International Journal of Advanced Science and Technology
Vol. 29, No. 12s, (2020), pp. 2184-2189

Clustering Techniques in Medical Analysis using Deep Representations

Bh. Prashanthi[1], Ganta Sowjanya[2]

Gokaraju Rangaraju Institute of Engineering and Technology, Department of CSE, Hyderabad, India; E-mail: bhupathi.prashanthi@gmail.com
Gokaraju Rangaraju Institute of Engineering and Technology, Department of CSE, Hyderabad, India; E-mail: sowjanyaganta1742@gmail.com

Abstract

The advancement of image processing techniques in the area of clustering using traditional features is still not getting the human expected results. Image content plays a major role in most of the current realtime applications. Processing of images to get user requested analysis is one of the promising domains nowadays. The medical analysis is still in the infancy stage with this traditional image features. To overcome this current advancement in the area of deep learning uses deep features instead of traditional features. Several techniques were proposed using deep clustering in the medical field. Different applications are using different deep clustering techniques in this domain. This paper aims to address various deep clustering techniques used in medical analysis and also discuss the interesting insights which can be useful for current medical analysis.

Keywords: *Deep Clustering, Deep Features, Image Processing, Medical Analysis, Traditional Features.*

1. Introduction

Most of the current availability of a large amount of data using for analysis is available in an unstructured format and also it useful for medical analysis. Storing and analyzing such data is one of the most demanding areas of research. Efficient clustering techniques suitable for various medical applications is a challenging task nowadays. Finding user queries from large image databases is an area of power and speed. Understanding user requirements and explaining them to mean is a difficult and challenging task. A large number of images are uploaded and stored on the internet all the time to develop high-speed techniques quickly. Advanced visual effects and end-user content-based communication have opened up opportunities for information and services to be transformed through image content. Today, effectively capturing relevant questions from high-speed and accurate information is challenging. With advances in image processing, image plays an important role in many applications and real-time fields. Maintenance and recovery of active audio-visual techniques are of particular interest. Finding the right image from the database of organized images is a difficult task. In response to the demand for latency and retention of images obtained on the Internet at high speed and accuracy-related to user queries and requests.

The clustering of images as a means of a group of images has not yet been organized into critical groups according to the nature of the photograph without prior knowledge. There is ample room to unwind with the small space of the machine and the view of the old man. Extracting state-of-the-art images from a variety of heterogeneous image data is a daunting challenge in Content-Based Image Retrieval (CBIR) [1]. CBIR is still having trouble finding useful articles from important news sources. The two problems with capturing these elegant images are the search space and the elegance of the elegant space. Recent advances in deep clustering technologies have shown that it is appropriate to integrate images by integrating high-quality information from the center to the center space. Alternatively, embedded algorithms that use these integrators have more time and space to process memory and data. This study highlights several approaches in which automation is used to achieve the complexity of creating similar images and large files to achieve its vision.

1.1 Image Clustering

The idea of image clustering is to break the larger groupings of images so that images in the same group are useful in comparison. It provides unwanted physical tension with a large collection of images and thus finds many useful applications. For example, articles on web searches with forums and photoshops are very useful for users. Also, the performance of an image in a large image area can be greatly improved by regrouping images instead of confusing individual images. Image collection also aims to organize a large number of images into groups, so that the images in the middle group have the same meaning and the images in the middle group have different meanings. It is used to solve image segmentation, graphical rendering, search space minimization, and gap reduction issues. Visual symbols are placed in semantic concepts to create a psychological meaning that reduces the semantic gap. Research to address the lag in semantic production extends to semantic coherence and focuses on the direct mapping of visual signals to the semantic brain. The important and fast quality of the search is an important aspect of image collection and large image storage. In general, Team Four is a two-tier approach. First, from the pre-order level, each image is assigned to a group and a single group. At the second level, the image in each group is assigned to more than one group, indicating that a single image may have multiple semantic meanings [2]. The main purpose of obtaining images is to obtain information about the content. Users are interested in the content of group tags related to photos. The main purpose of image collection is to obtain a detailed group map of images so that the cluster group can provide reliable information for storage in the image collection. Image integration allows the implementation of efficient retrieval algorithms and the creation of easy-to-use image database environments. A good number of method combinations treat different distance functions to measure the similarity between the negative pixels in the data area. The distance-based paradigms in handmade construction make the process of opening up important groups.

1.2 Deep Image clustering

A deep neural network requires a lot of information. With advances in computational computing, previously managed methods have been successfully taught in large-scale data sets [3]. Researchers in this field have identified millions of already labeled samples and larger data sets such as ImageNet. Also, video-based composite data acts as a driver of deep neural networks for event-detected events. The marking remains a mandatory task in all cases to exploit the capabilities of controlled learning algorithms, but the task of tagging data is tedious, difficult, slow, and costly [4]. To overcome these brief opportunities, much work has focused on examining non-descriptive image data representations. Exercises managed with a predefined set of elements require expanding and generalizing the model for real-time data. Many semi-managed and uncontrolled learning methods have been introduced to reduce data labeling efforts. Untrained efforts to study unaffected and easy-to-use images for real-time data generalization requirements. Dimension Reduction (DR) is another widely used processing method for cultivating and using DNN to study free signals from data space to unmatched locations. These ideologies regard their DNN as existing processing space and are created from subsequent integration. The hypothesis is that data obtained from these DNNs can be obtained as clusters. DNNs who study the data are not required to submit acceptable data at a low level because the advertising promotion objectives are not taken into account during the study [4].

The effectiveness and stability of the clustering algorithm depend on the type of input provided by the user. Reduced size and in-depth training practices are widely used in Integration technology to convert high-quality data into an interactive environment for performance. Adopting a web map that allows the data to be changed by more closely related group processes, there is now a great need for more DNN. The use of integrated and standard versions instead of the full-scale version is a very complex and efficient use of data analysis [5]. The level of the built-in location is based on the storage location changed from the original location. Data objects in the hidden space must contain elements of the group. Finding a friendly version of the library and packing templates that can teach you difficult results is important for the team. The highest goal is to use a nonlinear distribution of objects that divide groups. Daily, increasing the size of data requires storage and fast access methods [6]. The major flowering systems are facing credit and memory charges. The arrangement of flower arrangements may not be necessary for large clusters. The power of real-time image-detection applications will prompt scientists to seek powerful methods such as distributed resources for rapid deployment.

2. Literature Survey

This chapter contains previous articles that suggested the idea of the proposed measure. The research articles in the literature speak to discuss and present past and present work on visual topics. Shortages in areas related to the proposed studies are identified as part of this study.

2.1 Traditional Image clustering

The traditional clustering paradigm is generally divided into hierarchical and fragment-based methods [7]. Partition-based approaches require cluster centers and use measurement relationships to execute all data to the exact point of the center. Agglomerative grouping is a hierarchical grouping algorithm that combines a large number of small groups that are formed in an initial group state [8]. After the training, the network is brought together and mapped from the input room into the cluster-friendly secret room [9]. Another popular category of methodology-based groups is the division into groups based on similarities and groups based on characteristics. The similarity-based grouping is based on the distance matrix, the closest neighboring matrix, which calculates the distance between each set of closest samples. The popular method for grouping similarities is spectral clustering (SC), in which the Laplace spectrum of the distance matrix is brought to a lower dimension level before grouping. K-means is a popular feature-based grouping method that tries to divide samples into groups K to reduce the sum of the groups of quadratic errors. K meanings support the idea of grouping information samples that are distributed to identical centers [10]. The predominant and widespread method for uncontrolled learning is grouping with K-Means. The K-Mean grouping is a simple approach in which input features are divided into different classes based on their similarities [10]. Several group algorithms in the literature can be separated from different perspectives, e.g. B. hierarchical, centroid-based, graphically based, sequential (temporary)regression models. and models of sub-cluster groups [2].

In other words, clustering has two modes, output, and discrimination. The algorithms such as the k-means from the GMM [11] are optimally introduced using geometry and estimation curves and illustrate the method in stages by distributing input data. Unlike the clustering algorithm, the discriminant method identifies the partitioning by separating the atmosphere as a data distribution. Information-theoretic [12], Maxmargin [13], and spectral graph [14] are standard algorithms for discriminating discrimination model. Because the model of discrimination has some differences in data distribution and classifications immediately, it is first proposed to produce better results than the corresponding increase. However, the training may be by removing or removing it in an unwanted local minimum [15].

2.2 Deep Clustering

To get accurate, positive, and accurate results from human experience and figurative masterwork, it is important to think positively. Low-level features, obtained from higher levels, have been shown to achieve optimum results and to use their closely related features in the study technology. Many different design tools have been developed to draw a wide variety of materials. The challenging nature of the 1980s and DNNs has been forgotten for some time because its implications include high rates of incarceration and easy access to the economy. Although the DNN has received considerable respect for the plan for adopting a more efficient approach to learning for so-called deep-seated warming (DBN) by Hinton et al. [16]. The algorithm solves the aforementioned shortcomings by providing a good introduction to a network with a formal and informal education system. As a result, different DNNs have demonstrated their competence and performance in studies and have emerged for ease of use in many areas [17].

Song et al. [18] developed a method for aggregating data based on automated code by plotting reconstruction errors and reducing the distance to the learning point between the data and the assembly point, respectively. In this optimization process, you update the data presentation and make it a team member. This method requires a trade-in resource for hackers to be included in the training list. Automatic adjustment is a link to a size reduction icon that attempts to read compressors that are printable by correctly performing reconstruction errors [19]. Recently, many personality differences have been shown to detect violence and severe symptoms. Automatic map editing rarely clings to hidden objects to encode abnormal early visual cortex [19]. Denoising Autoencoder converts the corrupt output to a proxy secret server and tries to recover the original parameter from a proxy server

[20]. Hinton et al. [21] developed a high-precision first-order algorithm for the most experienced. The Deep Embedding Network (DEN) [17] uses a multilayer exchange model to create horizontal models. As a result of the same situation, another work [4] uses a deep physical network as a function of constant change and maintains a class-based structure using ancient information.

3. Deep Clustering in Medical Analysis

Nowadays deep clustering is useful in formulating disease-causing patterns for various diseases.

3.1 Discussion and comparative analysis

This chapter aims to discuss various deep clustering approaches using in the medical analysis along with their performance. Table 1. Compares some of the prominent approaches for various diseases using different datasets.

Table 1. Comparison of various Deep clustering methods in Medical Analysis

S. No	Method	Dataset	Performance	Advantage
1	Deep unsupervised single cell clustering [22]	Single-cell RNA sequencing data	Accuracy=95	Analysis of a living cell atlas living organisms
2	VaDER [23]	362 *de novo* PD patients	Purity=0.9	Alzheimer's patients and Parkinson's disease patients in small groups characterized by clinical progression profiles.
3	Combination of unsupervised and supervised learning for predicting patient outcomes [24]	IrishHip Fracture Database (IHFD)	Precision= 82.1, Recall= 87.6, Accuracy= 81.1	Improves the accuracy of predicting patient outcomes.
4	Deep Clustering Networks [25]	Multi-center data of T1-weighted MRI volumes of 70 NASH patients	Sensitivity= 0.78, Specificity= 0.52, F-Score= 0.78.	To find predictive clusters with a stable ranking
5	Deep convolutional autoencoder Clustering [26]	UCI machine learning repository	Calinski-Harabasz Scores= 7703	To learn informative representations of health-related tweets.
6	Rough K-means clustering (RKM) [27]	Machine-learning repository and Kaggle	Accuracy=100 (Kidney) Accuracy=80.55(Diabetes)	To predict chronic diseases

			Accuracy=97.53(Breast Cancer)	

3.2 Research Scope in Medical Analysis using Deep Representations

This paper analyses and compares various deep clustering methods used in the medical analysis of different diseases. Based on the above analysis, there is a research scope for formulating clusters using compact deep representations [28] [29]. Current Deep learning models are occupying a large amount of storage space. To overcome this, compact model development without degrading the performance grasp high attention for researchers. Deep clustering models need to be deployed in mobile devices in the future for widespread use of them for commercial real-time applications.

4. Conclusion

This paper explores different clustering methods and discusses the advantages and disadvantages. Different clustering methods have been used to extract useful models and thus information from these different sources. The choice of data and clustering methods is an important task in medical analysis and requires knowledge of the field. The main focus of this deep clustering study is on how integration strategies are implemented in the field of medicine. Each set is ideal for other treatment plans. You need to choose the right combination that works for your job from the existing clustering methods using deep clustering. Future research directions for effective utilization of deep clustering models in medical analysis with their scope of research discussed above will be useful for finding solutions for existing medical applications.

References

1. Eakins, John, and Margaret Graham. "Content-based image retrieval." (1999).
2. Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. "Data clustering: a review." *ACM computing surveys (CSUR)* 31, no. 3 (1999): 264-323.
3. Yang, Jianwei, Devi Parikh, and Dhruv Batra. "Joint unsupervised learning of deep representations and image clusters." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5147-5156. 2016.
4. Yang, Bo, Xiao Fu, Nicholas D. Sidiropoulos, and Mingyi Hong. "Towards k-means-friendly spaces: Simultaneous deep learning and clustering." *arXiv preprint arXiv:1610.04794*(2016).
5. Donahue, Jeff, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. "Decaf: A deep convolutional activation feature for generic visual recognition." In *International conference on machine learning*, pp. 647-655. 2014.
6. Kolla Morarjee, T VenuGopal, “Semantic Image Clustering with Global Average Pooled Deep Convolutional Autoencoder”, Volume 8(4), pp.3492-3497, Helix2018, Web of Science, ESCI Indexed, ISSN: 2319-5592.DOI 10.29042/2018-3561-3566.
7. Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. "Data clustering: a review." *ACM computing surveys (CSUR)* 31, no. 3 (1999): 264-323.
8. Kurita, Takio. "An efficient agglomerative clustering algorithm using a heap." *Pattern Recognition* 24, no. 3 (1991): 205-209.
9. Aljalbout, Elie, Vladimir Golkov, Yawar Siddiqui, and Daniel Cremers. "Clustering with Deep Learning: Taxonomy and New Methods." *arXiv preprint arXiv:1801.07648* (2018)
10. He, Kaiming, Fang Wen, and Jian Sun. "K-means hashing: An affinity-preserving quantization method for learning binary compact codes." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2938-2945. 2013.
11. Fraley, Chris, and Adrian E. Raftery. *Model-based clustering, discriminant analysis*. density estimation. Technical Report 380, University of Washington, Department of Statistics, 2000.
12. Li, Haifeng, Keshu Zhang, and Tao Jiang. "Minimum entropy clustering and applications to gene expression analysis." In *Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings. 2004 IEEE*, pp. 142-151. IEEE, 2004.
13. Zhao, Bin, Fei Wang, and Changshui Zhang. "Efficient multiclass maximum margin

clustering." In *Proceedings of the 25th international conference on Machine learning*, pp. 1248-1255. ACM, 2008.
14. Ng, Andrew Y., Michael I. Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm." In *Advances in neural information processing systems*, pp. 849-856. 2002.
15. Dizaji, Kamran Ghasedi, Amirhossein Herandi, Cheng Deng, Weidong Cai, and Heng Huang. "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization." In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 5747-5756. IEEE, 2017.
16. Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." *Neural computation* 18, no. 7 (2006): 1527-1554.
17. Huang, Peihao, Yan Huang, Wei Wang, and Liang Wang. "Deep embedding network for clustering." In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 1532-1537. IEEE, 2014.
18. Song, Chunfeng, Feng Liu, Yongzhen Huang, Liang Wang, and Tieniu Tan. "Auto-encoder based data clustering." In *Iberoamerican Congress on Pattern Recognition*, pp. 117-124. Springer, Berlin, Heidelberg, 2013.
19. Lee, Honglak, Chaitanya Ekanadham, and Andrew Y. Ng. "Sparse deep belief net model for visual area V2." In *Advances in neural information processing systems*, pp. 873-880. 2008.
20. Vincent, Pascal, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders." In *Proceedings of the 25th international conference on Machine learning*, pp. 1096-1103. ACM, 2008.
21. Hinton, Geoffrey E. "Training products of experts by minimizing contrastive divergence." *Neural computation* 14, no. 8 (2002): 1771-1800.
22. Srinivasan, Suhas, Nathan T. Johnson, and Dmitry Korkin. "A Hybrid Deep Clustering Approach for Robust Cell Type Profiling Using Single-cell RNA-seq Data." *bioRxiv* (2019): 511626.
23. Marek, Kenneth, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey et al. "The parkinson progression marker initiative (PPMI)." *Progress in neurobiology* 95, no. 4 (2011): 629-635.
24. Elbattah, Mahmoud, and Owen Molloy. "Clustering-Aided approach for predicting patient outcomes with application to Elderly Healthcare in Ireland." In *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
25. Perkonigg, Matthias, Daniel Sobotka, Ahmed Ba-Ssalamah, and Georg Langs. "Unsupervised deep clustering for predictive texture pattern discovery in medical images." *arXiv preprint arXiv:2002.03721* (2020).
26. Gencoglu, Oguzhan. "Deep representation learning for clustering of health tweets." *arXiv preprint arXiv:1901.00439* (2018).
27. Aldhyani, Theyazn HH, Ali Saleh Alshebami, and Mohammed Y. Alzahrani. "Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms." *Journal of healthcare engineering* 2020 (2020).
28. Kolla Morarjee, T VenuGopal, “Binary Quantized Deep Embedding for Semantic Image Clustering”,Vol 119, No. 16, pp. 1171-1180,International Journal of Pure and Applied Mathematics(IJPAM) 2018, UGC Approved, Scopus Indexed, ISSN: 1311-8080.
29. Kolla Morarjee, T VenuGopal, “Product Quantized Deep Embedding for Semantic Image Clustering”, pp. 309-315, Vol. 6, Issue 3, International Journal of Research in Electronics and Computer Engineering (IJRECE)2018, UGC Approved, ISSN: 2348-2281.