A PROPOSAL OF DEEP FUZZY CLUSTERING BY MEANS OF THE SIMULTANEOUS APPROACH

Claudia Rampichini [1] and Maria Brigida Ferraro[1]

[1] Department of Statistical Sciences, University of Rome "La Sapienza", (e-mail: claudia.rampichini@uniroma1.it, mariabrigida.ferraro@uniroma1.it)

ABSTRACT: Classical clustering methods may suffer from the presence of high dimensional or complex data. In this scenario, deep clustering can be useful to overcome such problems. The main idea is to use a neural network to reduce the input's complexity and apply a clustering algorithm to the reduced space. Our method consists in combining a neural network with the fuzzy *k*-means clustering algorithm. In particular, the proposal links the encoder part of an autoencoder neural network to a new layer, in which the membership degree values are calculated, and jointly optimizes the method by minimizing the fuzzy *k*-means objective function. Furthermore, to avoid the problem of collapsing centers, a penalization term is added. The adequacy of the proposal is evaluated by means of benchmark datasets.

KEYWORDS: deep clustering, neural networks, fuzzy *k*-means.

1 Introduction and background

Recent improvements in deep learning techniques have led to a new research field called deep clustering that shows new opportunities for conventional clustering to overcome problems with high-dimensional data. The idea of deep clustering is to learn latent features of training data using a deep neural network (DNN) and apply clustering methods to the resulting data representation. There exist two different deep clustering approaches: sequential and simultaneous. In the former, clustering algorithms are applied to the learned DNN representation, while in the latter, deep representation learning and clustering objectives are jointly optimized. Clustering approaches that are combined with deep learning models include *k*-means, graph clustering, spectral clustering, Gaussian mixture model, and many others, however, few studies focus on deep fuzzy clustering. One of the most famous models in the simultaneous approach is the deep embedded clustering method (DEC) that was proposed by Xie *et al.* (2016). This method simultaneously learns feature representations with stacked autoencoders and cluster assignments with soft *k*-means, minimizing a joint loss function. Later, some more complex deep fuzzy clustering methods have been proposed. The main differences are in the use of convolutional networks and more complex structures for loss functions (see, for example, Feng *et al.*, 2020, and Zhang *et al.*, 2020).

Starting by considering the problem of clustering a set of n points $\{\mathbf{x}_i \in X\}_{i=i}^{n}$ into k clusters, each represented by a centroid $\boldsymbol{\mu}_g$, $g = 1,...,k$, the DEC model consists in transforming the input data by a non-linear mapping $f_\theta : X \to Z$, where θ are learnable parameters and Z is the latent feature space where clustering is performed. Moreover, the Kullback–Leibler (KL) divergence loss between a centroid-based probability distribution and an auxiliary target distribution is used as the objective function:

$$L = KL(P||Q) = \sum_{i=1}^{n} \sum_{g=1}^{k} p_{ig} log\left(\frac{p_{ig}}{q_{ig}}\right). \tag{1}$$

In (1), q_{ig} is a Student's t-distribution used as a kernel to measure the similarity between embedded point $\mathbf{z}_i$ and centroid $\boldsymbol{\mu}_g$:

$$q_{ig} = \frac{(1+||\mathbf{z}_i - \boldsymbol{\mu}_g||^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{g'}(1+||\mathbf{z}_i - \boldsymbol{\mu}_{g'}||^2/\alpha)^{-\frac{\alpha+1}{2}}} \tag{2}$$

and it can be interpreted as the probability of assigning sample i to cluster g. Moreover, $\mathbf{z}_i = f_\theta(\mathbf{x}_i) \in Z$ corresponds to $\mathbf{x}_i \in X$ after embedding, α represents the degrees of freedom of the Student's t distribution. The auxiliary target distribution p_{ig} is calculated as the ratio between $\frac{q_{ig}^2}{f_g}$ and $\sum_{g'=1}^{k'} \frac{q_{ig'}^2}{f_{g'}}$ where $f_g = \sum_{i=1}^{n} q_{ig}$ are soft cluster frequencies; k-means is used only to initialize cluster centers. Their final model consists of the encoder part of the DNN and an additional layer in which the probability of assigning sample i to cluster g is calculated.

Starting from this model, we propose a simultaneous deep fuzzy clustering method in which the fuzzy k-means algorithm is involved. The idea for this proposal stems from noticing that the use of fuzzy clustering algorithms is often related to the initialization of cluster centers only, moreover, few works in the literature deal with these clustering algorithms. Additionally, we test our method on images improving upon traditional clustering methods which often poorly cluster this kind of data. The main reason is the difficulty of obtaining reliable similarity measures in high-dimensionality space but deep clustering methods have shown impressive performance in image clustering tasks.

2 Proposed method

The main idea is to replace the KL divergence loss with the fuzzy k-means objective function, hence

$$\arg\min_{U,C} \sum_{i=1}^{n} \sum_{g=1}^{k} u_{ig}^{m} \|\mathbf{z}_i - \boldsymbol{\mu}_g\|^2, \tag{3}$$

$$\text{s.t. } u_{ig} \in [0,1], \quad i = 1,...,n \quad and \quad g = 1,...,k; \quad \sum_{g=1}^{k} u_{ig} = 1, \quad i = 1,...,n.$$

Similar to the work of Xie *et al.* (2016), we create a deep autoencoder neural network and keep only the encoder part. Then we link a new layer to this part which computes the membership degrees values u_{ig} as follows

$$u_{ig} = \frac{1}{\sum_{j=1}^{k} \left(\frac{\|\mathbf{z}_i - \boldsymbol{\mu}_g\|}{\|\mathbf{z}_i - \boldsymbol{\mu}_j\|} \right)^{\frac{2}{m-1}}}. \tag{4}$$

In this way, by minimizing the fuzzy k-means loss function, we jointly optimize the cluster centers and DNN parameters. Since in the optimization process, the encoder part may lead, in an attempt to reduce the initial data, to the collapse of all the points into a single cluster, we introduce a penalization term, which is the absolute value of the sum of the pairwise differences. Additionally, to ensure that the two terms are on the same scale, we normalized them by dividing the first term by the product of the number of training examples used in one iteration hence batch size (b) and the number of cluster centers (k), and the second term by the product of the batch size and itself. Hence the new loss function takes the following form

$$\arg\min_{C,U} \frac{1}{bk} \sum_{i=1}^{n} \sum_{g=1}^{k} u_{ig}^{m} \|\mathbf{z}_i - \boldsymbol{\mu}_g\|^2 - \frac{1}{b^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{u}_{ig}^{m} |\mathbf{z}_i - \mathbf{z}_j|. \tag{5}$$

3 Results

We evaluate the proposed method on different benchmark datasets: Mnist, Fashion-Mnist and Cifar10. The first dataset consists of 70.000 black and white images of handwritten digits of 28× 28 pixel size, the second is a dataset of 70.000 Zalando's article images of 28× 28 pixel size and the last consists of 60.000 different colour images of 32×32 pixel size. The accuracy results achieved by the standard k-means, k-means in the embedding space (AE+k-means), DEC and our method are reported in Table 1.

Table 1. *Comparison of the accuracy level achieved by different methods on Mnist, Fashion-Mnist and Cifar10 datasets.*

Method	Mnist	Fashion-Mnist	Cifar10
k-means	53.5	47.4	22.9
AE+k-means	81.8	57.9	80.1
DEC	84.3	51.7	30.1
Our method	93.4	62.3	31.3

The results show the potential of the proposed method. In particular, on the Mnist dataset, we achieve an accuracy of 93.4% against 53.5% obtained with k-means and 84.3% with DEC; also on the Fashion-Mnist the accuracy of our method is higher than the others. On the Cifar10, the accuracy of our proposal is in line with the value reached by DEC but far from that of AE+k-means; this is probably related to the more complex dataset with colour images.

4 Concluding remarks

The new deep clustering method jointly learns feature representations with a deep autoencoder neural network and clusters assignments with fuzzy k-means by minimizing a loss function constructed in accordance with the chosen fuzzy algorithm. The results show margins of improvement with respect to the classical clustering methods and DEC model. Our future research will focus on the study of the different segmentation techniques for colour images.

References

FENG Q., CHEN L., CHEN P.C.L., & L., GUO. 2020. Deep Fuzzy Clustering-A Representation Learning Approach. *IEEE Transaction on Fuzzy Systems.*, **28**.

XIE, J., GIRSHICK R., & FARHADI, A. 2016. Unsupervised Deep Embedding for Clustering Analysis. *Proc. 33rd International Conference on Machine Learning.*, **48**, 478–487.

ZHANG R., LI X., ZHANG H., & F., NI. 2020. Deep Fuzzy K-Means With Adaptive Loss and Entropy Regularization. *IEEE Transaction on Fuzzy Systems.*, **28**.