2017 International Conference on Mathematics, Modelling and Simulation Technologies and Applications (MMSTA 2017)
ISBN: 978-1-60595-530-8

Spark Distributed Medical Image Segmentation Method Based on KFCM-CV Model

Jing-yi LIANG and Yu-Cui GUO

Beijing University of Posts and Telecommunications, Beijing 100876, China

Keywords: Spark, Medical image segment, KFCM-CV.

Abstract. Medical image segmentation is great significance for the study of medical diagnosis and pathology. In order to process medical images effectively and quickly, a medical image processing method based on Spark is proposed in this paper. Firstly, a model for medical image segment based on improved fuzzy kernel clustering and CV model is built, then the medical images are processed in batches by using this model on Spark. Simulation experiments show that the method proposed in this paper achieves efficient and accurate segmentation.

Introduction

With the arrival of the era of big data and the massive growth of medical image scale, it is an urgent need for medical image processing to realize computer-aided diagnosis by efficient processing of medical images. Spark provides better data reuse mechanism, strong data set abstraction and rich programming tools for medical image processing. It solves the high latency defects of Hadoop architecture by the introduction of flexible data set concept instead of MapReduce. It provides three kinds of API: Scala, Java, Python and realizes the integration with other programs, and through the global resources to effectively allocate tasks and resources, improve the efficiency of image processing[1]. This paper presents a parallel medical image processing scheme based on Spark. Firstly, build image segmentation algorithm, then implement the combination of improved fuzzy kernel clustering and CV model and then read the image data through the Spark distributed, and distributed running algorithm program, to achieve a high speed of medical image segmentation.

Medical image segmentation is a process based on a particular meaning or uniformity principle, segment the medical image into several meaningful and mutually disjoint regions, The resulting sub regions conform to the requirements of some consistency (or uniformity) and connectivity[2]. Medical image segmentation is a very important and special field, and it is an important premise for visualization, medical diagnosis and treatment plan formulation. Traditional FCM algorithm is easy to be affected by noise, Zhang Li [3]et al. proposed a kernel clustering algorithm on the basis of FCM, Mapping low dimensional space data into high dimensional space, make data feature extraction and amplification. Thus the clustering can be realized more accurately. However, the kernel clustering segmentation is not easy to obtain smooth segmentation boundaries and closed segmentation regions, In order to overcome these shortcomings, this paper combines KFCM with CV model, and proposes a KFCM-CV segmentation method for medical images. The level set method was proposed by Osher and Sethian in 1988[4], The method uses zero level set of higher dimensional function to represent active contour, The level set function is evolved under the control of partial differential equations until the zero level set evolves to the target edge of the image. Chan and Vese simplify the Mumford-Shah model[5], and establish a variational level set model, namely the CV model. However, the model is sensitive to noise and inaccurate in edge location. The images were pre-segmented using fuzzy kernel clustering algorithm, and the results will be used as the initialization data of CV model, by combining the two models, overcome the inherent shortcomings of the two models, obtained accurate segmentation results.

After the establishment of KFCM-CV model, this paper uses Spark to process image data in a distributed way, Through the Spark distributed running KFCM-CV segmentation algorithm, The final results are stored in a distributed file system. Finally, the simulation experiments are carried out, and the experimental results show that the proposed method can realize parallel processing of medical images, and is suitable for large-scale medical image segmentation.

Model Building

Kernel Fuzzy C Means Algorithm

Clustering analysis is a branch of unsupervised learning. Traditional fuzzy C means clustering model based on Euclidean distance. Therefore, the accuracy of algorithm segmentation depends largely on the distribution of samples[8]. To this end, many scholars have proposed a fuzzy C means clustering algorithm based on kernel function[9]. The original input data of KFCM algorithm is mapped to high dimensional space by kernel function, The linear non separable parts in low dimensional space can be linearly divided into high dimensional feature space by nonlinear mapping[10]. It is more robust to noisy images.

The following function for clustering:

$$f=\sum_{n=1}^{N}\sum_{m=1}^{c}u_{mn}^{l}\left\|\phi(i_n)-\phi(v_m)\right\|^2 \tag{1}$$

In the formula: ϕ is the nonlinear mapping for high-dimensional feature space.

$$\left\|\phi(i_n)-\phi(v_i)\right\|^2=K(i_n,i_n)+K(v_i,v_i)-2K(i_n,v_i) \tag{2}$$

In the formula: $K(x,y)=\phi^T(x)\phi(y)$ is the transvection of kernel function. The Gauss kernel function is as follows:

$$K(x,y)=\exp\left(-\frac{\left\|x-y\right\|^2}{\sigma^2}\right) \tag{3}$$

According to equation (2) (3) the objective function for KFCM:

$$f=\sum_{n=1}^{N}\sum_{m=1}^{c}u_{mn}^{l}(1-K(i_n,v_m)) \tag{4}$$

The membership matrix function is:

$$u_{mn}=\frac{(1-K(i_n,v_m))^{-1/(l-1)}}{\sum_{j}^{c}(1-K(i_n,v_j))^{-1/(l-1)}} \tag{5}$$

The fuzzy clustering central iteration function is shown in formula (6):

$$u_{mn}=\frac{\sum_{n=1}^{N}u_{mn}^{l}(1-K(i_n,v_j))^{i_n}}{\sum_{j}^{c}u_{mn}^{l}(1-K(i_n,v_j))} \tag{6}$$

CV Model

Let Ω be the image region to be segmented, and C as active contour, define the following energy functional

$$E_{cv}(C,c_1,c_2)=\mu L(c)+\lambda_1\int_{\Omega_1}\left|u_0(x,y)-c_1\right|^2dxdy+\lambda_2\int_{\Omega_2}\left|u_0(x,y)-c_2\right|^2dxdy \tag{7}$$

μ is the length control factor of the curve C, λ_1,λ_2 are internal and external weight coefficients of curves. ϕ is curve C corresponding to the level set function, $C=\phi_0$ (zero level set of ϕ), introduce the Heaviside function and Dirac function:

$$H_\varepsilon(x)=\frac{1}{2}\left[1+\frac{2}{\pi}\arctan(\frac{x}{\varepsilon})\right],\ \delta(z)=\frac{d}{dz}H(z) \tag{8}$$

Make the Li's does not need to re-initialize the level set Energy penalty item added to CV model, The energy functional is obtained:

$$\begin{aligned}E_{cv}(\phi,c_1,c_2)&=\mu\int_\Omega\left|\nabla H(\phi(x,y))\right|dxdy+\lambda_1\int_\Omega\left|u_0(x,y)-c_1\right|^2H(\phi(x,y)dxdy\\&+\lambda_2\int_\Omega\left|u_0(x,y)-c_2\right|^2\left[1-H(\phi(x,y)dxdy\right]+a\int_\Omega\frac{1}{2}(1-\left|\nabla\phi\right|^2)dxdy\end{aligned} \tag{9}$$

By using Euler Lagrange method solve the equation (9), the following partial differential equations can be obtained:

$$\frac{\partial\phi}{\partial t}=\alpha\left[\Delta\phi-div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right]+\delta(\phi)\left[\mu div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)-\lambda_1(u_0(x,y)-c_1)^2-\lambda_2(u_0(x,y)-c_2)^2\right] \tag{10}$$

The upper form is discretized by the finite difference method, and the numerical solution is obtained.

Segmentation Algorithm of KFCM - CV Model

In the light of the disadvantages of traditional CV model's energy weight coefficient λ_1,λ_2 has great influence on the segmentation accuracy and efficiency, sensitivity to initial position etc., Li Xuanping[6] et al proposed a fuzzy clustering collaborative region active contour model (FCM-CV), The energy weight coefficient is set as:

$$\lambda_1=f_1(\phi)=\frac{\int_\Omega\left(1-H(\phi)\right)dx\,dy}{N_xN_y} \tag{11}$$

$$\lambda_2=f_2(\phi)=\frac{\int_\Omega\left(1-H(\phi)\right)dx\,dy}{N_xN_y} \tag{12}$$

N_x and N_y are image pixels. Thus, the influence of area factor on energy function is eliminated. On this basis, this paper made the following improvements:

1) fuzzy kernel clustering is used to replace fuzzy kernel clustering, which increases the robustness of noisy images;
2) when calculating the initial level set, the number of pixels belonging to each cluster center is calculated, and the ratio between the image and the number is taken as the energy coefficient of the CV model.

Specific as follows:

In the process of fuzzy kernel clustering, in this paper, the pixels of the image are clustered into 2 categories by default, That is, c=2; It is assumed that the segmentation target contains the class I pixels in the fuzzy clustering segmentation results, The boundary of the region composed of I pixels can be set as zero level set initial contour. According to the result of KFCM segmentation, The initialization level set of CV model can be expressed as:

$$\phi_0(x,y)=p\ u_{ij}\geq 0.5;-p\ u_{ij}<0.5 \tag{13}$$

Where: $j = x * N_y + y$

The pixels of the first and second classes are calculated at the same time, The pixel number of the first class is Class1, and the pixel number of the second class is class2, then the number of pixels is:

class1=class1+1 if $u_{1j} \geq 0.5$ else class2=class2+1(14)

According to the number of pixels belonging to the first and second classes in the kernel clustering results, the initial values of λ_1, λ_2 are set as the values of the number of pixels in the kernel clustering results:

$$\lambda_1 = \frac{class2}{N_x N_y}$$

$$\lambda_2 = \frac{class1}{N_x N_y}$$

We default that the first pixel in the upper left corner of the image belongs to pixels outside the curve, determined

the membership function of the first class of kernel clustering results u_{12},if u_{12}>0.5, illustrate the first class of cluster results is outside the curve, then $\lambda_1 = \frac{class1}{N_x N_y}$ $\lambda_2 = \frac{class2}{N_x N_y}$, otherwised λ_1, λ_2 keep the initial value.

KFCM-CV collaborative segmentation algorithm steps:

1) set up the number of kernel clustering classes was 2, segmentation of images using KFCM model, initial segmentation results are obtained: Membership function u_{1j}, u_{2j} of each pixel $u_0(x, y)$.
2) The initialization level set of CV model is determined according to formula (13)
3) According to the formula (14) calculate the total number of pixels belonging to various types
4) Judge whether the first class is outside the curve and calculate the coefficient of energy weight
5) minimum the energy functional is and the final level set is calculated.

Parallel Processing Based on Spark

Based on the above model, we use Spark platform to realize the distributed processing of medical images.

Traditional medical image segmentation is usually run on a single machine, long running time in the case of a large number of medical images, inefficiency. Spark is a distributed computing framework implemented on the basis of MapReduce, Spark inherits the advantages of Hadoop's MapReduce, but it is more efficient than Hadoop. Spark pioneered the concept of abstract elastic data set RDD, making Spark very efficient when dealing with iterative, interactive, streaming data[7].

In this paper, the medical image segmentation process based on spark is shown in figure 1:

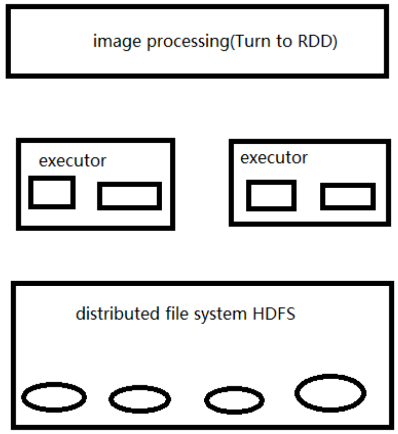

Figure 1. The medical image segmentation process based on spark.

The process of running Spark is as follows:

1) give the path address parameter of the incoming image file
2) distributed read the image, turn the image into RDD
3) distribute Excutor, and distribute run the spark algorithm program
4) store the running results into the HDFS file system

Experimental Results and Discussions

Figure (2) is a medical image to be segmented, The Figure (3) - (5) is the CV model, the KFCM-DRLSE model and the model based on this paper, It can be seen that only the model of this paper can achieve accurate segmentation.

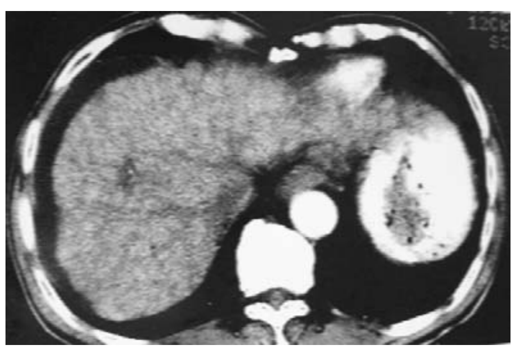

Figure 2. Original image.

Figure 3. Segmentation results of CV model.

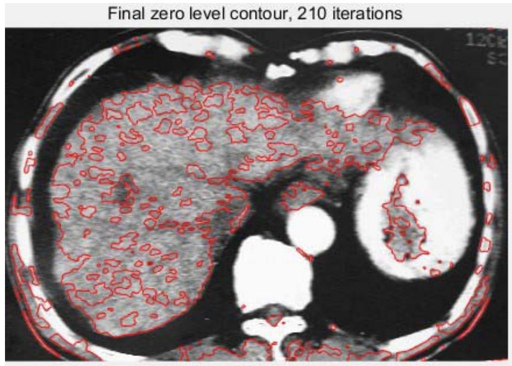

Figure 4. Segmentation results of KFCM-DRLSE model.

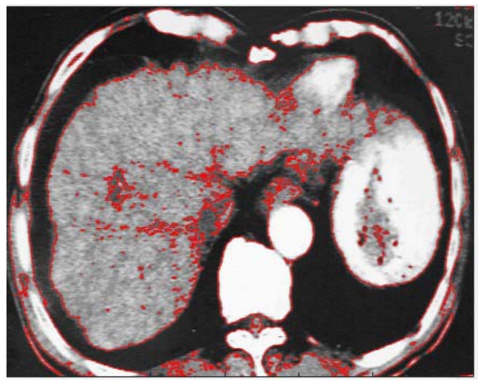

Figure 5. Model segmentation results in this paper.

Summary

This paper presents KFCM-CV Spark distributed model, the model does not need to re initialize the level set, insensitive to initial position. It combines the advantages of clustering model and CV model effectively. The influence of image segmentation target and area ratio on image segmentation is fully considered, effective segmentation can be achieved. The model is run by Spark distributed, high speed and effective segmentation can be achieved. However, the input data-medical image is distributed only in this paper, based on spark distributed algorithm will be improved in future studies.

References

[1] Lan Li, Wang Junfeng, Tang Peng. Research on parallel medical image processing based on [J]. Spark Journal of Sichuan University (Natural Science Edition), 2017, 54 (01): 65-70.(in Chinese)

[2] Yang Hongzhe. Level set theory and its application in medical image segmentation [D]. Beijing Institute of Technology, 2014. (in Chinese)

[3] Zhang Li, Zhou Weida, Jiao Li Cheng. Kernel clustering algorithm [J]. Journal of Computer Science, 2002, 25 (6): 587-590. (in Chinese)

[4] Osher S., Sethian J.A. Fronts propagating with curvature-dependent speed [J]. Journal of Computational Physics, 1988, 79: 12- 49.

[5] Li C.H. M., Xu C.H. Y., Gui C.H. F., et al. Distance regularized level set evolution and its application to image segmentation [J]. IEEE Trans. on Image Processing, 2010, 19 (12): 3243- 3254.

[6] Li Xuanping, Wang Xue. Fuzzy clustering collaborative region active contour model medical image segmentation [J]. Journal of Instruments and Meters, 2013, 34 (04): 860-865. (in Chinese)

[7] Wang Xun, Feng Rui. Design of massive image retrieval system based on [J]. Microcomputer Application of Spark, 2015, 31 (11): 11-13+17+2. (in Chinese)

[8] Li Wenjing. Research and implementation of medical image segmentation algorithm based on fuzzy clustering [D]. Northeastern University, 2014. (in Chinese)

[9] Liu Yajing, Song Yuqing, Liao Dingan, Xia Qianqian. Kernel fuzzy clustering and variational level set image segmentation method based on [J]. Application Research of Computers, 2013,30 (11): 3510-3513. (in Chinese)

[10] Zhang Fei. Study on MR [D]. of the Shaanxi Normal University Medical Image Segmentation Based on Fuzzy C Means Clustering, 2014. (in Chinese)