Breast Ultrasound Images Clustering Analysis Using Deep Clustering Method

Cheng Huang and Jinrong Cui(✉)

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
tweety1028@163.com

Abstract. With the expanding volume of medical data, the supervised algorithms have the difficulty in obtaining labels of the medical data. To address this problem, many researchers have introduced unsupervised learning algorithms to analyze medical data and to mine the potential information among different samples. Simultaneously, with the brisk growing of deep learning, the association of unsupervised algorithms and deep learning is becoming further intimate. In this paper, we will use clustering algorithm which is commonly used in unsupervised learning to analyze breast ultrasound images. The Improved Deep Embedded Clustering (IDEC) algorithm will be introduced in this paper, which is broad used in the current deep clustering methods to do clustering analysis. Meanwhile we have added the graph embedding strategies in IDEC algorithm to enhance the clustering performance. The algorithm includes two phases: pre-training phase to capture hidden features and fine- tuning phase to enhance clustering result. We have exerted our algorithm on breast ultrasound images dataset to perform clustering analysis, and the experiment results indicate that the clustering performance of this algorithm is better than other traditional clustering algorithms.

Keywords: Unsupervised learning · Clustering analysis · Deep clustering · Medical data

1 Introduction

To derive the hidden information of image, supervised learning algorithms based on deep learning exert fundamental algorithms and large-scale labeled materials for a range of image analysis [1]. Nevertheless, In the area of medical image analysis, manual description of medical image requires a series of medical and computer knowledge. In recent years, the medical image depository continues to enlarge. But there has not been an equivalent development in the number of labeled samples. Furthermore, the sources of medical image samples are extremely broad such as organs or human tissues images on X-ray and a range of images on Magnetic Resonance (MR), Faced with the variety of complex medical materials, the learning cost is a primary problem too.

S. Wang et al. (Eds.): IoTCare 2021, LNICST 415, pp. 321–330, 2022.
https://doi.org/10.1007/978-3-030-94182-6_23

In several paper, a series of approaches have been used to address these challenges including method based on one-shot learning. This algorithm uses some relatively small-scale labeled materials to train and deep model. Introducing transfer learning exert images from others general dataset as a feature learning phase [2]. Some approaches exert unsupervised learning algorithm to capture the feature of the medical data, where the purpose is to acquire medical image information. However, only exerting unsupervised learning algorithm to capture the hidden information of the medical data is not enough, and more unsupervised learning methods should be introduced to obtain better performance. Alternatively, the clustering method has been introduced to discriminate dissimilar samples. Clustering method plays an essential role in unsupervised learning methods. In image segmentation [3–5] and unsupervised image classification [6–9] areas, many clustering methods and models has been used.

Normally, medical image materials have the characteristics of large quantity, multiple data sources and complex picture structure. Breast ultrasound image is one of the typical image types. For breast ultrasound images, many researches have done a series of analysis [10–12]. Our purpose is to introduce an unsupervised learning method to obtain the main features from breast ultrasound images. Then exerting a clustering algorithm distinguishes these images and maximize the separation of images between different categories. Therefore, we introduced an improved Deep Embedding Clustering (IDEC) [13] algorithm that combines a deep model and a clustering algorithm. In order to enhance clustering performance, we join graph embedding strategy to maintain the neighbor relationship between samples. Meanwhile, we experimented on a breast ultrasound images dataset to verify the accuracy of proposed method. Final we have compared the proposed algorithm with other traditional clustering algorithms. Experimental results show that proposed algorithm performs better than other traditional Clustering algorithms. This may provide new thoughts for the area of medical image analysis.

2 Relate Works

2.1 Deep Clustering

Recently, a series of clustering methods have been published, and which introduces the deep models to learn main features representations and uses these hidden representations to enhance clustering performance. Deep clustering network (DCN) [14] was proposed in a framework which include two part: dimensional reduction and K-means clustering. Deep embedded clustering (DEC) [15] was introduced in a deep stacked autoencoder (SAE) [16] to optimize the feature extraction model by pre-train dataset, and then a novel clustering objective function was proposed in [17] and which is based on KL divergence. DEC optimized through iteratively self-training. These Clustering methods have displayed superiority beyond the tradition algorithms. While, these approaches are rarely exerted in the area of medical data analysis. This article applies the deep clustering method to medical image analysis, which can achieve better clustering performance than traditional methods, and can be suitable adapted to large-scale datasets.

2.2 AutoEncoder

Autoencoder is a neural network model of self-supervised learning, which include encodes and decodes. Samples were inputted into encoder and were outputted from decoder, the training strategies was introduced by comparing the similarity between input samples and decoded samples. The principal part of the stacked autoencoder is fully connected neural networks, which has a hidden layer z to capture the key features of the input sample. the function of encoder and decoder is $\overline{x} = F_E(F_D(x))$.

Where $F_E()$ indicates encoder mapping to learn hidden features z,$F_D()$ indicates decoder mapping to reconstruct input sample x, $\overline{x}$ indicates reconstruct sample for x.

Following are two general applications of autoencoders.

Under-complete autoencoder. The purpose of undercomplete autoencoder is to capture the key information of input sample, thus, it designs networks that keep the dimensions of hidden layer less than input layer.

Denoising autoencoder. The purpose of denoising autoencoder is to recover the input sample that destroyed by noise. The denoising autoencoder has better anti-noise performance than autoencoder, and the recovered images look clearer than autoencoder.

In this paper, under-complete autoencoder is introduced and the reconstruction loss function is used so that the autoencoder can learn the essential features of the input sample in pre-train phase.

2.3 Improved Deep Embedding Clustering

Improved Deep Embedded Clustering (IDEC) can initialize the hidden features with pre-training phase which using encoder and decoder. The decoder is removed after pre-training phase. In fine-tuning phase the remaining encoder is used to minimize objective function:

$$L{=}L_r{+}L_c \tag{1}$$

where L_r and L_c are reconstruction loss and clustering loss defined by Mean Squared Error (MSE) and KL divergence:

$$L_r{=}\sum_{i=1}^{n}\|x_i - F_D(z_i)\|_2^2 \tag{2}$$

$$L_c = KL(P||Q) = \sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{3}$$

In formula (2), L_r is so-call reconstruction loss which means the similarity between input sample and decoded sample. In formula (3), q_{ij} indicate the probability that the sample x_i can be clustered as clustering group j calculated by Student's t-distribution:

$$q_{ij} = \frac{(1+\|z_i-\mu_i\|^2)^{-1}}{\sum_j(1+\|z_i-\mu_i\|^2)^{-1}} \tag{4}$$

And p_{ij} in (2) indicate the target distribution derived by q_{ij}:

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{n} q_{ij}}{\sum_{j=1}^{c} (q_{ij}^2 / \sum_{i=1}^{n} q_{ij})} \tag{5}$$

3 The Proposed Method

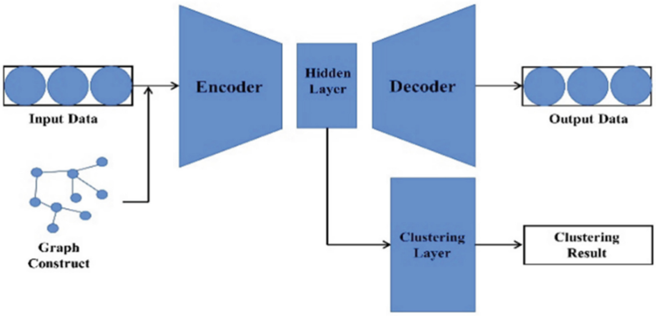

Fig. 1. The net construction of improved deep embedding clustering with graph embedding

Definition. Given a dataset {X} include n unlabel samples. The aim of clustering is to cluster n unlabel samples into k categories. In Fig. 1 we have gave the net construction of the improved deep embedding clustering with graph embedding. Generally, the dimension of samples is high-dimension. Thus, IDEC algorithm introduces an under-complete autoencoder to reduce the dimension of high-dimensional samples. Meanwhile, the key information and features of high-dimensional samples are also learned during the phase of dimension reduction. The hidden features are defined z. Using hidden features z can reduce the overhead of the clustering phase.

3.1 Graph Embedding Strategy

Inspired by a hypothesis in the area of subspace learning that if a couple of samples in the high-dimension space are near, then the low-dimensional features extracted from them should be nearby in the low-dimension space [18]. To satisfy this intention, we institute neighbor relationships by following neighbor graph strategy:

$$\min \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| z_i - z_j \right\|_2^2 G_{i,j} \tag{6}$$

where $G \in R^{n \times n}$ indicate the matrix of nearest neighbor graph relationship from { X} . The graph matrix G computed as follows:

$$G_{i,j} = \begin{cases} 1, \ x_i \in \Delta(x_j) \ or \ x_i \in \Delta(x_j) \\ 0, \qquad\qquad otherwise \end{cases} \tag{7}$$

where $\Delta(x_i)$ indicates the nearest neighbor of sample x_i under the Euclidean distance metric.

We impose this graph strategy on both pre-training and fine-tunning layers. The loss of pre-training layer is defined as follow:

$$L = L_r + \gamma L_g \tag{8}$$

where L_r and L_g defined as formulas (2) and (6) respectively, γ is a trade-off coefficient.

3.2 Fine-Tunning Layer and Initialization

After pre-training phase, encoder can capture hidden features of input samples. However, the hidden features that are only training trained by pre-training phase, and may not perform so well on the clustering task. Thus, we save the weight of model trained in the pre-training phase and remove the decoder part from the autoencoder. Then encoder part is connected with fine-tunning layer. Fine-tunning layer includes tunning and clustering part. The purpose of fine-tunning layer is to train an encoder that is able to which can capture features of clustering while obtaining enhanced clustering performance. Before training of fine-tunning layer, we initialize clustering centers by employing k-means. The loss of fine-tunning layer includes three parts: reconstruction loss, clustering loss and graph embedding loss as follow:

$$L = L_r + \gamma_1 L_c + \gamma_2 L_g \tag{9}$$

where L_r,L_c and L_g defined as same as formula (8) and (6).γ_1 and γ_2 is the trade-off coefficient of clustering loss and graph embedding loss respectively.

3.3 Parameter Updating

In this paper, there are three types of parameters to update: the weights of autoencoder, cluster centers and the parameter of target distribution. The clustering centers are updated as parameters, which is convenient to further extract the features suitable for clustering. Meanwhile, updating target distribution make the clustering group distribution more obvious. The whole method is displayed in Algorithm 1.

3.4 Method Summary

In general, inheriting the excellent properties of IDEC and graph strategy, Our proposed method has the potential to obtain the hidden representation of input samples while preserving the neighbor relationship among samples. By this way, the performance of clustering is further improved.

Algorithm 1: Improved Deep Embedding Clustering with Graph Embedding

Input: Data $\{X\}$; Parameter α; Maximum iterations: $Maxiter$
Update interval: T; Batch size: B_s; Stopping threshold ξ;
Output: Clustering result: $label$.
Initialization: Construct k-nearest neighbor graphs as formula (4); Initialize weights of autoencoder, Cluster centers μ according to Section 3.2
For $iter \epsilon (0,1,2 \ldots Maxiter)$ **do**
  if $iter \% T = 0$ **then**
    Compute hidden features points of input data $z_i = F_E(x_i)$
    Compute soft target P, Q according to (4), (5)
    Reserve last labels $s_{last} = s_i$; Update predicted labels
    $s_i = arg \max_j q_{ij}$
    if $sum(s_i \neq s)/n$ **then**
    Stop
End for

4 Experiment

4.1 Dataset

We impose the proposed method on breast ultrasound images which is published in MedMNIST [19]. The image dataset includes 780 ultrasound gray images of breast. The shape of origin images is 28 × 28 pixel. In this experiment, the shape of images will be converted into a 784-dimensional vector. The specific of this dataset is shown in Fig. 2 and Table 1.

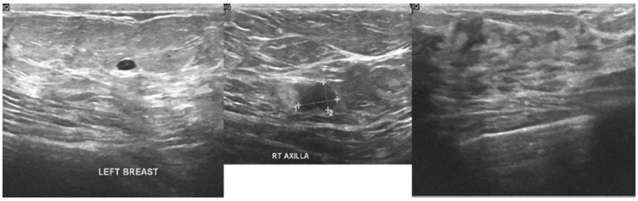

Fig. 2. Data set sample display

Table 1. The detail of BreastMNIST dataset

Name	Example number	Classes	Dimension
BreastMNIST	780	2	784

4.2 Experiment Work

Compared Methods. To demonstrate the effectiveness of our proposed algorithm, we compare our algorithm with a series of baseline. The mainly compared method is IDEC and DEC. When the reconstruction loss remains at zero, DEC can be regarded as a special situation of IDEC. The AE + k-means which means a two-stage processing includes pre-training autoencoder phase which like IDEC pre-training phase and performing k-means after pre-training phase.

Parameters Setting. The encoder and decoder network are combined into an under-complete autoencoder which is stacked by four layers fully connected neural network with neural numbers [a, 500, 500, 1000, 10] and [10, 1000, 500, 500, a], where a means the dimension of input samples. We introduce nonlinearity function 'ReLU' [20] as the activation function for all fully connected layers. In the pre-training phase and the fine-tuning phase, we use the optimizer 'Adam' [15] with learning rate $\lambda = 0.0003$ at pre-training phase and $\lambda = 0.0005$ at fine-tunning phase (We did experiments with learning rates ranging from 0.0001 to 0.0005 on fine-tunning phase) to optimization. We also use the coefficient of clustering loss in IDEC setting $\gamma_1 = 0.1$ and set the coefficient of graph embedding loss $\gamma = 0.001$,$\gamma_2 = 0.001$ in pre-training phase and fine-tunning phase. The threshold of stop training is assigned to $\delta = 0.1\%$.

Evaluation Metric. We introduce unsupervised clustering accuracy (ACC) as evaluation metric [13].

4.3 Experiment Result

Experimental results on the breast ultrasound dataset are demonstrated in Tables 2. In order to ensure the validity of the experimental results, we run pre-training 400 epochs with 256 batch size among compared methods which require pre-training phase. After 400 epochs pre-training phase, fine-tunning phase also keep 256 batch size. In order

Table 2. Comparison of experiment result in accuracy (%)

Methods	ACC
k-means	55.00%
AE+ k-means	60.90%
DEC	62.05%
IDEC	63.08%
IDEC with graph embedding (Ours)	**72.05%**

to better describe the influence of super parameters on the experiment, we adjusted the learning rate from 0.001 to 0.005 and the graph loss coefficient from 0.001 to 0.005, respectively.

Table 3. Adjusting different learning rate test accuracy (%). Fixing learning rate at 0.0005 test accuracy (%) under different graph embedding loss coefficients.

Learning rate	ACC	Graph loss	ACC
0.0001	56.54%	0.001	70.90%
0.0002	55.26%	0.002	71.28%
0.0003	69.36%	0.003	71.95%
0.0004	68.72%	0.004	72.05%
0.0005	70.90%	0.005	71.54%

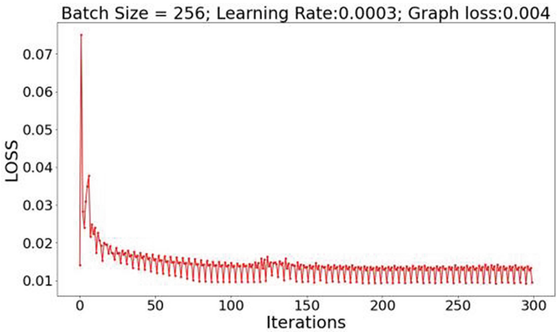

Fig. 3. Losses during fine-tunning phase on BreastMNIST. Note: Learning rate is 0.0005 and graph embedding loss is 0.004

We can notice the following results from Table 2, 3 and Fig. 3, 4: 1) IDEC with graph embedding notably outperforms the other compared methods on the breast ultrasound databases. 2) By modifying the learning rate, ACC has increased significantly. By modifying graph embedding loss coefficient, ACC kept a relatively stable level. 3) The

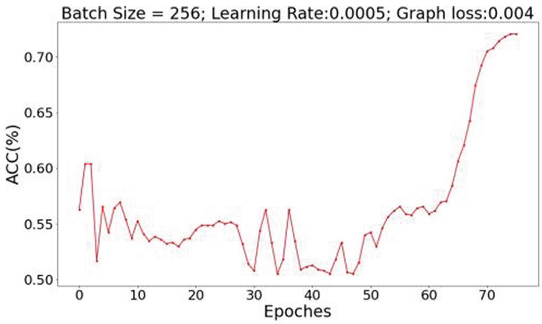

Fig. 4. Accuracies during fine-tunning phase on BreastMNIST. Note: Learning rate is 0.0005 and graph embedding loss is 0.004

loss of IDEC with graph embedding shows an oscillating trend which constantly oscillate between 0.02 and 0.01. 4) The clustering accuracy of IDEC with graph embedding fluctuates at first, and gradually rises after training for multiple epochs, and reaching the threshold at 72.05%.

5 Conclusion

In this paper, we proposed a novel method called IDEC with graph embedding. This method can be used for cluster of breast ultrasound images. Based on antoencoder and KL divergence, method introduces the graph embedding strategy to reserve the neighbor relationship of input samples. The excellent performance of IDEC with graph embedding are validated on BreastMNIST dataset. The IDEC with graph embedding may contribute extra insight for the unsupervised learning of medical image analysis.

Acknowledgment. This work supported by the Opening Project of Guangdong Province Key Laboratory of Computational Science at the Sun Yat-Sen Universty 2021011.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE, Piscataway, NJ (2016)
2. Ahn, E., Kim, J., Kumar, A., Fulham, M., Feng, D.: Sparsity-based convolutional kernel network for unsupervised medical image analysis. arXiv preprint arXiv:1807.05648 (2018)
3. Vishnuvarthanan, G., Rajasekaran, M.P., Subbaraj, P., Vishnuvarthanan, A.: An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images. Appl. Soft Comput. **38**, 190–212 (2016)
4. Moriya, T., et al.: Unsupervised segmentation of 3D medical images based on clustering and deep representation learning. In: Gimi, B., Krol, A. Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, p. 1057820. SPIE, Bellingham (2018)

5. Ahn, E., et al.: Saliency-based lesion segmentation via background detection in Dermoscopic images. IEEE J. Biomed. Health Inform. **21**(6), 1685–1693 (2017)
6. Moriya, T., et al.: Unsupervised pathology image segmentation using representation learning with spherical k-means. In: Gurcan,M N., Tomaszewski J E. Medical Imaging 2018: Digital Pathology, p. 1058111. SPIE, Bellingham (2018)
7. Zhang, Z., Liu, L., Shen, F., Shen, H.T., Shao, L.: Binary multi-view clustering. IEEE Trans. Pattern Anal. Mach. Intell. **41**(7), 1774–1782 (2018)
8. Zhang, Z., et al.: Highly- economized multi-view binary compression for scalable image clustering. In: 15th European Conference on Computer Vision, pp. 717–732. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01258-8_44
9. Gao, G., Yang, J., Jing, X.Y., Shen, F., Yang, W., Yue, D.: Learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn. **66**, 129–143 (2017)
10. Rahmawaty, M., Nugroho, H.A., Triyani, Y., Ardiyanto, I., Soesanti, I.: Classification of breast ultrasound images based on texture analysis. In: 1st International Conference on Biomedical Engineering , pp. 1–6. IEEE, Piscataway, NJ (2016)
11. Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L.: Medical breast ultrasound image segmentation by machine learning. Ultrasonics **91**, 1–9 (2019)
12. Moon, W.K., Lee, Y.W., Ke, H.H., Lee, S.H., Huang, C.S., Chang, R.F.: Computer- aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput. Methods Programs Biomed. **190**, 105361 (2020)
13. Guo, X., Gao, L., Liu, X., Yin, J.: Improved deep embedded clustering with local structure preservation. In: International Joint Conference on Artificial Intelligence, pp. 1753–1759. Morgan Kaufmann, Burlington (2017)
14. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: European Conference on Computer Vision (ECCV), pp. 139–156. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01264-9_9
15. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning (ICML), pp. 740–749. ACM, New York (2016)
16. Suk, H.-I., Lee, S.-W., Shen, D.: Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct. Funct. **220**(2), 841–859 (2013). https://doi.org/10.1007/s00429-013-0687-3
17. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(86), 2579−2605 (2008)
18. Kang, Z., Peng, C., Cheng, Q.: Clustering with adaptive manifold structure learning. In: International Conference on Data Engineering (ICDE), pp.79–82 (2017)
19. Yang, J., Shi R., Ni B.: MedMNIST classification decathlon: a lightweight automl benchmark for medical image analysis. arXiv preprintarXiv:2010.14925 (2020)
20. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. J. Mach. Learn. Res. **15**, 315–323 (2011)