Computers in Biology and Medicine 43 (2013) 73–83

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Effective FCM noise clustering algorithms in medical images

S.R. Kannan [a,*], R. Devi [a], S. Ramathilagam [b], K. Takezawa [c]

[a] Department of Mathematics, Pondicherry Central University, India
[b] Department of Mathematics, Periyar Government College, Tamil Nadu, India
[c] NARC, Tsukuba, Japan

ARTICLE INFO

Article history:
Received 19 May 2011
Accepted 21 October 2012

Keywords:
Noise clustering
Entropy FCM
Medical images
Segmentation

ABSTRACT

The main motivation of this paper is to introduce a class of robust non-Euclidean distance measures for the original data space to derive new objective function and thus clustering the non-Euclidean structures in data to enhance the robustness of the original clustering algorithms to reduce noise and outliers. The new objective functions of proposed algorithms are realized by incorporating the noise clustering concept into the entropy based fuzzy C-means algorithm with suitable noise distance which is employed to take the information about noisy data in the clustering process. This paper presents initial cluster prototypes using prototype initialization method, so that this work tries to obtain the final result with less number of iterations. To evaluate the performance of the proposed methods in reducing the noise level, experimental work has been carried out with a synthetic image which is corrupted by Gaussian noise. The superiority of the proposed methods has been examined through the experimental study on medical images. The experimental results show that the proposed algorithms perform significantly better than the standard existing algorithms. The accurate classification percentage of the proposed fuzzy C-means segmentation method is obtained using silhouette validity index.

1. Introduction

Diagnostic imaging is a precious tool in medicine nowadays. There are different medical image acquisition techniques such as magnetic resonance imaging, ultrasound, X-ray computer tomography, single photon emission tomography, positron emission tomography etc. MRI [19,22] is a superior, commonly used medical imaging technique. It provides detailed images of living tissues and has advantages over other imaging techniques for both brain and breast studies [12]. The registration of magnetic resonance images of the brain and breast has recently become an extensively used research tool [1,4] with the potential to enter usual medical use. In the analysis of medical images for diagnosing the diseases, segmentation is often essential as a preliminary stage. Successful numerical outcome in segmented medical image can assist physicians to study and detect the structure and function of the body in both health and disease. Accurate segmentation is necessary for detecting the diseased spot in medical images. Human error occurred while segmenting the medical images manually due to the intrinsic nature of the images. Also manual segmentation is a challenging and time consuming task. Therefore, computer aided segmentation is very significant to find out effective results in medical images. The brain and breast organs in human body have a convoluted structure and its computer aided segmentation is very essential for detecting normal and abnormal areas, in order to follow appropriate treatment. But, the medical images always comprise substantial ambiguity, unknown noise, partial volume effect and intensity inhomogeneity; these imaging artifacts degrade the segmentation process. The noises and other imaging artifacts cause serious misclassification and overlapping tissues when the computer aided analysis in medical image segmentation [6,11,26]. Selection of a suitable approach to a segmentation problem can be a complicated dilemma. Clustering is a popular unsupervised classification method and has found many applications in pattern classification and image segmentation. Clustering algorithm attempts to classify a voxel to a tissue class by using the notion of similarity to the class. Since there is no information given about the underlying data structure or the number of clusters, there is no single solution to clustering; neither is there a single similarity measure to differentiate all clusters. Therefore for this reason there is no theory which illustrates clustering uniquely. This paper has used the paradigm of fuzzy clustering which is based on the elements of fuzzy set theory for segmenting medical images. Fuzzy based clustering methods [2,3,9,10,21,30,32] have attracted more attention for image segmentation techniques, because they gathered more information from the image [26]. Because of Euclidian distance based objective function of standard FCM, it works well in clustering the noise free data and it fails to

* Corresponding author. Tel.: +919865707773.
E-mail address: srkannan_gri@yahoo.co.in (S.R. Kannan).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2012.10.002

cluster the dataset degraded by noise. In order to improve the robustness of the FCM algorithm [18,11,31,25], variants of FCM based algorithms are derived by replacing the squared-norm with other measures [7,29] to cluster more general datasets. Though some applications of FCM to real clustering problems have proved the good characteristic of this algorithm with respect to stability and partition quality, it is well known in the literature that FCM is not robust against outliers and noise [17,5,14,12]. Noise is a random error or a parasite, which is an invalid value but outlier is valid; an outlier is an observation of the data that deviates from other observations so much that it arouses suspicions that it was generated by a mechanism different from most part of the data [28]. Even a single outlier element can totally spoil the least squares estimate and thus the outcome of least square based clustering methods such as hard C-means (HCM) and the fuzzy C-means algorithm (FCM) [8]. Because of these reasons, Ohashi [24] made an effort to deal with the noise sensitivity in fuzzy C-means algorithms. Regrettably, his effort seems to be virtually unidentified in the clustering community. Dave proposed the idea of a noise cluster to deal with noisy data in the clustering approach known as NC [13] to modify the objective function directly. The NC clustering method of Dave characterizes the fixed noise distance δ, which is a constant parameter that cannot be executed in the clustering process in a fully automatic way for all datasets. Though some solutions [15,16] have been introduced in the literature to automatically choose the optimal value of noise distance, the estimation of this value is still an open-problem. In this work we introduce a robust technique for this noise distance to deal with the noisy data which has proven to be effective without prior knowledge about the data.

To eradicate the shortcomings of recent fuzzy C-means algorithms in data clustering and medical image segmentation, this paper presents three new robust fuzzy C-means methods based on kernel-induced distance, effective noise distance, entropy and an additional term for clustering the breast and brain medical images. The robust objective functions of proposed FCMs are primarily developed for improving the robustness to get meaningful clusters and robustness to noise and outliers. The additional term and entropy term have been introduced in the robust fuzzy C-means in order to deal with the uncertainty in large datasets and improve the strength of the clustering result. And we also introduce an effective cluster initialization algorithm for removing the blindness of random initialization in order to reduce the computational time of the clustering process. The proposed algorithms are compared with existing algorithms to explore the accuracy of our proposed approach through synthetic image which is corrupted by Gaussian noise, benchmark dataset and medical images.

The rest of the paper is organized as follows: Section 2 presents the notions of introducing the proposed approach. The effective proposed algorithms are presented in Section 3. Section 4 introduces the prototypes initialization method. The experimental results and validation of proposed algorithms are provided in Section 5. Finally, the conclusions are summarized in Section 6.

2. Notions

In this section we present some concepts that are needed later on. In particular we review some algorithms for fuzzy clustering with entropy and noise cluster concept especially in fuzzy C-means. The FCM clustering algorithm is defined in terms of the minimization problem. In the case of FCM, the solutions must satisfy the following constraints $u_{ik}\varepsilon[0,1]$, $1 \le i \le n$, $1 \le k \le c$.

$$\sum_{i=1}^{n} u_{ik} > 0, \quad k\varepsilon\{1,2,\cdots,c\} \tag{1}$$

$$\sum_{k=1}^{c} u_{ik} = 1, \quad 1 \le i \le n \tag{2}$$

Constraint (1) guarantees that no cluster is empty. Condition (2) ensures that the sum of membership degrees of each datum is equal to 1. The FCM method is defined in terms of the optimization of the following objective function:

$$J_{FCM}(U,V) = \sum_{i=1}^{n} \sum_{ki=1}^{c} (u_{ik})^m \|x_i - v_k\|^2 \tag{3}$$

where $m\varepsilon(1,\infty)$ is a weighting exponent that is called a fuzzifier. In fuzzy clustering analysis, the number of clusters and the level of fuzziness need to be identified before clustering. If the clustering algorithm does generate relatively different groups, then the analysis should be performed with upper and lower values of level of fuzziness since the information gain would be important. In information theory, the Shannon entropy is a measure of the uncertainty associated with the random variable. The concept of fuzzy entropy is an extension of Shannon entropy. The concept of fuzzy entropy was introduced in 1972 and after that many modifications [23] were introduced in original fuzzy C-means with the entropy term. The term entropy is similar to the average information content of a source and is defined as follows:

$$E(U) = -\sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \log u_{ik} \tag{4}$$

The entropy term attains its maximum value when there is a uniform distribution of membership grades (equal to $1/c$). Based on these reasons, entropy term is used in clustering algorithms to get the additional information of the data which provides better partition in the clustering result. Thus researchers like to employ some form of entropy in the objective function of the clustering algorithm. The entropy based FCM method is defined in terms of the optimization of the following objective function:

$$J_{EFCM}(U,V) = \sum_{i=1}^{n} \sum_{ki=1}^{c} u_{ik} d^2(x_i, v_k) + \gamma \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \log u_{ik} \tag{5}$$

where the parameter γ plays a similar role as the weighting exponent in FCM algorithm. This is the method that introduces fuzziness into the solution with a term based on entropy and a parameter γ ($\gamma > 0$) that forces the solution to be fuzzy. Here the $\gamma -> 0$, the second term becomes negligible and the algorithm yields a crisp solution. When $\gamma -> \infty$, the clustering process provides a single cluster. The way to solve EFCM is an iterative process; the data set partitioning is carried out through an optimization of the objective function shown above, with the update of membership u_{ik} and the cluster center v_k by

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\exp(d^2(x_i, v_k))}{\exp(d^2(x_i, v_j))} \right)^{1/\gamma} \right)^{-1} \tag{6}$$

$$v_k = \frac{\sum_{i=1}^{n} u_{ik} x_i}{\sum_{i=1}^{n} u_{ik}} \tag{7}$$

The main issue in FCM algorithm is sensitivity to noise and outliers. Data points which are not close to any cluster center are said to be noise or outliers. Outliers are data points which are significantly deviated from the remaining data points. Gathering the noise points as a separate cluster by a specific mathematical method is called noise clustering. The concept of noise clustering algorithm is the introduction of noise cluster that will contain noisy data points. Data points whose distance to all clusters exceeds a certain threshold are considered as noise. This distance is called noise distance. The objective function in the fuzzy C-means algorithm has one major drawback which is that it does