Structure Aware L1-Graph for Data Clustering

Shuchu Han, Hong Qin
Computer Science Department
Stony Brook University
Stony Brook, NY 11790

Abstract

In graph-oriented machine learning research, $\mathcal{L}_1$ graph is an efficient way to represent the connections of input data samples. Its construction algorithm is based on a numerical optimization motivated by Compressive Sensing theory. As a result, It is a nonparametric method which is highly demanded. However, the information of data such as geometry structure and density distribution are ignored. In this paper, we propose a Structure Aware (SA) $\mathcal{L}_1$ graph to improve the data clustering performance by capturing the manifold structure of input data. We use a local dictionary for each datum while calculating its sparse coefficients. SA-$\mathcal{L}_1$ graph not only preserves the locality of data but also captures the geometry structure of data. The experimental results show that our new algorithm has better clustering performance than $\mathcal{L}_1$ graph.

Introduction

$\mathcal{L}_1$ graph was proposed by Cheng et al. (Cheng et al. 2010), with an initial goal for Image Classification. It was presented as an alternative graph representation of input data samples for graph-oriented machine learning tasks. Comparing to existing popular graph representation methods such as k-nearest neighbor graph and ϵ-ball graph, the $\mathcal{L}_1$ graph was a parametric free method and the user didn't need to tune parameters like k or ϵ for best learning performance. Except for this predominant feature, experiment results also show that $\mathcal{L}_1$ graph has following three advantages:(1)robustness to noise; (2)sparsity; (3)datum-adaptive neighbors. (Cheng et al. 2010).

Although $\mathcal{L}_1$ graph has so many decent advantages, we notice that the construction algorithm of it is a pure numerical result from $\mathcal{L}_1$ minimization and based on Compressed Sensing theory. This numerical process brings the non-parametric characteristic to $\mathcal{L}_1$ graph but ignores the intrinsic physical information of input data. Recently, researchers in machine learning area propose several new $\mathcal{L}_1$ graph construction methods to improve the learning performance by exploiting the input data. For example, Zhou et al. propose k-nearest neighbor fused lasso graph by regularizing the sparse codes(or coefficients) of k-nearest neighbors (Zhou, Lu, and Peng 2013). Fang et al. add the auto-grouped effect to $\mathcal{L}_1$ graph by applying two sparse regularizations: Elastic net and Octagonal Shrinkage and Clustering Algorithm for Regression (OSCAR) (Fang et al. 2015). Han et al. use a reduced size dictionary to preserve the locality for data clustering applications (Han et al. 2015). Yang et al. use the Graph Laplacian regularization to improve the quality of $\mathcal{L}_1$ graph by exploiting the geometry structure information of input data (Yang et al. 2014).

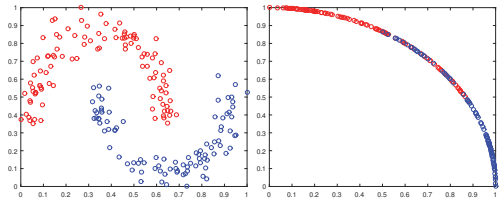

Figure 1: Dictionary normalization of *two moon* dataset. The red and blue points represent different clusters. Left: before normalization, right: after normalization. We can see that the neighborhood information is changed after normalization.

Another less attractive aspect of $\mathcal{L}_1$ graph construction algorithm is the *normalization* of dictionary. While calculating the sparse coefficient (or $\mathcal{L}_1$ minimization), it requires all dictionary atoms (or data sample) to have unit length. Usually, we use $\mathcal{L}_2$ normalization. This normalization process project all atoms to unit hypersphere and eliminates the *locality* information of data as show by figure 1. As we can see, the neighborhood information is changed after normalization.

In this paper, we propose a structure aware (SA) $\mathcal{L}_1$ graph to continue to improve the data clustering performance. Comparing to the strategy of adding regularization terms, we choose to search a local dictionary for each data sample while calculating the sparse coefficients. Unlike the method described in (Han et al. 2015) which use the k-nearest neighbor as dictionary, we select atoms following the intrinsic manifold structure of data. The advantage of our selection is that it not only preserves the locality, but also captures the geometry structure of data (figure 2). As pointed out by (Yang et al. 2014), in many real applications, high-dimensional data always reside on or close to an intrinsically low dimensional manifold embedded in the high-dimensional ambient space. This is the fundamental assumption of manifold learning and also emphasizes the importance of utilizing manifold struc-

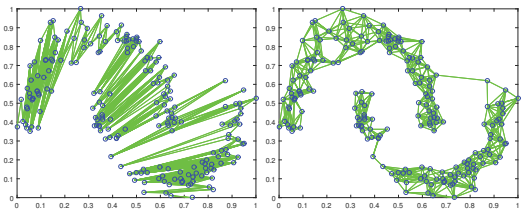

Figure 2: $\mathcal{L}_1$ graph (Left) and SA-$\mathcal{L}_1$ graph (Right,$K = 10$) of Two Moon dataset.

ture in learning algorithms. Our proposed algorithm has a user specific parameter k which leads to the lost of parametric-free characteristic. But our experiment results show that it increases the clustering performance and reduces the running time.

Algorithm

The basic idea of $\mathcal{L}_1$ graph is to find a sparse coefficient (or coding) for each data sample. Given dataset $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}]$, where $\boldsymbol{x_i} \in \mathbb{R}^m, i \in [1, \cdots, n]$ is a vector which represents a data sample. The sparse coefficient $\boldsymbol{\alpha}_i \in \mathbb{R}^{n-1}$ of $\boldsymbol{x_i}$ is calculated by following $\mathcal{L}_1$ minimization process.

$$\min_{\boldsymbol{\alpha_i}} \|\boldsymbol{\alpha_i}\|_1 \text{ subject to } \boldsymbol{x_i} = \boldsymbol{\Phi}^i \boldsymbol{\alpha_i}, \boldsymbol{\alpha_i} \geq 0. \quad (1)$$

We put constrain $\boldsymbol{\alpha_i} \geq 0$ here to let coefficients have physical meaning of *similarity*. In original $\mathcal{L}_1$ graph construction algorithm, the dictionary $\boldsymbol{\Phi}^i = [\boldsymbol{x_1}, \cdots, \boldsymbol{x_{i-1}}, \boldsymbol{x_{i+1}}, \cdots, \boldsymbol{x_n}]$. Here, we select K atoms $\hat{\boldsymbol{\Phi}}^i = [\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_K]$ from $\boldsymbol{\Phi}^i$ by using manifold ranking scores (Zhou et al. 2004) (Xu et al. 2011). The algorithm can be described as Algorithm 1.

We use the closed form solution to calculate the manifold ranking scores for all data samples:

$$\mathbf{F} = (I - \alpha S)^{-1}, \quad (2)$$

where S is the Graph Laplacian matrix and we use Gaussian Kernel here. Each column of F is the relative manifold ranking scores of data sample $\mathbf{x}_i$.

Experimental Results

To evaluate the performance of our proposed algorithm, we exam it through spectral clustering applications and compare it to different graphs: Gaussian similarity (GS) graph and $\mathcal{L}_1$ graph. Six UCI datasets are selected. The clustering performance is measured by Normalized Mutual Information(NMI) and Accuracy(AC). In our experiment setting, we select $\alpha = 0.99$ for manifold ranking, and K equals to 10%,20% and 30% percent of total number of data samples. Our experiment results show that SA-$\mathcal{L}_1$ graph has better clustering performance than $\mathcal{L}_1$ graph generally.

Acknowledgments

The authors wish to thank all anonymous reviewers for their helpful suggestions and critiques. This work is supported in part by National Natural Science Foundation of China (Grant No. 61190120, 61190121, 61190125, 61532002), National Science Foundation of USA (Grant No. IIS-0949467, IIS-1047715, and IIS-1049448).

Algorithm 1: SA-$\mathcal{L}_1$ graph

Input : Data samples $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}]$, where $\mathbf{x_i} \in \mathbf{X}$;
Parameter K;

Output: Adjacency matrix W of sparse graph.

Calculate the manifold ranking score matrix $\mathbf{F}$;
Normalize the data sample x_i with $\|x_i\|_2 = 1$;
for $\mathbf{x}_i \in X$ **do**
  Select top K atoms from $\mathbf{F}(i)$, and build $\hat{\boldsymbol{\Phi}}^i$;
  Solve: $\min_{\alpha_i} \|\alpha_i\|_1, \quad s.t. \quad x_i = \hat{\boldsymbol{\Phi}}^i \alpha_i, \alpha_i \geq 0$;
  $\mathbf{W}(i,:) = \alpha_i$;
end
return $\mathbf{W}$;

Name	Metric	$\mathcal{L}_1$	Gaussian graph			SA-$\mathcal{L}_1$ graph		
			K:10%	K:20%	K:30%	K:10%	K:20%	K:30%
BT	NMI	0.4055	0.4839	0.4749	0.5178	**0.5436**	0.5524	0.4702
	AC	0.5283	0.5189	0.5189	0.5377	**0.6604**	0.6321	0.5755
Wine	NMI	0.7717	0.8897	0.8897	0.8897	**0.9209**	0.8946	0.8043
	AC	0.9326	0.9719	0.9719	0.9717	**0.9775**	0.9663	0.9382
Glass	NMI	0.3794	0.3642	0.3763	0.2572	0.3746	**0.3998**	0.3715
	AC	0.4486	0.5140	0.5187	0.4439	0.4486	**0.4579**	0.4533
Soybean	NMI	0.6531	0.6509	0.7022	0.6884	0.6858	0.7096	**0.7192**
	AC	0.4984	0.4625	0.5505	0.5212	0.5179	0.5179	**0.5505**
Vehicle	NMI	0.1424	0.0802	0.0806	0.0814	0.1173	0.1127	**0.1651**
	AC	0.3747	0.3664	0.3676	0.3582	0.3818	0.3818	**0.3830**
Image	NMI	0.5658	0.5514	0.5454	0.5699	0.5034	**0.5877**	0.5694
	AC	0.6271	0.4752	0.5286	0.5505	0.5443	**0.6467**	0.6133

Table 1: Clustering performance of SA-$\mathcal{L}_1$-graph construction algorithms. $\mathcal{L}_1$-graph is the baseline.

References

Cheng, B.; Yang, J.; Yan, S.; Fu, Y.; and Huang, T. S. 2010. Learning with l1-graph for image analysis. *Tran. of Image Processing* 19(4):858–866.

Fang, Y.; Wang, R.; Dai, B.; and Wu, X. 2015. Graph-based learning via auto-grouped sparse regularization and kernelized extension. *IEEE Transactions on Knowledge and Data Engineering* 27(1):142–154.

Han, S.; Huang, H.; Qin, H.; and Yu, D. 2015. Locality-preserving l1-graph and its application in clustering. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, 813–818. ACM.

Xu, B.; Bu, J.; Chen, C.; Cai, D.; He, X.; Liu, W.; and Luo, J. 2011. Efficient manifold ranking for image retrieval. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 525–534. ACM.

Yang, Y.; Wang, Z.; Yang, J.; Wang, J.; Chang, S.; and Huang, T. S. 2014. Data clustering by laplacian regularized l1-graph. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 3148–3149.

Zhou, D.; Weston, J.; Gretton, A.; Bousquet, O.; and Schölkopf, B. 2004. Ranking on data manifolds. *Advances in Neural Information Processing Systems* 16:169–176.

Zhou, G.; Lu, Z.; and Peng, Y. 2013. L 1-graph construction using structured sparsity. *Neurocomputing* 120:441–452.