Regularized fuzzy $c$-means method for brain tissue clustering

Z. Hou $^{a,b,*}$, W. Qian $^a$, S. Huang $^b$, Q. Hu $^a$, W.L. Nowinski $^a$

$^a$ Biomedical Imaging Lab, Singapore Bioimaging Consortium, 30 Biopolis Street, #07-01, Matrix, Singapore 138671, Singapore

$^b$ Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore

Received 11 May 2004; received in revised form 16 September 2004

Available online 18 May 2007

Communicated by W. Pedrycz

Abstract

This paper presents a regularized fuzzy $c$-means clustering method for brain tissue segmentation from magnetic resonance images. A regularizer of the total variation type is explored and a method to estimate the regularization parameter is proposed.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Fuzzy $c$-means; Clustering; Regularization; Image segmentation; Spatial modeling; Noise level estimate

1. Introduction

Brain tissue segmentation from magnetic resonance (MR) head images is an important stage in the design of an automatic brain MR image processing system, which could be helpful in neurosurgical planning and clinical diagnosis. A number of techniques have been proposed. Joliot and Mazoyer (1993) used the thresholding method to segment brain MR images. Shareef et al. (1999) developed a technique from LEGION dynamics (locally excitatory globally inhibitory oscillator network) for medical image segmentation. Parametric methods (Liang et al., 1994; Schroeter et al., 1998) assume that the intensity distribution of a tissue follows some statistical distribution, then the process of tissue segmentation reduces to estimate the distribution parameters. Chiou and Hwang (1995) proposed a neural network-based active contour method to segment brain tissue, where the shape of object is often predefined from brain atlas or manual segmentation by expert. The Markov random field model has also received much attention in brain tissue segmentation (Rajapakse et al., 1997).

Of interest in this paper is the fuzzy $c$-means (FCM) clustering (Dunn, 1973; Bezdek, 1981) method which has widely been employed for brain tissue segmentation. Brandt et al. (1994) exploited the FCM method to estimate the volumes of the cerebrospinal fluid, white matter and gray matter in the MR images of hydrocephalic children. Philips et al. (1995) utilized the technique to differentiate boundaries of tumor from edema or hemorrhage. Banerjee et al. (1999) applied the method for segmenting the computed tomography and MR images of the human brain.

The FCM method has been modified in several ways. Li and Mukaidono (1995) proposed a maximum entropy approach to fuzzy clustering, which can be regarded as a kind of regularization of the crisp $c$-means clustering (MiyaMoto and Mukaidono, 1997). A quadratic regularized FCM method was presented in (MiyaMoto and Mukaidono, 1998).

Earlier FCM clustering methods take into account only the intensity information and the data are regarded as separated points without any connections. Needless to say, the intensity information alone is not sufficient to derive the segmentation map. Distant points belonging to different parts can have similar intensities. And the intensities of
2.1. FCM clustering

where the regularizer is the type of total variation.

In this paper, we will present another for a particular class is inversely related to the membership of one pixel/voxel penalty term so that the membership of one pixel/voxel to other methods. In this paper, we will present another regularized FCM clustering for brain tissue segmentation, where the regularizer is the type of total variation.

The paper is organized as follows. In Section 2, firstly, the theory of FCM is briefly reviewed. Then the basic regularization theory is introduced, followed with the generic framework of regularized FCM clustering as well as an example of the total variation type regularization. Section 3 illustrates the proposed regularized FCM method with an application to MR volume data for brain tissue clustering. Also a comparison with the robust FCM (Pham, 2001) is made and presented in this section. Finally, the paper is concluded in Section 4 with a discussion on regularization parameter selection and regularizer design.

2. Method

For completeness, we will briefly review the conventional FCM method. Let \( x = \{ x_1, x_2, \ldots, x_n \} \) denote an observed monochromatic intensity field, where \( n \) is the number of the data and \( x_i \) is the intensity at position \( i \). For convenience, we assume \( x_i \) to be a scalar. However, it is straightforward to apply the FCM clustering to the case of \( x_i \) as a vector, that is, to classify multi-channel data.

2.1. FCM clustering

Let \( c \ (1 < c < n) \) be an integer and \( v = (v_1, \ldots, v_c) \) be a vector of cluster prototypes. Then a fuzzy \( c \)-partition of \( x \) is defined by the \( n \times c \) matrix \( u = \{ u_{ki} \} \) subject to the following constraints:

\[
0 \leq u_{ki} \leq 1,
\]

\[
\sum_{i=1}^{c} u_{ki} = 1
\]

and

\[
0 \leq \sum_{k=1}^{n} u_{ki} < n.
\]

In the above definition, \( u \) is usually called the membership function, and \( u_{ki} \) describes the degree that \( x_k \) belongs to cluster \( i \) (prototype \( v_i \)). In the bi-element logic, the membership takes two values, 0 or 1, which means that \( x_k \) is either or not a member of cluster \( i \). Contrary to the bi-element logic, the fuzzy logic allows the membership to take multiple values, thus \( x_k \) can belong to any cluster with some degree. The last constraint guarantees that \( u \) is non-degenerate, thus the resulting partition is nontrivial.

The cost of the fuzzy partition \( u \) with prototype \( v \) can be defined as

\[
J(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ki}^2 (x_k - v_i)^2.
\]

The conventional fuzzy \( c \)-means (FCM) clustering algorithm (Bezdek, 1981) searches for the optimal partition \( \hat{u} \) and the optimal prototype \( \hat{v} \) by minimizing the cost subject to the constraints on \( u \), which in turn can be solved using the method of Lagrange multipliers. Let

\[
J_L = J(u, v) + \sum_{k=1}^{n} \lambda_k \left( 1 - \sum_{i=1}^{c} u_{ki} \right),
\]

where \( \lambda_k \) are the Lagrange multipliers. From the necessary conditions,

\[
\frac{\partial}{\partial u_{ki}} J_L = 0,
\]

\[
\frac{\partial}{\partial v_i} J_L = 0,
\]

where \( i = 1, \ldots, c \) and \( k = 1, \ldots, n \), we can derive the following equations:

\[
u_{ki} = \left[ \sum_{j=1}^{c} \frac{u_{kj}^2 (x_k - v_j)^2}{(x_k - v_i)^2} \right]^{-1}
\]

and

\[
v_i = \frac{\sum_{k=1}^{n} u_{ki}^2 x_k}{\sum_{k=1}^{n} u_{ki}^2}.
\]

With Eqs. (8) and (9), we can alternatively update \( u \) and \( v \) from an arbitrary partition till the iteration process converges.

The final stage of the FCM clustering is to interpret the membership function, which is usually a “hard” clustering by classifying a voxel to the cluster with the largest membership. It should be noted that this is an information-loss process. The original membership function conveys the partial volume effect that is very important in medical image segmentation.

2.2. Regularization theory and FCM clustering

A number of diverse problems from science to engineering can reduce to solve the equation:

\[
Az = y,
\]

where \( A : Z \to Y \) is the operator with domain \( \mathcal{D}_A \subseteq Z \), \( Z \) and \( Y \) are metric spaces, \( y \in Y \) is given and \( z \in \mathcal{D}_A \) is the variable to be solved. In practice, the measurement can be corrupted by noise:

\[
Az = y + \xi,
\]
where $\xi$ represents the noise process. One can obtain the solution by directly solving the above equations, however, this type of solution will often be unphysically meaningful due to the ill-posed nature of the problem. Tikhonov (1963) proposed to restrict the desired solution with a priori requirements as follows:

$$\min_z \|Az - y\| + \beta \Phi(z),$$

(12)

where $\beta \geq 0$ is the regularization parameter.

The purpose of stabilizing term $\Phi(z)$ is to restrict the admissible solution within the space of smooth functionals and the choice of $\Phi(z)$ is something of art. Very often the derivative is chosen for $\Phi(z)$, which makes the roughness of $z$ and is justified by the fact that the derivative of a smooth function should be bounded. The most popular is the Tikhonov’s stabilizing functionals as defined by

$$\Phi(z) = \int_{\Omega} \sum_{i=0}^{p} p_i(t) \left( \frac{d^2z}{dt^2} \right)^2 dt,$$

(13)

where $p_i(t)$ are nonnegative weighting functionals. It was shown that the stabilizing functionals of the Tikhonov type correspond to either interpolating or approximating splines (Terzopoulos, 1986).

Besides the Tikhonov regularizer, the stabilizer can also be defined in other ways, for example, the maximum entropy proposed to restrict the desired solution with $a$ priori requirements as follows:

$$\min_z \|Az - y\| + \beta \Phi(z),$$

where $\beta \geq 0$ is the regularization parameter.

The purpose of stabilizing term $\Phi(z)$ is to restrict the admissible solution within the space of smooth functionals and the choice of $\Phi(z)$ is something of art. Very often the derivative is chosen for $\Phi(z)$, which makes the roughness of $z$ and is justified by the fact that the derivative of a smooth function should be bounded. The most popular is the Tikhonov’s stabilizing functionals as defined by

$$\Phi(z) = \int_{\Omega} \sum_{i=0}^{p} p_i(t) \left( \frac{d^2z}{dt^2} \right)^2 dt,$$

(13)

where $p_i(t)$ are nonnegative weighting functionals. It was shown that the stabilizing functionals of the Tikhonov type correspond to either interpolating or approximating splines (Terzopoulos, 1986).

Besides the Tikhonov regularizer, the stabilizer can also be defined in other ways, for example, the maximum entropy regularizer (Engl et al., 1996)

$$\Phi(z) = -\sum_{i=1}^{N_i} z_i \log(z_i),$$

(14)

where the image data is normalized so that $\sum_{i=1}^{N_i} z_i = 1$; or the total variation regularizer (Rudin et al., 1992)

$$\Phi(z) = \| \nabla z \|_1,$$

(15)

where $\nabla$ denotes the gradient operator and $\| \cdot \|_1$ represents the $L_1$ norm. The total variation regularizer is very similar to the standard Tikhonov regularizer and in general, the $L_1$ norm could be better in preserving sharp edges. Interested readers can refer to (Tikhonov and Arsenin, 1977) for more on the regularization theory and to (Karl, 2000) for a review of its applications to image restoration.

As mentioned in Section 1, the regularization theory has been applied to the FCM clustering problem. Some examples are listed as follows:

1. Regularization by entropy (Li and Mukaidono, 1995)

$$J_{\text{ent}}(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} (x_k - v_i)^2 + \beta \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} \log u_{ki},$$

(16)

2. Regularization by quadratic term (Miyamoto and Mukaidono, 1998)

$$J_{\text{quad}}(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} (x_k - v_i)^2 + \beta \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki}^2,$$

(17)

(3) Pham’s method (Pham, 2001)

$$J_{\text{Pham}}(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki}^q (x_k - v_i)^2 + \beta \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki}^q \left[ \sum_{f \in N_k} \sum_{m \in M_f} u_{km} \right],$$

(18)

where $q > 1$ is the “fuzziness” parameter, $N_k$ denotes a neighbourhood of pixel/voxel $k$ and $M_i = \{1, \ldots, C\} \setminus \{i\}$.

Miyamoto and Umayahara (2000) has discussed the features of the first two regularizers. However, similar to the standard FCM method, they are still intensity-based. Thus, when the data is polluted by noise, their performance would degrade. To develop more robust algorithms, the spatial effects should be incorporated into the process of clustering as done by the Pham’s method (Pham, 2001).

In this study, we exploit a moving average filter as the regularizer which is described in the following

$$J_{\text{quad}}(u, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki} (x_k - v_i)^2 + \beta \sum_{k=1}^{n} \sum_{i=1}^{c} ((u_{ki} - \bar{u}_{ki})^2,$$

(19)

where $\bar{u}_{ki}$ denotes the average of the fuzzy membership of the neighbouring pixel/voxels. We can see that the regularizer defined in Eq. (19) measures the local variation of the membership functional and is in effect similar to the first order derivative constraint, or the total variation regularizer.

Using the Lagrange multiplier method, we can derive the updating equation of membership as follows

$$u_{ki} = \frac{1 - \sum_{j=1}^{c} \frac{\mu_{kj}^q}{(x_k - v_j)^2 + \beta}}{\sum_{j=1}^{c} \frac{1}{(x_k - v_j)^2 + \beta}},$$

(20)

Note that the regularization term does not contain $v$, thus the updating equation of prototypes is the same as Eq. (9).

Before proceeding to next section, it would be meaningful to understand the effect of $\beta$ on the proposed regularized FCM from two extreme cases: (1) if $\beta = 0$, then Eq. (20) reduces to Eq. (8), which means that the membership is totally determined by the intensity; (2) if $\beta \gg (x_k - v_i)^2$, then $u_{ki} = \bar{u}_{ki}$, which suggests that the membership can totally be determined from neighbours. In general, dependent on the value of $\beta$, the effect of the regularization will oscillate between these two extreme cases.

3. Results

To validate the proposed regularized FCM (regFCM) method, an investigation on brain tissue segmentation was carried out, based on both phantom data (Collins et al., 1998) and patient data. The programme is written
in ANSI C and run in a PC with Intel Pentium IV 3.06 GHz processor and 1 GB RAM.

3.1. Qualitative evaluation

The proposed method has been applied to segment patient data. Fig. 1 (on the left) is an axial spoiled gradient echo pulse (SPGR) scan of a volume data. Before the tissue segmentation, the brain MR volume is preprocessed to remove the skull/scalp, which is usually carried out through morphological operation and connected component analysis, with occasional manual edition, and contrast enhancement through histogram equalization. Fig. 1 (on the right) is the brain tissue after preprocessing.

Fig. 2 shows the surface rendering results of the segmented gray matter by the FCM (left) and the regFCM (right), where the regularization parameter is 400. The effect of noise is clearly visible on the FCM result which contains a number of pulse-like holes. Comparatively, the regFCM result is much smoother and is better in fissure delineation. Similar observation can also be found from the surface rendering results of the segmented white matter as presented in Fig. 3, where on the left is the FCM result and on the right is the regFCM result.

3.2. Quantitative evaluation

Quantitative validation was done through the phantom data which is computer-simulated and widely employed to investigate the performance of brain tissue segmentation algorithms. The phantom data used in this study is normal T1 weighted pulse sequences with three noise levels 3%, 5% and 7%. The data size is 181 × 217 × 181. The present method has been compared with the recently proposed robust FCM (RFCM) method (Pham, 2001).

The accuracy of the three clustering methods is summarized in Table 1, from which we can see the effect of noise on the segmentation result. It can be observed that at 3% noise level the RFCM and the regFCM methods only slightly outperform the FCM method, which suggests that the FCM method is still competitive against other two methods under light noise conditions. With the noise level

Fig. 1. On the left is an axial SPGR scan of a brain volume and on the right is the result after skull/scalp removal and contrast enhancement.

Fig. 2. The surface rendering results of the segmented gray matter by the FCM (left) and the regFCM (right) methods.
increasing to 7%, the accuracy of the FCM method decreases to around 90%, which is about 3% lower than the accuracy of the RFCM or the regFCM methods. As for the RFCM and the regFCM methods, the difference in accuracy between them is marginal.

Besides the accuracy, the speed is another issue concerning the performance of an algorithm. To that end, the computation cost among the three methods is compared and also given in Table 1. Because the FCM method can be formulated based on the gray level histogram of the data, the CPU time taken by the FCM method is significantly lower than those by the RFCM and regFCM methods. For the latter two methods, the regFCM method saves 30% the CPU time on average.

4. Discussion and conclusion

The regularization parameter is critical for the performance of regularization method and can be selected automatically, by the cross-validation method (Wahba, 1977), the L-curve method (Hansen, 1992), etc. Pham (2001) used the cross-validation method for parameter selection, where the CPU time is increased tens times. In general, the strength of the regularization relates to the noise level: the more sever the noise, the larger the regularization parameter. Moreover, if we know or can sufficiently estimate the power of the noise, then we can use this knowledge for the parameter selection, for example, by the discrepancy principle (Morozov, 1984). In our implementation, an empirical method is developed for parameter estimation, which is based on the estimate of noise level in the image.

To estimate the noise level, it is common to select a region of interest (ROI) which is relatively homogeneous and the variance of this ROI can be an estimate of the noise level. For that purpose, we first roughly segment the brain tissue by the standard FCM clustering, then the cluster variances are estimates of the noise level and are further used to derive the regularization parameter. In particular, for the phantom data we can compare the estimated cluster variance with the optimal regularization parameter as shown in Fig. 4, where the noise level ranges from 3% to 9%, the solid line corresponds to the estimated variance of WM (denoted as $S_{WM}$) and the dashed line represents the variance of CSF ($S_{CSF}$). We can see that the variance of WM can be used to predict the regularization parameter. Comparatively, the curve of $\beta$ vs $S_{CSF}$ departs away from linearity. This is due to the sinuous structure of the CSF, which in turn makes the segmentation of CSF more severely subject to the impact of partial volume effect. On the bottom of Fig. 4 is the plot of $S'$ against $S_{WM}$ ($S_{CSF} - S_{WM}$), which can also be used for parameter prediction. It should be pointed out that the standard FCM clustering is intensity-based, hence the implementation can be carried out on the histogram. Consequently, the above estimate would be very fast (<0.5 s in our study).

Fig. 5 presents an example of the impact of parameter on the performance of the regFCM method when applied to the phantom data with 5% noise level. One can see that for the broad range of the parameter (from 50 to 500) the segmented accuracy oscillates in a narrow extent (from 94% to 95%). This fact further justifies the feasibility of parameter selection by the above method.

In the standard FCM clustering, there is usually a “fuzziness” parameter as is also present in the RFCM method (Eq. (18)). This parameter is sometimes critical in cluster-

Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>3%</th>
<th>5%</th>
<th>7%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM</td>
<td>0.963</td>
<td>0.937</td>
<td>0.901</td>
</tr>
<tr>
<td>RFCM</td>
<td>0.964</td>
<td>0.948</td>
<td>0.934</td>
</tr>
<tr>
<td>regFCM</td>
<td>0.964</td>
<td>0.948</td>
<td>0.933</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>CPU time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM</td>
<td>&lt;0.5</td>
</tr>
<tr>
<td>RFCM</td>
<td>49.0</td>
</tr>
<tr>
<td>regFCM</td>
<td>33.5</td>
</tr>
</tbody>
</table>

Fig. 3. The surface rendering results of the segmented white matter by the FCM (left) and the regFCM (right) methods.
However, similar to earlier regularization methods (Eqs. (16) and (17)), the present formulation has set the parameter to a fixed value in order to avoid the nonlinearity arising in solving the membership function. This would be a drawback compared with respect to the RFCM method.

Regularized solution can be affected by Gibbs oscillations, which are produced by the truncation of Fourier series in the case of discontinuous functions. There have been efforts towards preserving the discontinuity in the design of the regularization model. Li (1995) presented a discontinuous adaptive smoothness model as defined in the Euler equation with constraints on adaptive interaction functionals. This result could provide guidance on the selection of the regularizers. Actually, the critical part in regularization methods is the design of the regularizer. The RFCM method is essentially a type of regularized FCM clustering. From our simulations for 3D brain tissue clustering, the proposed regularized FCM method could attain similar segmentation accuracy with smaller computation cost.

In summary, we have presented a regularized FCM clustering with the regularizer of the total variation type. The proposed regularized FCM method was validated through clustering the brain tissue from MR volume data and compared with the RFCM method. It turns out in our experiments that the proposed method could attain similar segmentation accuracy with about 30% reduction in computation cost.

Acknowledgements

The authors thank the reviewers for their very constructive comments. They also thank the Nagasaki University School of Medicine, Japan, for providing the patient data. This study is supported by the Biomedical Research Council, Agency for Science, Technology and Research, Singapore.

References


