Abstract:

Recently, there has been increasing interest to leverage the competence of neural networks to analyze data. In particular, new clustering methods that employ deep embeddings have been presented.

In this talk, we depart from centroid-based models and suggest a new framework, called Clustering-driven deep embedding with PAirwise Constraints (CPAC), for non-parametric clustering using a neural network. We present a clustering-driven embedding based on a Siamese network that encourages pairs of data points to output similar representations in the latent space. Our pair-based model allows augmenting the information with labeled pairs to constitute a semi-supervised framework. Our approach is based on analyzing the losses associated with each pair to refine the set of constraints.

We show that clustering performance increases when using this scheme, even with a limited amount of user queries.

We present state-of-the-art results on different types of datasets and compare our performance to parametric and non-parametric techniques.