New Approaches to Robust Gaussian Mixture Estimation for Brain MRI

Philippe SCHROETER, Jean-Marc VESIN
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
e-mail: schroep@ltssg4.epfl.ch

ABSTRACT

This paper presents two new methods for robust parameter estimation of mixtures in the context of MR data segmentation. The head is constituted of different types of tissue that can be modeled by a finite mixture of multivariate Gaussian distributions. Our goal is to estimate accurately the statistics of desired tissues in presence of other ones of lesser interest. These latter can be considered as outliers and can severely bias the estimates of the former. For this purpose, we introduce a first method, which is an extension of the EM-algorithm, that estimates parameters of Gaussian mixtures but incorporates an outlier rejection scheme which allows to compute the properties of the desired tissues in presence of atypical data. The second method is based on genetic algorithms and is well suited for estimating the parameters of mixtures of different kind of distributions. Experiments on synthetic and real MR data show that accurate estimates of the gray and white matters parameters are computed.

1 INTRODUCTION

Since more than two decades, medical radiology has developed imaging techniques to observe the inside of the human body. These techniques include today Magnetic Resonance Imaging (MRI), Computer Tomography (CT), echography and isotropic imaging methods. In this context, segmentation based volume representation of organs and their pathology allows a better analysis of lesions and better information exchange between physicians, in particular between radiologists and surgeons for the planning of various surgeries.

The aim of this study is to introduce new robust parameter estimation schemes of a mixture of normal components by using genetic algorithms [1] or by means of a modified version of the standard EM algorithm [2]. Our goal is to estimate accurately the statistics of the gray and white matters which are most of the time modeled by such a mixture in the relevant literature [3, 4]. We will take into account the presence of surrounding tissues, which can be considered as outliers to mixture components. In order to robustify the estimation scheme, we propose to add a class of noise with a uniform distribution to the mixture of normal components. The proposed algorithm increases the stability of the estimation process with respect to the number of classes and guarantees convergence for different initial settings. In the sequel, we consider only T_1-weighted images but the algorithms can easily be extended to support the multi-spectral nature of MR images.

2 ROBUST EM

In the case of finite mixture models, each data x_i is assumed to be drawn from a mixture of a finite number c of components $G_1, \ldots, G_c$ in some proportions $\omega_1, \ldots, \omega_c$. For modeling the different tissues of the brain, we consider that the data are drawn from c Gaussian components. Such a model can handle pure tissues but also mixed tissues (partial volume effects [3, 4]). Because of the limited resolution of MR images and because of the complexity of tissue boundaries, many voxels contain a mixture of tissues. The mixture model is expressed as follows:

$$p(x_i \mid \phi) = \sum_{k=1}^{c} \omega_k p_k(x_i \mid \theta_k) \qquad (1)$$

where $\phi = (\Omega^t, \Theta^t)^t = (\omega_1, \ldots, \omega_c, \theta_1^t, \ldots, \theta_c^t)^t$ is the vector of all unknown parameters and ω_k is the proportion of each component G_k, where $\sum_{k=1}^{c} \omega_k = 1$ and $\omega_k \geq 0$. In the particular case of normal component densities, $p_k(x_i \mid \theta_k)$ is expressed by

$$p_k(x_i \mid \theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x_i - \mu_k)^2\right) \qquad (2)$$

Assuming that the observations x_i are independent, the problem is to estimate the different parameters of the mixture. The standard way to do this is to minimize the log-likelihood function

$$L(\phi) = \sum_{i=1}^{n} \log \sum_{k=1}^{c} \omega_k p_k(\mathbf{x}_i \mid \theta_k) \qquad (3)$$

derived from Eq. (1), using the so-called EM algorithm [2].

In many practical situations, a fraction of data does not belong to any of the c components G_k and thus does not fit to the mixture. Such observations can strongly bias the estimation of the parameters ϕ and need to be given reduced weight in the estimation process. Here, we propose to relax this problem by adding a class of noise to the mixture. For this purpose, we add a new component in the mixture, say G_0, that approximatively models the atypical observations. We assume that the outliers are drawn from a uniform distribution p_0. Thus, in this case, the mixture model becomes

$$p(x_i \mid \Phi) = \sum_{k=1}^{c} \omega_k p_k(x_i \mid \theta_k) + \omega_0 p_0 \qquad (4)$$

where $\Phi = (\omega_0, \phi^t)^t$ is the vector of all unknown parameters. Substituting Eq. (4) to Eq. (1) yields a modified log-likelihood function

$$L_n(\Phi) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{c} \omega_k p_k(\mathbf{x}_i \mid \theta_k) + \omega_0 p_0 \right) \qquad (5)$$

to be minimized.

The effect of adding a class of noise can be understood intuitively as follows. Suppose that x_i is an observation atypical of the mixture. In this case, we have $p(x_i \mid \phi) \approx 0$ and the contribution of the observation x_i to the log-likelihood function is high since $\log(p(x_i \mid \phi)) \rightarrow -\infty$. In the case of normal mixtures, we try to minimize - log of the likelihood function. Thus it is easy to see that each outlier will tend to increase the negative log-likelihood function by a large amount. This effect is much reduced when adding a class of noise. In this case $p(x_i \mid \Phi) \approx \omega_0 p_0$, which reduces the effect of the outliers on the negative log-likelihood function.

Addition of G_0 to the mixture modifies only the E step in the EM algorithm, namely the computation of the local probabilities h_{ki} that point x_i belongs to class G_k. Unfortunately, we observed in experiments that this modification greatly impaired the convergence behavior. The maximization of $L_n(\Phi)$ (5) is in consequence taken care of by the genetic algorithm (see Section 3). However, we obtain an EM solution, called EM-N, to the maximization problem of $L_n(\Phi)$ in the binary case where an observation x_i either belongs to the component G_0 or to the mixture composed of the $G_1, \ldots, G_c$ normal components. It means that the values of h_{0i} will be equal to 1 if $x_i \in G_0$ and $h_{ki} = 0$ for $k = 1, \ldots, c$. In the binary case, we thus need to define an outlier rejection scheme where each outlier fully belongs to the class of noise. For detecting the outliers, in the case of Normal components, we simply use the Mahalanobis distance d_{ki} and decide that an observation is atypical of each component G_k if $d_{ki} > \alpha, k = 1, \ldots, c$, where in the univariate case, α can take values between 1 and 2.5.

3 GENETIC ALGORITHM

Genetic algorithms (GA) are a class of robust stochastic search and optimization procedures based on the Darwinian theory of evolution, and whose basic principles were first introduced by Holland in 1962 [5]. Since then, they have been employed with success in a variety of applications ranging from combinatorial optimization to image enhancement. A GA typically consists of a population of suitably encoded solutions (called chromosomes) to the problem at hand, an evaluation function allowing to rank the solutions, operators permitting to create new solutions starting from already existing ones, and an evolution strategy in order to create successive populations (i.e. generations). A description of the GA and of its operators (crossover and mutation) can be found in [5]. Here, We chose to use the scheme introduced in [6], that is, to consider as chromosomes the usual parameter vectors, and to adapt the operators of mutation and crossover. Mutation must remain a mean for a population to explore the parameter space, and crossover must constitute a way to merge the good characteristics of two chromosomes. This is why we defined mutation of a chromosome $\vec{v}_d$ into a chromosome $\vec{v}_a$ by:

$$\vec{v}_d \longmapsto \vec{v}_a = \vec{v}_d + \nu, \qquad (6)$$

where ν is a random vector with independent Gaussian components, whose variances must be appropriately chosen. Crossover of two chromosomes $\vec{v}_f$ and $\vec{v}_m$ yielding a chromosome $\vec{v}_s$ can be defined in the following way:

$$\vec{v}_f, \vec{v}_m \longmapsto \vec{v}_s = (1 - \lambda)\vec{v}_f + \lambda \vec{v}_m, \qquad (7)$$

where λ is a random variable uniformly distributed in $[0, 1]$.

We applied this type of GA to the estimation problem at hand in two ways: First, minimization of the log-likelihood defined by Eq. (5) and second, minimization by a direct least-squares fit of the model density function to the data histogram. Either of these two approaches was successful, with only a few parameters to tune and a good robustness with respect to initial conditions.

In order to differentiate between the different estimation schemes, we introduce the following acronyms: GA for the genetic algorithm (Gaussian mixture only), GA-N with addition of the class of noise, HGA and HGA-N for the same algorithms but applied to the histogram.

4 EXPERIMENTS ON SYNTHETIC DATA

In this section we will apply the EM and genetic algorithms to synthetic data in order to show their efficiency to estimate the parameters of a normal mixture. In the presence of outliers, the need of the modified model with a class of noise will be demonstrated. In the first row of experiments, the parameters are estimated by minimizing the log-likelihood function (Eq. (5)). We have also

considered the approach (see for instance [4]) consisting in directly fitting a mixture model to the histogram of the data set in a least-squares sense.

The different parameters of the synthetic model are set so as to approximate the characteristics of the real data on which the parameter estimation schemes will be applied. A typical set of data $X_1 = \{x_{11}, x_{12}, \ldots, x_{1n}\}$ contains n observations x_{1i} drawn from a mixture of c Gaussian components G_k with scalar means μ_k and variances σ_k^2. A percentage ω_0 of outliers is substituted to the normal observations. They are drawn from a uniform population with a range equal to double the range the uncorrupted data set. We also consider a second set, say X_2, which is simply the histogram of X_1. Here, we set the parameters as follows ($c = 3$): $(\omega_0, \omega_1, \omega_2, \omega_3) = (0.1, 0.21, 0.37, 0.32)$, $(\mu_1, \mu_2, \mu_3) = (120, 220, 280)$ and $(\sigma_1, \sigma_2, \sigma_3) = (40, 25, 20)$.

As observed in [2], the convergence with the EM algorithm is slow and is typically slowed by a poor choice of $\phi^{(0)}$. It may be even the case that the sequence of estimates $\phi^{(k)}$ diverges in cases where the likelihood is unbounded on the edge of the parameter space. For the EM schemes, we give initial estimates $\phi^{(0)}$ and start by computing the E step. The same strategy will be used for the real MR data since we have some *a priori* knowledge of the expected values of the parameters. The genetic algorithm is somewhat less sensitive to the initial conditions since one needs only to specify a range of values for each parameter. The bounds for the different parameters define a convex space, called the acceptance domain, with the condition that the solution is element of this space. In theory, the algorithm can escape from any local minimum since at any mutation we add a Gaussian perturbation which has an infinite support. Here, we chose the initial chromosomes so as to cover regularly the acceptance domain. The GA guarantees that at each iteration the values of the log-likelihood function cannot increase.

Table 1 shows the estimates of the last two modes obtained with the different estimation schemes applied 30000 data. The histogram of this mixture can be observed in Figure 1 (dotted line) together with the estimated mixtures superimposed in solid line. For comparison purposes, one of the initial chromosome of the GA algorithm was taken to be $\phi_{(0)}$. Furthermore, the acceptance domain was defined by the following boundary values: $\mu_{min} = 50, \mu_{max} = 350, \sigma_{min} = 5, \sigma_{max} = 50, \omega_{min} = 0.05$ and $\omega_{max} = 0.9$.

As expected, the simple EM and GA estimation schemes broke down. In case of the EM estimation, the third Gaussian of the model was estimated to have excessive mean and standard deviation, which is due to the fact that it tried to approximate the class of noise. Thus, the second Gaussian had to approximate two components of the mixture (upper left of Figure 1). Also, in case of the GA, the means and standard deviations of two components adjusted in such a way as to incorporate the outliers.

Φ	ω_2	ω_3	μ_2	μ_3	σ_2	σ_3
Φ	0.37	0.32	220.0	280.0	25.0	20.0
$\Phi^{(0)}$	0.33	0.33	170.0	320.0	20.0	15.0
EM	0.71	0.03	249.1	465.1	41.3	56.1
EM-N	0.36	0.34	218.9	279.6	25.1	21.0
GA	0.56	0.17	240.1	300.0	37.9	44.9
GA-N	0.32	0.36	216.7	277.1	23.9	22.5
HGA	0.35	0.37	216.1	277.6	24.7	23.3
HGA-N	0.36	0.32	217.9	278.7	26.2	21.1

Table 1: Comparison between the estimated parameters for the different estimation schemes in the case of $c = 3$ normal components corrupted by a certain amount of noise ($\omega_0 = 0.1$).

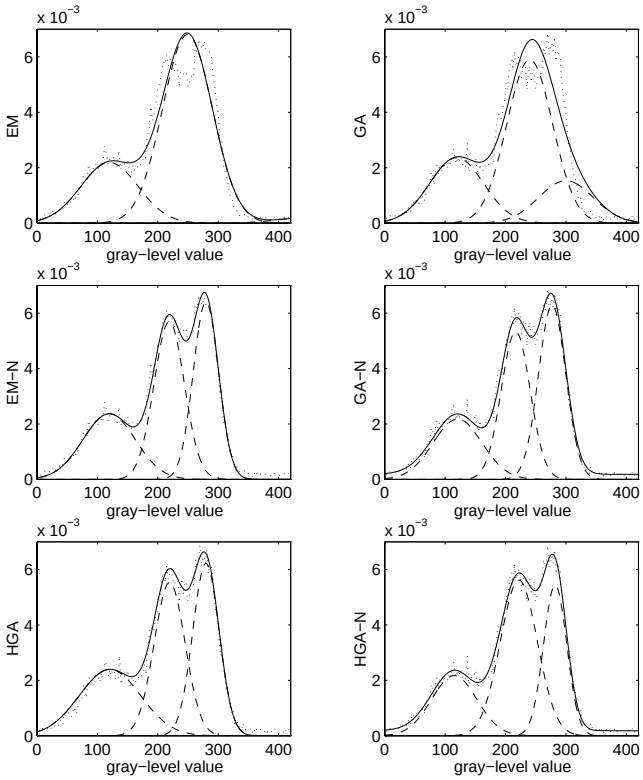

Figure 1: Histogram of the data obtained with the parameters of Table 1 (dotted line) and the results obtained with the different estimation schemes. The Gaussian components are drawn in dashed line and the mixture in solid line.

The EM-N, HGA, GA-N and HGA-N converged to correct estimates. One should note that the GA-N and HGA-N estimated accurately the class of noise whereas the binary EM-N algorithm is somewhat weaker. However, this does not influence much the estimation of the means and standard deviations.

5 EXPERIMENTS ON MR DATA

We applied the robust estimation schemes to real MR brain data for estimating the gray (GM) and white (WM) matter statistics. Here, the set of data on which the estimation is applied is obtained by performing a crude segmentation (volume of interest (VOI)) of the brain that includes mostly the GM and WM. This step can be effectuated automatically by using some *a priori* knowledge of the approximate position of the brain in a subset of frames. This operation eliminates part of the surrounding tissues of the brain but does not need to be precise since robust estimations will be used henceforward. Robust estimation allows to skip a manual segmentation of the brain as used in [4] for estimating the GM and WM statistics.

We applied the different estimation schemes to more than 150 sets of T_1-weighted MR data coming from patients of ages ranging from 20 to 65 years. For illustration purposes, five frames of one set were selected, corresponding to about 115000 data points (within the mask of the head). Due to the low inter-frame variability of MR brain data, this number of frames was deemed sufficient to represent the global statistics of the GM and WM. The different estimation results are displayed in Table 2 for $c = 4$ components. Figure 2 shows the mixtures densities obtained with the EM-N, GA-N, HGA and HGA-N schemes superimposed on the histogram applied to the data set.

Φ	μ_3	μ_4	σ_3	σ_4
Φ	304.2	395.3	37.4	29.3
$\Phi^{(0)}$	330.0	430.0	40.0	25.0
EM-N	307.9	400.4	53.3	34.1
GA-N	303.7	397.3	41.3	34.0
HGA	308.3	397.1	45.8	35.1
HGA-N	304.2	393.4	42.2	34.6

Table 2: Comparison between the estimated parameters resulting from the estimation schemes applied on a subset of 5 frames of a T_1-weighted MR data set with $c = 4$ components. ϕ are the parameters estimated within the manually segmented masks of the GM and WM.

In order to compare the estimation results, we determined manually two masks corresponding to the GM and the WM, within which we estimated the means and standard deviations. These values do not represent exactly the actual statistics of the GM and WM since inter-tissues boundaries are very irregular and not well defined. Also, the *partial volume effect* is not taken into account. However, these values are sufficiently well estimated for comparison between the different methods.

6 CONCLUSION

In this paper, robust estimation schemes were introduced for estimating parameters of mixtures of Gaussian components. These schemes allow to compute accurate estimates of specific tissues from MR data, which are in our case the statistics of the gray and white matters.

For all the 150 data sets, we observed that the modes of the histogram corresponding to the GM and WM were very well fitted. In addition, these estimation schemes were incorporated into a 3-D brain segmentation algorithm. Visual inspection of the 3-D results were found very satisfactory by physicians. Also, from a quantitative view point, the documented decrease of total brain and GM volumes with age was effectively observed.

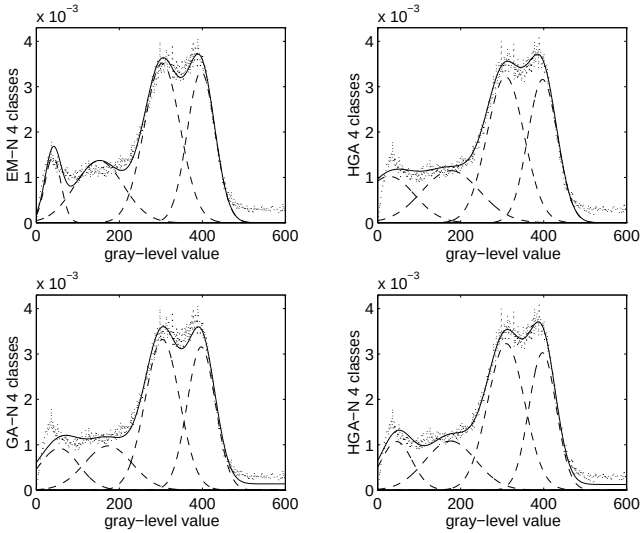

Figure 2: Estimation results obtained with the EM-N, HGA, GA-N and HGA-N schemes (4 classes).

References

[1] J.H. Holland. *Adaptation in Natural and Artificial Systems*. Univ. of Michigan Press, Ann Arbor, MI, 1975.

[2] G.J. McLachlan and K.E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, Inc., New York, 1988.

[3] Z. Liang, J.R. MacFall, and D.P. Harrington. "Parameter estimation and tissue segmentation from multispectral MR images". *IEEE Transactions on Medical Imaging*, Vol. 13, pp. 441–449, September 1994.

[4] P. Santago and H.G. Gage. "Quantification of MR brain images by mixture density and partial volume modeling". *IEEE Transactions on Medical Imaging*, Vol. 12, pp. 566–574, September 1993.

[5] J.H. Holland. "Outline for a theory of adaptive systems". *J. ACM*, Vol. 3, pp. 297–314, July 1962.

[6] J.M. Vesin. "Genetic algorithms in the continuous space for recursive adaptive filter design". In *Proc. 7th European Sig. Proc. Conf. EUSIPCO-94*, pp. 732–735, Edinburgh, U.K., September 1994.