Leveraging local data structure for multi-view correlation analysis with graph-structured associations

Akifumi Okuno[†,††] and Hidetoshi Shimodaira[†,††]

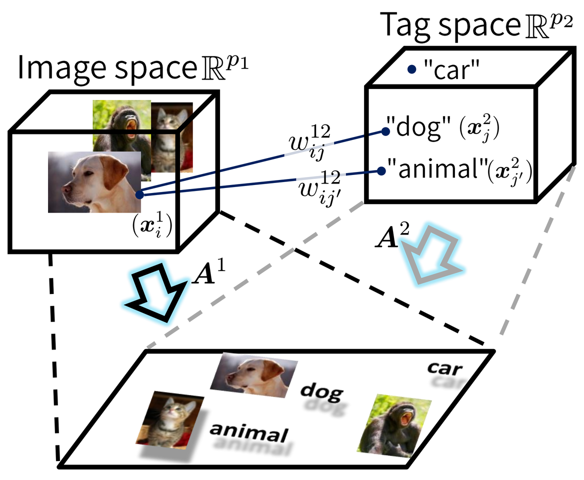
[†]Graduate School of Informatics, Kyoto University [††]RIKEN Center for AI Project

▷ Abstract

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

$$w_{ij}^{de} > 0 \Rightarrow (\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d \approx (\boldsymbol{A}^e)^\top \boldsymbol{x}_j^e,$$

by extending Canonical Correlation Analysis (CCA). The dimensionality $K \leq \min_d\{p_d\}$ of $\boldsymbol{y}_i^d := (\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d$ is chosen so that the transformed vectors $\{\boldsymbol{y}_i^d\}$ geometrically reflect the association structure well.



Ex. 1 Let $\{\boldsymbol{x}_i^1\}_{i=1}^{n_1} \subset \mathbb{R}^{p_1}$ be image feature vectors, $\{\boldsymbol{x}_j^2\}_{j=1}^{n_2} \subset \mathbb{R}^{p_2}$ be tag vectors, and $w_{ij}^{12} = 1$ if the image $\boldsymbol{x}_i^1$ is tagged with the word $\boldsymbol{x}_j^2$ and $w_{ij}^{12} = 0$ if not. Linear transformation $\mathbb{R}^{p_d} \ni \boldsymbol{x}_i^d \mapsto (\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d \in \mathbb{R}^K$ makes associated vectors closer in the common subspace $\mathbb{R}^K$.

These graph-structured associations $\{w_{ij}^{de}\}$ are generally specified by external knowledge, i.e. annotation by human effort in Ex. 1. In practice, each of these obtained associations is given as binary value which only tells us whether two vectors have an association, but not the weight of the association. Since CDMCA can consider the non-binary weights of associations, binary associations do not fully utilize the method. Thus, in this study, we aim at giving weights to the associations before applying CDMCA by leveraging local data structure.

▷ CDMCA

Let $\tilde{w}_{ij}^{de} := w_{ij}^{de} / \sum_{d=1}^{d} \sum_{e=1}^{D} \sum_{i=1}^{n_d} \sum_{j=1}^{n_e} w_{ij}^{de}$. CDMCA maximizes

$$\text{(CDMCA)} \qquad \sum_{d=1}^{D} \sum_{e=1}^{D} \sum_{i=1}^{n_d} \sum_{j=1}^{n_e} \tilde{w}_{ij}^{de} \langle (\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d, (\boldsymbol{A}^e)^\top \boldsymbol{x}_j^e \rangle,$$

with respect to $\boldsymbol{A} = ((\boldsymbol{A}^1)^\top, \cdots, (\boldsymbol{A}^D)^\top)^\top$ satisfying a quadratic constraint $\boldsymbol{I}_K = \sum_{d=1}^{D} \sum_{i=1}^{n_d} (\sum_{e=1}^{D} \sum_{j=1}^{n_e} \tilde{w}_{ij}^{de}) [(\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d][(\boldsymbol{A}^d)^\top \boldsymbol{x}_i^d]^\top$, then we obtain $\hat{\boldsymbol{A}}$. The solution can efficiently be computed by eigenvalue decomposition.

▷ Related methods

Let $D = 2$, CDMCA with $w_{ij}^{11} = w_{kl}^{22} = 0, v_{ij} := w_{ij}^{12}$ maximizes

$$\text{(Eq. 1)} \qquad \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} v_{ij} \langle (\boldsymbol{A}^1)^\top \boldsymbol{x}_i^1, (\boldsymbol{A}^2)^\top \boldsymbol{x}_j^2 \rangle.$$

If $n_1 = n_2 =: n$ and $w_{ij} = \delta_{ij}$, this method reduces to Canonical Correlation Analysis (Hotelling, 1936; CCA) which is widely used in the field of Statistics, Image Processing, etc. Locality Preserving CCA (Sun et al., 2007; LPCCA), A LPCCA (Wang et al., 2013, ALPCCA) and Multi-view Diffusion Maps (Dhillon, 2015; DvDM) incorporate Locality

$$s_{ij}^d := \exp(-\|\boldsymbol{x}_i^d - \boldsymbol{x}_j^d\|_2^2 / t_d)$$

into CCA. They maximize the loss function (Eq.1) with weights

$$\begin{aligned}
&\text{(2-view CDMCA)} && v_{ij} = w_{ij}^{12},\\
&\text{(CCA)} && v_{ij} = \delta_{ij},\\
&\text{(LPCCA)} && v_{ij} = s_{ij}^1 s_{ij}^2,\\
&\text{(ALPCCA)} && v_{ij} = 1 + s_{ij}^1 + s_{ij}^2,\\
&\text{(MvDM)} && v_{ij} = \sum_{k=1}^{n} s_{ik}^1 s_{kj}^2,
\end{aligned}$$

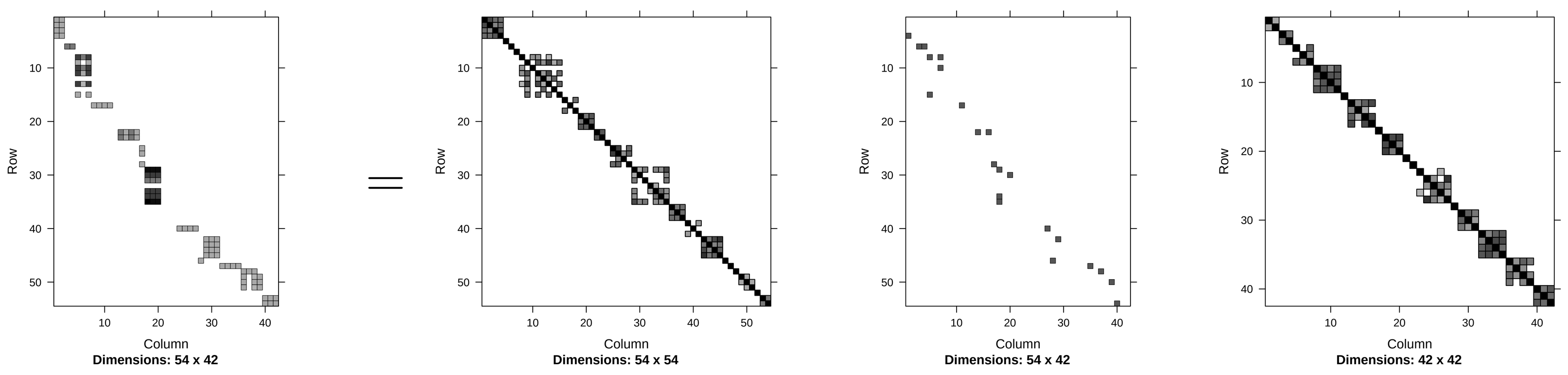
where $t_1, t_2 > 0$ are tuning parameters, namely, LPCCA, ALPCCA and MvDM attach the importance to association between $\boldsymbol{x}_i^1$ and $\boldsymbol{x}_j^2$, by leveraging geometric structure of data vectors in each domain.
If $D = 1$ CDMCA reduces to General Framework for Graph Embedding (Yan et al., 2007; GFGE), which includes Locality Preserving Projections (He et al., 2004; LPP) that uses s_{ij}^1 instead of w_{ij}^{11}.

▷ Proposed method

Under the setting of CDMCA, graph-structured association $\{w_{ij}\}$ is known, unlike CCA whose association is one-to-one: $w_{ij} = \delta_{ij}$. As well as LPCCA, ALPCCA and MvDM, we propose a weight which incorporates Locality into given strength of association as

$$\text{(Proposal)} \qquad v_{ij} = \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} s_{ik}^1 w_{kl} s_{lj}^2.$$

This equation can be written as matrix form $\boldsymbol{V} = \boldsymbol{S}^1 \boldsymbol{W}^{12} \boldsymbol{S}^2$, where $\boldsymbol{V} = (v_{ij}), \boldsymbol{S}^1 = (s_{ij}^1), \boldsymbol{W} = (w_{ij}), \boldsymbol{S}^2 = (s_{ij}^2)$, and it reduces to the weight of MvDM if $n_1 = n_2 =: n$ and $w_{ij} = \delta_{ij}$.



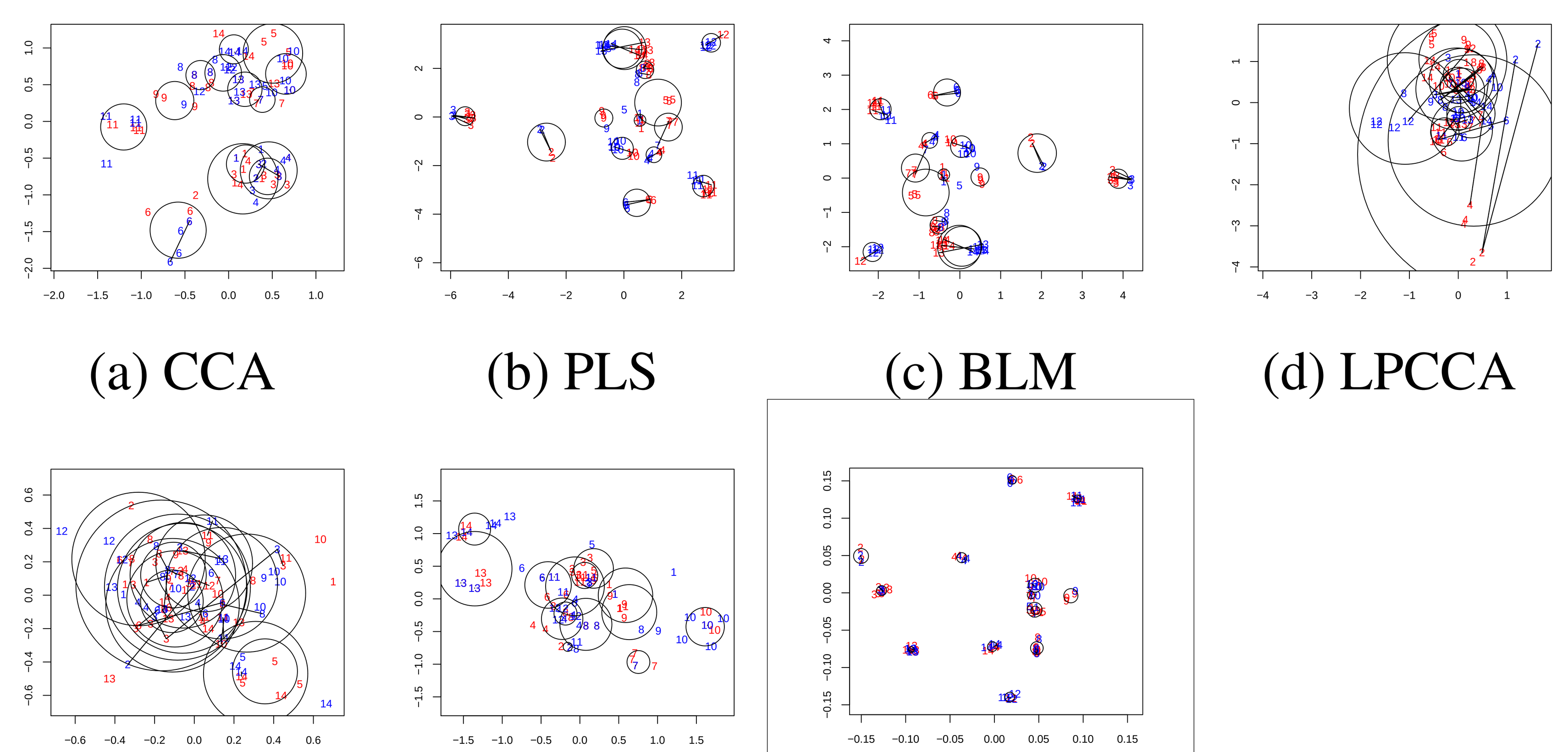
$\boldsymbol{\nu}$ $\quad$ $\boldsymbol{S}^1$ $\quad$ $\boldsymbol{W}$ $\quad$ $\boldsymbol{S}^2$

Thanks to Locality, $\boldsymbol{\nu} = (\nu_{ij})$ contains more information than given associations $\boldsymbol{W}$.

		1-to-1	Many-to-many
	1-view	(None)	PCA, GFGE
	2-view	CCA	CvGE
	>2-view	MCCA	CDMCA, HIMFAC
With Locality	1-view	(None)	LPP
	2-view	LPCCA, ALPCCA	
	>2-view		**(Proposal)**

▷ Comparison with other methods

2-dimensional unified representations obtained through each of CCA, Partial Least Squares (Wold, 1966; PLS), Bilinear Models (Sharma et al., 2011; BLM), LPCCA, ALPCCA, CDMCA ($D = 2$) and proposed method are plot in the following.



(a) CCA (b) PLS (c) BLM (d) LPCCA (e) ALPCCA (f) CDMCA ($D = 2$) (g) Proposal

The accuracy of classification conducted with our proposed method can be higher than remaining methods.

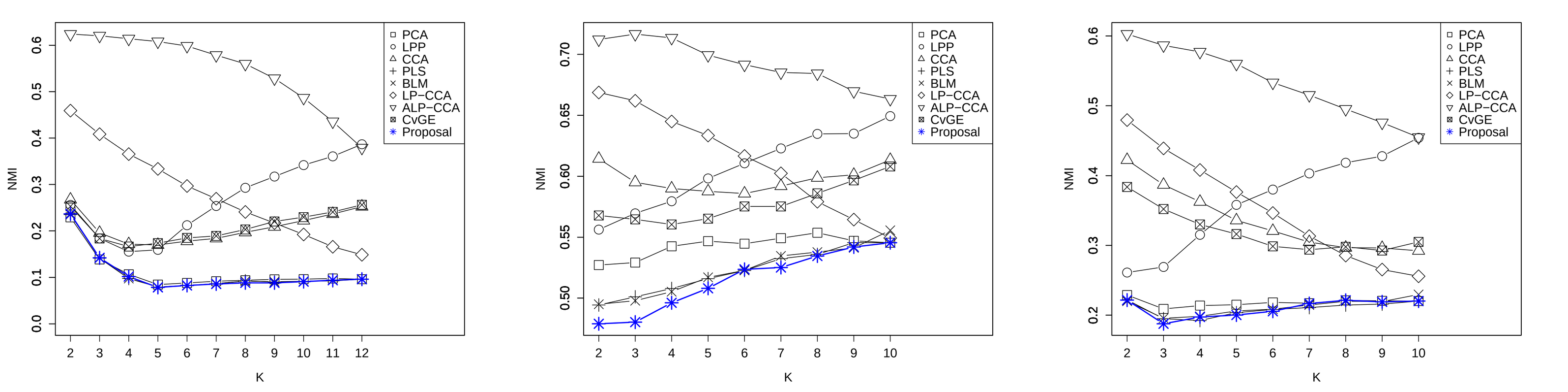
▷ Numerical Experiment

- **Synthetic Dataset** for $k, l \in [n_0], i \in [n_d], j \in [n_e], d, e \in [D]$,

$$\boldsymbol{x}_{(k-1)n_d+i}^d \mid \boldsymbol{z}_k \sim N_{p_d}[\boldsymbol{B}^d \boldsymbol{z}_k, \sigma_d^2 \boldsymbol{I}_{p_d}],$$
$$w_{kn_d+i, ln_e+j}^{de} = \begin{cases} \delta_{kl} & \text{w.p. } \alpha \\ 0 & \text{otherwise.} \end{cases},$$

where $\boldsymbol{z}_k \sim N_{p_0}[\boldsymbol{0}, (2^{-|i-j|})_{ij}]$ and $[n] := \{1, 2, \ldots, n\}$.

- **Evaluation** Normalized Mutual Information (NMI) for transformed vectors in view-1 with Hierarchical Clustering. Smaller value is better.
- **Result** We take averages over 500 times experiments.



Setting 1: $n_0 = 5, p_0 = 5, p_1 = p_2 = 20, n_1 = n_2 = 7, \sigma_1 = \sigma_2 = 1, \alpha = 0.1$ Setting 2: $n_0 = 5, p_0 = 5, p_1 = p_2 = 10, n_1 = n_2 = 10, \sigma_1 = \sigma_2 = 1.5, \alpha = 0.2$ Setting 3: $n_0 = 5, p_0 = 5, p_1 = p_2 = 10, n_1 = n_2 = 5, \sigma_1 = \sigma_2 = 0.8, \alpha = 0.2$