Linear Manifold Regularization with Adaptive Graph for Semi-supervised Dimensionality Reduction

Kai Xiong, Feiping Nie, Junwei Han

Northwestern Ploytechnical University, Xi'an, 710072, P. R. China

Motivations & Contributions

◆ **Two Issues of graph-based methods**

1. Most existing graph-based methods perform dimensionality reduction and graph construction separately.
2. One usually constructs a graph by KNN criteria. However, due to the noise and redundant information in the original data, such a pre-defined graph has no clear structure.

◆ **Our Contributions**

1. We performe dimensionality reduction and graph construction simultaneously. Both the optimal graph and the projection matrix can be obtained.
2. We take the rank constraint to learn a structured graph. The learned graph is also sparse.
3. A simple algorithm is derived for the problem with rank constraint. Extensive experiments demonstrate its effectiveness.

Notations & Background

➢ Notations: $X \in \mathbb{R}^{d\times n}$ is the data matrix, where the first l samples are labeled and the last u samples are unlabeled. $Y \in \mathbb{R}^{c\times n}$ is the label matrix defined as y_{ji}=1 if x_i has label j and y_{ji}=0, otherwise. $W \in \mathbb{R}^{d\times c}$ is the projection matrix. L is the graph Laplacian and S is the similarity matrix. The i-th column of matrix M is m_i, and the (i, j)-th entry is m_{ij}.

◆Linear Manifold Regularization (LMR)

We take linear Laplacian regularized least squares (LapRLS/L) as an example to briefly introduce LMR.

$$\min_{W,b} \gamma_A \|W\|_F^2 + \gamma_I Tr(W^T XLX^T W) + \frac{1}{l}\sum_{i=1}^{l}\|W^T x_i + b - y_i\|^2.$$

◆Adaptive Neighbor Learning (ANL)

We consider the probabilistic neighbors to learn the similarity matrix S, and we simply have s_{ii}=0.

$$\min_{s_i^T \mathbf{1}=1, 0\le s_{ij}\le 1} \sum_{i,j=1}^{n} (\|x_i - x_j\|_2^2 s_{ij} + \gamma s_{ij}^2).$$

Formulation of LMRAG

➢ Incorporate ANL into LMR. Take the rank constratint as a structure constraint.

$$\min_{\substack{W,b,S\ge 0, S^T\mathbf{1}=\mathbf{1},\\ rank(L)=n-c}} \sum_{i,j=1}^{n}\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \gamma\|S\|_F^2 + \beta\|W\|_F^2 + \alpha Tr(W^T X + b\mathbf{1}^T - Y)U(W^T X + b\mathbf{1}^T - Y)^T$$

➢ Transform the rank constratint.

$$\sum_{i=1}^{c}\sigma_i = 0 \Leftrightarrow rank(L) = n-c$$

$$\sum_{i=1}^{c}\sigma_i = \min_{FF^T=I_c} Tr(FLF^T)$$

➢ The final objective.

$$\min_{W,b,S\ge 0, S^T\mathbf{1}=\mathbf{1}, FF^T=I_c} \sum_{i,j=1}^{n}\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \gamma\|S\|_F^2 + \beta\|W\|_F^2 + \alpha Tr(W^T X + b\mathbf{1}^T - Y)U(W^T X + b\mathbf{1}^T - Y)^T + 2\lambda Tr(FLF^T)$$

➢ U is a diagonal matrix that picks out the labeled data. σ_i is the i-th smallest eigenvalue of L. F is a kind of embedding of X.

➢ The first term of the objective function is acturally conducting ANL in the projected space.

➢ The multiplicity of eigenvalue zero of graph Laplacian L is equal to the number of connected components in the graph. Thus we can take the rank constraint as a structure constraint.

➢ The graph and the projection matrix can be both optimized.

Optimization

Update F with W,b, S fixed ⇒

$$\min_{FF^T=I_c} Tr(FLF^T) \quad (1)$$

Update W,b with F, S fixed ⇒

$$\min_{W,b} \sum_{i,j=1}^{n}\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \beta\|W\|_F^2 + \alpha Tr(W^T X + b\mathbf{1}^T - Y)U(W^T X + b\mathbf{1}^T - Y)^T \quad (2)$$

Update S with W,b, F fixed ⇒

$$\min_{S\ge 0, S^T\mathbf{1}=\mathbf{1}} \sum_{i,j=1}^{n}\|W^T x_i - W^T x_j\|_2^2 s_{ij} + \gamma\|S\|_F^2 + 2\lambda Tr(FLF^T) \quad (3)$$

$$\Rightarrow \min_{s_i\ge 0, s_i^T\mathbf{1}=1}\left\|s_i + \frac{1}{2\gamma}d_i\right\|_2^2, \left(d_{ij} = \|W^T x_i - W^T x_j\|_2^2 + \lambda\|f_i - f_j\|_2^2\right)$$

➢ the optimal F is formed by the eigenvectors of L corresponding to the first c smallest eigenvalues.

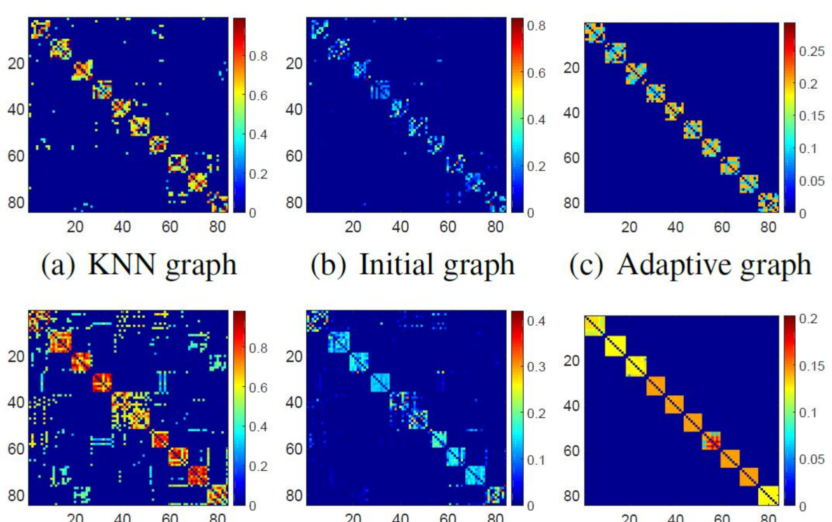
➢ The optimal W and b can be obtained by setting the derivatives of the objective equal to zero, respectively.

➢ It is independent to learn s_i for each sample. The quadratic problem with linear constraint has closed form solution. [Huang et al., 2015]

➢ Initialization: We learn an initial graph by ANL. The parameter γ can be initialized adaptively, and λ can be tuned dynamically.

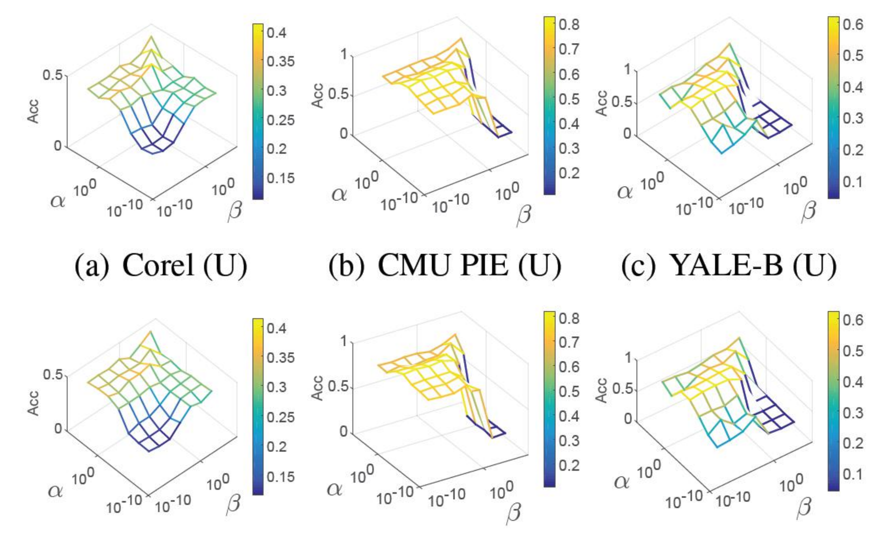
Experimental Results

➢ Comparison methods: (1) SDA [Cai et al., 2007]. (2) TR-FSDA [Huang et al., 2012]. (3) SSDL [Gao et al., 2015]. (4) FME [Nie et al., 2010], and (5) LapRLS/L [Sindhwani et al., 2005].



(a) KNN graph (b) Initial graph (c) Adaptive graph (d) KNN graph (e) Initial graph (f) Adaptive graph

Figure 1. Practical illustrations



(a) Corel (U) (b) CMU PIE (U) (c) YALE-B (U) (d) Corel (T) (e) CMU PIE (T) (f) YALE-B (T)

Figure 2. Parameter effect

Table 1 Classification Accuracy (Part of the results)

Dataset	Method	1 labeled sample		2 labeled sample		3 labeled sample	
		Unlabeled	Test	Unlabeled	Test	Unlabeled	Test
Corel	SDA	25.44 ± 3.42	25.42 ± 2.81	34.86 ± 3.67	34.58 ± 2.71	39.61 ± 4.68	38.39 ± 1.43
	TR-FSDA	25.44 ± 4.46	25.70 ± 2.97	33.83 ± 3.16	33.46 ± 3.27	38.07 ± 4.00	38.55 ± 3.56
	SSDL	26.75 ± 1.93	27.20 ± 2.84	34.89 ± 2.55	34.08 ± 2.26	38.32 ± 1.79	37.62 ± 2.74
	FME	23.65 ± 3.13	24.44 ± 2.56	30.15 ± 3.16	31.83 ± 4.46	33.79 ± 1.06	32.60 ± 1.26
	LapRLS/L	26.90 ± 1.60	26.60 ± 3.65	33.83 ± 2.34	34.31 ± 0.55	40.49 ± 1.42	39.94 ± 3.52
	LMRAG	**27.86 ± 3.49**	**27.73 ± 3.00**	**36.12 ± 2.56**	**36.46 ± 1.06**	**41.13 ± 1.28**	**41.27 ± 2.08**
COIL-20	SDA	69.75 ± 3.46	68.42 ± 2.35	77.85 ± 2.87	76.70 ± 3.29	82.23 ± 2.50	81.79 ± 2.74
	TR-FSDA	69.54 ± 2.36	68.53 ± 2.01	76.52 ± 3.11	76.69 ± 3.62	83.00 ± 1.18	82.77 ± 3.23
	SSDL	65.29 ± 2.54	64.56 ± 2.53	75.11 ± 2.04	75.56 ± 2.08	79.69 ± 3.37	80.12 ± 1.53
	FME	69.36 ± 4.12	69.35 ± 2.28	78.48 ± 1.95	76.98 ± 2.38	84.38 ± 1.74	84.35 ± 2.70
	LapRLS/L	68.68 ± 4.29	66.38 ± 1.69	75.85 ± 1.29	75.98 ± 2.67	79.69 ± 1.33	79.30 ± 1.61
	LMRAG	**70.75 ± 1.80**	**70.12 ± 2.57**	**79.70 ± 2.73**	**77.84 ± 3.09**	**84.58 ± 1.86**	**85.00 ± 1.45**
JAFFE	SDA	91.62 ± 1.76	88.06 ± 4.05	95.94 ± 2.84	97.36 ± 1.41	98.52 ± 1.55	99.22 ± 1.88
	TR-FSDA	87.84 ± 3.02	86.05 ± 7.25	96.88 ± 3.66	96.28 ± 2.76	98.15 ± 3.21	99.22 ± 1.34
	SSDL	83.24 ± 3.65	84.65 ± 3.97	94.69 ± 2.61	94.26 ± 2.72	98.89 ± 1.01	98.29 ± 1.77
	FME	80.27 ± 8.23	82.48 ± 5.20	92.19 ± 5.18	90.23 ± 3.54	94.81 ± 4.22	94.57 ± 1.98
	LapRLS/L	86.49 ± 6.12	86.51 ± 4.98	95.63 ± 5.11	94.57 ± 3.84	**99.26 ± 2.69**	98.29 ± 1.01
	LMRAG	**97.84 ± 2.80**	**98.45 ± 1.55**	**99.38 ± 2.09**	**98.45 ± 1.45**	**99.26 ± 1.66**	**99.69 ± 0.43**
CMU PIE	SDA	31.53 ± 3.71	32.29 ± 1.68	68.40 ± 2.31	68.38 ± 1.78	77.47 ± 2.32	77.57 ± 2.63
	TR-FSDA	18.98 ± 0.92	22.56 ± 1.37	67.55 ± 2.85	67.51 ± 1.35	79.27 ± 1.79	78.13 ± 1.30
	SSDL	53.78 ± 1.98	53.17 ± 2.35	70.28 ± 2.83	70.69 ± 2.10	77.49 ± 1.14	78.14 ± 1.11
	FME	53.49 ± 1.47	52.26 ± 1.24	69.92 ± 2.17	69.06 ± 1.36	78.06 ± 2.39	77.19 ± 1.92
	LapRLS/L	53.31 ± 2.19	52.80 ± 2.68	69.15 ± 2.09	68.63 ± 1.77	77.35 ± 2.33	76.57 ± 2.20
	LMRAG	**61.30 ± 2.29**	**61.29 ± 1.08**	**72.61 ± 2.50**	**72.61 ± 2.71**	**82.42 ± 1.06**	**81.93 ± 1.15**

References & Acknowledgments

[Huang et al., 2015] A new simplex sparse learning model to measure data similarity for clustering, IJCAI, 2015.

[Cai et al., 2007] Semisupervised discriminant analysis, ICCV, 2007.

[Huang et al., 2012] Semisupervised dimension reduction using trace ratio criterion, TNNLS, 2012.

[Gao et al., 2015] A novel semi-supervised learning for face recognition, Neurocomputing, 2015.

[Sindhwani et al., 2005] Linear manifold regularization for large scale semi-supervised learning, ICML, 2005.

[Nie et al., 2010] Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction, TIP, 2010.

◆**Acknowledgements:** This work was supported in part by the National Science Foundation of China under Grants 61522207 and 61473231.

Poster Presenter: Kai Xiong (bearkai1992@gmail.com). Research Interest: ML.

I am supposed to get my master's degree in March next year, and I'm looking for a job now.