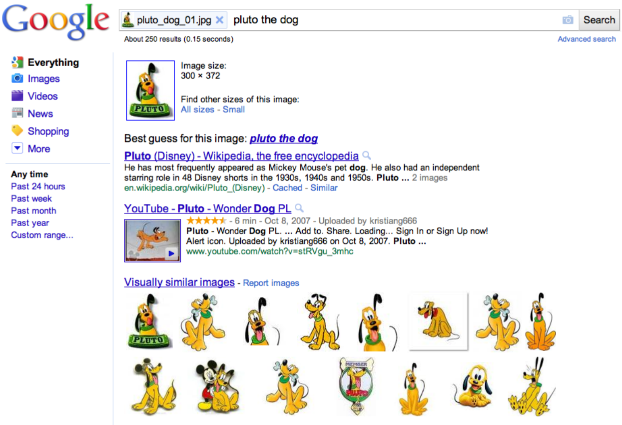
Learning Multi-View Neighborhood Preserving Projections

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

ICML Bellevue, 30th June 2011

An Image Content-Based Retrieval System

- Several features can be used to describe content of the images, such as Color, Text, shape, . . .

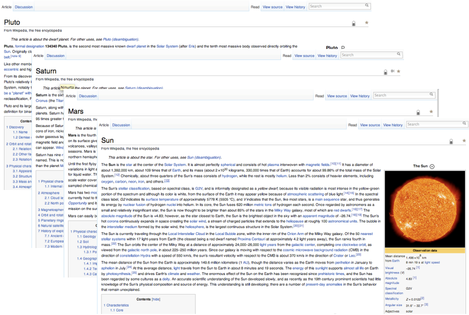


A Cross-Media Retrieval System

- Query objects and objects in the database have different representations.

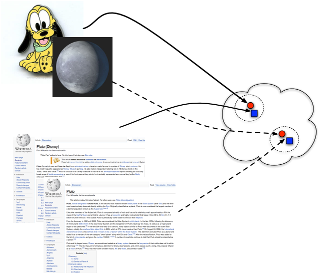
QUERY :

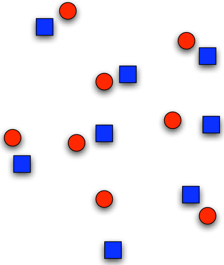
DATABASE :



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 $\mathcal{S}_{x_i}$ 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 $\{\mathcal{S}_{x_i}\}_{i=1}^m$.

Assumption:

- A linear parameterization of the functions $g_1^w(x_i) := \langle w_1, \phi(x_i) \rangle$ 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:

$$\underbrace{\sum_{i,j=1}^{m} L^{i,j}(w_1, w_2, x_i, y_j, S_{x_i})}_{\text{The Wisdom Loss}} + \underbrace{\eta\Omega(w_1) + \gamma\Omega(w_2)}_{\text{The Regularizer}}$$

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:

$$\underbrace{\sum_{i,j=1}^{m} L^{i,j}(w_1, w_2, x_i, y_j, \mathcal{S}_{x_i})}_{\text{The Wisdom Loss}} + \underbrace{\eta\Omega(w_1) + \gamma\Omega(w_2)}_{\text{The Regularizer}}$$

The Wisdom Loss Function

$$\underbrace{L^{i,j}(w_1, w_2, x_i, y_j, \mathcal{S}_{x_i})}_{\text{The Wisdom Loss}} = \underbrace{\frac{\mathbf{I}_{[\![y_j \in \mathcal{S}_{x_i}]\!]}}{2} \times L_1^{i,j}}_{\text{The Friends Loss}} + \underbrace{\frac{\left(1 - \mathbf{I}_{[\![y_j \in \mathcal{S}_{x_i}]\!]}\right)}{2} \times L_2^{i,j}}_{\text{The Enemies Loss}}$$

with

$$L_1^{i,j} = \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|_{\mathrm{Fro}}^2$$

$$L_2^{i,j}(\beta_d) = \max(0, 1 - \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|_{\mathrm{Fro}}^2).$$

The Wisdom Loss Function

$$\underbrace{L^{i,j}(w_1, w_2, x_i, y_j, \mathcal{S}_{x_i})}_{\text{The Wisdom Loss}} = \underbrace{\frac{\mathbf{I}_{[\![y_j \in \mathcal{S}_{x_i}]\!]}}{2} \times L_1^{i,j}}_{\text{The Friends Loss}} + \underbrace{\frac{\left(1 - \mathbf{I}_{[\![y_j \in \mathcal{S}_{x_i}]\!]}\right)}{2} \times L_2^{i,j}}_{\text{The Enemies Loss}}$$

with

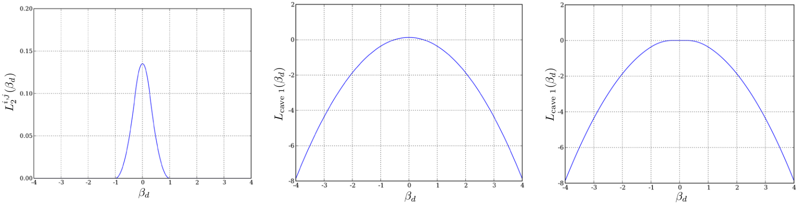
$$L_1^{i,j} = \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|^2_{\mathrm{Fro}}$$

~~$$L_2^{i,j}(\beta_d) = \max(0, 1 - \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|^2_{\mathrm{Fro}})$$~~

$$L_2^{i,j}(\beta_d) = \begin{cases} -\frac{1}{2}\beta_d^2 + \frac{a\lambda^2}{2}, & \text{if } 0 \le |\beta_d| < \lambda \\ \frac{|\beta_d|^2 - 2a\lambda|\beta_d| + a^2\lambda^2}{2(a-1)}, & \text{if } \lambda \le |\beta_d| \le a\lambda \\ 0, & \text{if } |\beta_d| \ge a\lambda, \end{cases}$$

where $\beta_d = \|g_1^{w_1}(x_i) - g_2^{w_2}(y_j)\|_{\mathrm{Fro}}$.

What So Special with the Wisdom Loss?



$$L_2^{i,j}(\beta_d) = L_{\mathsf{cv}}^1(\beta_d) - L_{\mathsf{cv}}^2(\beta_d)$$

Back to the Objective Function

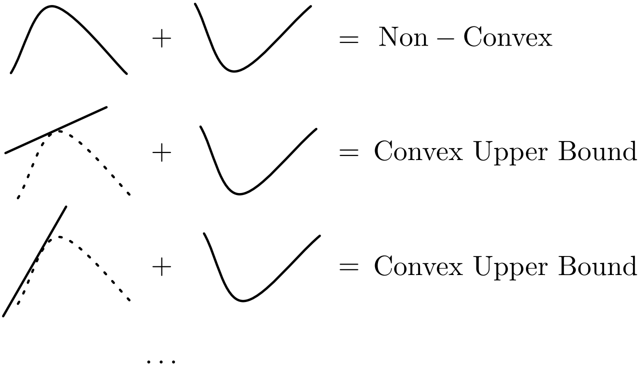
Recall:

- Objective = Enemies term + Friends term + Regularizers

$$\underbrace{\frac{\left(1-\mathbf{I}_{[\![y_j\in\mathcal{S}_{x_i}]\!]}\right)}{2}\times(L^1_{\text{cv}}}_{\text{Concave Function}}\underbrace{-L^2_{\text{cv}})+\frac{\mathbf{I}_{[\![y_j\in\mathcal{S}_{x_i}]\!]}}{2}\times L^{i,j}_1+\eta\Omega(w_1)+\gamma\Omega(w_2)}_{\text{Convex Functions}}$$

Optimization

The ConCave-Convex Procedure (CCCP):



Algorithm

Algorithm A–Multi-View Neighborhood Preserving Projection

Assume: $g_1^w(x_i) := \langle w_1, \phi(x_i)\rangle$ and $g_2^w(y_i) := \langle w_2, \psi(y_i)\rangle$
Input: $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_m\}$, $\{\mathcal{S}_{x_i}\}_{i=1}^m$, 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: $g_1^w(x_i) := \langle w_1, \phi(x_i)\rangle$ and $g_2^w(y_i) := \langle w_2, \psi(y_i)\rangle$
Input: $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_m\}$ and $\{\mathcal{S}_{x_i}\}_{i=1}^m$
Output: w_1^{PCA} and w_2^*
Initialize w_2
Solve CCCP w.r.t. w_2 while fixing $w_1 = w_1^{\mathrm{PCA}}$

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.

Results for A Retrieval Task

Algorithm A v. Baselines (PCA and CCA) for Color Query - Color Database (accuracy $\pm$ std):

Method	#dim	5-NN	10-NN	30-NN
Original	64	31.4±2.52	31.3±3.87	30.4±3.55
PCA	10	28.9±2.25	30.1±2.35	29.4±3.08
	50	31.3±3.12	31.4±3.46	30.3±2.99
CCA	10	24.8±3.86	24.7±3.42	24.0±3.78
	50	29.9±3.43	28.2±2.78	26.6±4.06
Ours	10	26.4±4.33	27.6±3.39	27.4±3.54
	50	30.0±3.90	29.5±2.81	30.2±3.98

Results for A Retrieval Task

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

Method	#dim	5-NN	10-NN	30-NN
Original	300	32.2±2.37	33.2±3.18	30.2±4.00
PCA	10	29.6±1.99	30.2±3.18	29.9±2.84
	50	31.8±3.30	32.8±3.33	30.2±4.05
CCA	10	16.7±1.88	17.7±2.48	19.1±2.00
	50	19.4±3.14	21.7±3.91	20.6±3.08
Ours	10	31.4±3.92	32.9±3.16	33.4±3.62
	50	34.0±2.76	35.4±2.83	34.2±1.67

Results for A Cross-Retrieval Task

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

Method	#dim	5-NN	10-NN	30-NN
PCA	10	9.3±1.66	9.3±2.03	10.0±2.31
	50	9.4±1.17	10.7±1.38	10.5±2.04
CCA	10	15.4±4.27	15.8±4.53	15.9±4.59
	50	16.2±4.83	16.8±5.27	18.2±6.30
Ours	10	**18.6±2.07**	**18.9±2.28**	**18.7±2.21**
	50	**20.4±3.43**	**20.4±2.88**	**21.8±3.21**

Results for A Cross-Retrieval Task

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

Method	#dim	5-NN	10-NN	30-NN
PCA	10	8.2±2.54	9.2±3.35	9.4±3.36
	50	8.6±2.65	9.8±2.47	9.8±3.33
CCA	10	12.5±2.98	13.8±2.36	13.8±2.82
	50	13.2±1.77	13.2±2.32	13.4±2.62
Ours	10	**19.0±3.63**	**20.8±3.52**	**22.0±3.98**
	50	**22.6±2.07**	**22.9±1.93**	**22.4±4.30**

Results for A Cross-Retrieval Task

Algorithm A v. Algorithm B (accuracy ± std):

Color Query - SIFT Database

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

SIFT Query - Color Database

Ours Type A	10	19.0±3.63	20.8±3.52	22.0±3.98
	50	22.6±2.07	22.9±1.93	22.4±4.30
Ours Type B	10	18.8±3.59	19.1±3.14	19.4±3.71
	50	27.8±4.27	26.8±4.28	27.0±3.09

Nonlinearity and Multiple Views

Kernelization:

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

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

ANU NICTA IST AUSTRIA Institute of Science and Technology

- 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.

Thanks

Thank you