NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 38 (2007) 677–695

Diffusion-based spatial priors for imaging

L.M. Harrison,[a,*] W. Penny,[a] J. Ashburner,[a] N. Trujillo-Barreto,[b] and K.J. Friston[a]

[a]The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London, WC1N 3BG, UK
[b]Cuban Neuroscience Centre, Havana, Cuba

Received 25 October 2006; revised 3 July 2007; accepted 16 July 2007
Available online 8 August 2007

We describe a Bayesian scheme to analyze images, which uses spatial priors encoded by a diffusion kernel, based on a weighted graph Laplacian. This provides a general framework to formulate a spatial model, whose parameters can be optimized. The application we have in mind is a spatiotemporal model for imaging data. We illustrate the method on a random effects analysis of fMRI contrast images from multiple subjects; this simplifies exposition of the model and enables a clear description of its salient features. Typically, imaging data are smoothed using a fixed Gaussian kernel as a pre-processing step before applying a mass-univariate statistical model (e.g., a general linear model) to provide images of parameter estimates. An alternative is to include smoothness in a multivariate statistical model (Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J., 2005. Bayesian fMRI time series analysis with spatial priors. *Neuroimage* 24, 350–362). The advantage of the latter is that each parameter field is smoothed automatically, according to a measure of uncertainty, given the data. In this work, we investigate the use of diffusion kernels to encode spatial correlations among parameter estimates. Nonlinear diffusion has a long history in image processing; in particular, flows that depend on local image geometry (Romeny, B.M.T., 1994. Geometry-driven Diffusion in Computer Vision. Kluwer Academic Publishers) can be used as adaptive filters. This can furnish a non-stationary smoothing process that preserves features, which would otherwise be lost with a fixed Gaussian kernel. We describe a Bayesian framework that incorporates non-stationary, adaptive smoothing into a generative model to extract spatial features in parameter estimates. Critically, this means adaptive smoothing becomes an integral part of estimation and inference. We illustrate the method using synthetic and real fMRI data.

Keywords: Diffusion kernel; Weighted graph Laplacian; Spatial priors; Gaussian process model; fMRI; General linear model; Random effects analysis

Introduction

Functional MRI data are typically transformed to a three-dimensional regular grid of voxels in anatomical space, each containing a univariate time series of responses to experimental perturbation. The data are then used to invert a statistical model, e.g., general linear model (GLM), after a number of pre-processing steps, which include spatial normalization and smoothing (i.e., convolving the data with a spatial kernel). In mass-univariate approaches (e.g., statistical parametric mapping), a statistical model is used to extract features from the smoothed data by treating each voxel as a separate observation. Model parameters, at each voxel, are estimated (Friston et al., 2002) and inference about these parameters proceeds using SPMs or posterior probability maps (Friston and Penny, 2003). Smoothing the data ensures the maps of parameter estimates are also smooth. This can be viewed as enforcing a smoothness prior on the parameters. The current paper focuses on incorporating smoothness into the statistical model by making smoothness a hyperparameter of the model and estimating it using empirical Bayes. This optimizes the spatial dependencies among parameter estimates and has the potential to greatly enhance spatial feature detection.

Recently Penny et al. (2005) extended the use of shrinkage priors on parameter estimates (Penny et al., 2003), which assume spatial independence, to spatial priors in a statistical model of fMRI time series. They developed an efficient algorithm using a mean-field approximation within a variational Bayes framework. The result is a smoothing process that is incorporated into a generative model of the data, where each parameter is smoothed according to a measure of uncertainty in that parameter. The advantage of a mean-field approximation is that inversion of a requisite spatial precision matrix is avoided. The advantage of a Bayesian framework is that the evidence for different spatial priors can be compared (MacKay, 2003). Other Bayesian approaches to spatial priors in fMRI include those of Gossl et al. (2001); Woolrich et al. (2004); and more recently Flandin and Penny (2007).

There are two main departures from this previous work on spatiotemporal models in the current method. The first is that we use a Gaussian process prior (GPP) over parameter estimates. Spatial correlations are then encoded using a covariance matrix instead of precisions (cf. Penny et al., 2005). The second is that the covariance

* Corresponding author. Fax: +44 207 813 1445.
E-mail address: l.harrison@fil.ion.ucl.ac.uk (L.M. Harrison).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.07.032

matrix is the Green's function of a diffusive process, i.e., a diffusion kernel, which encodes the solution of a diffusion equation involving a weighted graph Laplacian. This has the advantage of providing a full spatial covariance matrix and enables inference with regards to the spatial extent of activations. This is not possible using a mean-field approximation that factorizes the posterior distribution over voxels. The result is an adaptive smoothing that can be spatially non-stationary, depending on the data. This is achieved by allowing the local geometry of the parameter field to influence the diffusion kernel (smoothing operator). This is important as stationary smoothing reveals underlying spatial signal at the expense of blurring spatial features. Given the convoluted spatial structure of the cortex and patchy functional segregation, it is reasonable to expect variability in the gradient structure of a parameter field. The implication is that the local geometry of activations should be preserved. This can be achieved with a nonlinear smoothing process that adapts to local geometric 'features'. A disadvantage is the costly operation of evaluating matrix exponentials and inverting potentially large covariance matrices, which the mean-field approach avoids. However, many approximate methods exist (MacKay, 2003; Rasmussen and Williams, 2006) that can ameliorate this problem, e.g., sparse GPPs (see discussion and Quinonero-Candela and Rasmussen, 2005).

The paper is organized as follows. First, we discuss background and related approaches, before giving an outline of the theory of the method. We start with the model, which is a two-level general linear model (GLM) with matrix-variate density priors on GLM parameters. We focus on reducing the model to the specification of covariance components, in particular, the form of covariance and its hyperparameters. We then look at the form of the spatial priors using graph Laplacians and diffusion kernels. We then describe the **EM** algorithm that is used to update hyperparameters of covariance components, which embody empirical spatial priors. The edge preserving quality of diffusion over a weighted graph is demonstrated using synthetic data and then applied to real fMRI data. The illustrations in this paper use 2D spatial images, however, the method can be easily extended to 3D, subject to computational resources, which would be necessary to analyze a volume of brain data. We perform a random effects (between subjects) analysis (Penny and Holmes, 2003) on a sample of contrast images from twelve subjects. This means that we consider a scalar field of parameter estimates encoding the population response. However, the nonlinear diffusion kernels described here can be extended to fields of vectors and matrices (ChefD'Hotel et al., 2004; Zhang and Hancock, 2006b). This paper concludes with comments on outstanding issues and future work.

Background

The current work draws on two main sources in the literature; diffusion-based methods in image processing and Gaussian process models (GPM). The image processing community has been using diffusion models for many years, e.g., for the restoration of noisy images (Knutsson et al., 1983). For overviews, from the perspective of scale-space theories, see Romeny (1994, 2003). These models rest on the diffusion equation, which is a nonlinear partial differential equation describing the density fluctuations in an ensemble undergoing diffusion; $\dot{\mu}=\nabla\cdot D(\mu)\nabla\mu$, where μ can be regarded as the density of the ensemble (e.g., image intensity) and D is the diffusion coefficient. Generally, the diffusion coefficient depends on the density, however, if D is a constant, the equation reduces to the 'classical heat equation'; $\dot{\mu}=D\nabla^2\mu$, where $\nabla^2\equiv\Delta$ is the Laplacian operator (second-order spatial derivative). A typical use in image processing is to de-noise an image, where the noisy image is the initial condition, $\mu(t=0)$ and a smoothed, de-noised, image is the result of integrating the heat equation to evaluate the diffused image at some time later; $\mu(t)$. In particular, Perona and Malik (1990) used nonlinear diffusion models to preserve the edges of images using an image dependent diffusion term, $D=D(\nabla\mu)$. The dependence on this spatial gradient has the effect of reduced diffusion over regions with high gradient, i.e., edges. Later formulations of nonlinear diffusion methods include those of Alvarez et al. (1992) and Weickert (1996). Of particular relevance to the method presented here are graph-theoretic methods, which use graph Laplacians (Chung, 1991). These have been used recently to adaptively smooth scalar, vector and matrix-valued images (Zhang and Hancock, 2005). Graphical methods provide a general formulation on arbitrary graphs, which is easy to implement. There are also many useful graph-based algorithms in the literature, e.g., image processing on arbitrary graphs (Grady and Schwartz, 2003) and, more generally, graph partitioning to sparsify and solve large linear systems (Spielman and Teng, 2004).

Gaussian process models also have a long history. A Gaussian process prior (GPP) is a collection of random variables, any finite number of which have a joint Gaussian distribution (MacKay, 2003; Rasmussen and Williams, 2006). As such it is completely specified by a mean and covariance function. This is a very flexible prior as it is a prior over a function, which can be used to model general data, not just images. Given a function over space, this function is assumed to be a sample from a Gaussian random field specified by a mean and covariance, which can take many forms, as long as it is positive semi-definite.

Diffusion methods in image processing and covariance functions in GPMs furnish the basis of a spatial smoothing operator; however, the emphasis of each approach is different. One main difference is that a GPM is a statistical model from which inferences and predictions can be made (MacKay, 1998). The objective is not solely to smooth data, but to estimate an optimal smoothing operator, which is embedded in a model of how the data were generated. Graphical models in machine learning (Jordan, 1999) provide a general and easy formulation of statistical models. The similar benefits of graph-based diffusion methods in image processing further motivates the use of graph-theoretic approaches to represent and estimate statistical images, given functional brain data.

The relation between models of diffusion and GPPs is seen when considering a random variable as a diffusive process, which locally is a Gaussian process. We can see this by comparing the Green's function of the classical heat equation, used in early diffusion methods in image processing (Romeny, 1994) and the squared exponential (SE) covariance function used in GPMs (Rasmussen and Williams, 2006). In two dimensions, (u_k, u_l), where subscripts indicate location in the domain and D is a scalar;

$$\dot{\mu}=D\Delta\mu$$

$$\mu(t+\tau)=K(\tau)\mu(t)$$

$$K(u_k,u_l;\tau)=\frac{1}{4\pi D\tau}\exp\left(-\frac{(u_k-u_l)^T(u_k-u_l)}{4D\tau}\right) \quad (1)$$

where $K(\tau)$ is the Green's function (solution) of the diffusion equation that represents the evolution of a solution over time. The