Adaptive Affinity Matrix for Unsupervised Metric Learning

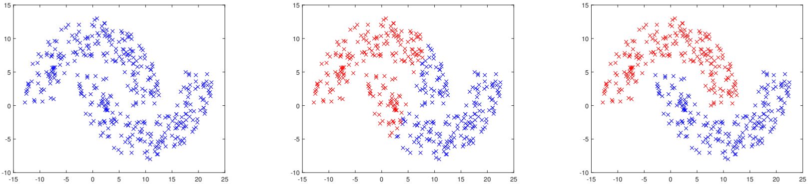
Yaoyi Li, Junxuan Chen, Yiru Zhao and Hongtao Lu

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering,
Shanghai Jiao Tong University,
P.R.China

July 2016

Background: Spectral Clustering

- Spectral clustering: nonlinear feature reduction.
- The distribution of real data does not always obey uniform or gaussian.
- Spectral clustering can preserve the local neighborhood information.



(a) (b) (c)

Background: Spectral Clustering

- Spectral clustering demonstrates a splendid performance on many challenge data sets.
- Objective function:

$$y = \arg\min_{y^T D y = 1} \sum_{i,j}^{n} w_{ij} \|y_i - y_j\|_2^2 ,$$

  where w_{ij} is the similarity between data sample x_i and x_j (a.k.a. affinity graph).
- Shortcomings of spectral clustering
  - Out-of-sample extension is not straightforward
  - Cubic time complexity
  - Sensitive to the affinity graph

Background: Locality Preserving Projections

- Locality Preserving Projections (LPP) [HN04] is the linear approximation of Laplacian Eigenmap.
- Locality Preserving Projections conducts dimensionality reduction by solving the optimization problem:

$$a = \underset{a^T XDXa=1}{\arg\min} \sum_{i,j}^{n} w_{ij} \| a^T x_i - a^T x_j \|_2^2 \ ,$$

- The superiority of LPP
  - Explicit projection for out-of-sample extension
  - Complexity is reduced

Motivation

- The performance of spectral clustering methods highly depends on the robustness of the affinity graph.
- Some weighting methods like $k-$NN heat kernel will be corrupted by noises.
- Our goal:
  - Learn a robust affinity graph by optimization efficiently.
  - Optimize the linear projection and affinity graph simultaneously.

Related Works

- Dominant Neighbors [PP07] reduces the noise of the affinity matrix by maximal cliques.
- Consensus k-NNs [PK13] builds affinity graph by consensus information.
- ClustRF-Strct [ZLG14] constructs an affinity graph via the clustering random forests.
- CAN and PCAN [NWH14] learn data similarity and cluster structure simultaneously.

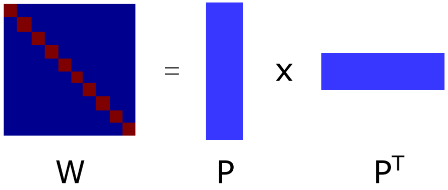
AdaAM: Assumption

- Assumption 1: The affinity matrix W is a positive semidefinite matrix. Hence we have,

$$W = PP^T .$$

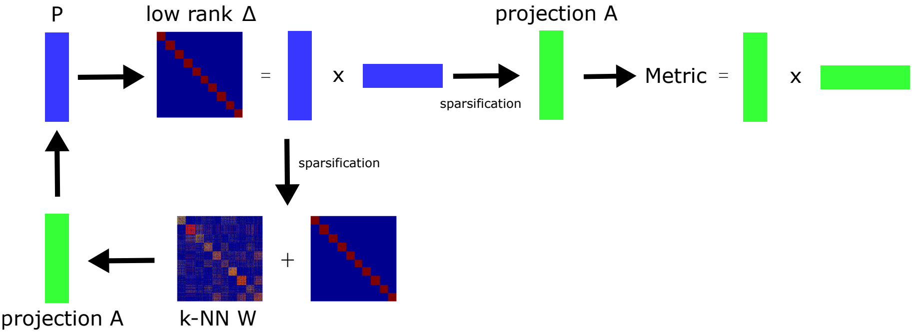
  This assumption also appeared in [CC11]

- Assumption 2: The ideal affinity matrix W is a low rank matrix (1 for the sample in the same class and 0 for the others).



AdaAM: Diagram

- A glance of our algorithm



AdaAM: Intermediate Affinity Matrix Δ

- Let Δ be the intermediate affinity matrix, and assume $\Delta = PP^T$. Compute P by solving optimization problem

$$\min_{P^TP=I} tr(X^T(D_\Delta - PP^T)X)$$
$$\Rightarrow \min_{P^TP=I} tr(X^T D_\Delta X) + tr(X^T(-PP^T)X)$$

  similar to spectral clustering

- When X is normalized with zero mean, we have $D_\Delta = \mathbf{0}$. The above problem is equivalent to

$$P = \arg\max_{P^TP=I} \; tr(P^T X X^T P)$$

AdaAM: Final Adaptive Affinity Matrix

- With the intermediate affinity matrix Δ, we can solve the following problem for a linear projection A:

$$A = \arg\min_{A^T A = I} tr(A^T X^T (L + L_\Delta) X A)$$

  $L + L_\Delta$ is the combination of the Laplacian of k-NN heat kernel and the intermediate affinity matrix.

- With the linear projection A, we can rewrite the affinity optimization problem and update matrix P ($D_\Delta = \mathbf{0}$ still holds).

$$P = \arg\max_{P^T P = I} tr(P^T X A A^T X^T P)$$

Experiments

- We evaluate the proposed approach on five image data sets
  - UMIST, COIL20, USPS, MNIST, ExYaleB
- We impose the same parameter selection criteria on all the algorithms in our experiments.
  - the size of neighborhood $k = \mathrm{Round}(\log_2(n/c))$
  - projected dimension is the same as the number of classes
- We denote 10 times of k-Means as a round and select the clustering result with the minimal within-cluster sum as the result of each round of k-Means.

Accuracy

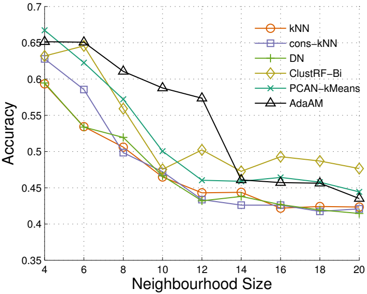
- 100 rounds *k*-Means to each algorithms for the evaluation of the performance.

Table: Clustering accuracy on image data sets(%)

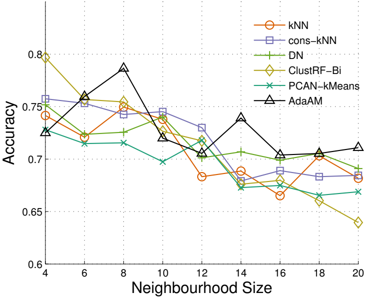
	AdaAM		***k*-NN**		**Cons-*k*NN**		**DN**		**ClustRF-Bi**		**PCAN-*k*Means**		**PCAN**
	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	Avg	Max	
UMIST	**66.06**	**75.65**	58.16	65.39	60.27	69.22	59.15	66.96	64.63	74.44	53.79	56.52	55.30
COIL20	74.72	**87.29**	71.89	81.18	75.53	84.31	71.95	82.01	**76.50**	85.07	72.28	83.75	81.74
USPS	**69.36**	**69.61**	68.25	68.35	68.21	68.34	68.08	68.31	58.74	65.90	64.04	67.95	64.20
MNIST	**60.84**	**61.34**	48.13	48.27	47.88	48.00	49.72	49.76	51.93	52.03	58.93	58.98	59.83
ExYaleB	**54.36**	**57.87**	24.17	26.76	25.63	28.75	24.21	27.42	23.10	26.43	25.74	27.63	25.89

Accuracy

- 10 rounds *k*-Means for the experiment of the sensitivity to the neighborhood size



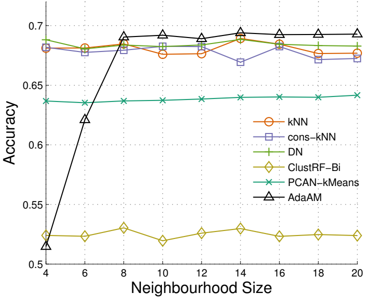
(d) UMIST



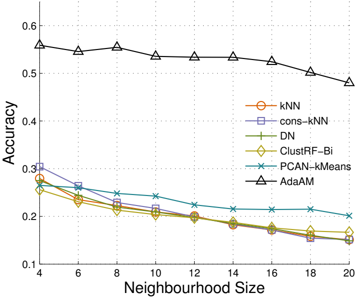
(e) COIL20

Figure: Comparison between different with different of neighborhood size *k*

Accuracy



(a) USPS



(b) ExYaleB

Figure: Comparison between different with different of neighborhood size k

- Requires more information from the pairwise similarity.
- For small k, sometimes does not perform well.

Time Consumption

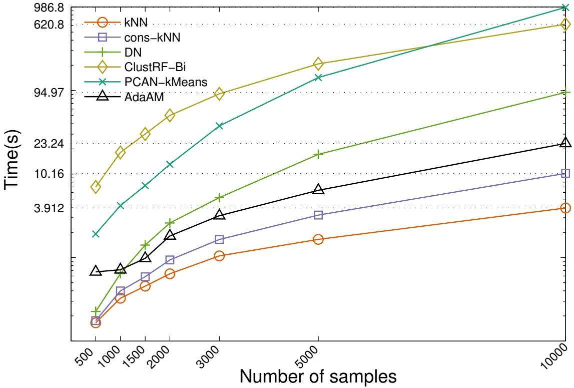


Figure: Time consumption of six approaches with different number of data instances

Conclusion & Future Work

- Conclusion
  - We present a novel affinity learning approach for unsupervised metric learning.
  - The affinity matrix is learned from the same framework of spectral clustering.
  - The affinity learning can be reduced to a singular value decomposition problem.
  - We employ the low rank trick to make our approach more efficient.
- Future Work
  - A better way to learn the parameter of sparsification
  - A better way to fuse low rank Δ and k-NN W.
  - More applications

Thanks

Thanks for your Attention.

References

[CC11] Xinlei Chen and Deng Cai, *Large scale spectral clustering with landmark-based representation.*, AAAI, 2011.

[HN04] Xiaofei He and Partha Niyogi, *Locality preserving projections*, NIPS, vol. 16, 2004, p. 153.

[NWH14] Feiping Nie, Xiaoqian Wang, and Heng Huang, *Clustering and projected clustering with adaptive neighbors*, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 977–986.

[PK13] Vittal Premachandran and Ramakrishna Kakarala, *Consensus of k-nns for robust neighborhood selection on graph-based manifolds*, Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 1594–1601.

[PP07] Massimiliano Pavan and Marcello Pelillo, *Dominant sets and pairwise clustering*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **29** (2007), no. 1, 167–172.

[ZLG14] Xiatian Zhu, Chen Change Loy, and Shaogang Gong, *Constructing robust affinity graphs for spectral clustering*, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014, pp. 1450–1457.