Information-Based Cost Function for a Bayesian MRI Segmentation Framework

David Cárdenas-Peña[1(✉)], Alvaro A. Orozco[2], and Germán Castellanos-Dominguez[1]

[1] Universidad Nacional de Colombia, Manizales, Colombia
dcardenasp@unal.edu.co
[2] Universidad Tecnológica de Pereira, Pereira, Colombia

Abstract. A new information-based cost function is introduced for learning the conditional class probability model required in the probabilistic atlas-based brain magnetic resonance image segmentation. Aiming to improve the segmentation results, the α-order Renyi's entropy is considered as the function to be maximized since this kind of functions has been proved to lead to more discriminative distributions. Additionally, we developed the model parameter update for the considered function, leading to a set of weighted averages dependant on the α factor. Our proposal is tested by segmenting the well-known *BrainWeb* synthetic brain MRI database and compared against the log-likelihood function. Achieved results show an improvement in the segmentation accuracy of $\sim 5\%$ with respect to the baseline cost function.

Keywords: Magnetic resonance imaging · Atlas-based segmentation · Entropy-based optimization

1 Introduction

Techniques of Brain Magnetic Resonance (MR) imaging play a significant role in many medical applications like: *i*) Identification of differences among functional brain structures along the time or space, which may help to model evolution of pathologies (as dementia, Alzheimer, and schizophrenia) [1], *ii*) Building realistic conductivity head models enhancing activity reconstruction accuracy [2,3], *iii*) Extraction of spatial characteristics (as size, shape, and place) allowing to build representative anatomical models of populations [4]. Since most of the above applications demand quantitative analysis and objective interpretation of the properties of brain structures, a reliable and accurate segmentation of the brain regions of interest has to be carried out from the MRIs.

However, region segmentation tasks are far from being an easy task due to the presence of image artifacts and inherent magnetic properties of each structure [5]. In order to overcome these problems, the atlas-based segmentation is usually employed. Here, an intensity template is registered to a target image, in such a way that the resulting spatial transformation allows to propagate information to the coordinates of the target image space [6]. Commonly, this information

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 548–556, 2015.
DOI: 10.1007/978-3-319-23231-7_49

is provided in the form of probabilistic atlases for structures like white matter, gray matter, and cerebrospinal fluid. Subsequently, the Atlas information is combined with tissue classification approaches, where voxels are assigned to a tissue class according to their intensity. To this end, each tissue distribution has to be accurately modeled, for instance by single or a mixture of Gaussians, but using tissue probability maps to weigh the classification according to the Bayes rule.

In general, to map the atlas information and to find the tissue model parameters, a maximum likelihood estimation is carried out by minimizing an introduced cost function. The most common employed function is the negative log-likelihood of the entire voxel set joint probability while expecting each tissue class model to measure the density of voxels within an intensity range [7]. Then, class conditional distribution can be used as a voxel-wise discriminant function so that the classification is performed using the maximum likelihood principle. Nevertheless, the above cost function relies upon the assumption that tissue properties do not overlap significantly; this supposition is far from being realistic.

Bearing the above in mind, this work discusses the use of the Renyi's α entropies as a new cost function for learning each tissue parameters. This sort of function has been proved to lead to more discriminative distributions while maximizing the information entropy inside the provided target image [8]. Additionally, we develop the model update of the tissue parameter for the considered function. Finally, our proposal is tested by segmenting the well-known *BrainWeb* synthetic brain MRI database. For the purpose of comparison, we also use the baseline log-likelihood function as a cost function. As a result, obtained results of segmentation accuracy, measured in terms of the Dice index similarity, show that our proposal outperforms the log-likelihood ($\sim 5\%$).

2 Materials and Methods

Let $\boldsymbol{X}=\{x_r\in\mathbb{R}:r\in\Omega\}$ be a scalar image, where the value r indexes the spatial elements (spels). The probability of obtaining a spel with intensity x_r, given that it belongs to the class $c\in C$, can be written as $P(x_r|l_r=c,\theta)=f_{rc}(\theta)$, where $l_r\in[1,C]$ is the label associated to the r-th spel, f_{rc} is a predefined probability model for the class c and evaluated at x_r, s.t. $\int f_{rc}(\theta)dr=1;\forall c\in[1,C]$, and θ is the set of the model parameters. In practice, $f_{rc}(\theta)$ is assumed as either parametric (Gaussian or Mixture of Gaussians) or non-parametric (Parzen-based) model. Moreover, the prior probability of any spel belonging to the c-th class, regardless of its intensity, can be provided by a spatial varying probability atlas as $P(l_r = c)=b_{rc}\in[0,1]$, s.t. $\sum_{c=1}^{C} b_{rc}=1$.

As a result, the probability of x_r, given set of model parameters, can be obtained by the Bayes theorem as follows:

$$\begin{aligned} P(x_r) &= \sum\nolimits_{c=1}^{C} P(x_r, l_r = c) = \sum\nolimits_{c=1}^{C} P(x_r|l_r = c,\theta)P(l_r = c) \\ &= \sum\nolimits_{c=1}^{C} f_{rc}(\theta)b_{rc} \end{aligned} \tag{1}$$

Consequently, evary image can be segmented into C classes using the *maximum a posteriori* (MAP) criterion, expressed as l_r^*=arg max$_{\forall c}\, P(l_r = c|x_r)$.

In the most common scheme, the set of model parameters θ is found by maximizing the probability of the entire set of voxels under the assumption of having just independent voxels, yielding:

$$P(\boldsymbol{X}) = \prod_{r\in\Omega} \sum\nolimits_{c=1}^{C} f_{rc}(\theta) b_{rc} \tag{2}$$

The probability defined in Equation (2) is equivalent to minimize the cost function, known as the negative log-likelihood of the entire voxel set joint probability, as follows:

$$\mathcal{L}(\boldsymbol{X}) = -\sum_{r\in\Omega} \log\left(\sum\nolimits_{c=1}^{C} f_{rc}(\theta) b_{rc}\right) \tag{3}$$

2.1 Information-Based Cost Function

Instead of using the common log-likelihood as the cost function, we introduce the amount of information contained in the image $\boldsymbol{X}$. To this end, we consider maximizing the α-order Renyi's entropy with respect to the set of parameters θ as follows:

$$\max_{\theta} \mathcal{H}_\alpha(\boldsymbol{X}) \equiv \min_{\theta} \frac{-1}{1-\alpha} \log\left(\int_{\Omega} P^\alpha(x_r)\right) \tag{4}$$

Since the aim of the current work is to evaluate the cost function for estimating the model parameter, we will assume that each tissue class is described by the normal distribution, $\mathcal{N}(,)$, that is:

$$f_{rc}(\theta) = \gamma_c \mathcal{N}(x_r|\mu_c, \sigma_c^2), \tag{5}$$

where μ_c and σ_c^2 are the class mean and variance, respectively. $\gamma_r \in [0,1]$ is the prior probability of any voxel, irrespective of its intensity, to belong to the c-th tissue, and it is subject to $\sum_{c=1}^{C} \gamma_c = 1$. Consequently, the parameter set becomes $\theta = \{\gamma_r, \mu_r, \sigma_c^2\}_{c=1}^{C}$.

2.2 Optimization Framework

For the optimization, we use the Expectation-Maximization (EM) algorithm that attempts to minimize a given energy function $\mathcal{E}$ w.r.t. the parameters θ and a newly introduced distribution $Q = \{q_{rc} \in [0,1]; \forall r \in \Omega, c \in [1, C]\}$:

$$-\mathcal{H}_\alpha \leq \mathcal{E} = -\mathcal{H}_\alpha + \sum_{c=1}^{C} \mathcal{D}_\alpha\left(q_{rc}||P(l_r = c|x_r)\right) \tag{6}$$

This new energy function works as an upper bound on the proposed cost function and it is composed of two terms. The first one consider only the α-order

Renyi's entropy, while the second term, $\mathcal{D}_\alpha(\|)$, is the α-order Renyi's divergence between the posterior probability and the introduced distribution Q. Therefore, the problem in Equation (6) can be rewritten as:

$$\min_{\{\theta,Q\}} \mathcal{E} = \frac{-1}{1-\alpha} \log\left(\int_\Omega p^\alpha(x_r)\right) + \sum_{c=1}^{C} \int_\Omega \frac{1}{\alpha-1} \log\left(\frac{q_{rc}^\alpha}{P^{\alpha-1}(l_r=c|x_r)}\right) \quad (7)$$

For the M-step, the α-divergence is minimized w.r.t. Q as:

$$\min_Q \mathcal{E} \equiv \min_Q \sum_{c=1}^{C} \int_\Omega \frac{1}{\alpha-1} \log\left(\frac{q_{rc}^\alpha}{P^{\alpha-1}(l_r=c|x_r)}\right)$$

that yields to the solution:

$$q_{rc} = P^{(\alpha-1)/\alpha}(l_r=c|x_r), \quad (8)$$

under the assumption that the following restriction holds: $\sum_{c=1}^{C} q_{rc}^{\alpha/(\alpha-1)}=1$.

For the E-step, the energy function $\mathcal{E}$ is minimized w.r.t. the parameters θ. Given the result of the M-step, the second term in the cost function is at minimum zero whenever relation in Equation (8) holds. Additionally, given the following expression:

$$P(x_r) = \frac{P(x_r, l_r=c)}{P(l_r=c|x_r)} = \sum_{c=1}^{C} q_{rc}^{\alpha/(\alpha-1)} \frac{P(x_r, l_r=c)}{P(l_r=c|x_r)},$$

then, the optimization for the E-step is rewritten as:

$$\min_\theta \frac{-1}{1-\alpha} \log\left(\sum_{c=1}^{C} \sum_{r\in\Omega} \frac{P^\alpha(x_r|l_r=c)}{q_{rc}^\alpha}\right)$$

yielding,

$$\min_\theta \frac{-1}{1-\alpha} \log\left(\sum_{c=1}^{C} \sum_{r\in\Omega} \frac{f_{rc}^\alpha b_{rc}^\alpha}{q_{rc}^\alpha}\right) \quad (9)$$

Taking into account that $\alpha\in[0,1]$, the minimization of the function in Equation (9) is equivalent to maximize the argument of the log function as follows:

$$V(\theta) = \sum_{c=1}^{C} \sum_{r\in\Omega} \left(\frac{f_{rc}(\theta) b_{rc}}{q_{rc}}\right)^\alpha \quad (10)$$

Finally, the M-step assigns new parameter values θ, in such a way that the derivatives of V with respect to parameters are zero:

$$\frac{dV(\boldsymbol{X})}{d\theta} = \alpha \sum_{r\in\Omega} \left(\frac{b_{rc}}{q_{rc}}\right)^\alpha f_{rc}^{(\alpha-1)} \frac{df_{rc}}{d\theta}$$

By differentiating Equation (10) with respect to the means μ_c, we obtain the following expression:

$$\frac{dV(\boldsymbol{X})}{d\mu_c} = \alpha \sum_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha} \frac{(x_r - \mu_c)}{\sigma_c}, \tag{11}$$

that is solved for $dV/d\mu_c$=0, resulting in the next updating rule:

$$\mu_c^{(n+1)} = \frac{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha} x_r}{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha}} \tag{12}$$

Likewise, the derivative of V with respect to the variance parameters obtained is follows:

$$\frac{dV(\boldsymbol{X})}{d\sigma_c^2} = \frac{\alpha}{2} \sum_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha} \left(\frac{(x_r - \mu_c)^2}{\sigma_c^2} - 1 \right). \tag{13}$$

Hence, the variance is updated in accordance to the following rule:

$$(\sigma_c^2)^{(n+1)} = \frac{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha} \left(x_r - \mu_c^{(n+1)} \right)^2}{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha}} \tag{14}$$

Following the above derivative scheme, the attained updating function for the prior parameter γ_c is given by:

$$\gamma_c^{(n+1)} = \frac{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha} \mathcal{N}(x_r | \mu_c, \sigma_c^2)}{\sum\limits_{r \in \Omega} \left(\frac{b_{rc} f_{rc}}{q_{rc}} \right)^{\alpha}} \tag{15}$$

3 Experimental Setup

3.1 Image Database Description

A simulated MRI set was used as test data that had been generated with the Internet connected MRI Simulator at the McConnell Brain Imaging Centre in Montreal publicly available at [1]. The pre-computed simulated MRI volumes for normal brain database was employed with the following parameters: $T1$ image modality, $1\,mm \times 1\,mm \times 1\,mm$ voxel size, $\{0, 1, 3, 5, 7, 9\}\%$ noise (relative to the brightest tissue in the images), and intensity non-uniformity (INU) values of 40%. The $T1$ image was simulated as a spoiled FLASH Figure 1.

[1] http://brainweb.bic.mni.mcgill.ca/brainweb/ [9].

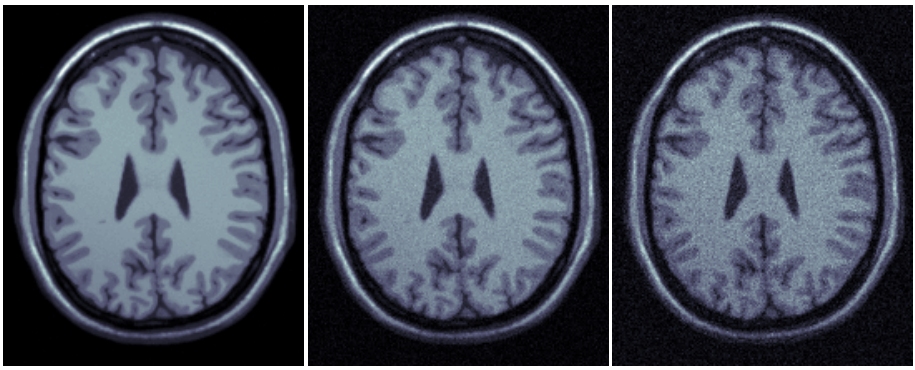

Fig. 1. Simulated MRI sample for the different noise intensities (left: 1%, center: 5%, right: 9%) for two different coronal slices

3.2 Evaluation of Performed Segmentation

Our proposed information-based cost function is employed for segmenting the images of the dataset into five regions. Namely, white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), skull (SK), and scalp (SC). To this end, we make use of the prior probability atlas provided by the SPM software [7]. Since the scope of the work is to evaluate only the cost function for improving MRI partitioning, we will consider only a fixed affine atlas mapping to the target images. Though it is important noting that there are iterative deformable mapping schemes for enhancing the image segmentation. The resulting segmentation is attained using the MAP criterion at each voxel. The accuracy is measured in terms of the average dice index similarity index ($\kappa \in [0, 100]\%$), expressed as follows:

$$\kappa = \frac{1}{C} \sum_{c=1}^{C} \frac{2 \times TP_c}{2 \times TP_c + FP_c + FN_c}, \tag{16}$$

being TP_c the number of true positives, FP_c false positives, and FN_c false negatives, for the c-th tissue.

Firstly, we analyze the influence of the α factor in the optimization process. In Figure 2, the introduced information-based cost function versus the number of iterations is depicted for several α values. As expected, the relation between the entropy order is $\mathcal{H}_\alpha(\boldsymbol{X}) < \mathcal{H}_\beta(\boldsymbol{X}); \forall 0 < \beta < \alpha < 1$. This inequation means that the larger the entropy order, the smaller the entropy value. Moreover, we get that the EM algorithm converges faster for the case of smaller orders.

The evaluation of the influence of the entropy on the segmentation accuracy is given in the Figure 3 showing the curves of DI versus α for the considered noise intensities. As seen, the α factor leads the segmentation results so that for very small or very large values the amount of misclassifications is greater

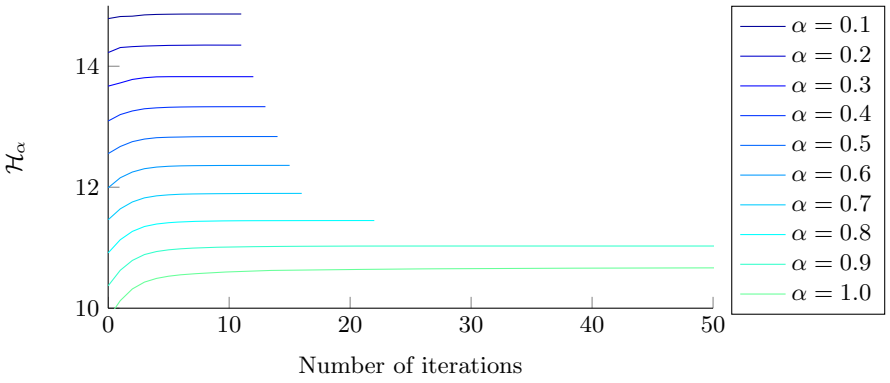

Fig. 2. α-order Renyi's entropy versus the number of iteration for the optimization procedure, for several α values and a given image in the dataset

than for values of the mid range. Moreover, the highest segmentation accuracy is achieved at α=0.5. It is also important noting that the algorithm performance decreases as the noise level increases. This result may be explained mainly due to variations in the tissue distribution because of the high noise intensity.

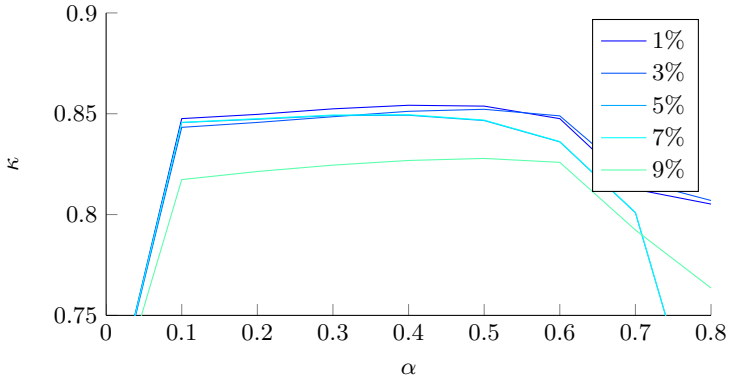

Fig. 3. Average Dice similarity index versus the entropy order for available image noise intensities.

Finally, we compare attained results against the well-known log-likelihood cost function in Equation (3). The achieved segmentation accuracy is computed for optimal α=0.5 for each considered structure. As shown in Table 1, the proposed Renyi's entropy outperforms the other compared cost function – the baseline log-likelihood.

Table 1. Dice index for each structure at optimal $\alpha = 0.5$

	Proposed Entropy					Baseline Log-likelihood				
Noise	1%	3%	5%	7%	9%	1%	3%	5%	7%	9%
Average	85.38	85.22	84.66	84.69	82.78	81.08	80.25	79.94	79.73	77.88
SC	88.61	89.28	86.86	86.92	88.11	84.43	84.97	82.17	82.87	83.15
SK	67.73	67.90	68.40	68.46	68.46	63.16	63.28	64.17	64.26	63.65
CSF	73.03	73.15	71.66	71.57	68.59	68.52	68.60	67.30	67.28	64.35
GM	89.60	88.50	89.73	89.72	84.15	84.88	84.18	85.39	84.89	79.27
WM	90.22	89.12	90.51	90.53	85.77	85.72	85.08	85.85	86.11	80.99

4 Concluding Remarks

In the current paper, we have discussed the use of information-based measures into the parameter optimization scheme for MRI segmentation. In particular, we introduce the α-order Renyi's entropy as a new cost function for finding the tissue distribution parameters under the assumption of normally distributed classes. Additionally, we have developed the model of updating equations for an EM-based optimization using the considered function. As a result, parameters are updated from weighted averages (see Equations (12), (14) and (15)), where the influence of the r-th voxel for each parameter is $(b_{rc}f_{rc}/q_{rc})^{\alpha}$.

As seen in Figure 2, we have proved the relationship between two different entropy orders. We show that in the range $[0, 1]$, the larger the order, the smaller the information measure. In fact, the maximum possible value for the Renyi's entropy is achieved when $\alpha = 0$, corresponding to $\mathcal{H}_0 = -\log\left(\frac{1}{|\Omega|}\right)$. Additionally, we found a proportional relationship between the order and the algorithm convergence iterations. The above is due to the influence of α in the probability values. As α tends to zero, the entropy tends to weight all the events more evenly, regardless their probability, i.e., $(b_{rc}f_{rc}/q_{rc})^{\alpha} \to 1; \forall r \in \Omega, c \in [1, C]$. On the other hand, for large α values, the entropy is determined by events with higher probabilities.

Regarding the segmentation accuracy, from Figure 3, we found the optimal order at $\alpha = 0.5$. Additionally, we show that the larger the noise intensity, the larger the number of misclassifications. Here, it has to be highlighted that Parzen-based estimation of the class conditional probability distribution may overcome this issue. Nevertheless, such a test is out of the scope of the current paper. Then, we compared our proposal against the log-likelihood as the baseline approach. Achieved results for the optimal entropy order, in Table 1, show that our scheme outperforms the baseline, since the obtained parameters for the entropy function are more discriminative than those for the log-likelihood.

Finally, as future work two main research lines are proposed. As there are iterative schemes integrating the mapping parameters of the prior image distribution atlases into the optimization process, we plan to extend the entropy cost function for finding a proper model update for such mapping parameters. Next, we have performed all of the experiments under the assumption of normally

distributed classes. Therefore, we will test our approach for the class conditional models as mixture of Gaussians or Parzen.

Acknowledgments. This work was supported by *Programa Nacional de Formación de Investigadores "Generación del Bicentenario"*, 2011/2012 funded by COLCIENCIAS and the research project number 111065740687, both funded by COLCIENCIAS.

References

1. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage **21**(1), 46–57 (2004)
2. Valdés-Hernández, P.A., von Ellenrieder, N., Ojeda-Gonzalez, A., Kochen, S., Alemán-Gómez, Y., Muravchik, C., Valdés-Sosa, P.A.: Approximate average head models for EEG source imaging. Journal of neuroscience methods **185**(1), 125–32 (2009)
3. Strobbe, G., Cárdenas-Peña, D., Montes-Restrepo, V., Van Mierlo, P., Castellanos-Dominguez, G., Vandenberghe, S.: Selecting volume conductor models for EEG source localization of epileptic spikes: preliminary results based on 4 operated epileptic patients. In: First International Conference on Basic and Clinical multimodal Imaging (BaCI), vol. c (2013)
4. Ericsson, A., Aljabar, P., Rueckert, D.: Construction of a patient-specific atlas of the brain: application to normal aging. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 480–483. IEEE, May 2008
5. Cardenas-Pena, D., Martinez-Vargas, J.D., Castellanos-Dominguez, G.: Local binary fitting energy solution by graph cuts for MRI segmentation. In: Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society, Conference, vol. 2013(2), 5131–5134, January 2013
6. Lötjönen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage **49**(3), 2352–2365 (2010)
7. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage **26**(3), 839–851 (2005)
8. Erdogmus, D., Xu, D., Hild, Kenneth, I.: Classification with EEC, divergence measures, and error bounds. In: Information Theoretic Learning, pp. 219–261 (2010)
9. Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Pike, G.B., Evans, A.C.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. NeuroImage **5**, 425 (1997)