Constrained Spectral Clustering Network with Self-Training

Xinyue Liu, Shichong Yang, Linlin Zong*

Dalian University of Technology

yangshichong.dut@gmail.com, {xyliu, llzong}@dlut.edu.cn

Abstract

Deep spectral clustering networks have shown their superiorities due to the integration of feature learning and cluster assignment, and the ability to deal with non-convex clusters. Nevertheless, deep spectral clustering is still an illposed problem. Specifically, the affinity learned by the most remarkable SpectralNet is not guaranteed to be consistent with local invariance and thus hurts the final clustering performance. In this paper, we propose a novel framework of Constrained Spectral Clustering Network (CSCN) by incorporating pairwise constraints and clustering oriented fine-tuning to deal with the ill-posedness. To the best of our knowledge, this is the first constrained deep spectral clustering method. Another advantage of CSCN over existing constrained deep clustering networks is that it propagates pairwise constraints throughout the entire dataset. In addition, we design a clustering oriented loss by selftraining to simultaneously finetune feature representations and perform cluster assignments, which further improve the quality of clustering. Extensive experiments on benchmark datasets demonstrate that our approach outperforms the state-of-the-art clustering methods.

Motivation

- Deep Clustering

Deep clustering combines feature learning and cluster assignment to surpass the shallow methods. However, unsupervised clustering is inherently an ill-posed problem due to its data-driven nature.

- Deep Constrained Clustering

Deep constrained clustering, which employs a small amount of pairwise constraints as prior knowledge, can release the ill-posedness and significantly improve the clustering performance. However, most existing methods do not consider constraint propagation to make use of unlabeled data.

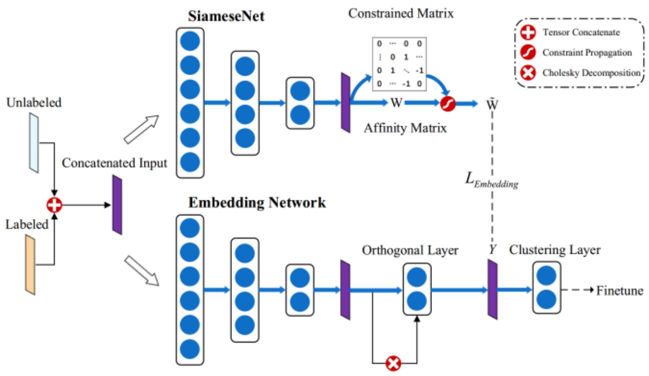
- Spectral Clustering

Spectral clustering achieves superior performance than other clustering methods due to the elegant theory and its ability to handle non-convex clusters.

- Deep Spectral Clustering

The affinity learned by the SpectralNet is not guaranteed to be consistent with local invariance and thus hurts the final clustering performance.

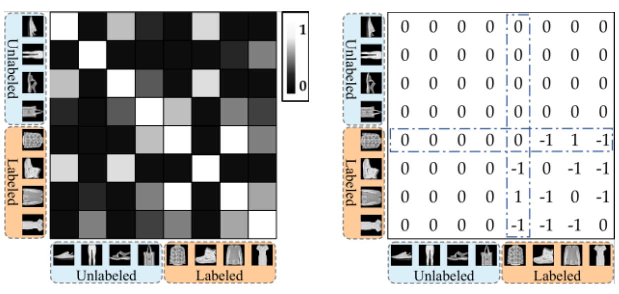
Proposed Approach



We concatenate unlabeled and labeled batches as new inputs. Constraint propagation is performed after SiameseNet training is completed. The adjusted W is used to guide the embedding network to obtain discriminative features. We apply an orthogonal layer to avoid trivial solutions, and the clustering layer to finetune feature space.

- Constraints and their Propagation

Affinity Matrix | Constraint Matrix



- Unsupervised Building Block SpectralNet

$$L_c = P\|g_\theta(x_i) - g_\theta(x_j)\|^2 + (1-P)\max(c - \|g_\theta(x_i) - g_\theta(x_j)\|, 0)^2 \qquad L_{Embedding} = \frac{1}{m^2}\sum_{i,j=1}^{m} W_{i,j}\|y_i - y_j\|^2$$

- Constraint propagation and affinity adjustment

$$\tilde{W}_{ij} = \begin{cases} 1-(1-F^*_{ij})(1-W_{ij}), & F^*_{ij} \geq 0; \\ (1+F^*_{ij})W_{ij}, & F^*_{ij} < 0. \end{cases} \qquad \begin{aligned} F_{ij}(t+1) &= \sum_s W_{is}F_{sj}(t)/\sqrt{D_{ii}D_{ss}} \\ &= (SF(t))_{ij}. \end{aligned}$$

- Clustering Oriented Fine-tuning

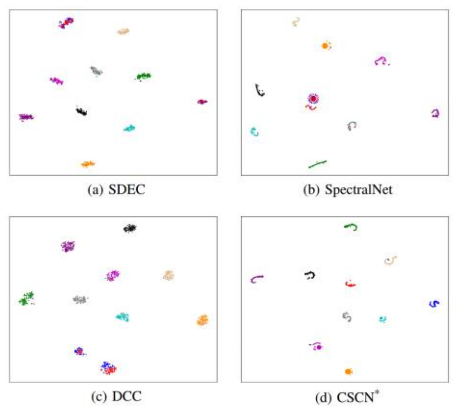
$$L_f = KL(P\|Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}}$$

Experiment

TABLE I

COMPARISON OF CLUSTERING PERFORMANCE IN TERMS OF ACC AND NMI. THE BEST RESULTS ARE IN BOLD.

Model	MNIST		Fashion-MNIST		REUTERS-10K	
	ACC	NMI	ACC	NMI	ACC	NMI
AE+k-means	0.7548	0.7297	0.5260	0.5630	0.6728	0.3885
COP-Kmeans	0.7913	0.7392	0.5555	0.5804	0.7145	0.4358
PCKmeans	0.7758	0.7349	0.5514	0.5737	0.7117	0.4378
DEC	0.8653	0.8370	0.5720	0.6274	0.7047	0.4571
IDEC	0.8810	0.8652	0.5955	0.6341	0.7536	0.4956
SpectralNet	0.8309	0.8929	0.6577	0.7034	0.7195	0.5135
SDEC	0.8651	0.8367	0.5985	0.6328	0.7368	0.4976
DCC	0.8958	0.8782	0.6389	0.6309	0.7870	0.6386
CSCN*	0.9486	0.9286	0.6836	0.7092	0.8504	0.6447
CSCN	**0.9500**	**0.9303**	**0.6883**	**0.7127**	**0.8528**	**0.6484**



(a) SDEC (b) SpectralNet (c) DCC (d) CSCN*

Fig. 3. Visualization of clustering results on subset of MNIST. Different colors represent different clusters. Note clusters colored by blue and red (digits 4 and 9), they are totally mixed together in other methods while still clearly separable in our CSCN*.

Conclusion

- We propose Constrained Spectral Clustering Network (CSCN), which incorporates pairwise constraints and clustering oriented fine-tuning.
- Constraints are propagated to protect local invariance and guide spectral embedding.
- The clustering loss is optimized to finetune the feature space and perform cluster assignments simultaneously.