Robust Deep Spectral Clustering

Andrei Rykov[0000−0003−1008−7421]
Supervised by Sibylle Hess[0000−0002−2557−4604]

Eindhoven University of Technology, Eindhoven, the Netherlands
a.g.rykov@student.tue.nl
s.c.hess@tue.nl

Keywords: Deep Clustering · Spectral Clustering · Graph Kernel.

Clustering analysis is a crucial component of data exploration and understanding, enabling the identification of meaningful groups within datasets. Traditional clustering methods face limitations in handling large and high-dimensional datasets. To address these challenges, deep clustering methods leverage the power of deep learning to learn informative representations for clustering.

An intuitive and straightforward way to gain a low-dimensional embedding of a dataset is given by autoencoders. The autoencoder objective can further be regularized towards clustering-friendly embeddings. A popular strategy is to add a simple clustering objective such as the one of k-means to the reconstruction loss. Although models based on this idea demonstrate a suitable performance on some benchmarks, non-convex clustering remains an unsolved problem. The k-means objective has many local optima that have very similar objective function values. Correspondingly, the objective to find an embedding that enables an accurate reconstruction while clustering well is often too ambiguous to detect even clearly cut clusters. Figure 1 shows an example of a clustering-steered embedding computed by Deep K-Means (DKM) [1], failing to extract the ground truth clusters.

Another deep clustering approach is to learn a deep embedding that optimizes the spectral clustering objective. Spectral clustering (SC) is a graph clustering method that can be seen as an instance of kernel k-means. Given a weighted adjacency matrix that represents similarities of the data points, the objective of SC is to cut clusters such that the weights of cut edges are minimized. This is formalized as Ratio Cut objective [3]

$$\min \mathcal{L}_{RC}(Z) = \sum_{1\leq i,j\leq n} S_{ij}||Z_{\cdot i} - Z_{\cdot j}||_2^2 \quad \text{s.t. } Z^\top Z = I.$$

Spectral clustering is optimized by applying k-means on the top eigenvectors of the graph Laplacian matrix. The embedding given by the eigenvectors has a clear connection to the connected components in the graph from which the suitability of the method is derived [3]. In fact, optimizing the objective of k-means on the eigendecomposition is indeed approximating the Ratio Cut objective [2].

A drawback of spectral clustering models is that an affinity matrix S has to be chosen, which heavily influenced the obtained clustering. Several approaches

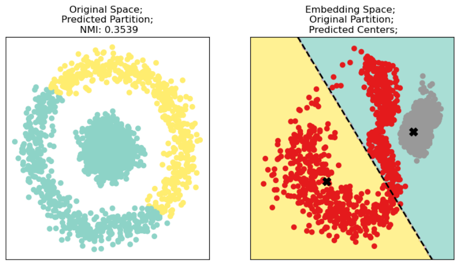

Fig. 1: The clustering of DKM on the two circles dataset (left) and the corresponding embedding (right). Ground truth clusters are highlighted in red and grey, and cluster centers as black crosses. Predicted clusters colored in yellow and green.

Table 1: Mean NMI train and test values for the best configurations of each considered method on MNIST and FashionMNIST.

Model	MNIST		fMNIST	
	Train	Test	Train	Test
RSM*	68.69	69.99	59.93	58.62
SC	74.10	–	64.04	–
SN	69.00	55.14	61.36	45.08
DKM	82.53	83.76	55.42	55.24

* indicates proposed model

have been proposed to solve this problem [4, 5]. The most straightforward option is to come up with robust kernel function. SpectralNet (SN) tried to learn it using separate Siamese Network [6].

We propose Robust Spectral Map (RSM), that is less sensitive to the choice of the affinity matrix RSM optimizes the objective $\mathcal{L} = \mathcal{L}_{RC} + \lambda\mathcal{L}_{pen}$, where $\mathcal{L}_{pen} = ||Z^T Z - I||_F^2$, and makes two main modifications compared to SpectralNet.

1. We compute the affinity matrix S depending on each batch as a randomized graph kernel, where the bandwidth of the kernel serves as a randomly sampled parameter from a predefined range. The variance in the batch-kernels is expected to maintain information about the underlying manifold while filtering out noisy connections throughout the training process.
2. We replace SN's orthogonal layer with the penalty term $\mathcal{L}_{pen}$ to enforce the orthogonal property of the learned space. As a result, optimization is simple as the whole model is jointly optimized by stochastic gradient descent.

We show experimentally that the proposed model performs consistently in terms of Normalized Mutual Information (NMI) compared to the closest competitors (see Table 1) - around 70% and 60% NMI on MNIST and FashionMNIST training subsets respectively. In addition, RSM manages to generalize results to unseen data and shows better preservation of orthogonal properties than other deep learning competitors. However, we found that the produced embedding space is less similar to top eigenvectors comparing to SpectralNet. Therefore, we conclude that clustering-friendly embeddings do not necessarily need to approximate eigendecomposition precisely.

A qualitative invspection of the mean images reveals that spectral-based models extract clusters as meaningful as those extracted by Deep K-Means though demonstrating a lower NMI score on the MNIST subset.

References

1. Fard, M., Thonet, T. & Gaussier, E. Deep k-means: Jointly clustering with k-means and learning representations. *Pattern Recognition Letters*. 138, pp. 185-192 (2020)
2. Hess, S., Duivesteijn, W., Honysz, P., & Morik, K.: The spectacl of nonconvex clustering: A spectral approach to density-based clustering. *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 3788-3795) (2019).
3. Von Luxburg, U. A tutorial on spectral clustering. *Statistics And Computing*. 17, pp. 395-416 (2007)
4. Yang, X., Deng, C., Zheng, F., Yan, J. & Liu, W. Deep spectral clustering using dual autoencoder network. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 4066-4075 (2019)
5. Affeldt, S., Labiod, L. & Nadif, M. Spectral clustering via ensemble deep autoencoder learning (SC-EDAE). *Pattern Recognition*. 108, pp. 107-522 (2020)
6. Shaham, U., Stanton, K., Li, H., Nadler, B., Basri, R. & Kluger, Y. SpectralNet: Spectral clustering using deep neural networks. *ArXiv Preprint ArXiv:1801.01587*. (2018)