SPECTRALNET: SPECTRAL CLUSTERING USING DEEP NEURAL NETWORKS-REPRODUCTION

Richard Cepka
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Slovakia, Bratislava, 842 48
cepka.richard@gmail.com

ABSTRACT

SpectralNet results reproduction (link: https://arxiv.org/pdf/1801.01587.pdf) on MNIST dataset. Our implementation can be found on https://github.com/CepkaR/SpectralNet_reconstruction.

1 Introduction

Spectral clustering (SC) is an unsupervised method, which is used in applications, such as: data clustering, community detection, image segmentation, parallel computing, etc. It became well-known after being published in [1], [3] and [4]. A great advantage of SC is that it does not require a certain cluster form and it can cluster also non-convex clusters. It has been shown that outperform other standard clustering methods, such as k-means, etc. Vanilla SC is based on the solution of the following relaxed optimization task:

$$\begin{aligned} \min_{Y \in \mathbb{R}^{n \times k}} \quad & tr(Y^T(D-W)Y) \\ \text{s.t.} \quad & Y^TY = I. \end{aligned} \tag{1}$$

The solution is the k smallest eigenvectors of matrix $D - W = L$ (Laplacian matrix), where the solution arises from the Rayleigh-Ritz theorem. As the solution of (1) is a solution of a relaxed problem, due to $\mathbb{R}^{n \times k}$, it is necessary to obtain a discrete solution that indicates the affiliation to the particular clusters. For a re-construction of the indicator vectors, we can use k-means. Thus, SC can be understood as a projection f from the data space $\mathcal{X}$ to the space $\mathcal{Y}$ and the subsequent clustering in this projected space. The space $\mathcal{Y}$ is the row space of the matrix Y, while the columns of Y consist of k (number of clusters) smallest eigenvectors of the Laplacian matrix.

SC is not easy to scale and a great disadvantage is that it cannot cluster out of sample data. The work [2] presents a SpectralNet which should do away with these problems. It is a neural network that can approximate projection f. It was shown that this method can generalize to new data and thanks to stochastic training it is easy to scale.

2 Methods

- **SpectralNet:** As SC arises from the solution constrained optimization task (1), in SpectralNet we will consider objective function from (1) as loss function, and the required constraint will be achieved by orthogonal layer mentioned in [2]. The weights are specially trained by Cholesky decomposition, such as $W_{ortho} = \sqrt{m}(R^{-1})^T$, where R is obtained from the orthogonal layer input. The SpectralNet is trained as an orthogonal step and standard gradient step respectively. The details are presented in [2] article.
- **Autoencoder:** Autoencoders (AE) are used for coding data into the code space where they acquire a better representation for clustering. In MNIST dataset, as claimed in the original paper [2], code space is obtained from pretrained AE [5]. In the case of Fashion-MNIST dataset we train our own vanilla AE.
- **Siamese network:** In SC, as well as in SpectralNet training it is necessary to set a matrix of similarities W. The most common approach is to set W by means of the k-NNG approach, where $w_{ij} = e^{-\frac{||x_i - x_j||^2}{2\sigma}}$ and σ is

a median of distances among the k nearest neighbours (σ_{med}). As k-NN is not a symmetric relation, matrix W, which has been set by k-NNG approach, is not symmetric either, thus matrix W will be symmetrized as $W \leftarrow \frac{W+W^T}{2}$. Euclidean distance can be replaced by Siamese network, which can better identify the structure in the data. In general Siamese network is trained by supervised fashion, that is problem, because our task is unsupervised. The authors of [2] have demonstrated that the Siamese network trained by the self-supervised approach, where proximity of vectors is approximated by k-NN method, will significantly improve the results of clustering.

3 Experiments

This part offers the reconstructed results on the MNIST dataset from the paper [2] and presents new results on the Fashion-MNIST dataset. The results will be compared with SC and k-means. All parameters have been left the same, as in article [2], except for the optimization algorithm. Instead of RMSprop we chose Adam, due to the instability in the orthogonal layer while calculating Cholesky decomposition.

Table 1: Experiments results.

	MNIST		Fashion-MNIST	
Algorithm	ACC	MNI	ACC	MNI
k-means	.5419	.4848	.4758	.5122
k-means (Siamese dist.)	.6640	.6912	**.5857**	.5886
k-means (code space)	.7710	.6836	.5652	.6035
k-means (code space, Siamese dist.)	**.9533**	**.8926**	.5850	**.6286**
SC	.6628	.7410	.5270	.6004
SC(Siamese dist.)	.7025	.7518	.4867	.6464
SC(code space)	.8057	.8431	**.6052**	.6753
SC(code space,Siamese dist.)	**.8253**	**.8773**	.5983	**.6854**
SpectralNet	.6088	.6584	.5161	.5863
SpectralNet (Siamese dist.)	.6469	.7631	.5639	.6246
SpectralNet (code space)	.8002	.8234	**.5903**	.6646
SpectralNet (code space, Siamese dist.)	**.8261**	**.8786**	.5883	**.6697**

The red marked results denote significant difference with original paper [2].

As seen in Table 1, SpectralNet thoroughly copied the SC results. All three algorithms, which we compared in our work, achieved better results in the code space and with the use of Siamese distance. The results, marked red in Table 1 represent much lower results than those in the original study [2]. And we have not managed to reproduce the state of art result on the MNIST dataset in NMI metrics. Thus we can say that we did not manage to achieve such a high efficiency growth after applying Siamese distance. For the given configurations SpectralNet (Siamese dist.) and SpectralNet (code space, Siamese dist.) we carried out the experiments three more time, but the results did not significantly differ from those presented in Table 1.

At the end we divided the MNIST dataset into a training set (60000) and testing set (10000), where for the testing set SpectralNet reached ACC: .6145 and MNI: .6710. Thus SpectralNet is able to both generalize to a new data and cluster out of sample data.

References

[1] Ng, A., Jordan, M., Weiss, Y.: *On spectral clustering: analysis and an algorithm*, In T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14. MIT Press, 2002

[2] Shaham, U., Stanton, K., Li, H., ...: *SpectralNet: Spectral Clustering using Deep Neural Networks*, International Conference on Learning Representations, 2018

[3] Shi, J., Malik, J.: *Normalized cuts and image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888-905, 2000

[4] von Luxburg, U.: *A Tutorial on Spectral Clustering*, Statistics and Computing 17 (4), 2007

[5] Zheng, Y., Tan, H., Tang, B., ...: *Variational deep embedding: A generative approach to clustering.* arXiv preprint arXiv:1611.05148, 2016