Learning Coupled Feature Spaces for Cross-modal Matching

Kaiye Wang, Ran He, Wei Wang, Liang Wang, Tieniu Tan

{kaiye.wang, rhe, wangwei, wangliang, tnt}@nlpr.ia.ac.cn

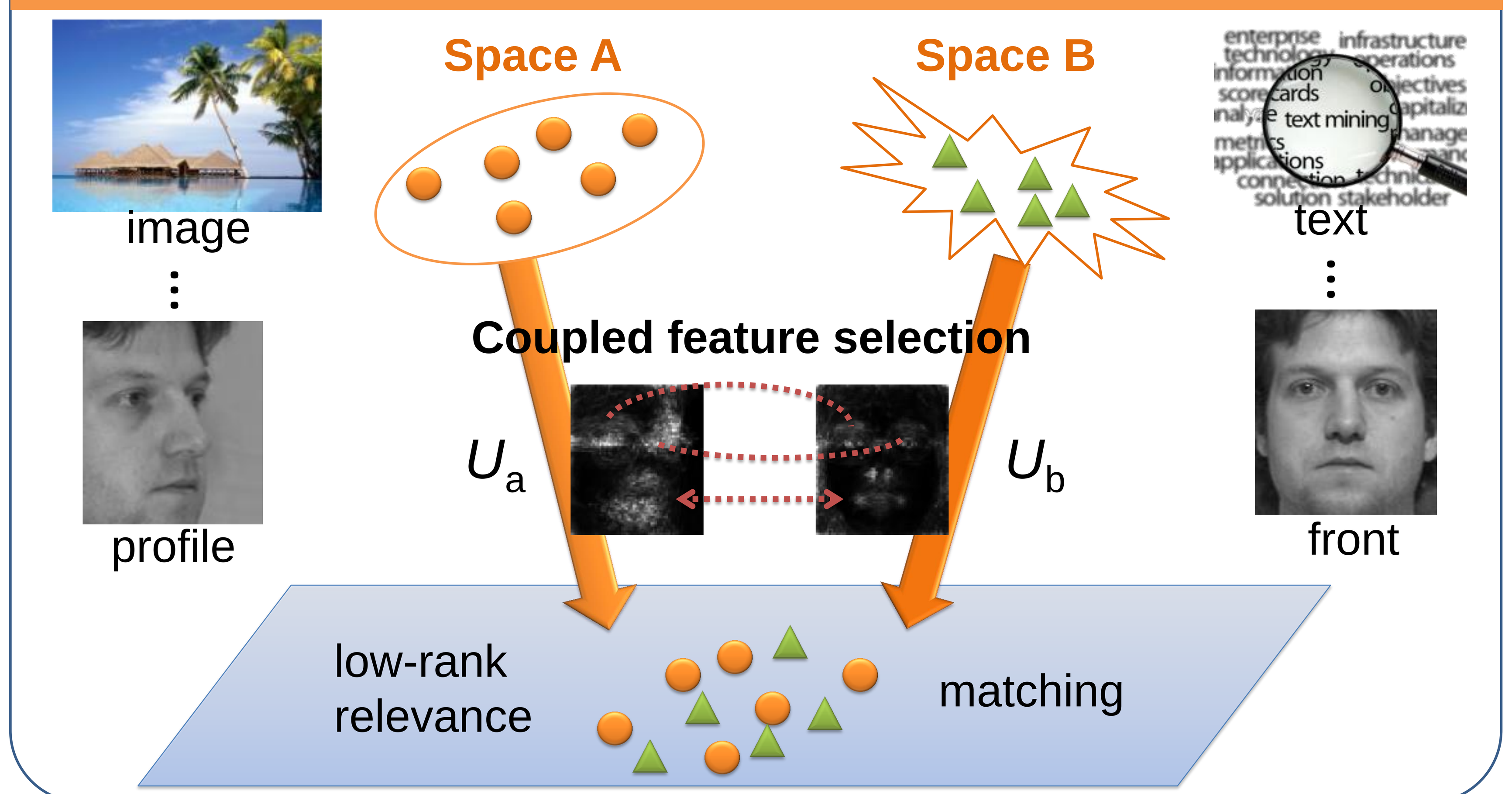
Abstract

Goal: match data from different modalities.

Challenge: bridge the heterogeneity gap.

Contribution: we propose a general regularization framework for cross-modal matching problem, which jointly performs common subspace learning and coupled feature selection.

Overview



Our method

LCFS = common subspace learning
+ coupled feature selection

Objective function

$$\min_{\mathbf{U}_a,\mathbf{U}_b} \frac{1}{2}\left(\left\|\mathbf{X}_a^T\mathbf{U}_a - \mathbf{Y}\right\|_F^2 + \left\|\mathbf{X}_b^T\mathbf{U}_b - \mathbf{Y}\right\|_F^2\right)$$

Coupled linear regression

$$+\lambda_1\left(\left\|\mathbf{U}_a\right\|_{21} + \left\|\mathbf{U}_b\right\|_{21}\right) + \lambda_2\left(\left\|[\mathbf{X}_a^T\mathbf{U}_a \ \ \mathbf{X}_b^T\mathbf{U}_b]\right\|_*\right)$$

The L21 norm

The trace norm

Our method continued..

Coupled linear regression: learn two projection matrices for mapping two different modal data into a common space.

The L21 norm: select the relevant and discriminative features on two feature spaces simultaneously.

The trace norm: enhance the relevance of different modal data with similar relationship.

Reformulation for the trace norm:

$$\frac{\lambda_2}{2}\left(tr(\mathbf{U}_a^T\mathbf{X}_a\mathbf{S}^{-1}\mathbf{X}_a^T\mathbf{U}_a) + tr(\mathbf{U}_b^T\mathbf{X}_b\mathbf{S}^{-1}\mathbf{X}_b^T\mathbf{U}_b) + tr(\mathbf{S})\right)$$

$$\mathbf{S} = (\mathbf{X}_a^T\mathbf{U}_a\mathbf{U}_a^T\mathbf{X}_a + \mathbf{X}_b^T\mathbf{U}_b\mathbf{U}_b^T\mathbf{X}_b + \mu_i\mathbf{I})^{\frac{1}{2}}$$

Algorithm 1: Iterative Algorithm for Learning Coupled Feature Spaces (LCFS)

Input: $\mathbf{X}_a \in \mathbb{R}^{d1\times n}$, $\mathbf{X}_b \in \mathbb{R}^{d2\times n}$ and $\mathbf{Y} \in \mathbb{R}^{n\times c}$

Output: $\mathbf{U}_a \in \mathbb{R}^{d1\times c}$ and $\mathbf{U}_b \in \mathbb{R}^{d2\times c}$

Set $t = 0$. Initialize $\mathbf{U}_a$ and $\mathbf{U}_b$ as zero matrix.

repeat

1. Compute $\mathbf{V}Diag(s_k)\mathbf{V}^T$ as the eigenvalue decomposition of $(\mathbf{X}_a^T\mathbf{U}_a\mathbf{U}_a^T\mathbf{X}_a + \mathbf{X}_b^T\mathbf{U}_b\mathbf{U}_b^T\mathbf{X}_b)$.
2. Set $\mathbf{S}^{-1} = \mathbf{V}Diag(1/\sqrt{s_k+\mu})\mathbf{V}^T$.
3. Compute p_i^t and q_i^t according to
4. Compute $\mathbf{U}_a^t$ and $\mathbf{U}_b^t$ by solving the two linear system problems in
5. $t = t + 1$

until Converges

$$\begin{cases} p_i = \dfrac{1}{2\sqrt{\left\|\mathbf{u}_a^i\right\|_2^2 + \varepsilon}} \\ q_i = \dfrac{1}{2\sqrt{\left\|\mathbf{u}_b^i\right\|_2^2 + \varepsilon}} \end{cases}$$

$$\begin{cases} (\mathbf{X}_a\mathbf{X}_a^T + \lambda_1\mathbf{P} + \lambda_2\mathbf{X}_a\mathbf{S}^{-1}\mathbf{X}_a^T)\mathbf{U}_a = \mathbf{X}_a\mathbf{Y} \\ (\mathbf{X}_b\mathbf{X}_b^T + \lambda_1\mathbf{Q} + \lambda_2\mathbf{X}_b\mathbf{S}^{-1}\mathbf{X}_b^T)\mathbf{U}_b = \mathbf{X}_b\mathbf{Y} \end{cases}$$

Experimental results

Evaluation: MAP, PS curve

Compared Methods:
CCA, PLS, BLM (CVPR'11): similar pairs
GMLDA, GMMFA (CVPR'12): similar pairs + label

Results on Pascal image-tag data

20 classes, 2808 / 2841 training/testing samples
Image: 512-dim Gist, Text: 399-dim word frequency

Experimental results continued..

Methods	Image query	Text query	Average
PCA+PLS	0.2757	0.1997	0.2377
PCA+BLM	0.2667	0.2408	0.2538
PCA+CCA	0.2655	0.2215	0.2435
PCA+GMMFA	0.3090	0.2308	0.2699
PCA+GMLDA	0.2418	0.2038	0.2228
LCFS	**0.3438**	**0.2674**	**0.3056**

Table 1. Comparison of MAP for different methods

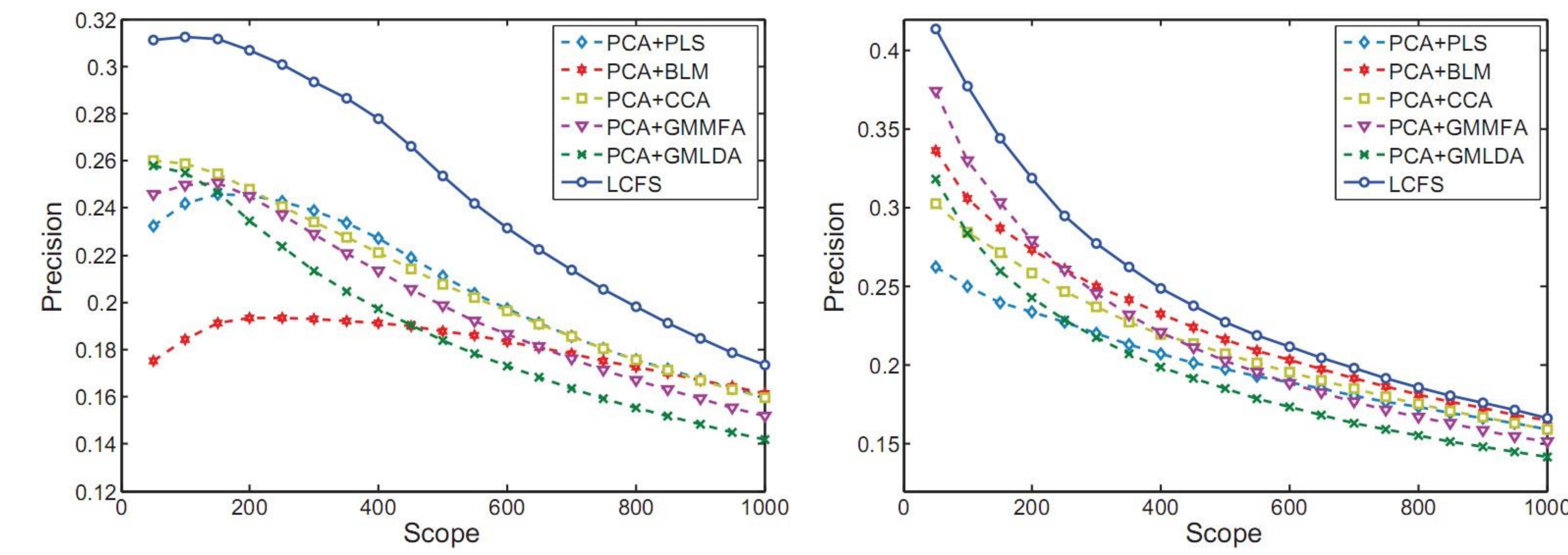


Figure 1. Precision-scope curves of different methods. **Left:** Image as query, **Right:** Text as query

Results on Wikipedia image-text data

10 classes, 1300 / 1566 training/testing samples
Image: 128-dim bags of SIFT, Text: 10-dim LDA

Methods	Image query	Text query	Average
PLS	0.2402	0.1633	0.2032
BLM	0.2562	0.2023	0.2293
CCA	0.2549	0.1846	0.2198
GMMFA	0.2750	0.2139	0.2445
GMLDA	0.2751	0.2098	0.2425
LCFS	**0.2798**	**0.2141**	**0.2470**

Table 2. Comparison of MAP for different methods

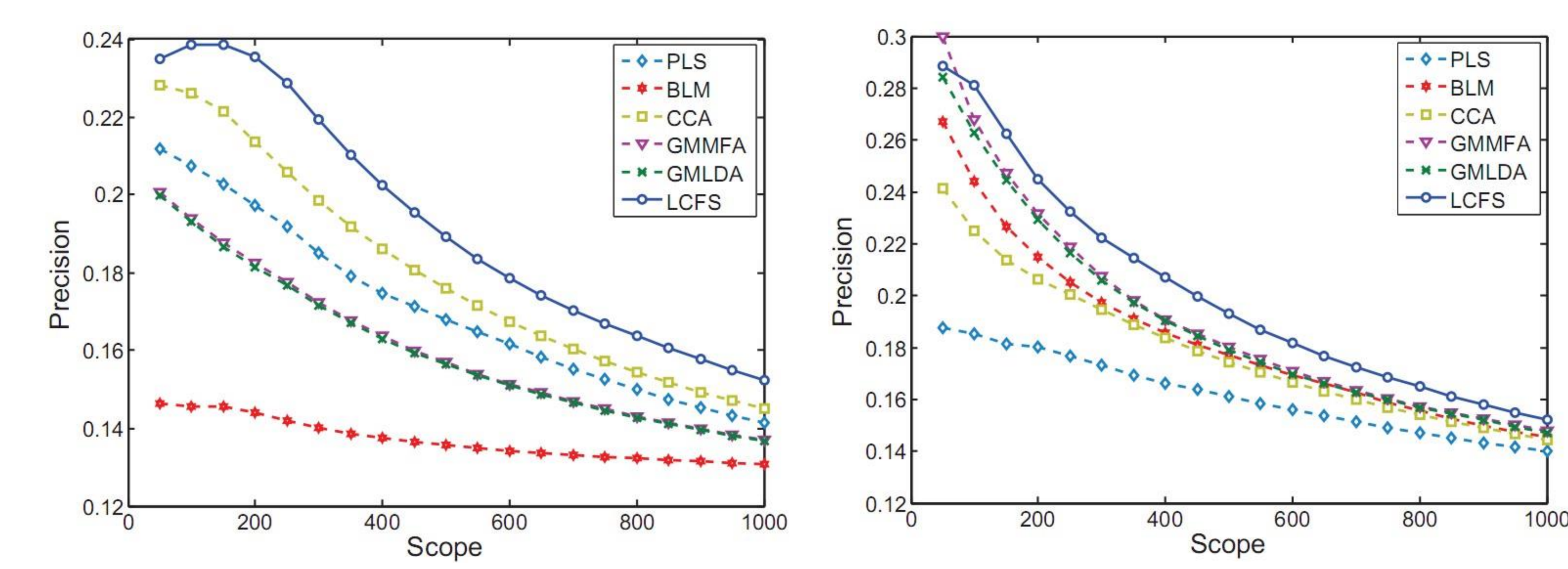


Figure 2. Precision-scope curves of different methods. **Left:** Image as query, **Right:** Text as query