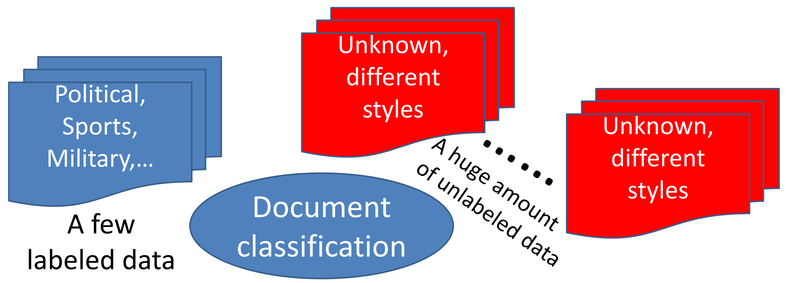
Northeastern University

Low-Rank Transfer Subspace Learning

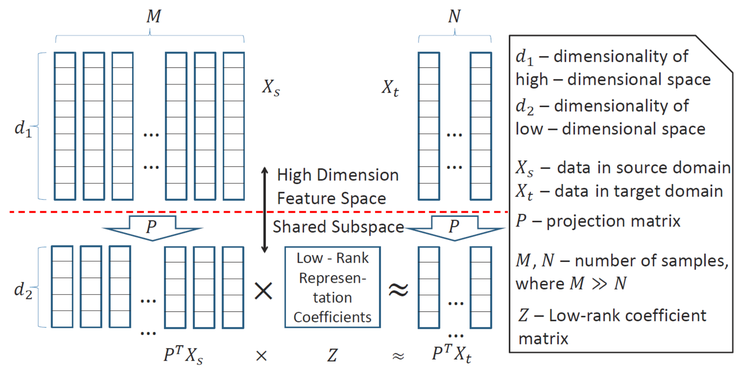
Ming Shao, Carlos Castillo, Zhenghong Gu, Yun Fu
Electrical and Computer Engineering, Northeastern University

Motivation



- One common assumption in classification problems is the training/testing consistency of the data.
- This cannot be always satisfied, especially in complex applications common in many areas, e.g., web document classification, sentiment analysis, image annotation, face recognition.
- How to apply previous well-labeled data to a huge amount of unseen data with possibly different distributions?

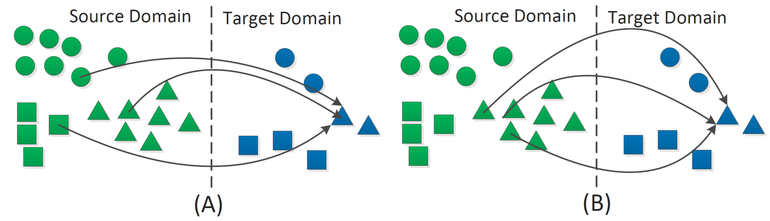
Key Idea & Contributions



Contributions

We present a novel method for transfer learning via low-rank representation, which we call it low-rank transfer subspace learning (LTSL).

Problem Formulation



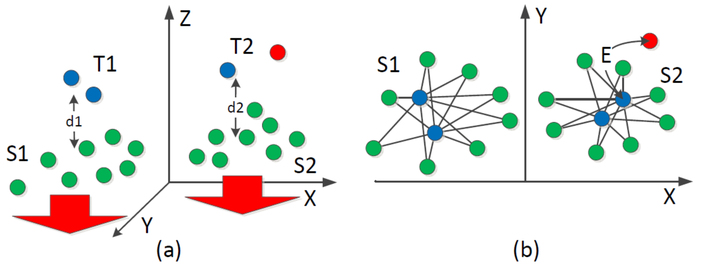
- A given data set is seldom well described by a single subspace, rather, data are more likely lying in several subspaces.
- Suppose we adopt source data to linearly represent target data to achieve the purpose of knowledge transfer.
- For over-complete source data that span the entire feature space, however, we could always obtain trivial solutions.
- The correct way might be using only a few data in the source domain within an appropriate subspace to reconstruct a specific target data, as shown in the above figure.

Problem Formulation and Solution

- ***Basic Problem****: Given test data Xt from a union of multiple subspaces in the target domain and training data Xs from a union of multiple subspaces in the source domain, the goal is to find a discriminative subspace P where test data can be linearly represented by the data from some subspace Si in the source domain with rank constraint, i.e.,*

$$P, Z = \arg\min_{P,Z} F(P, X_s) + \text{rank}(Z),$$
$$\text{s.t., } P^T X_t = P^T X_s Z,$$

- *where F(P,Xs) is a general subspace learning function and rank(Z) is the constraint imposed on the knowledge transfer from Xs to Xt in the projected subspace P.*



- If there are target data that can not be well-explained by the source data in the learned subspace *P*, we add an error term *E:*

$$P, Z, E = \arg\min_{P,Z,E} F(P, X_s) + \text{rank}(Z) + \lambda \|E\|_{2,1},$$
$$\text{s.t., } P^T X_t = P^T X_s Z + E.$$

Problem Solution and Experiments

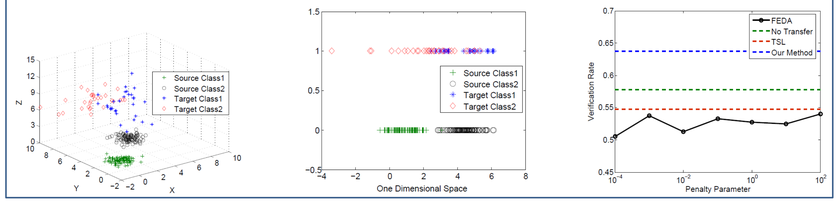
- The former problem can be solved by augmented Lagrangian multipliers (ALU). We therefore reformulate the previous problem into:

$$\begin{aligned} \mathcal{L} &= F(P, X_s) + \|J\|_* + \lambda \|E\|_{2,1} \qquad (6) \\ &+ < Y_1, P^T X_t - P^T X_s Z - E > + < Y_2, Z - J > \\ &+ \frac{\mu}{2}(\|P^T X_t - P^T X_s Z - E\|_F^2 + \|Z - J\|_F^2), \end{aligned}$$

- Many well-established subspace learning methods can be incorporated into our framework, and the difference among these methods lie in the term *F(P, Xs)*. We show their corresponding formulations in next table:

Method	Objective $F(P, X_s)$	Gradient $\nabla F(P, X_s)$
PCA	$\arg\min_P \mathbf{Tr}\,(-P^T \Sigma P + (P^T P - I)\Lambda)$	$2(-\Sigma P + P\Lambda)$
LDA	$\arg\min_P \mathbf{Tr}\,(P^T S_w P + (P^T S_b P - I)\Lambda)$	$2(S_w P + S_b P\Lambda)$
LPP[2]	$\arg\min_P \mathbf{Tr}\,(P^T X_s \mathcal{L} X_s^T P + (P^T X_s D X_s^T P - I)\Lambda)$	$2(X_s \mathcal{L} X_s^T P + X_s D X_s^T P\Lambda)$
NPE	$\arg\min_P \mathbf{Tr}(P^T X_s (I-W)^T (I-W) X_s^T P + (P^T X_s X_s^T P - I)\Lambda)$	$2(X_s (I-W)^T (I-W) X_s^T P + X_s X_s^T P\Lambda)$
MFA[3]	$\arg\min_P \mathbf{Tr}(P^T X_s (D-W) X_s^T P + (P^T X_s (D_P - W_P) X_s^T P - I)\Lambda)$	$2(X_s (D-W) X_s^T P - X_s (D_P - W_P) X_s^T P\Lambda)$
DLA[4]	$\arg\min_P \mathbf{Tr}(P^T X_s L X_s^T P + (P^T P - I)\Lambda)$	$(X_s L X_s^T + X_s L^T X_s^T)P + 2P\Lambda$

- We show three groups of experiments to validate the proposed method, i.e., synthetic data, Yale B to CMU PIE (Y2P) face recognition, and kinship verification.
- **Group 1: Synthetic data, 2 classes in both source and target domains**



Experiments and Conclusions

- **Group 2: Kinship verification, UB KinFace database**

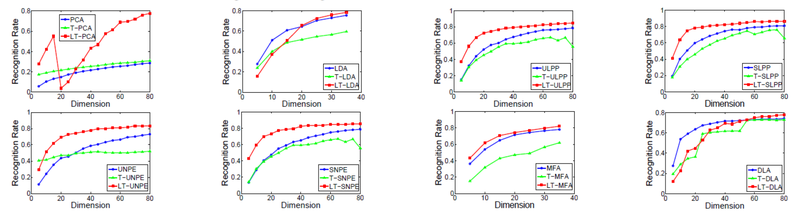
BEST RESULTS AND DIMENSIONS OF KINSHIP VERIFICATION.

Method	PCA	SLPP	ULPP	SNPE	UNPE	MFA	DLA
No Transfer	53.98%(11)	55.00%(9)	57.74%(11)	53.26%(9)	54.26%(21)	52.74%(17)	54.74%(35)
TSL	54.78%(25)	54.02%(3)	54.02%(11)	50.74%(9)	53.26%(9)	52.24%(3)	53.98%(39)
Our Method	**56.57%**(19)	**57.17%**(17)	**63.72%**(11)	**54.60%**(11)	**58.80%**(3)	**54.50%**(35)	**55.00%**(33)

- **Group 3: Face recognition, from Yale B to CMU PIE**

BEST RESULTS AND DIMENSIONS OF PROBLEM Y2P.

Method	PCA	SLPP	ULPP	SNPE	UNPE	LDA	MFA	DLA
No Transfer	28.6%(80)	80.7%(80)	78.3%(80)	78.6%(80)	73.1%(80)	75.4%(35)	78.1%(35)	74.1%(80)
TSL	30.9%(80)	75.7%(75)	67.0%(65)	67.0%(65)	52.1%(80)	59.6%(35)	62.0%(35)	72.8%(80)
Our Method	**77.6%**(80)	**86.1%**(75)	**84.6%**(80)	**85.2%**(75)	**83.5%**(80)	**78.4%**(35)	**82.2%**(35)	**77.8%**(80)



Conclusions

In this paper, we proposed a novel framework towards transfer subspace learning.

- **Essentially, we utilize the low-rank constraint to bridge the source and target domain in the low-dimensional space.**
- **In addition, many well-established methods can be incorporated into our framework to form their low-rank transfer versions.**
- **Extensive experimental results on Yale B, CMU PIE and UB KinFace databases sufficiently validate the effectiveness of our method on solving cross-database learning problems.**