Adv Big Data

Professor Gu

Xueyan Bai

The Summary of Deep Learning Approach to Clustering

The article is trying to build a new deep clustering network that leverages the discriminative power of information-theoretic divergence measures, which have been shown to be effective in traditional clustering. The proposal of the article is to set up a novel loss function that incorporates geometric regularization constraints to avoid degenerate structures of the resulting clustering partition.

There are a few methods have been proposed to exploit deep learning architectures for clustering, including CatGAN and AAE, which are based on the idea of adversarial networks, and DEC, which is the most closely relative method to the approach of the article. DEC is founded on traditional clustering approaches. The first step is to analysis data based on Student t-distribution, and then, to optimize the parameters based on soft assignments to target distribution. The benefit of the DEC is that the method is performed by minimizing KL divergence. However, the disadvantage is that the effectiveness of DEC requires explicit feature design to handle complex image data.

The fundamental part of this project is to set up the loss function that allow the network to learn from gradient descent the intrinsic cluster structure in the input data. The propose of the model is to compute the dissimilarity between PDFs to set up each cluster and optimize clusters by maximizing the divergence between their PDFs. There are several different formulations of divergence measures exist, but the article focuses on Cauchy-Schwarz (CS) divergence, which can be used in multi-cluster problems by averaging the pairwise divergence over all pairs of cluster PDFs. Consider distinct PDFs, the CS divergence is defined as:

Since are unknown, the article follows the data-driven approach and approximate the PDFs using a Parzen window estimator, configured with a Gaussian kernel having bandwidth . Define so that the element

Where is the Euclidean distance between data point . The article assumes that there is a n x k cluster assignment matrix representing the crisp cluster assignment of data point q to cluster . Thus:

where

In order to enforce closeness to a corner of the simplex, the article define an additional term for the loss function:

*.*

Hence, the final clustering loss function is:

Where:

The article used MNIST dataset, SEALS dataset and Reuters dataset to test the algorithm. To evaluate the partition quality obtained after training, there are two different supervised measures to consider: Normalized mutual information (NMI) and unsupervised clustering accuracy (ACC) .

Define below as:

where I and H denote mutual information and entropy functions.

Comparing the baseline algorithms: k-NN approach (ITC-kNN) and Deep Embedded Clustering (SEC) and Local Discriminant Models and Global Integration (LDMGI) with the article algorithms DDC and DDC-VOTE (DDC-VOTE is to run 20 times of DDC and report the results of a voting scheme of the top three runs according to the unsupervised loss function), the result shows that ITC algorithm could not be evaluated on MNIST and large datasets in general. This shows an important advantage of DDC regards to previous clustering approaches based on the CS divergence. Except the SEALS-3 dataset, DDC-VOTE has better performance than DDC.