangle \)

**Input:** \( X = \{x_1, \ldots, x_m\} \) and \( Y = \{y_1, \ldots, y_m\} \) and \( \{S_{x_i}\}_{i=1}^m \)

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

Initialize \( w_2 \)

Solve CCCP w.r.t. \( w_2 \) while fixing \( w_1 = w_1^{\text{PCA}} \)
Experimental Setup

Experimentations on a image retrieval task

**Israeli-Images dataset:**
- 1000 images with 11 categories;
- View 1: global color descriptors;
- View 2: local SIFT descriptors.

**Baselines:**
- Principal Component Analysis (PCA);
- Canonical Correlation Analysis (CCA).

**Performance metric:**
- \( k \)-Nearest Neighbor classification metric.
Algorithm A v. Baselines (PCA and CCA) for Color Query - Color Database (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>64</td>
<td>31.4±2.52</td>
<td>31.3±3.87</td>
<td>30.4±3.55</td>
</tr>
<tr>
<td>PCA</td>
<td>10</td>
<td>28.9±2.25</td>
<td>30.1±2.35</td>
<td>29.4±3.08</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>31.3±3.12</td>
<td>31.4±3.46</td>
<td>30.3±2.99</td>
</tr>
<tr>
<td>CCA</td>
<td>10</td>
<td>24.8±3.86</td>
<td>24.7±3.42</td>
<td>24.0±3.78</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>29.9±3.43</td>
<td>28.2±2.78</td>
<td>26.6±4.06</td>
</tr>
<tr>
<td>Ours</td>
<td>10</td>
<td>26.4±4.33</td>
<td>27.6±3.39</td>
<td>27.4±3.54</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>30.0±3.90</td>
<td>29.5±2.81</td>
<td>30.2±3.98</td>
</tr>
</tbody>
</table>
Results for A Retrieval Task

Algorithm A v. Baselines (PCA and CCA) for SIFT Query - SIFT Database (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>300</td>
<td>32.2±2.37</td>
<td>33.2±3.18</td>
<td>30.2±4.00</td>
</tr>
<tr>
<td>PCA</td>
<td>10</td>
<td>29.6±1.99</td>
<td>30.2±3.18</td>
<td>29.9±2.84</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>31.8±3.30</td>
<td>32.8±3.33</td>
<td>30.2±4.05</td>
</tr>
<tr>
<td>CCA</td>
<td>10</td>
<td>16.7±1.88</td>
<td>17.7±2.48</td>
<td>19.1±2.00</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>19.4±3.14</td>
<td>21.7±3.91</td>
<td>20.6±3.08</td>
</tr>
<tr>
<td>Ours</td>
<td>10</td>
<td>31.4±3.92</td>
<td>32.9±3.16</td>
<td>33.4±3.62</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>34.0±2.76</td>
<td>35.4±2.83</td>
<td>34.2±1.67</td>
</tr>
</tbody>
</table>
## Results for A Cross-Retrieval Task

Algorithm A v. Baselines (PCA and CCA) for **Color Query - SIFT Database** (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>10</td>
<td>9.3±1.66</td>
<td>9.3±2.03</td>
<td>10.0±2.31</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>9.4±1.17</td>
<td>10.7±1.38</td>
<td>10.5±2.04</td>
</tr>
<tr>
<td>CCA</td>
<td>10</td>
<td>15.4±4.27</td>
<td>15.8±4.53</td>
<td>15.9±4.59</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>16.2±4.83</td>
<td>16.8±5.27</td>
<td>18.2±6.30</td>
</tr>
<tr>
<td>Ours</td>
<td>10</td>
<td><strong>18.6±2.07</strong></td>
<td><strong>18.9±2.28</strong></td>
<td><strong>18.7±2.21</strong></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td><strong>20.4±3.43</strong></td>
<td><strong>20.4±2.88</strong></td>
<td><strong>21.8±3.21</strong></td>
</tr>
</tbody>
</table>
Results for A Cross-Retrieval Task

Algorithm A v. Baselines (PCA and CCA) for SIFT Query - Color Database (accuracy ± std):

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>10</td>
<td>8.2±2.54</td>
<td>9.2±3.35</td>
<td>9.4±3.36</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>8.6±2.65</td>
<td>9.8±2.47</td>
<td>9.8±3.33</td>
</tr>
<tr>
<td>CCA</td>
<td>10</td>
<td>12.5±2.98</td>
<td>13.8±2.36</td>
<td>13.8±2.82</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>13.2±1.77</td>
<td>13.2±2.32</td>
<td>13.4±2.62</td>
</tr>
<tr>
<td>Ours</td>
<td>10</td>
<td>19.0±3.63</td>
<td>20.8±3.52</td>
<td>22.0±3.98</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>22.6±2.07</td>
<td>22.9±1.93</td>
<td>22.4±4.30</td>
</tr>
</tbody>
</table>
Results for A Cross-Retrieval Task

Algorithm A v. Algorithm B (accuracy ± std):
Color Query - SIFT Database

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours Type A</td>
<td>10</td>
<td>18.6±2.07</td>
<td>18.9±2.28</td>
<td>18.7±2.21</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>20.4±3.43</td>
<td>20.4±2.88</td>
<td>21.8±3.21</td>
</tr>
<tr>
<td>Ours Type B</td>
<td>10</td>
<td>24.2±2.59</td>
<td>24.9±2.72</td>
<td>26.3±2.82</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>30.0±3.20</td>
<td>29.2±3.12</td>
<td>30.2±3.42</td>
</tr>
</tbody>
</table>

SIFT Query - Color Database

<table>
<thead>
<tr>
<th>Method</th>
<th>#dim</th>
<th>5-NN</th>
<th>10-NN</th>
<th>30-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours Type A</td>
<td>10</td>
<td>19.0±3.63</td>
<td>20.8±3.52</td>
<td>22.0±3.98</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>22.6±2.07</td>
<td>22.9±1.93</td>
<td>22.4±4.30</td>
</tr>
<tr>
<td>Ours Type B</td>
<td>10</td>
<td>18.8±3.59</td>
<td>19.1±3.14</td>
<td>19.4±3.71</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>27.8±4.27</td>
<td>26.8±4.28</td>
<td>27.0±3.09</td>
</tr>
</tbody>
</table>
Nonlinearity and Multiple Views

Kernelization:

- Representer Theorem,
  \[ w_1 = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot), \quad \text{and} \quad w_2 = \sum_{j=1}^{m} \beta_i l(y_j, \cdot), \]
  for a positive-definite kernel \( k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \) and a kernel \( l : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R} \).

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.
Take Home Messages

- We address the problem of projecting data in different representations into a shared space with a structure;

- We formulate an objective function that maps together matching samples and pushes apart non-matching samples;

- We show that this resulting objective function can be efficiently optimized using the convex-concave procedure (CCCP);

- Our proposed approach has a direct application for cross-media and content-based retrieval tasks.
Take Home Messages

• We address the problem of projecting data in different representations into a shared space with a structure;

• We formulate an objective function that maps together matching samples and pushes apart non-matching samples;

• We show that this resulting objective function can be efficiently optimized using the convex-concave procedure (CCCP);

• Our proposed approach has a direct application for cross-media and content-based retrieval tasks.
Take Home Messages

- We address the problem of projecting data in different representations into a shared space with a structure;

- We formulate an objective function that maps together matching samples and pushes apart non-matching samples;

- We show that this resulting objective function can be efficiently optimized using the convex-concave procedure (CCCP);

- Our proposed approach has a direct application for cross-media and content-based retrieval tasks.
Take Home Messages

- We address the problem of projecting data in different representations into a shared space with a structure;

- We formulate an objective function that maps together matching samples and pushes apart non-matching samples;

- We show that this resulting objective function can be efficiently optimized using the convex-concave procedure (CCCP);

- Our proposed approach has a direct application for cross-media and content-based retrieval tasks.
Thank you