



Multiple Frame CT Image Sequencing Big Data Batch Clustering Method

Xiao-yan Wang, Guo-hui Wei, Zheng-wei Gu,
Ming Li^(✉), and Jin-gang Ma

Shandong University of Traditional Chinese Medicine, Jinan, China
lmzhj@126.com

Abstract. CT image diagnosis technology had developed rapidly in China. In clinical testing, large-scale data gradually existed in the form of sequences. Data clustering was the integration of different substances in an image according to certain properties, but there were still many problems in the use of commonly used data clustering methods in medicine. The current sequencing big data clustering method analysis was still in the research stage and had very important significance. This paper proposed a large-scale batch clustering method based on multi-frame CT images, which was compared with the traditional clustering method, and hoped to provide assistance for clinical applications.

Keywords: Multi-frame CT · Sequencing data · Clustering method

1 Introduction

With the development of CT image technology, the application in clinical detection has been extensive in recent years. However, under actual circumstances, CT technology is affected by many objective conditions, and the needs of clinical testing must also be considered. Data imaging is somewhat defective. How to collect large quantities of data and reconstruct CT images completely and accurately is of great significance for clinical testing, both theoretical and practical [1, 2]. In recent years, although many relatively mature data clustering methods have been studied, the application is also very extensive. For example, filtered back projection data clustering method and iterative image sequencing data clustering method, but the calculation amount is very large, the clustering integrity of the data is low, and the imaging effect is not ideal [3, 4]. Therefore, neither of these methods can be applied to image sequencing big data batch clustering.

In recent years, the technical requirements for medical imaging are very high. In clinical testing, large quantities of data exist in the form of sequences [5]. Because the human body's organizational structure is quite complex, it is difficult to deal with some large-structure images with large data volume. Therefore, many methods of data clustering are not suitable for clinical imaging detection [6]. At present, many medical sequence imaging data clustering methods are in the exploration and research stage. In recent years, many medical scholars have studied data clustering [6], proposed a fuzzy clustering method (CFN), and tested it in medical sequence imaging, but the results

show that this method is very sensitive to the initial value of the data. In order to improve this deficiency, the researchers have proposed many ways to improve [8]. For example, based on the genetic algorithm of CFN, although this method can satisfy the shortcomings of the CFN method, the speed of clustering data is very slow [9]. In addition, researchers have proposed a density-based data clustering method, through the density distribution model as a prerequisite for data clustering. If the initial values set are small, the adjacent density is combined, but the accuracy of this method is relatively low. In order to improve the accuracy of the image, a data clustering method based on texture features has been proposed. This method is to generate a matrix according to the texture of the collected data [10]. The accuracy of this method is very high, but the matrix after data generation is sparse, so that there is an influence between each data, affecting the speed of clustering and imaging quality [11]. In general, the above several methods have their own problems, the difference in imaging will directly affect the judgment of certain conditions, not to be underestimated. By analyzing and summarizing the previous research results and experience, this paper proposes a DD-TV clustering method based on multi-frame CT image sequencing data, and achieves a good clustering effect.

In this paper, through the analysis of multi-frame CT image imaging principle, the multi-frame CT image clustering method is studied, and a new method of clustering data with incomplete projection data (DD-TV method) is proposed, which has a high value.

2 Application Analysis of Multi-frame CT Image in Data Batch Clustering

2.1 Principle Analysis of Multi-frame CT Imaging

Medical imaging technology has developed rapidly in clinical testing and is widely used, thus further enhancing the importance of image analysis and research [12]. To some extent, the accuracy of the image is an important basis for clinicians to judge the condition. Medical data clustering plays an important role in the analysis of images. It is also an imaging method under computer simulation technology. With the aid of computer simulation technology, the data is collected, analyzed and studied, and finally imaged [13].

Especially in recent years, with the increasing demand for imaging diagnostic applications, medical image data is more in clinical form in the form of large-scale sequence images [14]. At the same time, the anatomical structure of the human body is very complicated, which makes some traditional clustering methods have some inapplicability in processing such medical sequence images with very complicated structure and very large data volume [15, 16].

In the process of CT image reconstruction based on computer simulation technology, many researchers tend to explore the image reconstruction technology, and have formed a better reconstruction method [17]. For example, image reconstruction based on data analysis, simple operation, small amount of calculation, and fast image reconstruction speed are widely used in clinical detection [18]. However, this method requires very high data integrity. If the projected data is incomplete, it will cause artifacts, and if it is serious, it may cause partial image distortion [19, 20].

The principle of multi-frame CT imaging is shown in Fig. 1 below.

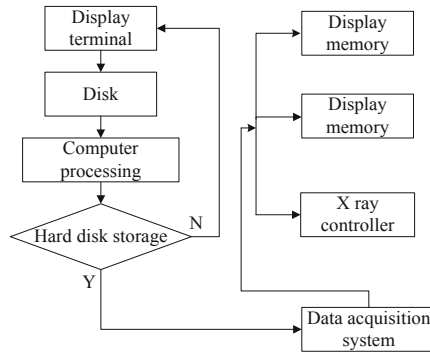


Fig. 1. Principle of multi-frame CT imaging

The basic principle of CT image is to scan a layer of a certain thickness of the inspection part of the human body with an X-ray beam. After the X-ray is received by the detector and transformed into visible light, it is transformed from an photoelectric converter into an electrical signal, which is then converted into a digital signal by an analog/digital converter and fed into a computer for processing. The processing of image formation is like dividing the selected plane into several cuboids with the same volume, which are called voxels. The X-ray attenuation coefficient or absorption coefficient of each voxel is calculated and arranged into a matrix, that is, a digital matrix. The digital matrix can be stored on disk or cd-rom. A CT image is made by converting each digit in the digital matrix into a small square of gray varying from black to white through a digital/analog converter.

2.2 Multi-frame CT Image Sequencing Big Data Bulk Clustering Method Analysis

The traditional multi-frame CT image sequencing big data batch clustering method directly processes the original data. However, the differences between the data samples are particularly small, the speed of clustering is very slow, and the phenomenon of

repeated clustering is prone to occur, which increases the difficulty of clinical detection to some extent. Therefore, this paper proposes the DD-TV clustering method by spatially correlating the data of medical image sequencing. This method can convert the data of the original space into the feature space, which not only greatly reduces the amount of data, but also improves the speed of clustering. More importantly, the transformed data can display its own attributes in the feature space, greatly improving the accuracy of clustering.

The flow chart of multi-frame CT image reconstruction is shown in the following Fig. 2:

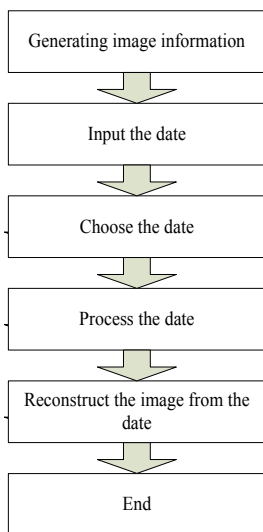


Fig. 2. Multi-frame CT image reconstruction flowchart

Multi-frame CT image program:

```

Begin
clear all;
clc
load matlab;
%reconstruct
Figure(1)
title('original image')
imshow(P);

%the size of image
N = 258;
%coordinate of image
center= floor((N + 1)/2);
xleft = -center + 1;
x = (1:N) - 1 + xleft;
x= repmat(x,N,1);

ytop = center - 1;
y= (N:-1:1).' - N + ytop;
y = repmat(y, 1, N);

%angle change
theta1 = theta1*pi/180;
costheta = cos(theta1);
sintheta = sin(theta1);

len = size(R1,1);
ctrIdx = ceil(len/2);
img = zeros(N,class(R1));
for i = 1:length(theta1)
proj = R1(:,i);
t = x.*costheta(i) + y.*sintheta(i);
a = floor(t);
img=img+(t-a).*proj(a+1+ctrIdx+(a+1-t)).*proj(a+ctrIdx);
img = img*pi/(2*length(theta1));

%show reconstruction image
    Figure(2);
    imshow(img,[]);
end

```

A flowchart of a DD-TV data clustering method based on a multi-frame CT image is shown in Fig. 3:

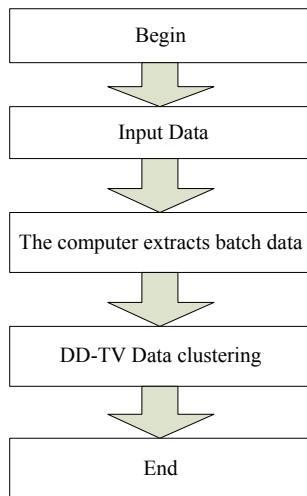


Fig. 3. DD-TV data clustering method flow chart

DDTV data clustering procedure:

```
price = [1.1,1.2,1.3,1.4,10,11,20,21,33,34]
```

```
increase = [1 for i in range(10)]
```

```
data = np.array([price,increase],dtype='float32')
```

```
class Myhcluster():
```

```
def __init__(self):
```

```
print('start system clustering')
```

```
"""The start function of the system clustering method, there are two input
parameters of the input variable and the distance calculation method"""
```

```
def prepare(self,data,method='zx'):
```

```
if method == 'zx'
```

```
self.zx(data)
```

```
"""Center of gravity method for system clustering"""
```

```
def zx(self,data):
```

```
token = len(data[0,:])
```

```
flu_data = data.copy()
```

```
classifier = [[] for i in range(len(data[1,:]))]
```

```
LSdist = np.array([0 for i in range(token ** 2)],
dtype='float32').reshape([len(data[0, :]), token])
```

```
index = 0
```

```
while token > 1:
```

```
"""Calculating the distance matrix"""
```

```
for i in range(len(data[0,:])):
```

```
for j in range(len(data[0,:])):
```

```
LSdist[i,j]=round(((flu_data[0,i]-flu_data[0,j])**2+(flu_data[1,i]- flu_data[1,j])*
*2)**0.5,4)
```

```
end
```

```
#Implementation of System Clustering
```

```
> rm(list=ls())
```

```
> a <- Sys.time()
```

```
> price <- c(1.1,1.2,1.3,1.4,10,11,20,21,33,34)
```

```
> increase <- rep(1,10)
```

```
> data <- data.frame(price,increase)#generate sample data frame
```

```
> d<- dist(data)#create sample distance array
```

```
> hc <- hclust(d,'centroid')#systematic clustering with center of gravity>
```

```
cbind(hcmerge,hcmerge,hcheight)# demonstrate the classification process
```

First, we first detect the distance-driven operation of the image orthographic projection. We select the circular plate with uniform size as the pseudo-true image. The distribution of the projection values obtained by the distance-driven operation of the image orthographic projection is shown in Fig. 4 The horizontal axis represents the detector, and the number of detectors is 120; the vertical axis represents the projection value detected by the corresponding detector. From the Fig. 4, we can find that the projection value calculated by the distance projection of the image is almost stable, and there is no fluctuation.

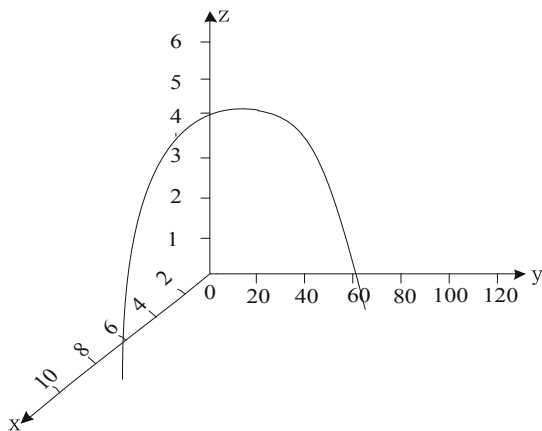


Fig. 4. Distance-driven orthographic projection

Let the image be initialized to W , and set the number of times of the DD-TV method to $i, i = 1, W[i] = 0$.

Constrain the multi-frame CT image, and set the constrained image to $p[n]$, that is: if $p[n] \geq 0, p[n] = W[i]$, otherwise, $p[n] = 0$.

The speed algorithm finds the minimum value of the objective function, and the number of times is $t, t = 1, 2, \dots, N_t$. When $t = 1$,

$$P[n, t] = P[n] \tag{1}$$

When the number of DD-TV method is $i = i + 1$,

$$W[i + 1] = W[i] \tag{2}$$

$$N = i_1g_1 + i_2g_2 \tag{3}$$

The DD-TV algorithm flow chart is shown in Fig. 5.

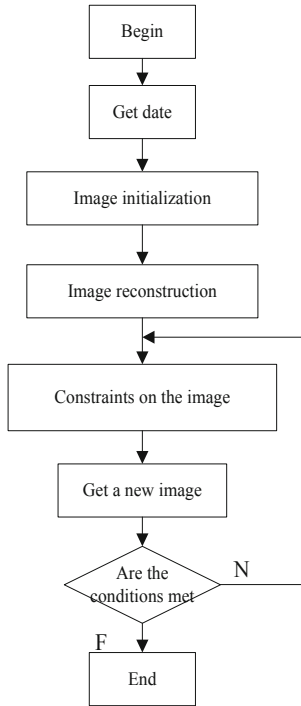


Fig. 5. DD-TV algorithm flow chart

3 Experimental Design and Result Analysis

In order to verify the application effect of the method of multi-frame CT image sequencing big data clustering, a simulation experiment was carried out. The experimental operating system is Windows 10, and the simulation software is matlab 7.0. The traditional method is selected as the experimental comparison method to comprehensively compare the performance of different methods. The experimental data are shown in Fig. 6.

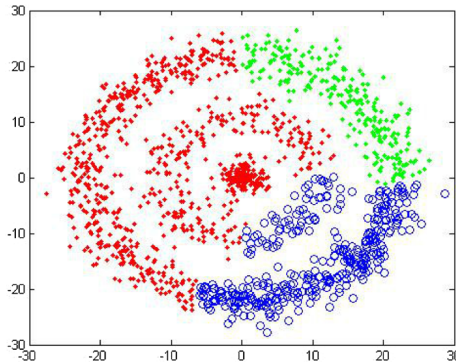


Fig. 6. Experimental data

The results are shown in Table 1, Fig. 7 and Fig. 8.

Table 1. Performance comparison

Method	Precision	Speed
Traditional method	91.86%	321.71 s
DD-TV method	95%	168.3 s

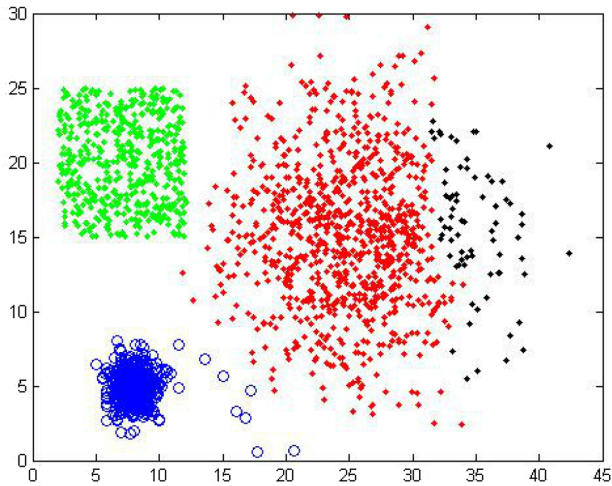


Fig. 7. Clustering effect of traditional method

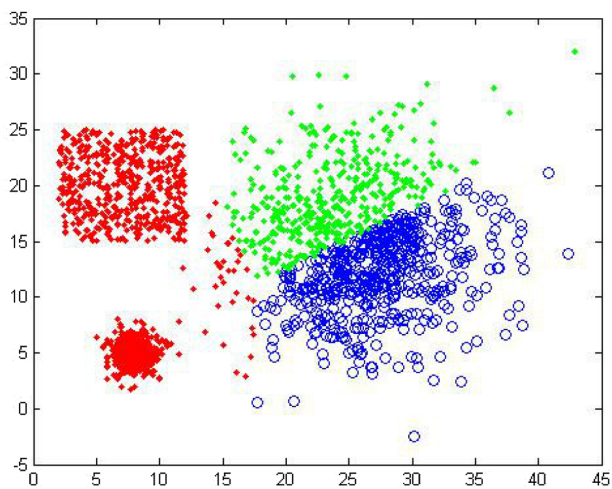


Fig. 8. Clustering effect of DD-TV method

The study found that the DD-TV method not only has high clustering accuracy, but also greatly improves the speed of image data processing.

The DD-TV method does not need to store a large number of data matrix systems, but calculates the individual data in the matrix sequence at all times, saving a lot of time and space. At the same time, this method does not need to go through the entire detector and image, just cyclically calculate the data on the projection line. Compared with the traditional iterative data clustering method, the overall speed is improved, and combined with the principle of data standardization optimization, it has an important impact on the quality of reconstructed images.

With the development of technology, the way of CT scanning is very different from the past, and multi-frame images can be obtained in a short time. Because the scanning time is very short, a movie image can be obtained, which can prevent image artifacts after motion. For the research and analysis of DD-TV data clustering based on multi-frame CT images, there are also many shortcomings in the experiment. In this paper, the image is reconstructed only under 2D conditions, and image reconstruction under 3D conditions can be achieved in later studies. Moreover, other optimization methods can be used in subsequent studies when optimizing the objective function, making the calculation more accurate. It is hoped that the research in this paper can provide some help for clinical testing.

4 Conclusion

Along with the development of image processing technology, image feature analysis method is used for medical diagnosis has become one of the important tool of case analysis, common medical images with CT image, ultrasound image and X-ray image, etc., for case analysis and diagnosis of CT image which provides a powerful data support and intuitive image, in the CT image acquisition, the need for more frame scan and gathering, finally realizes the sequencing batch clustering large data, analysis of CT image can improve the efficiency and capacity, is of great significance in the aspect of doctors for patients. On the basis of studying the characteristics of traditional clustering methods, a multi-frame CT image sequencing method is proposed. The simulation results show that this method has high clustering accuracy and low clustering time. The method proposed in this paper provides a new idea for the further development of multi-frame CT image sequencing big data clustering technology, which improves the efficiency and ability of CT image analysis and plays an important role in the accurate acquisition of patients' conditions by doctors.

References

1. Bricq, S., Collet, C.H., Armspach, J.P.: Unifying framework for multimodal brain MRI segmentation based on hidden markov chains. *Med. Image* **23**(12), 22–25 (2017)
2. Gilbert: Ultrasonic detection based on computer simulation technology, **36**(4), 94–95 (2017)
3. Liew, A.W.C., Yan, H.: An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation. *IEEE Trans. Med. Imaging* **11**(2), 60–62 (2017)

4. Zhang, D.-Q., Chen, S.-C.: A novel kemelized fuzzy c-means algorithm with application in medical image sementation. *Artif. Intell. Med.* **19**(2), 37–38 (2017)
5. Chuang, K.S., Tzang, H.L., Chen, T.-J.: Fuzzy c-means clustering with spatial information for image segmentation. *Comput. Med. Imaging Graph.* **3**(10), 123–124 (2017)
6. Liu, S., Pan, Z., Cheng, X.: A novel fast fractal image compression method based on distance clustering in high dimensional sphere surface. *Fractals* **25**(4), 1740004 (2017)
7. Tang, H., Dillenseger, J.L., Bao, X.D., Luo, L.M.: A vectorial image soft segmentation method based on neighborhood weighted gaussian mixture model. *Comput. Med. Imaging Graph.* **42**(12), 34–36 (2017)
8. Raghuvira Pratap, A., Vani, K.S., Rama Devi, J., KnageeswaraRao, V.A.: An efficient density based improved K-medoids clustering algorithm. *Int. J. Adv. Comput. Sci. Appl.* **3** (9), 25–27 (2017)
9. Chabat, F., Yang, G.-Z., Hansell, D.-M.: Obstructive lung diseases: texture classification for differentiation at CT. *Radiology* **5**(34), 57–58 (2017)
10. Kuo, W.-F., Lin, C.-Y., Sun, Y.-N.: MR images segmentation. In: Computer Society Conference on Computer Vision and Pattern Recognition, vol. 27, no. 5, pp. 67–69 (2017)
11. Yang, J., Xie, Y., Guo, Y.: Panel data clustering analysis based on composite PCC: a parametric approach. *Cluster Comput.* **34**(5), 1–11 (2018)
12. Huang, S., Ren, Y., Xu, Z.: Robust multi-view data clustering with multi-view capped-norm K-means. *Neurocomputing* **311**, 197–208 (2018)
13. Bo, Y.: The data clustering based dynamic risk identification of biological immune system: mechanism, method and simulation. *Cluster Comput.* **7330**, 1–14 (2018). <https://doi.org/10.1007/s10586-018-1960-2>
14. Payne, W.V., Heo, J., Domanski, P.A.: A data-clustering technique for fault detection and diagnostics in field-assembled air conditioners. *Int. J. Air-Cond. Refrig.* **26**, 1850015 (2018)
15. Das, P., Das, D.K., Dey, S.: A new class topper optimization algorithm with an application to data clustering. *IEEE Trans. Emerg. Top. Comput.* **PP**(99), 1 (2018)
16. Ma, R., Angryk, R.A., Riley, P., et al.: Coronal mass ejection data clustering and visualization of decision trees. *Astrophys. J. Suppl.* **236**(1), 14 (2018)
17. Karpinski, P., Patai, A., Hap, W., et al.: Multilevel omic data clustering reveals variable contribution of methylator phenotype to integrative cancer subtypes. *Epigenomics* **10**(3), 1289–1299 (2018). epi-2018-0057
18. Fan, W., Bouguila, N., Du, J.X., et al.: Axially symmetric data clustering through Dirichlet process mixture models of Watson distributions. *IEEE Trans. Neural Netw. Learn. Syst.* **PP** (99), 1–12 (2018)
19. Popova, O., Kuznetsova, E., Sazonova, T.: Algorithms of data clustering in assessing the transport infrastructure of the region. In: MATEC Web of Conferences, vol. 170, no. 3, pp. 123–130 (2018)
20. Sriadi, S., Gultom, S., Martiano, M., et al.: K-means method with linear search algorithm to reduce Means Square Error (MSE) within data clustering. In: IOP Conference Series Materials Science and Engineering, vol. 434, no. 3, p. 012032 (2018)