Generalized Correlation Analysis for High-dimensional Multi-view Data

Songcan Chen

Joint work with Xiaohong Chen, Jingjing Gu and Xudong Zhou

Nanjing University of Aeronautics and Astronautics
Multi-view scenario

Object. People

View1

View2

View3

More …
Images retrieved for the text query: “at phoenix sky harbor on July 6, 1997. 757-2s7, n907 wa phoenix suns taxis past n902aw teamwork America west America west 757-2s7, n907wa phoenix suns taxis past n901aw Arizona at phoenix sky harbor on July 6, 1997.” The actual match is the middle picture in the first row. one object has 2 views (text & image)
The demosaicking problem deals with the reconstruction of a full color image $f$ from its partial sampling $m$. 

[Yac2004]
Detecting pixels correlated to sound by CCA on audio & video

[ESE2005] (CVPR05), [TTS2007](NIPS07) (Audio↔ Video)
Sensor Localization

Estimate physical coordinates of any sensors by using their signal strengths [PKYC2006]
In this case, CCA is more suitable for clear description of manifold than PCA, and result shows that CCA outperforms PCA in estimation accuracy. [SC2007] (Image $\leftrightarrow$ Pose)
Illustration of multi-view metric learning problem [ZCS2011]:
(a) Pose alignment of two observation sets in the same dim. space;
(b) Low-high resolution face matching in two different dim. spaces.
The basic overview of the proposed method. 
**W_x** and **W_y** are learnt projection matrices using PLS on X and Y. 

[SJ2011](CVPR11)
Image Feature Integration

The visual patterns descriptions LBP, GIST, CENTRIST, DoG-SIFT and HOG of three samples Images from Caltech 101 data.

[CNHK2011](CVPR2011)
Cross-view Action Recognition

Exemplar frames from IXMAS multi-view data set. Each row shows one action viewed from different angles.

[LSKS2011](CVPR2011)
Other Multi-view Scenarios

Multi-lingual Translation: [LS2006, AUG2009, KAG2010]

Multi-label Learning [ZS2011]

......
Motivation

- How to effectively analyze such multi-view data
  1) dimensionality reduction
  2) classification
  3) clustering, ...

- Popular Solution:
  Incorporating the prior information into the traditional canonical correlation analysis
According to the paired information between different views, classify them into two categories.
Outline

- Correlation Analysis for **Paired** Multi-view Data
- (Generalized) Correlation Analysis for **Semi-paired** Multi-view Data
- Ongoing & Future works
Outline

- Correlation Analysis for Paired Multi-view Data
- (Generalized) Correlation Analysis for Semi-paired Multi-view Data
- Ongoing & Future works
According to the **class label information** available or not, subclassify paired multi-view data into three categories:

\[
X = (x_1, x_2, \cdots, x_t, x_{t+1}, x_{t+2}, \cdots, x_n)
\]

\[
Y = (y_1, y_2, \cdots, y_t, y_{t+1}, y_{t+2}, \cdots, y_n)
\]

**Unsupervised**

\[
X = (x_1, x_2, \cdots, x_t, x_{t+1}, x_{t+2}, \cdots, x_n)
\]

\[
Y = (y_1, y_2, \cdots, y_t, y_{t+1}, y_{t+2}, \cdots, y_n)
\]

**Supervised**

\[
X = (x_1, x_2, \cdots, x_t, x_{t+1}, x_{t+2}, \cdots, x_n)
\]

\[
Y = (y_1, y_2, \cdots, y_t, y_{t+1}, y_{t+2}, \cdots, y_n)
\]

**Semi-supervised**
Dimensionality Reduction Methods

(1) Paired and **unsupervised** scenario
    ----- CCA, LPCCA, LCA

(2) Paired and **supervised** scenario
    ----- DCCA

(3) Paired and **semi-supervised** scenario
    ----- MVSSDR
CCA Illustration

Given $X$ set $\{x_1, \ldots, x_n\} \in \mathbb{R}^p$, and $Y$ set $\{y_1, \ldots, y_n\} \in \mathbb{R}^q$, CCA aims to simultaneously seek $w_x \in \mathbb{R}^p$, $w_y \in \mathbb{R}^q$, to ensure the correlation is maximized [Bor99].
CCA formulation

\[
\max_{w_x, w_y} \frac{\text{cov}(w_x^T x, w_y^T y)}{\sqrt{\text{var}(w_x^T x)} \sqrt{\text{var}(w_y^T y)}}
\]

\[
\max_{w_x, w_y} w_x^T X Y^T w_y
\]

s.t. \quad w_x^T X X^T w_x = 1, \quad w_y^T Y Y^T w_y = 1

where \( X = [x_1, \ldots, x_n], Y = [y_1, \ldots, y_n] \)

\[
\begin{pmatrix} X Y^T \\ Y X^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} X X^T \\ Y Y^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix}
\]

A Generalized Eigenvalue Problem (GEP)
CCA Equivalent Formulations

\[
\begin{align*}
\text{max} & \quad w_x^T X Y^T w_y \\
\text{s.t.} & \quad w_x^T X X^T w_x = 1 \\
& \quad w_y^T Y Y^T w_y = 1
\end{align*}
\]

\[
\begin{align*}
\text{max} & \quad w_x^T \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)(y_i - y_j)^T \cdot w_y \\
\text{s.t.} & \quad w_x^T \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)(x_i - x_j)^T = 1 \\
& \quad w_y^T \sum_{i=1}^n \sum_{j=1}^n (y_i - y_j)(y_i - y_j)^T = 1
\end{align*}
\]

\[
\begin{align*}
\min_{w_x, w_y} & \quad \sum_{i=1}^n \left\| w_x^T (x_i - \bar{x}) - w_y^T (y_i - \bar{y}) \right\|^2 \\
\text{s.t.} & \quad \sum_{i=1}^n \left\| w_x^T (x_i - \bar{x}) \right\|^2 = 1, \quad \sum_{i=1}^n \left\| w_y^T (y_i - \bar{y}) \right\|^2 = 1
\end{align*}
\]

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^n w_x^T (x_i - \bar{x})(y_i - \bar{y})^T w_y \\
\text{s.t.} & \quad \sum_{i=1}^n w_x^T (x_i - \bar{x})(x_i - \bar{x})^T = 1 \\
& \quad \sum_{i=1}^n w_y^T (y_i - \bar{y})(y_i - \bar{y})^T = 1
\end{align*}
\]

Basis of our LPCCA

Basis of our LCA
LPCCA: Locality Preserving CCA

- **Motivation:** CCA is linear and insufficient to reflect the nonlinear correlation between two **manifold** sets X and Y
- **Solution:** Use **local** correlation to replace **global** correlation of CCA
- **Objective:**

\[
\max_{w_x, w_y} \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij}^x (x_i - x_j) S_{ij}^y (y_i - y_j)^T \cdot w_y
\]

s.t. \(\sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij}^{x2} (x_i - x_j)(x_i - x_j)^T \cdot w_x = 1\)

\(w_x^T \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij}^{y2} (y_i - y_j)(y_i - y_j)^T \cdot w_y = 1\)

\[
S_{ij}^x = \begin{cases} 
\exp \left(-\frac{||x_i - x_j||^2}{t_x}\right), & \text{if } x_j \in \text{LN}(x_i) \text{ or } x_i \in \text{LN}(x_j) \\
0, & \text{otherwise}
\end{cases}
\]

Note: “LN” denotes **local** neighborhoods. [SC2007]
Experiment of LPCCA

Error distribution histograms
**LCA: Locality Correlation Analysis**

**CCA:**
\[
\max_{w_x, w_y} \sum_{i=1}^{n} w_x^T (x_i - \bar{x})(y_i - \bar{y})^T w_y \\
\text{s.t.} \sum_{i=1}^{n} w_x^T (x_i - \bar{x})(x_i - \bar{x})^T w_x = 1 \\
\sum_{i=1}^{n} w_y^T (y_i - \bar{y})(y_i - \bar{y})^T w_y = 1
\]

**LCA:**
\[
\max w_x^T \cdot \sum_{i=1}^{n} \left( x_i - \sum_{x_j \in \text{ne}(x_i)} S_{ij}^X x_j \right) \left( y_i - \sum_{y_j \in \text{ne}(y_i)} S_{ij}^Y y_j \right)^T \cdot w_y \\
\text{s.t.} \ w_x^T \cdot \sum_{i=1}^{n} \left( x_i - \sum_{x_j \in \text{ne}(x_i)} S_{ij}^X x_j \right) \left( x_i - \sum_{x_j \in \text{ne}(x_i)} S_{ij}^X x_j \right)^T \cdot w_x = 1 \\
\ w_y^T \cdot \sum_{i=1}^{n} \left( y_i - \sum_{y_j \in \text{ne}(y_i)} S_{ij}^Y y_j \right) \left( y_i - \sum_{y_j \in \text{ne}(y_i)} S_{ij}^Y y_j \right)^T \cdot w_y = 1
\]

Use local means to substitute global means! [GC2011]
LCA: Locality Correlation Analysis

Denote

\[ M_{xy} = I - C^x - C^y + C^x \circ C^y \]
\[ M_{xx} = I - 2C^x + C^x \circ C^x \]
\[ M_{yy} = I - 2C^y + C^y \circ C^y \]

LCA can be formulated as

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T XM_{xy} Y^T w_y \\
\text{s.t.} & \quad w_x^T XM_{xx} X^T w_x = 1 \\
& \quad w_y^T YM_{yy} Y^T w_y = 1
\end{align*}
\]

Which can be showed in Matrix Form (GEP)

\[
\begin{pmatrix}
X M_{xy} Y^T \\
Y M_{yx} X^T \\
YM_{yx} X^T
\end{pmatrix}
\begin{pmatrix}
w_x \\
w_y
\end{pmatrix}
= \lambda
\begin{pmatrix}
XM_{xx} x^T \\
YM_{yy} Y^T
\end{pmatrix}
\begin{pmatrix}
w_x \\
w_y
\end{pmatrix}
\]
Localization accuracy vs # training samples
Selective Materials on this case

- Too many, just name a few:


[Che06&07] Songcan Chen, CCA with applications, MLA06&07’ Slides
(1) Paired and unsupervised scenario
    ----- CCA, LPCCA, LCA

(2) Paired and supervised scenario
    ----- DCCA

(3) Paired and semi-supervised scenario
    ----- MVSSDR
DCCA: Discriminant CCA

- **Motivation**: in CCA, the correlation between \((x_i, y_i)\) is insufficient to discriminate between classes.

- **Solution**: takes correlation as similarity metric, aims to maximize the within-class correlation, minimize the between-class correlation.

- **Objective**:

  \[
  \max_{w_x, w_y} \frac{w_x^T (C_w - \eta C_b) w_y}{(w_x^T C_{xx} w_x \cdot w_y^T C_{yy} w_y)^{1/2}}
  \]

  \[
  C_w = \sum_{i=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} x_k^{(i)} y_l^{(i)^T},
  C_b = \sum_{i=1}^{c} \sum_{j=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \sum_{j \neq i} x_k^{(i)} y_l^{(j)^T}
  \]

  [SCY08]
DCCA: Discriminant CCA

- **within-class correlation**
  \[
  C_w = \sum_{i=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_i} x_k^{(i)} y_l^{(i)T} = XAY^T
  \]

- **between-class correlation**
  \[
  C_b = \sum_{i=1}^{c} \sum_{j=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} x_k^{(i)} y_l^{(j)T}
  \]

- **Objective function**
  \[
  \max_{w_x, w_y} \frac{w_x^T(C_w - \eta C_b)w_y}{(w_x^T C_{xx} w_x \cdot w_y^T C_{yy} w_y)^{1/2}}
  \]
# DCCA Results

### Handwritten numeral recognition (Multiple Feature database)

<table>
<thead>
<tr>
<th>#</th>
<th>X</th>
<th>Y</th>
<th>Recognition accuracy using different methods</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>DCCA PR1</td>
<td></td>
<td>CCA PR1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>mfeat_fac</td>
<td>mfeat_fou</td>
<td>0.9560</td>
<td>0.9813</td>
<td>0.8673</td>
<td>0.8785</td>
</tr>
<tr>
<td>2</td>
<td>mfeat_fac</td>
<td>mfeat_kar</td>
<td>0.9752</td>
<td>0.9789</td>
<td>0.9603</td>
<td>0.9598</td>
</tr>
<tr>
<td>3</td>
<td>mfeat_fac</td>
<td>mfeat_mor</td>
<td>0.9077</td>
<td>0.9302</td>
<td>0.7596</td>
<td>0.7656</td>
</tr>
<tr>
<td>4</td>
<td>mfeat_fac</td>
<td>mfeat_pix</td>
<td>0.9718</td>
<td>0.9752</td>
<td>0.9472</td>
<td>0.9476</td>
</tr>
<tr>
<td>5</td>
<td>mfeat_fac</td>
<td>mfeat_zer</td>
<td>0.9589</td>
<td>0.9772</td>
<td>0.8542</td>
<td>0.8623</td>
</tr>
<tr>
<td>6</td>
<td>mfeat_fou</td>
<td>mfeat_kar</td>
<td>0.9393</td>
<td>0.9687</td>
<td>0.8969</td>
<td>0.9195</td>
</tr>
<tr>
<td>7</td>
<td>mfeat_fou</td>
<td>mfeat_mor</td>
<td>0.8089</td>
<td>0.8278</td>
<td>0.7567</td>
<td>0.7633</td>
</tr>
<tr>
<td>8</td>
<td>mfeat_fou</td>
<td>mfeat_pix</td>
<td>0.9373</td>
<td>0.9662</td>
<td>0.8270</td>
<td>0.8431</td>
</tr>
<tr>
<td>9</td>
<td>mfeat_fou</td>
<td>mfeat_zer</td>
<td>0.8367</td>
<td>0.8543</td>
<td>0.8239</td>
<td>0.8351</td>
</tr>
<tr>
<td>10</td>
<td>mfeat_kar</td>
<td>mfeat_mor</td>
<td>0.8928</td>
<td>0.9253</td>
<td>0.7857</td>
<td>0.8158</td>
</tr>
<tr>
<td>11</td>
<td>mfeat_kar</td>
<td>mfeat_pix</td>
<td>0.9493</td>
<td>0.9497</td>
<td>0.9643</td>
<td>0.9641</td>
</tr>
<tr>
<td>12</td>
<td>mfeat_kar</td>
<td>mfeat_zer</td>
<td>0.9383</td>
<td>0.9638</td>
<td>0.9081</td>
<td>0.9211</td>
</tr>
<tr>
<td>13</td>
<td>mfeat_mor</td>
<td>mfeat_pix</td>
<td>0.8799</td>
<td>0.9100</td>
<td>0.7263</td>
<td>0.7602</td>
</tr>
<tr>
<td>14</td>
<td>mfeat_mor</td>
<td>mfeat_zer</td>
<td>0.7943</td>
<td>0.8097</td>
<td>0.7258</td>
<td>0.7452</td>
</tr>
<tr>
<td>15</td>
<td>mfeat_pix</td>
<td>mfeat_zer</td>
<td>0.9310</td>
<td>0.9544</td>
<td>0.8232</td>
<td>0.8398</td>
</tr>
</tbody>
</table>

Note: PR1 and PR2 correspond to features in parallel and in serial, respectively.
Dimensionality Reduction Methods

(1) Paired and unsupervised scenario
    ----- CCA, LPCCA, LCA

(2) Paired and supervised scenario
    ----- DCCA

(3) Paired and semi-supervised scenario
    ----- MVSSDR
MVSSDR: Multi-view Semi-supervised Dimensionality Reduction

Inspired by the success of SSDR[ZZC2007], Hou et al. utilize must-link and cannot-link constraints to realize dimensionality reduction for semi-supervised and fully-paired multi-view data.

Fig. 1. The basic procedure of MVSSDR.
They believe that each view pattern can be approximated by the consensus pattern through linear transformation. Thus, the optimization objective of MVSSDR is

$$\min J(W_1, \ldots, W_l, P_1, \ldots, P_l, Y) = \sum_{v=1}^{l} \| W_v^T X^{(v)} - P_v Y \|^2 - \lambda \sum_{v=1}^{l} \text{tr}(W_v^T X^{(v)} L^{(v)} (X^{(v)})^T W_v^T)$$

where

$$L^{(v)} = D^{(v)} - S^{(v)} , \quad S^{(v)}_{ij} = \begin{cases} \frac{1}{n^2} + \frac{\alpha_v}{n_C^{(v)}}, & \text{if } (x_i^{(v)}, x_j^{(v)}) \in C_v, \\ \frac{1}{n^2} - \frac{\beta_v}{n_M^{(v)}}, & \text{if } (x_i^{(v)}, x_j^{(v)}) \in M_v, \\ \frac{1}{n^2}, & \text{otherwise,} \end{cases}$$
CCA, LPCCA, LCA, DCCA and MVSSDR are designed all for paired multi-view data, which are not so effective for semi-paired multi-view data.

In fact, Semi-paired multi-view data is widely existed in real world applications, such as multimodal image classification.
Some improved algorithms of CCA that can effectively deal with such **semi-paired** scenario have emerged, we name them generalized correlation analysis (GCA).
Outline

- Correlation Analysis for Paired Multi-view Data
- (Generalized) Correlation Analysis for Semi-paired Multi-view Data
- Ongoing & Future works
Dimensionality Reduction Methods

(1) Semi-paired and **unsupervised** scenario
   -- SemiCCA, Regularized KCCA, PPLCA, NCA

(2) Semi-paired and **supervised** scenario
   ----- DCCAM

(3) Semi-paired and **semi-supervised** scenario
   ----- S²GCA
Dimensionality Reduction Methods

(1) Semi-paired and unsupervised scenario
   -- SemiCCA, Regularized KCCA, PPLCA, NCA

(2) Semi-paired and supervised scenario
    ----- DCCAM

(3) Semi-paired and semi-supervised scenario
    ----- $S^2$GCA
SemiCCA [KKS2010](ICPR2010)

\[
\hat{X} = (x_1, x_2, \ldots, x_N, x_1', x_2', \ldots, x_{N_x}')
\]

\[
\hat{Y} = (y_1, y_2, \ldots, y_N, y_1', y_2', \ldots, y_{N_y}')
\]

Motivation: incorporate the additional unpaired samples for DR
Corresponding objective function **inversely derived by us**:

\[
\begin{align*}
\text{max} \quad & \sum_{i=1}^{N} \beta w_x^T (x_i - \bar{x})(y_i - \bar{y})^T w_y + \frac{1 - \beta}{2} \left[ w_x^T \tilde{C}_{xx} w_x + w_y^T \tilde{C}_{yy} w_y \right] \\
\text{s.t.} \quad & \beta \left[ w_x^T C_{xx} w_x + w_y^T C_{yy} w_y \right] = 2
\end{align*}
\]
Regularized KCCA [BLG2008]

\[
\tilde{X} = (x_1, x_2, \ldots, x_N, x_1', x_2', \ldots, x_{N_x}')
\]

\[
\tilde{Y} = (y_1, y_2, \ldots, y_N, y_1', y_2', \ldots, y_{N_y}')
\]

**Motivation:** meaningful directions are not only those that have high correlation but also capture the manifold structure of the data.
RKCCA:

$$\max_{\alpha, \beta} \frac{\alpha^T K_{\hat{X}X} K_{YY} \beta}{\sqrt{\alpha^T (K_{\hat{X}X} K_{XX} + R_{\hat{X}}) \alpha \beta^T (K_{\hat{Y}Y} K_{YY} + R_{\hat{Y}}) \beta}}$$

where

$$R_{\hat{X}} = \varepsilon_{\hat{X}} K_{\hat{XX}} + \frac{\gamma_{\hat{X}}}{m_{\hat{X}}} K_{\hat{XX}} L_{\hat{X}} K_{\hat{XX}}$$

$$L_{\hat{X}} = D_{\hat{X}} - W_{\hat{X}}$$

$$K_{XX} = \phi(X)^T \phi(X) \quad K_{\hat{XX}} = \phi(\hat{X})^T \phi(X) \quad K_{\hat{XX}} = \phi(\hat{X})^T \phi(\hat{X})$$

linearized

$$\max_{w_x, w_y} \frac{w_x^T \tilde{X} \tilde{Y}^T w_y}{\sqrt{w_x^T \left( \tilde{X} \tilde{X}^T + \varepsilon_X I + \frac{\gamma_X}{n_x} XL X^T \right) w_x \sqrt{w_y^T \left( \tilde{Y} \tilde{Y}^T + \varepsilon_Y I + \frac{\gamma_Y}{n_y} YL Y^T \right) w_y}}}$$

local information
\[
\text{RKCCA GEP}
\]
\[
\begin{align*}
\text{max} & \quad \mathbf{w}_x^T \tilde{X}\tilde{Y}^T \mathbf{w}_y \\
\text{s.t.} & \quad \mathbf{w}_x^T \left( \tilde{X}\tilde{X}^T + \varepsilon_x \mathbf{I} + \frac{\gamma_x}{n_x^2} \mathbf{X} \mathbf{L}_x \mathbf{X}^T \right) \mathbf{w}_x = 1 \\
& \quad \mathbf{w}_y^T \left( \tilde{Y}\tilde{Y}^T + \varepsilon_y \mathbf{I} + \frac{\gamma_y}{n_y^2} \mathbf{Y} \mathbf{L}_y \mathbf{Y}^T \right) \mathbf{w}_y = 1
\end{align*}
\]
\[
\text{Matrix Form}
\]
\[
\begin{bmatrix}
0 & \tilde{X}\tilde{Y}^T \\
\tilde{Y}\tilde{X}^T & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{w}_x \\
\mathbf{w}_y
\end{bmatrix} = \lambda
\begin{bmatrix}
\tilde{X}\tilde{X}^T + \varepsilon_x \mathbf{I} + \frac{\gamma_x}{n_x^2} \mathbf{X} \mathbf{L}_x \mathbf{X}^T & 0 \\
0 & \text{Diagonal}
\end{bmatrix}
\begin{bmatrix}
\mathbf{w}_x \\
\mathbf{w}_y
\end{bmatrix}
\]
PPLCA: Partially-paired Locality Correlation Analysis [GSC2011]

\[ \hat{X} = (x_1, x_2, \cdots, x_N, x'_1, x'_2, \cdots, x'_{N_x}) \]

\[ \hat{Y} = (y_1, y_2, \cdots, y_N, y'_1, y'_2, \cdots, y'_{N_y}) \]

**Motivation:** LCA neglects the neighbor information of unpaired data of each view data.

**Solution:** Incorporate the neighbor information into the definition of those means which involve unpaired data!
**PPLCA**

\[
\max_{\hat{\mathbf{w}}_x, \hat{\mathbf{w}}_y} \hat{\mathbf{w}}_x^T \sum_{i=1}^{n} \left( x_i - \sum_{j=1}^{n+ps} \hat{S}_{ji} x_j \right) \left( y_i - \sum_{j=1}^{n+pl} \hat{S}_{ji} y_j \right) \hat{\mathbf{w}}_y \\
\text{s.t.} \hat{\mathbf{w}}_x^T \sum_{i=1}^{n} \left( x_i - \sum_{j=1}^{n+ps} \hat{S}_{ji} x_j \right) \left( x_i - \sum_{j=1}^{n+ps} \hat{S}_{ji} x_j \right) \hat{\mathbf{w}}_x = 1 \\
\hat{\mathbf{w}}_y^T \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{n+pl} \hat{S}_{ji} y_j \right) \left( y_i - \sum_{j=1}^{n+pl} \hat{S}_{ji} y_j \right) \hat{\mathbf{w}}_y = 1
\]

**Matrix form**

\[
\begin{pmatrix}
\hat{X} \hat{G}_{XY} \hat{Y}^T \\
\hat{Y} \hat{G}_{YX} \hat{X}^T
\end{pmatrix}
\begin{pmatrix}
\hat{\mathbf{w}}_x \\
\hat{\mathbf{w}}_y
\end{pmatrix} = \lambda \begin{pmatrix}
\hat{X} \hat{G}_{XX} \hat{X}^T \\
\hat{Y} \hat{G}_{YY} \hat{Y}^T
\end{pmatrix}
\begin{pmatrix}
\hat{\mathbf{w}}_x \\
\hat{\mathbf{w}}_y
\end{pmatrix}
\]
Experiments in PPLCA: Fire alarming WSN (1)
Experiments in PPLCA:
Fire alarming WSN (2)
Experiments in PPLCA:
HKUST sensor network(1)
Experiments in PPLCA: HKUST sensor network (2)
NCA: Neighborhood Correlation Analysis

- **Motivation:** In semi-paired situation, CCA tends to overfitting due to limited paired data.

- **Solution:** consider the correlations not only between paired samples but also between each sample in one view and its between-view local neighbors in the other view.
Construction of between-view neighborhood relationships

\[ S_{ij}^{XY} = \sum_{k=1}^{p} S_{ik}^x \times S_{kj}^y \] [SGL2010]

NCA Formulation & its two variants

**NCA**

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T X S^{XY} Y^T w_y \\
\text{s.t.} & \quad w_x^T X D^R X^T w_x = 1, \quad w_y^T Y D^C Y^T w_y = 1.
\end{align*}
\]

\[
D_{ii}^R = \sum_j S_{ij}^{XY}, \quad D_{jj}^C = \sum_i S_{ij}^{XY}
\]

**Two Variants**

**LRNCA**

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T X S^{XY} Y^T w_y \\
\text{s.t.} & \quad w_x^T (X D^R X^T + \gamma_x X L_x X^T) w_x = 1 \\
& \quad w_y^T (Y D^C Y^T + \gamma_y Y L_y Y^T) w_y = 1.
\end{align*}
\]

**PRNCA**

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T X S^{XY} Y^T w_y + \eta \left( \frac{1}{2} w_x^T C_{xx} w_x + \frac{1}{2} w_y^T C_{yy} w_y \right) \\
\text{s.t.} & \quad w_x^T X D^R X^T w_x + w_y^T Y D^C Y^T w_y + \eta \left( w_x^T w_x + w_y^T w_y \right) = 1.
\end{align*}
\]
## Experiments on WebKB

<table>
<thead>
<tr>
<th></th>
<th>CCA and its variants</th>
<th>NCA and its variants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCA</td>
<td>SemiLRCCA</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>76.80 ± 6.7</td>
<td>79.20 ± 4.1</td>
</tr>
<tr>
<td>Link</td>
<td>76.45 ± 8.9</td>
<td>77.00 ± 8.1</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>77.15 ± 7.9</td>
<td>81.70 ± 2.8</td>
</tr>
<tr>
<td>Link</td>
<td>77.95 ± 8.0</td>
<td>81.05 ± 7.6</td>
</tr>
<tr>
<td>15%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>81.95 ± 5.7</td>
<td>82.15 ± 6.5</td>
</tr>
<tr>
<td>Link</td>
<td>79.35 ± 8.6</td>
<td>83.05 ± 8.8</td>
</tr>
</tbody>
</table>

LR- Laplacian Regularized and PR- PCA Regularized
## Experiments on MFD(part)

<table>
<thead>
<tr>
<th></th>
<th>CCA and its variants</th>
<th>NCA and its variants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCA</td>
<td>SemiLRCCA</td>
</tr>
<tr>
<td>1</td>
<td>Fac</td>
<td>79.39 ± 2.8</td>
</tr>
<tr>
<td></td>
<td>Kar</td>
<td>56.23 ± 3.7</td>
</tr>
<tr>
<td>2</td>
<td>Fac</td>
<td>69.45 ± 3.7</td>
</tr>
<tr>
<td></td>
<td>Zer</td>
<td>58.99 ± 3.5</td>
</tr>
<tr>
<td>3</td>
<td>Fou</td>
<td>62.79 ± 4.1</td>
</tr>
<tr>
<td></td>
<td>Kar</td>
<td>59.72 ± 4.2</td>
</tr>
<tr>
<td>4</td>
<td>Fou</td>
<td>49.09 ± 5.1</td>
</tr>
<tr>
<td></td>
<td>Zer</td>
<td>62.60 ± 3.6</td>
</tr>
<tr>
<td>5</td>
<td>Kar</td>
<td>47.87 ± 3.7</td>
</tr>
<tr>
<td></td>
<td>Zer</td>
<td>63.45 ± 3.9</td>
</tr>
<tr>
<td>6</td>
<td>Pix</td>
<td>60.74 ± 4.1</td>
</tr>
<tr>
<td></td>
<td>Zer</td>
<td>65.35 ± 2.6</td>
</tr>
</tbody>
</table>
Dimensionality Reduction Methods

(1) Semi-paired and unsupervised scenario
   --SemiCCA, Regularized KCCA, PPLCA, NCA

(2) Semi-paired and supervised scenario
    ----- DCCAM

(3) Semi-paired and semi-supervised scenario
    ----- S²GCA
Motivation: consider the paired and unpaired information together

Objective:

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T XAY^T w_y \\
\text{s.t.} & \quad w_x^T X X^T w_x = 1, \quad w_y^T Y Y^T w_y = 1
\end{align*}
\]

\[A = \begin{bmatrix} I_{m \times n_t} & \cdots & I_{m \times n_t} \\ \vdots & \ddots & \vdots \\ I_{m \times n_t} & \cdots & I_{m \times n_t} \end{bmatrix} \in \mathbb{R}^{m \times n_t} \]
### Experiments of DCCAM

#### The recognition accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>RA-I</th>
<th>RA-II</th>
<th>RA-III</th>
<th>RA-IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCCAM</td>
<td>0.9437</td>
<td>0.9441</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>DCCA-I</td>
<td>0.9407</td>
<td>0.9339</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CCA-I</td>
<td>0.8979</td>
<td>0.9045</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>PLS-I</td>
<td>0.8985</td>
<td>0.9062</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>DCCA-II</td>
<td><strong>0.9509</strong></td>
<td><strong>0.9480</strong></td>
<td><strong>0.9522</strong></td>
<td><strong>0.9486</strong></td>
</tr>
<tr>
<td>CCA-II</td>
<td>0.9062</td>
<td>0.9149</td>
<td>0.9251</td>
<td>0.9292</td>
</tr>
<tr>
<td>PLS-II</td>
<td>0.9043</td>
<td>0.9117</td>
<td>0.9232</td>
<td>0.9277</td>
</tr>
<tr>
<td>CCA-III</td>
<td>0.8942</td>
<td>0.9008</td>
<td>0.8962</td>
<td>0.8987</td>
</tr>
<tr>
<td>PLS-III</td>
<td>0.8947</td>
<td>0.9042</td>
<td>0.8976</td>
<td>0.9028</td>
</tr>
</tbody>
</table>
Dimensionality Reduction Methods

(1) Semi-paired and unsupervised scenario
--SemiCCA, Regularized KCCA, PPLCA, NCA

(2) Semi-paired and supervised scenario
----- DCCAM

(3) Semi-paired and semi-supervised scenario
----- $S^2$GCA
Semi-supervised and semi-paired generalized correlation analysis

Paired data

Unpaired data

X-view data

Y-view data

Semi-paired CCA

Semi-supervised Learning
S²GCA: semi-supervised and semi-paired generalized correlation analysis

Objective:

\[
\max_{w_x, w_y} \ w_x^T \tilde{C}_{xy} \ w_y + \frac{\eta}{2} \left[ w_x^T \left( S_{rlb}^X - S_{rlw}^X \right) w_x + w_y^T \left( S_{rlb}^Y - S_{rlw}^Y \right) w_y \right]
\]

s.t. \( w_x^T \tilde{C}_{xx} w_x + w_y^T \tilde{C}_{yy} w_y = 1 \)

\[
\begin{bmatrix}
\eta \left( S_{rlb}^X - S_{rlw}^X \right) & \tilde{C}_{xy} \\
\tilde{C}_{xy}^T & \eta \left( S_{rlb}^Y - S_{rlw}^Y \right)
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix}
= \lambda
\begin{bmatrix}
\tilde{C}_{xx} & 0 \\
0 & \tilde{C}_{yy}
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix}
\]

Experiments in SSL(1)

Experimental Benchmark Datasets [KKS2010]
Experiments in SSL(2)
Experiments in SSL(3)
<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>CCA</th>
<th>DCCA</th>
<th>SCCA</th>
<th>SemiCCA</th>
<th>SemiLRCCA</th>
<th>S²GCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fac</td>
<td>46.25 ± 0.5</td>
<td>10.88 ± 0.3</td>
<td>54.76 ± 0.4</td>
<td>72.38 ± 2.8</td>
<td>88.72 ± 0.3</td>
<td>85.27 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>Fac</td>
<td>61.79 ± 1.0</td>
<td>11.19 ± 0.3</td>
<td>54.31 ± 0.9</td>
<td>75.94 ± 9.3</td>
<td>88.31 ± 0.5</td>
<td>88.82 ± 0.8</td>
</tr>
<tr>
<td>3</td>
<td>Fac</td>
<td>57.62 ± 0.9</td>
<td>10.37 ± 0.3</td>
<td>36.71 ± 0.8</td>
<td>59.28 ± 0.4</td>
<td>81.78 ± 1.0</td>
<td>83.68 ± 0.5</td>
</tr>
<tr>
<td>4</td>
<td>Fac</td>
<td>44.33 ± 0.6</td>
<td>9.93 ± 0.1</td>
<td>75.99 ± 0.2</td>
<td>58.86 ± 0.6</td>
<td>89.48 ± 0.1</td>
<td>79.11 ± 0.3</td>
</tr>
<tr>
<td>5</td>
<td>Fac</td>
<td>54.63 ± 0.6</td>
<td>10.26 ± 0.1</td>
<td>51.17 ± 0.1</td>
<td>61.0 ± 0.3</td>
<td>88.44 ± 0.6</td>
<td>90.37 ± 0.5</td>
</tr>
<tr>
<td>6</td>
<td>Fou</td>
<td>50.13 ± 0.3</td>
<td>47.21 ± 1.0</td>
<td>66.31 ± 1.3</td>
<td>77.27 ± 0.3</td>
<td>69.89 ± 0.2</td>
<td>78.48 ± 1.0</td>
</tr>
<tr>
<td>7</td>
<td>Fou</td>
<td>54.05 ± 1.0</td>
<td>42.91 ± 0.7</td>
<td>44.23 ± 0.3</td>
<td>74.67 ± 0.2</td>
<td>56.36 ± 0.4</td>
<td>78.40 ± 0.1</td>
</tr>
<tr>
<td>8</td>
<td>Fou</td>
<td>48.91 ± 0.4</td>
<td>47.21 ± 1.0</td>
<td>67.55 ± 0.4</td>
<td>77.46 ± 0.2</td>
<td>71.99 ± 0.1</td>
<td>78.45 ± 0.2</td>
</tr>
<tr>
<td>9</td>
<td>Fou</td>
<td>52.80 ± 0.5</td>
<td>47.21 ± 1.0</td>
<td>63.87 ± 0.3</td>
<td>76.63 ± 0.3</td>
<td>68.75 ± 0.2</td>
<td>78.99 ± 0.4</td>
</tr>
<tr>
<td>10</td>
<td>Kar</td>
<td>59.18 ± 0.5</td>
<td>58.59 ± 0.5</td>
<td>54.71 ± 0.6</td>
<td>84.22 ± 0.2</td>
<td>72.95 ± 1.2</td>
<td>81.86 ± 0.4</td>
</tr>
<tr>
<td>11</td>
<td>Kar</td>
<td>66.83 ± 0.4</td>
<td>75.78 ± 0.8</td>
<td>85.75 ± 0.7</td>
<td>84.22 ± 0.2</td>
<td>90.31 ± 0.1</td>
<td>84.92 ± 0.7</td>
</tr>
<tr>
<td>12</td>
<td>Kar</td>
<td>66.39 ± 0.9</td>
<td>75.78 ± 0.8</td>
<td>82.84 ± 0.2</td>
<td>84.22 ± 0.2</td>
<td>79.82 ± 0.1</td>
<td>89.46 ± 0.4</td>
</tr>
<tr>
<td>13</td>
<td>Mor</td>
<td>65.44 ± 0.6</td>
<td>65.21 ± 0.6</td>
<td>67.73 ± 0.3</td>
<td>65.86 ± 0.6</td>
<td>63.81 ± 0.9</td>
<td>70.36 ± 0.6</td>
</tr>
<tr>
<td>14</td>
<td>Mor</td>
<td>65.44 ± 0.6</td>
<td>65.21 ± 0.6</td>
<td>67.73 ± 0.3</td>
<td>65.88 ± 0.6</td>
<td>63.84 ± 0.9</td>
<td>70.17 ± 0.5</td>
</tr>
<tr>
<td>15</td>
<td>Pix</td>
<td>36.87 ± 0.5</td>
<td>10.38 ± 0.2</td>
<td>30.35 ± 0.4</td>
<td>73.08 ± 0.2</td>
<td>82.58 ± 0.1</td>
<td>67.28 ± 0.6</td>
</tr>
</tbody>
</table>
## Experiments in WebKB

<table>
<thead>
<tr>
<th>Labeled %</th>
<th>fullext</th>
<th>inlinks</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CCA</strong></td>
<td>74.11±8.6</td>
<td>81.09±0.7</td>
</tr>
<tr>
<td><strong>DCCA</strong></td>
<td>90.32±0.63</td>
<td>90.99±0.2</td>
</tr>
<tr>
<td><strong>SCCA</strong></td>
<td>71.35±17.1</td>
<td>78.00±0.7</td>
</tr>
<tr>
<td><strong>SemiCCA</strong></td>
<td>91.41±1.3</td>
<td>90.78±1.6</td>
</tr>
<tr>
<td><strong>SemiLRCCA</strong></td>
<td>88.21±1.5</td>
<td>85.85±1.1</td>
</tr>
<tr>
<td><strong>S²GCA</strong></td>
<td>92.88±2.2</td>
<td>93.60±0.9</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CCA</strong></td>
<td>78.57±1.5</td>
<td>80.91±3.1</td>
</tr>
<tr>
<td><strong>DCCA</strong></td>
<td>87.96±1.5</td>
<td>87.03±0.6</td>
</tr>
<tr>
<td><strong>SCCA</strong></td>
<td>76.70±5.7</td>
<td>79.14±0.4</td>
</tr>
<tr>
<td><strong>SemiCCA</strong></td>
<td>94.61±0.1</td>
<td>90.23±2.8</td>
</tr>
<tr>
<td><strong>SemiLRCCA</strong></td>
<td>90.48±0.5</td>
<td>85.58±1.4</td>
</tr>
<tr>
<td><strong>S²GCA</strong></td>
<td>94.61±0.9</td>
<td>92.67±3.6</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CCA</strong></td>
<td>81.70±1.2</td>
<td>83.47±4.0</td>
</tr>
<tr>
<td><strong>DCCA</strong></td>
<td>80.57±2.1</td>
<td>83.24±6.4</td>
</tr>
<tr>
<td><strong>SCCA</strong></td>
<td>77.35±4.7</td>
<td>82.27±0.6</td>
</tr>
<tr>
<td><strong>SemiCCA</strong></td>
<td>95.31±0.1</td>
<td>92.69±1.6</td>
</tr>
<tr>
<td><strong>SemiLRCCA</strong></td>
<td>91.64±0.5</td>
<td>87.58±1.2</td>
</tr>
<tr>
<td><strong>S²GCA</strong></td>
<td>95.56±0.2</td>
<td>93.52±1.1</td>
</tr>
</tbody>
</table>
# Experiments in Ad Data

## Accuracy on Ads dataset (X-view)

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>CCA</th>
<th>DCCCA</th>
<th>SCCA</th>
<th>SemiCCA</th>
<th>SemiLRCCA</th>
<th>S2GCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alt</td>
<td>cap</td>
<td>32.85 ± 0.3</td>
<td>27.49 ± 0.2</td>
<td>32.98 ± 0.3</td>
<td>33.33 ± 0.3</td>
<td>34.34 ± 0.2</td>
<td>37.72 ± 3.4</td>
</tr>
<tr>
<td>2</td>
<td>alt</td>
<td>ancurl</td>
<td>29.21 ± 0.9</td>
<td>27.49 ± 0.2</td>
<td>25.64 ± 0.2</td>
<td>31.90 ± 0.3</td>
<td>34.28 ± 0.3</td>
<td>36.28 ± 1.7</td>
</tr>
<tr>
<td>3</td>
<td>alt</td>
<td>orig</td>
<td>28.13 ± 1.1</td>
<td>27.49 ± 0.2</td>
<td>25.64 ± 0.2</td>
<td>32.23 ± 0.3</td>
<td>33.48 ± 0.4</td>
<td>37.63 ± 0.2</td>
</tr>
<tr>
<td>4</td>
<td>alt</td>
<td>url</td>
<td>28.30 ± 1.0</td>
<td>27.49 ± 0.2</td>
<td>25.64 ± 0.2</td>
<td>32.64 ± 0.2</td>
<td>33.67 ± 0.6</td>
<td>37.71 ± 3.3</td>
</tr>
<tr>
<td>5</td>
<td>cap</td>
<td>ancurl</td>
<td>15.57 ± 0.01</td>
<td>15.59 ± 0.01</td>
<td>15.34 ± 0.1</td>
<td>16.28 ± 0.01</td>
<td>16.47 ± 0.01</td>
<td>16.74 ± 0.01</td>
</tr>
<tr>
<td>6</td>
<td>cap</td>
<td>orig</td>
<td>15.65 ± 0.01</td>
<td>15.59 ± 0.01</td>
<td>15.34 ± 0.1</td>
<td>16.29 ± 0.01</td>
<td>16.48 ± 0.01</td>
<td>16.65 ± 0.01</td>
</tr>
<tr>
<td>7</td>
<td>cap</td>
<td>url</td>
<td>15.63 ± 0.01</td>
<td>15.59 ± 0.01</td>
<td>15.34 ± 0.1</td>
<td>16.41 ± 0.01</td>
<td>16.48 ± 0.01</td>
<td>16.69 ± 0.01</td>
</tr>
<tr>
<td>8</td>
<td>ancurl</td>
<td>orig</td>
<td>55.54 ± 12.1</td>
<td>56.01 ± 17.4</td>
<td>44.34 ± 0.5</td>
<td>61.89 ± 13.0</td>
<td>65.92 ± 7.8</td>
<td>72.43 ± 9.5</td>
</tr>
<tr>
<td>9</td>
<td>ancurl</td>
<td>url</td>
<td>55.53 ± 15.8</td>
<td>56.01 ± 17.4</td>
<td>43.88 ± 0.4</td>
<td>62.12 ± 12.7</td>
<td>66.72 ± 7.6</td>
<td>72.39 ± 10.0</td>
</tr>
<tr>
<td>10</td>
<td>orig</td>
<td>url</td>
<td>62.06 ± 5.5</td>
<td>57.67 ± 0.9</td>
<td>49.95 ± 0.7</td>
<td>62.78 ± 1.0</td>
<td>69.10 ± 0.3</td>
<td>73.58 ± 0.3</td>
</tr>
</tbody>
</table>
## Summary of the related works

<table>
<thead>
<tr>
<th>Methods</th>
<th>Paired information</th>
<th>Class information</th>
<th>Structure information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paired</td>
<td>Semi--paired</td>
<td>Unlabeled</td>
</tr>
<tr>
<td>CCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>DCCA</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCCAM</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>LPCCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>LCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>SemiCCA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RKCCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>PPLCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>NCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>S³GCA</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Where the methods in blur are ours
Observations from the above works

Summary: From CCA $\rightarrow$ GCA  
$\rightarrow$ General DR framework!

Alternative ways *slightly* deviated from CCA:
From *Instance-level* pairing $\rightarrow$ *Semantic-level* pairing

Semantic-level pairing

- Label [SCY08] [SCY09]
- Must-link and Cannot-link
- Manifold Alignment [WS2011]
Ongoing work: Existing

(1) Classification: [HM2011](ICML2011), [CNHK2011] (CVPR2011)

(2) Clustering: [KD2011](ICML2011)

Common Characteristics:

1) Design a classifier or clusterer for each view;
2) Perform joint learning!
Ongoing work: Ours

- GCA as a preprocessing

- Benefits:
  1. Enlarged samples in the common low-dimensional space
  2. Desirably better generalization
  3. Single classifier or clusterer for all the views
Future work (1)

• MV Multi-instance learning

[WNH2011]
Future work (2)

- MV Multi-task learning [HL2011](ICML2011)

Realize multi-task multi-view learning for complicated learning problems with both feature heterogeneity and task heterogeneity.


Reference (3)

Thanks!

Q&A