UNSUPERVISED FEATURE SELECTION BY MANIFOLD REGULARIZED SELF-REPRESENTATION

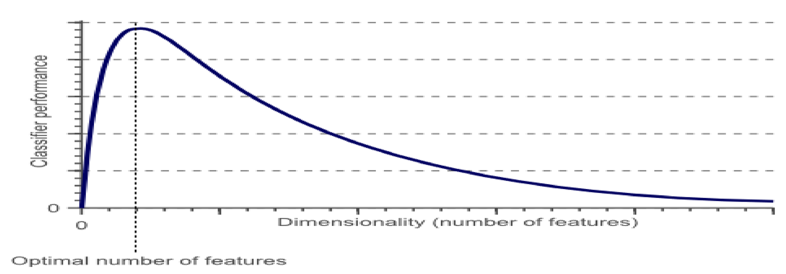
Siqi Liang[1], *Qian Xu*[2], *Pengfei Zhu*[2], *Qinghua Hu*[2], *Chongqing Zhang*[2]

[1]*School of Mathematical Sciences, Nankai University, Tianjin, China*

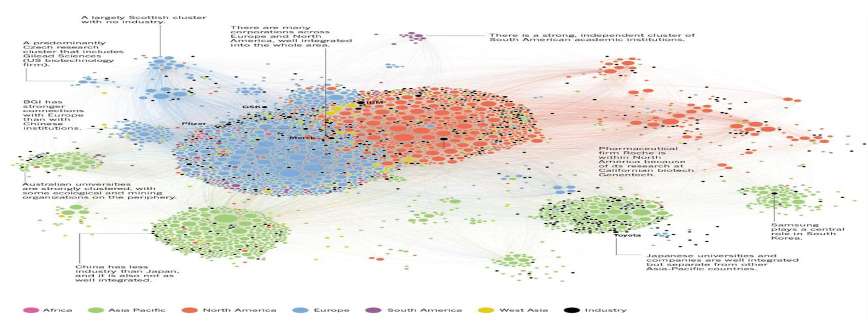
[2]*School of Computer Science and Technology, Tianjin University, Tianjin, China*

Introduction

In the past few years, high dimensional data emerge in many domains. Unsupervised feature selection has been proven to be an efficient technique in mitigating the curse of dimensionality. Moreover, the self-similarity property of objects, which assumes that a feature can be represented by the linear combination of its relevant features, has been successfully used in unsupervised feature selection. In this paper, we propose an algorithm that consider both the self-representation property and the manifold structure.



Curse of Dimensionality



Visualized High-Dimensional Data

Algorithm

The model discussed above is actually convex, but both the loss function and the regularization terms are non-smooth. In this section, we solve the optimization of MRSR using Iterative Reweighted Least-Squares (IRLS) algorithm.

Given the current estimation W^t , we define the diagonal weighting matrice G_L^t, G_R^t by $g_{:,i}^t = 1/2\|x_i - x_i W^t\|_2$, $g_{R,i}^t = 1/2\|w_j^t\|_2$.

In order to avoid the overflow error, a sufficiently small value ϵ is introduced by defining by $g_{:,i}^t = 1/\max(\|x_i - x_i W^t\|_2, \epsilon)$ and $g_{R,i}^t = 1/\max(\|w_j^t\|_2)$.The formal algorithm is stated in Algorithm 1.

Algorithm 1 MRSR based unsupervised feature selection

Input:
Training data $\mathbf{X} \in \mathbb{R}^{n\times m}$
Parameters λ_0 and λ_1
Set $t = 0$ and initialize $\mathbf{G}_L^t$ and $\mathbf{G}_R^t$;
Compute the Laplacian matrix $\mathbf{L}$;
repeat
$\mathbf{W}^{t+1} = (\mathbf{X}^T\mathbf{G}_L^t\mathbf{X} + \lambda_0\mathbf{X}^T\mathbf{L}\mathbf{X} + \lambda_1\mathbf{G}_R^t)^{-1}\mathbf{X}^T\mathbf{G}_L^t\mathbf{X}$;
update $\mathbf{G}_L^{t+1}$ and $\mathbf{G}_R^{t+1}$, $t = t+1$;
until Converge.
Calculate feature weights $v_i = ||\mathbf{w}_i||_2$
Output:
$v_i, i = 1, 2, \ldots, m$

Experimental Results

Dataset	Laplacian	MCFS	UDFS	SPEC	RSR	MRSR
warpPIE10P	20.6	42.1	**47.9**	36.1	43.4	41.4
warpAR10P	21.1	22.7	44.2	45.6	32.8	**44.8**
pixraw10P	50.7	**87.6**	69.2	48.1	63.9	85.1
orlraw10P	40.1	**78.4**	72.3	37.81	60.1	76.7
TOX-171	40.4	40.3	40.3	38.8	42.3	**50.0**
CLL-SUB-111	37.4	50.9	50.3	50.9	50.6	**54.4**
Average	35.1	53.7	54.0	42.9	48.9	**58.7**

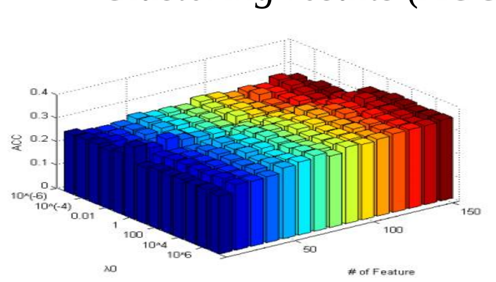
Clustering results (ACC)

Dataset	Laplacian	MCFS	UDFS	SPEC	RSR	MRSR
warpPIE10P	20.6	**54.6**	52.3	39.7	50.3	51.6
warpAR10P	20.2	20.0	**48.6**	48.0	34.7	47.9
pixraw10P	67.1	91.4	77.4	54.7	72.4	**92.3**
orlraw10P	49.4	84.4	78.1	44.5	67.4	**84.7**
TOX-171	11.9	11.8	11.4	9.8	14.8	**27.9**
CLL-SUB-111	2.9	19.7	14.9	19.9	19.4	**22.9**
Average	28.7	47.0	47.1	36.1	43.2	**54.6**

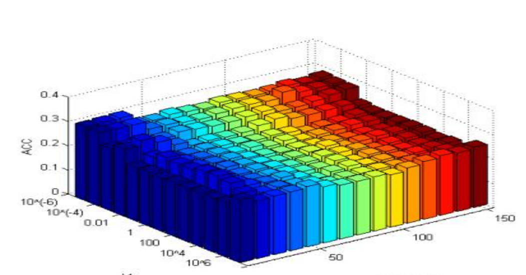
Clustering results (NMI)

Dataset	Laplacian	MCFS	UDFS	SPEC	RSR	MRSR
warpPIE10P	87.0	**99.1**	96.2	86.5	94.8	**99.1**
warpAR10P	63.7	74.4	83.2	74.7	57.5	**86.6**
pixraw10P	70.1	97.5	97.2	49.3	87.3	**99.0**
orlraw10P	45.7	91.6	92.9	67.1	71.7	**96.9**
TOX-171	54.7	**65.5**	56.9	54.1	57.2	63.5
CLL-SUB-111	62.3	58.1	**81.9**	59.5	66.3	66.0
Average	63.9	81.0	84.7	65.2	72.5	**85.2**

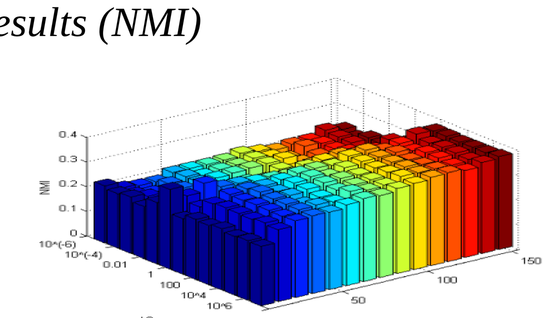
Classification rates(%)



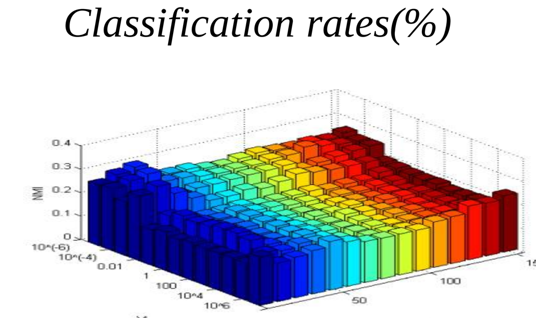
ACC($\lambda_1 = 0.001$)



ACC($\lambda_0 = 0.001$)



NMI($\lambda_1 = 0.001$)



NMI($\lambda_0 = 0.001$)

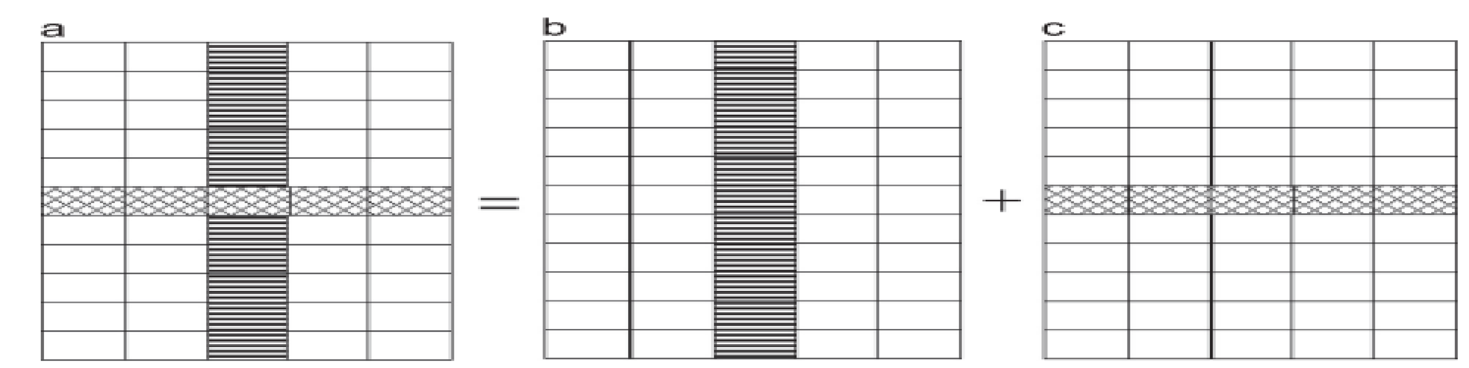
MRSR achieves the best performance in terms of clustering accuracy, NMI and classification accuracy among all the competing methods. Besides, we also investigate the sensitiveness of the parameters of MRSR. The experiment result indicates that our method is not very sensitive to the number of the features. Furthermore, the performance of MRSR is also not very sensitive to parameters λ_0 and λ_1.

MRSR

In RSR, data matrix X is used as the response matrix, and each feature can be represented by all the features with different representation coefficients. However, RSR did not take the structure information of unlabeled data into consideration. Motivated by the manifold learning, we further incorporate a manifold regularization term to preserve data similarity. According to the discussion above, now we have the following minimization problem:

$$\widehat{W} = \arg\min\|X - XW\|_{2,1} + \lambda_0 tr(W^T X^T L X W) + \lambda_1\|W\|_{2,1}$$

We call the above model Manifold Regularized Self-Representation (MRSR) for unsupervised feature selection.



A data matrix with outliers and redundant features. (a) Corrupted data matrix (b) Redundant features and (c) Outliers.

Dataset Summary

In this paper, we use six real-world datasets for extensive experiments. We use six real-world datasets for extensive experiments. There are 4 face image datasets (i.e., warpPIE10P, warpAR10P, pixraw10p, orlraws10P) and 2 microarray datasets (i.e., TOX-171 and CLL-SUB-111).

Datasets	Instances	Features	Classes	Domains
warpPIE10P	210	2420	10	Image, Face
warpAR10P	130	2400	10	Image, Face
pixraw10P	100	10000	10	Image, Face
orlraw10P	100	10304	10	Image, Face
TOX-171	171	5748	4	Microarray, Bio
CLL-SUB-111	111	11340	3	Microarray, Bio

Parameter Setting

- We fix k = 5 for all the datasets to specify the neighborhood size.
- We tune the bandwidth and two regularization parameters from $\{10^{-6}, 10^{-5}, 10^{-4}, \cdots, 10^{5}, 10^{6}\}$ and record the best result.
- For feature dimension, we set the number of features as $\{10, 20, 30, \cdots, 150\}$ and report the average results over different dimensions.
- The K-means clustering algorithm is performed on the selected features by different algorithms. The experiment is run for 20 times with different random initializations. The average results are reported for all the comparing algorithms.

Conclusion

In this paper, we proposed a manifold regularized self-representation (MRSR) model for unsupervised feature selection. The L2,1-norm is used to measure the self-representation residual to alleviate the impact of the outliers. The representation coefficients are also regularized by the L2,1-norm sparsity to select effective features. To maintain the sample similarity of the raw space in the reconstructed space, a manifold regularization is imposed on reconstructed samples. As a result, the most representative features which can reconstruct other features and preserve locality are selected. The experiment results validated the effectiveness of MRSR in terms of both the clustering and classification performances. In the future, we will extend MRSR tasks to multi-view or classification problems.

References: Tang, J.; Liu, H.; Guyon, I.; Elisseeff, A.; Dy, J.; Brodley, C.; Nie, F.; Huang, H.; Cai, X.; He, X.; Niyogi, P.; Wang, S.; Buades, A.; Coll, B.; Morel, J.; Elhamifar, E.; Vidal, R.; Zhu, P.; Zuo, W.; Zhang, L.; Hu, Q.; Shiu, S,; Liu, Y.; Liu, K.; Chen, J.; Zhou, J.; Ye, J.; Stojnic, M.; Parcaresh, F.; Hassibi, B.; Yang, Y.; Jafari, A.; Almasganj, F.; Ding, C.; Cai, D.; Zhao, Z.; Luo, J.; Tao, H.; Huang, Z.; Zhuang, Y.; Zha Y.; Shen, T.;