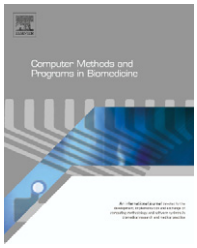
ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Bayesian mixture models of variable dimension for image segmentation

*Adelino R. Ferreira da Silva**

Electrical Engineering Department, Universidade Nova de Lisboa, Rua Dr. Bastos Goncalves, n. 5, 10A, 1600-898 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 28 January 2008
Received in revised form
28 May 2008
Accepted 30 May 2008

Keywords:
Bayesian estimation
Mixture models
Reversible jump MCMC
Dirichlet process mixture model
Magnetic resonance imaging
Automatic brain MRI classification

ABSTRACT

We present Bayesian methodologies and apply Markov chain sampling techniques for exploring normal mixture models with an unknown number of components in the context of magnetic resonance imaging (MRI) segmentation. The experiments show that by estimating the number of components using sample-based approaches based on variable dimension models the discriminating power of the estimated components is improved. Two different MCMC methods are compared to perform the segmentation of simulated magnetic resonance brain scans, the reversible jump MCMC model and the Dirichlet process (DP) mixture model. The preference given to the Dirichlet process mixture model is discussed.

1. Introduction

Statistically based classification methods have been widely used in image analysis and applied statistics [1,2]. In particular, finite mixture models have been adapted to image intensity estimation and classification by many researchers and practitioners. Normal mixture models provide a flexible environment for statistical modeling in the context of image clustering and segmentation. The expectation–maximization (EM) algorithm [3] is the most commonly used algorithm in non-Bayesian analysis of mixture models [4], namely in maximum likelihood computations. However, EM procedures for mixture modeling exhibit several important weaknesses which compromise the accuracy and robustness of the estimates. The EM optimization procedure requires prior knowledge of the number of components in the mixture. In maximum likelihood computations, the EM algorithm often fails to converge to the major mode of the likelihood. Besides, the rate of convergence can be very slow. Initialization is critical in EM computations, because the likelihood surface tends to have multiple nodes. Moreover, for certain types of mixtures the likelihood is unbounded and therefore the resultant likelihood estimator yields meaningless estimates [4,5].

Several methods have been proposed to overcome the limitations of the EM procedure. For example, model-based clustering has been used in many applications. Model-based clustering extends the EM procedure by combining maximum likelihood estimation of finite mixture models with model selection criteria [2]. To select the number of components and specify the model parameterization, the method relies on information criteria, such as the Bayesian Information Criterion (BIC) [6]. In image segmentation applications, the Bayesian model selection criterion takes into account spatial neighborhood information, and is termed pseudo-likelihood information criterion (PLIC) [7]. However, in this kind of applications model-based Gaussian fitting has been

* *Corresponding author*. Tel.: +351 217261320; fax: +351 212948532.
E-mail address: afs@fct.unl.pt.
0169-2607/$ – see front matter © 2008 Published by Elsevier Ireland Ltd.
doi:10.1016/j.cmpb.2008.05.010

found to be unstable and algorithmically non-robust [1]. Additional correction mechanisms of computational nature have been devised to optimize estimates, such as model-based cluster trees [1]. Nevertheless, the main problem lies in the fact that in model-based clustering the number of components and the mixture component parameters are estimated separately.

The methodology of fully Bayesian mixture modeling is the methodology of choice in applied statistics, because it enables the estimation of posterior distributions of objects of interest, such as model parameters and predictive densities and not just optimized estimates. The Bayesian methodology allows for probability statements to be made directly about the unknown parameters. Markov chain Monte Carlo (MCMC) methods offer a flexible family of techniques for Bayesian modeling and inference on mixtures of distributions [8]. MCMC methods in mixture models are preferable on grounds of accuracy and flexibility. Specifically, the *reversible jump Markov chain Monte Carlo* (rjMCMC) method can be used to sample mixture distributions with an unknown number of components. In contrast to EM-based procedures, in rjMCMC the number of components and the mixture component parameters are modeled jointly.

One common criticism against the Bayesian methodology is based on the subjective choice of the prior probabilities and its influence on the estimation of the mixture distributions [9]. A Bayesian non-parametric approach known as Dirichlet process (DP) mixture model has recently gained popularity for mixture modeling by Monte Carlo simulation [10–12]. Bayesian mixture models in which a Dirichlet process prior defines the mixing distribution are used to avoid critical dependence on parametric assumptions, and to robustify parametric models. By employing a prior distribution for mixing proportions, the Dirichlet process prior can handle a countable infinite number of mixture components. Several MCMC computational techniques are available for sampling from the posterior distribution of a Dirichlet process mixture model [11]. For instance, methods based on Gibbs sampling can easily be implemented for models using conjugate prior distributions.

Our motivation for the application of mixture models lies in the development of accurate procedures for the automatic segmentation of magnetic resonance (MR) images of the human brain. Accurate classification of MR images according to tissue type or region of interest has become a critical requirement in diagnosis and treatment planning. In view of its practical importance, automated magnetic resonance imaging (MRI) segmentation has been a highly researched area in recent years. Mixture models are often used in MRI for intensity estimation and classification purposes [13–16]. Clustering methods can be used to partition MR images into different tissue types, and estimate the defining parameters for each tissue class. Typically, healthy brain tissue is classified into three broad tissue classes: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) [14]. Since each class is assumed to be modeled by a single Gaussian distribution, with mean and variance as parameters, tissue classification amounts to the estimation of a three component Gaussian mixture. However, there are several aspects that contribute to make segmentation of MR images into different tissue classes, a challenging task. Due to limitations of the image acquisition process, image intensity values are typically corrupted by non-uniformities and noise [16]. Reliable estimates of the number of clusters in MR images, especially MR images of abnormal brains, may not be available a priori [17]. As a consequence of the finite resolution of the imaging process and the complexity of tissue boundaries, there is a partial volume effect which causes many voxels in MRI images to contain a mixture of more than one tissue type. The partial volume effect causes the distributions of the intensities to deviate from normal [18]. In addition, image noise produces speckled regions and ambiguous tissue boundaries, with direct impact on the accuracy of intensity-based classification procedures.

In this work, we present Bayesian methodologies and apply Markov chain sampling techniques for exploring normal mixture models with an unknown number of components in the context of MRI segmentation. We show that by estimating the number of components using sample-based approaches based on variable dimension models, the discriminating power of the estimated components is improved. This feature of simulating mixture models with an unknown number of components is often referred to as transdimensional simulation [19,20]. The second key feature of the proposed methods is that they are weakly informative about the parameters of the mixture. Based on experiments, we show that these features enhance model flexibility, yield more accurate cluster estimates, and improve segmentation. For comparison purposes, we have applied two different MCMC methods to the segmentation of simulated MR brain scans: the rjMCMC method and the DP mixture model. The results are compared with similar results from other well-known MRI segmentation methods.

The rest of the paper is organized as follows: Section 2.1 provides background material on finite mixture models. Section 2.2 reviews key aspects of the rjMCMC model. In Section 2.3, the Dirichlet process mixture model used in this work is introduced. Section 3 provides details of the experimental work, and Section 4 discusses the results. Finally, Section 5 draws some conclusions.

2. Bayesian mixture modeling

2.1. Finite mixtures

We consider the case where we wish to model data $y = (y_1, \ldots, y_n)$ as independent observations from a mixture with a specified number of components k,

$$f(y_i) = \sum_{j=1}^{k} \omega_j f_j(y_i), \quad i = 1, \ldots, n,$$

where the distributions $f_j (1 \le j \le k)$ are known up to a parameter and the mixing proportions $0 < \omega_j < 1$ satisfy $\sum_{j=1}^{k} \omega_j = 1$. When the f_j's are from a given parametric family, with unknown parameter θ_j, we obtain the parametric mixture model:

$$y_i \sim \sum_{j=1}^{k} \omega_j f(y_i|\theta_j). \tag{1}$$