Abstract
We have $D$-view data vectors $\{x_i^d\}_{i=1}^n \subset \mathbb{R}^p$ and specify a non-negative real value $w_{ij}^{(d)}$ called matching weight, which represents the strength of association between $x_i^d$ and $x_j^e$ for $i = 1, 2, \ldots, n_d; j = 1, 2, \ldots, n_e; d = 1, 2, \ldots, D$. Cross-Domain Matching Correlation Analysis (Shimodaira, 2016, Neural Networks; CDMCA) finds linear transformation matrices $A^d \in \mathbb{R}^{p \times K^d}$ ($d = 1, 2, \ldots, D$) so that

$$w_{ij}^{(d)} > 0 \Rightarrow (A^d)^\top x_i^d \approx (A^e)^\top x_j^e,$$

by extending Canonical Correlation Analysis (CCA). The dimensionality $K \leq \min\{n_d\}$ of $y_i^d := (A^d)^\top x_i^d$ is manually chosen so that the transformed vectors $\{y_i^d\}$ geometrically reflect the association structure well.

To improve CDMCA, we propose CDMCA based on Maximum Likelihood.

Let $\{x_i^d\}_{i=1}^n \subset \mathbb{R}^p$ be image feature vectors, $\{x_i^e\}_{i=1}^n \subset \mathbb{R}^p$ be tag vectors, and $w_{ij}^{(d)} = 1$ if the image $x_i$ is tagged with the word $x_j$ and $w_{ij}^{(d)} = 0$ if not. CDMCA finds linear transformation $\mathbb{R}^p \ni x \mapsto (A^d)^\top x \in \mathbb{R}^K$ so that associated vectors get closer in $\mathbb{R}^K$.

Although some studies have already shown the advantages of using CDMCA by application experiments, its theoretical aspect is still under-developed. In this presentation, we give a theoretical guarantee of CDMCA. Especially, we show following three results in this presentation.

(A) Original CDMCA requires strong assumptions to be consistent.
(B) We propose CDMCA based on Maximum Likelihood.
(C) Our proposed algorithm is consistent with weaker assumptions.

![Image]

**CDMCA**

Let $w_{ij}^{(d)} := w_{ij}^{(d)} / \sum_{i=1}^n \sum_{j=1}^n w_{ij}^{(d)}$. CDMCA minimizes

$$\sum_{d=1}^D \sum_{i=1}^{n_d} \sum_{j=1}^{n_e} \sum_{l=1}^{n_d} \sum_{m=1}^{n_e} w_{ij}^{(d)} \| (A^d)^\top x_i^d - (A^e)^\top x_j^e \|_2^2,$$

with respect to $A = ((A^1)^\top, \ldots, (A^D)^\top)$ satisfying a quadratic constraint

$I_K = \sum_{d=1}^D \sum_{i=1}^{n_d} \sum_{j=1}^{n_e} w_{ij}^{(d)} (A^d)^\top x_i^d (A^d)^\top x_j^e (A^e)^\top x_j^e = 0$,

then we obtain $A$. The solution can efficiently be computed by eigenvalue decomposition.

**Proposal of a novel probabilistic model**

Assuming that $K_\delta \leq \min\{n_d\}, \alpha \in (0, 1], A^d \in \mathbb{R}^{p \times K_\delta}$ are given. We propose a novel probabilistic model for $\{x_i^d\}_{i=1}^n \subset \mathbb{R}^p, \{u_{ij}^{(d)}\}_{i=1}^n$ as

$$K_\delta \leq \min\{n_d\}, \alpha \in (0, 1], A^d \in \mathbb{R}^{p \times K_\delta}$$

$$\begin{align*}
\pi_{ij}^{(d)} &= \text{Po}(\alpha, \exp(-\| (A^d)^\top x_i^d - (A^e)^\top x_j^e \|_2^2)), \\
\pi_{ij}^{(d)} &= \text{Po}(\alpha, \exp(-\| (A^d)^\top x_i^d - (A^e)^\top x_j^e \|_2^2)).
\end{align*}$$

![Image]

**Numerical Experiment**

We define the regularized error as $\varepsilon(A) = \frac{\| A A^\top - A A_e^\top B \|_F}{\| A \|_F}$. Sample average of the regularized error over 100 times experiments with $K_\delta = 3, K = 2, 3, 4$ are plotted, along with increasing sample size. (Black:CDMCA, Blue:Proposal, x-axis:n, y-axis:$\varepsilon(A_e)$.)

\[ \text{Note 1: } D = \{1, 2, 3, n_d, n_e\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}, \{x_p, y_p\}. \]

\[ \text{Note 2: } \text{Proposal method estimates }\alpha \text{ and } A \text{ simultaneously, while original CDMCA estimates only } A. \]