UDC 004.932:519.237.8:61 (045)

USING CLASTERING-BASED SEGMENTATION ALGORITHMS IN REGIONS OF INTEREST DETECTION ON THE MEDICAL IMAGES

O. Koba, PhD, ***O. Dobrina***

National Aviation University

misericorde@bk.ru

The comparative analysis of segmentation algorithms was shown in this article that are based on clustering and it was faced up the most profitable areas of use during region of interest detection in medical image processing area.

Keywords: segmentation, clustering, medical image processing, segmentaton algorithms.

Проведено порівняльний аналіз алгоритмів сегментації на базі кластеризації та розглянуто їх найбільш продуктивне застосування при виділенні регіонів інтересу при роботі з медичними зображеннями.

Ключові слова: сегментація, кластеризація, медичні зображення, алгоритми сегментації.

Introduction

At the current moment, the development of computer technology and data mining tools was increased to automate many areas of human endeavor, thus increase productivity, reduce errors, liberate people from monotonous work and etc. In recent decades, computers began to provide substantial assistance in problems related to pattern recognition, data mining, and in particular image processing. Image analysis is relevant for the such areas as data compression, document recognition, creation of databases of images, quality control, medical diagnosis and many others.

Most of the images derived from photographic and scanners are raster images that represent a rectangular grid of colored dots and gray values. The human brain, analyzing the received using eyes the image does not represent it as a matrix of dots. Instead of it, it integrates these points into homogeneous areas and continues to operate with the obtained objects. This process is called segmentation in humans and occurs at a subconscious level. In addressing the problem of computer image analysis and pattern recognition it should be implemented automatically. Segmentation is performed in the early stages of analyzing the image and quality of its implementation has a key impact on the speed, accuracy, and sometimes the opportunity for the further analysis.

Statement of the problem

The publications in the field of processing, segmentation and retrieval of objects on the images are very intense. A lot of works were presented by native (Andreev, Bayakovskiy, Bohuslav, Vezhnevets, Kazan, Sergeev) and foreign (Russell, J. Malik, M. Andretti, Yu Li, Yu-Jin Zhang, J. Shi, P. Viola, M. Jones) scientists. So far there were produced a lot of approaches to segment images, but none of them became universal. Depending on the origin, texture, quality, size of the object of the research, and many other options it makes sense to apply the different methods of segmentation. So the problem is not in finding the universal way of segmenting any kind of images, but rather to find the ways of segmenting those ones which give the most appropriate outcome in a particular area. As this article observes the medical images, the problem is to choose the most suitable algorithm for segmentation of the such types of images.

Research and publication analysis

During the article creation an overview of the research and publications was performed according to the browsing topic. According to the amount of the available material it can be concluded that the subject is very popular today. Special attention was paid to the algorithms that are based on clustering, because if clustering is the process of grouping points in space into groups, these groups may be the same objects or bodies of the medical image. Particular attention can be provided to the article [1], that gives the most complete information about the concept of clustering algorithms have been considered the main goals of clustering and its application. More attention to the types of algorithms was payer in [2], and in the work [3] that considers the use of algorithms clustering in machine learning. The most clear clustering methods are explained in the book [4] from the point of applications in various fields of human activity, and lectures. [5] For a more complete review of the subject area, refer to abstracts book [6], which provides a lot of examples of algorithms for segmentation, registration, etc. in the field of medical image processing.

Purposes

The main idea of this article is to give an overview and to perform the analysis of the existing clustering based segmentation algorithms and to determine the application of these methods in such areas of medical image processing, where they could give the more accurate result.

Clustering main ideas

Clustering is a procedure that allows grouping of the discrete number of points into clusters using the particular metrics (e.g. the distance between the points).

It is assumed that the clustering problem should be solved without any a priori information about the points belonging to the classes. The formal statement of the problem of clustering as follows. Let the given set of objects $X = (x_1, x_2,..., x_n)$ and IDs (names, labels) of the clusters $Y = (y_1, y_2,... y_k)$. For X it is determined some function of the distance between the objects $D(x, x')$.

Except these statements, the finite sample case studies $X_m = (x_1, x_2, ..., x_m)$ from the X array that should be divided into X_m clusters in such a way that each of them could contain the elements that are close the D metrics.

Also each object x_i from the X_m array should receive the number of the cluster y_j.

Then the problem will be to find the function f, which is any object x from set X assigns a cluster number y from a set Y, which itself is known in advance.

However, in most cases it is necessary to determine the optimal number of clusters based on the characteristics of problem.

Clustering allows to achieve the following objectives:

1. Improves the understanding of the data by identifying the structural groups.

2. Partitioning the sample into several groups of similar objects to simplify the further data processing and decision-making, applying to the cluster analysis method.

3. Allows you to store data compactly. Instead of storing the entire sample it can be left on a typical observation for each cluster.

4. Identification of the new custom objects that do not fall in any one cluster.

There are many clustering methods that can be classified into distinct and indistinct. Distinct (or crisp) clustering methods divide the original set of objects into several disjoint subsets. In this case, any object with belongs to only one cluster. Indistinct (or fuzzy) clustering methods allow the same object simultaneously belong to several (or even all) clusters, but with varying degrees. Fuzzy clustering in many situations a more “natural” than clear, for example, for facilities located on the border of clusters.

Clustering methods are also classified by whether specified number of clusters in advance or not. In the latter case, the number of clusters is determined during the execution of the algorithm based on the distribution of the input data.

Segmentation in a context of clustering direction

Segmentation is a division process of the image on some groups with taking into account the similarity of the pixels' characteristics. The main idea is the following: each pixel should be connected with the particular visual properties, such as brightness, color or texture. Within a single object or a part of an object, these attributes are changed relatively small, whereas when crossing the border from one object to another is usually a significant change in one or more of these attributes. The main problem is to find a type of partitioning the image into the particular set of pixels that these constraints were satisfied as much as possible.

The main idea of the segmentation in this context can be reduced to a clustering problem as follows. Put in correspondence to each pixel vector of its optical characteristics. Then, after selecting an appropriate metric in the space of vectors of their clustering will give some solution to the problem of segmentation. Pixels are close to each other and have similar characteristics, and also have priority to include them in the same cluster and, therefore, in the same segment of the image. Successful segmentation will depend both on the administered metrics, and on clustering algorithm.

***k*-means algorithm**

The main action of the algorithm is that it seeks to minimize the total deviation points from the centers of the clusters. The algorithm divides the set of the elements from a vector space when according to the already known number of clusters k. This method of clustering is significantly different from the other algorithmic techniques like association (tree clustering) and double-input association. Let's say that you already have a hypothesis on the number of clusters (from observations or variables). You can tell the system to create exact three cluster to make them as different as possible. This is exactly the type of tasks that is solved by the method of k-means algorithm. In general, the k-means method builds exactly k different clusters located at the large possible distances from each other.

The algorithm finishes its work when at some iteration there is no changes in the cluster. The number of such iterations is finite, and it depends on a finite set of possible partitions.

From the computational point of view, this method can be considered as analysis of variance “in reverse”. The program begins with k randomly selected clusters, and then changes the objects that belong to them, to minimize variability within the clusters and maximize variability between the clusters. This method is similar to the method of

analysis of variance (ANOVA) in the sense that the criteria of the significance in the analysis of variance compares intergroup variability of intra-checking the hypothesis that the average in the groups differ from each other. In clustering using the *k*-means algorithm program moves the objects (e.g. observation) of some groups (clusters) in order to get the most relevant results during the analysis of variance.

Advantages:

- simple implementation;
- intuitive understanding and the clearance of the algorithm.

Disadvantages:

- the number of clusters should be a known value;
- depending of the result from the initialization of the clusters centers;
- calculation complexity – as it was shown be David Arthur and Sergiy Vasylvitskiy, on some classes of the arrays the complexity of the algorithm could reach the value of $2^{\Omega(\sqrt{n})}$.

Such algorithms shouldn't be used when the potential number of the segmented objects can be large. Such methods can be used with combining with the expert interaction that can assign the places of the cluster centers, from which the segmentation process can be started. They can be used for lungs segmentation, heart segmentation and etc.

***C*-means fuzzy clustering algorithm**

The most popular fuzzy clustering algorithm is the *c*-means algorithm. It is a modification of the *k*-means method. The *c*-means fuzzy clustering methos allows to create the array of the vectors with the *p* power on the selected number of the fuzzy arrays. The feature of this method is using the belonging matrix *U* with the u_{ij} elements that determine belonging the *i*-element of the output array to *j*-cluster. Clusters are described using their 'mass centers' c_j — vectors of the same space to which the output vector array belongs to. The steps of the algorithm are:

1. At first you should select the initial fuzzy division of the *n* objects into *k* clusters using the selection of the belonging matrix *U* size of n_x *k*.

2. Using the *U* matrix, you should find the value of the fuzzy mistake criteria:

$$E^2(X,U)=\sum_{i=1}^{N}\sum_{k=1}^{K}U_{ik}\left\|x_i^{(k)}-c_k\right\|^2;$$

where c_k is a kind of a 'mass center' of the fuzzy cluster *k*: $c_k=\sum_{i=1}^{N}U_k x_i$, .

3. Then you should regroup the objects to reduce the value of the fuzzy criteria.

4. Return to step 2 until the changes in the *U* matrix will become minor.

Such algorithm can be not effective when the number of clusters is unknown at the beginning, or when one object should belong only to a single cluster.

Advantages:

- Blurring of the algorithm;
- Better convergency (comparing with *k*-means);
- simple implementation;
- intuitive understanding.

Disadvantages:

- calculation complexity;
- the number of clusters should be a known value;
- sensitive to the initial division.

Such method requires the qualitative input image — it is obvious from the last statement. The more accurate and qualitative result can be expected during bone segmentation, where the brightness can give the complete input information for the further division and refining.

Hierarchical methods

This group of methods is characterized by the consistent association of source elements and a corresponding reduction in the number of clusters. Among the hierarchical clustering algorithms it is possible to determine two types of them: ascending and descending algorithms. Descending algorithms work like in a down direction: at first all objects should be located on a single cluster that is divided further on more smaller clusters. The ascending algorithms are more commonly used that at the beginning of their work try to locate each object to the separate cluster, and then they are trying to unite these clusters to more huge until all objects will not be placed at the same cluster. According to this method the system of the nested division is created. The results of the such algorithms usually are presented like a tree.

For the calculation of the distance between the clusters can be used the single connection of the full connection.

Advantages:

- receiving the whole map of the divisions
- builds the optimal division

Disadvantages:

- it should be entered the *k* limitation*;*
- calculation convergency;
- clusters cannot intersect each other;
- calculation complexity;
- full-division system that can be redundant in a context of the problem.

You can notice that these types of algorithms can be used only in such types of cases, when the number of the segmented areas is rather small, they don't intersect each other and the scanning was performed with high quality.

Such objects can be non-intersected tumors or some small objects that can be received using micro- or nano-CT scanning.

Conclusions

This article provided an overview of the concept of clustering, clustering algorithms and their application in solving image segmentation problems.

A common disadvantage for all clustering based segmentation algorithms is that these models implicitly assume that the range of variation of characteristics within each object that is segmented is small according to the difference of characteristics of different objects.

Many classes of images are not satisfied with this statement. Besides the lighted parts of the objects (e.g. titanium implants on CT images of the bones) may significantly affect on the result of segmentation.

REFERENCES

1. *Jain A. K.* Data Clustering : A Review: ACM Computing Surveys / A. K. Jain, M. N. Murty, P. J. Flynn. — V. 31, No. 3, September 1999. — P. 264–323.

2. http://www.cs.csi.cuny.edu/~natacha/TeachFall_2008/GradCenter/Readings/Lec4Read/clustering.pdf- Fung's article about clustering

3. *Clarke Bertrand*. Zhang Principles and Theory for Data Mining and Machine Learning: Springer Science / Bertrand Clarke, Ernest Fokou'e, Hao Helen. — 2009. — P. 773.

4. *Mandel I. D*. Cluster analysis / I. D. Mandel. — M. : Finance and Statistics, 1988. — 176 c.

5. http://www.statsoft.ru/home/textbook/modules/stcluan.html — all information about the clustering analysis

6. *Dougherty G.* Medical Image Processing: Techniques and Applications: Springer Science / G. Dougherty. 2011. — P. 373.

Article received 25.02.13.