지구물리 제 4권 제 3호, P. 175-182 (2001년 9월)

Use of Training Data to Estimate the Smoothing Parameter for Bayesian Image Reconstruction

Lee, Soo-Jin
Department of Electronic Engineering Paichai University

요 약

본 논문에서는 의료영상의 응용분야로서 방출전산화단층 영상에 사용되는 베이지안 방법을 위한 Gibbs 사전정보의 평활 파라미터를 결정하는 문제를 다룬다. 특히, 광역 하이퍼파라미터(평활 파라미터)가 해의 편향과 분산의 균형을 조절하는 단순 평활사전정보(일명 멤브레인)를 연구 대상으로 한다. 본 논문에서 사용된 방법은 관측된 훈련데이터에 ML 방법을 적용한 하이퍼파라미터 추정법에 기반을 두며, 이러한 접근방법에 대한 동기에 대하여도 논한다. 멤브레인 사전정보를 위한 평활 파라미터의 경우 단순한 ML 추정법을 적용하여도 파라미터가 쉽게 추정될 수 있음을 보인다.

ABSTRACT : We consider the problem of determining smoothing parameters of Gibbs priors for Bayesian methods used in the medical imaging application of emission tomographic reconstruction. We address a simple smoothing prior (membrane) whose global hyperparameter (the smoothing parameter) controls the bias/variance tradeoff of the solution. We base our maximum-likelihood (ML) estimates of hyperparameters on observed training data, and argue the motivation for this approach. Good results are obtained with a simple ML estimate of the smoothing parameter for the membrane prior.

Introduction

Bayesian methods utilizing Gibbs priors have proven useful for problems in image restoration and reconstruction. [1,2] In our own application to medical image reconstruction [3-5] in emission computed tomography (ECT), the Bayes model is especially apt for two reasons: (1) Accurate statistical modelling, via the likelihood function, of the Poisson noise associated with gamma-ray projection data is essential since noise is severe, and (2) given the ill-posed nature of tomographic inversion and the poor quality of projection data, a regularization needed to stabilize maximum-likelihood (ML) solutions may be provided by MAP (maximum *a posteriori*) estimates that can be formulated using Bayesian approaches that incorporate suitable prior models. In addition, the likelihood model may conveniently incorporate a system model

needed to account for the physics of image formation. As in other domains, the prior models in Bayesian ECT most often model some notion of local smoothness or piecewise smoothness via Markov random field (MRF) models expressable as Gibbs energies. To date, much of the appeal of Bayesian approaches in ECT has been based on both modelling considerations, and the encouraging quality of results seen in many studies.

Computational costs have dampened enthusiasm, however. Algorithms that use Bayesian approaches to compute MAP or other image estimates in ECT are always iterative and costly to compute, but the relentless improvement in computing hardware has by now made such computation practical. A second computational problem stems from the fact that these algorithms contain free parameters whose values affect the nature of the solution. Most of these are easily dealt with, but in Bayes approaches, hyperparameters of the prior, i.e. the few (one or two) parameters that control the degree of smoothing and nature of discontinuities, are costly to compute in a principled way.

We now state the hyperparameter estimation problem: Given a likelihood and prior model, and given a realization of noisy projection data from a patient, compute some optimal estimate of the hyperparameters. The variety of approaches used to attack this problem in ECT include regularization methods and estimation-theoretic methods based on maximum likelihood. [6,7] All of these methods are characterized by severe computational cost, yet this is the only principled way to attack the problem as stated. The cost hinders not only application to clinical studies, but also slows down further development of Bayesian approaches and algorithms.

In this paper, we pursue a rather different approach in which (noiseless) training exemplars are used in place of noisy projection data for ML hyperparameter estimation. However, our motivation lies in the use of training set approaches for algorithm development.

Motivation for Traning Set Approach

Training exemplars have been used in a variety of problems for MRF parameter estimation, [8] The basic notion is simple: Given a known, noiseless training object denoted by a vector $\boldsymbol{f}$, hyperparameters $\boldsymbol{\lambda}$, and a prior probability model $\Pr(\boldsymbol{f}|\boldsymbol{\lambda})$, compute the ML estimate $\hat{\boldsymbol{\lambda}} = \arg\max_{\lambda}[\Pr(\boldsymbol{f}|\boldsymbol{\lambda})]$, and use $\hat{\boldsymbol{\lambda}}$ in subsequent reconstructions. (Note that here, $\boldsymbol{f}$ is a single realization of a random field modelled by the prior.) The ML problem tends to be robust since there are only one or two hyperparameters while the vector $\boldsymbol{f}$ may have thousands of components.

Two natural questions are: 1. What constitutes a proper training set, and how representative is it? Unlike the case in computer vision, "scenes" in medical imaging are quite stereotypical. One brain scan acquired under a very specific medical protocol looks quite similar to others acquired under the same protocol. Researchers utilize this fact to construct elaborate hardware and software "phantoms" that mimic, to varying degrees, the typical distribution of radionuclide within the relevant patient anatomy. Image formation from the phantom or set of phantoms is then simulated or physically acquired, and comparison of reconstructed image quality to the known phantom(s) then serves as a test of the reconstruction. The verisimilitude of the phantom in terms of anatomy and radionuclide distribution turns out to be important for a variety of applications. We have explored the use of autoradiography as a source of realistic phantoms [9] and we include such phantoms along with others, as shown in Fig. 1, in this paper.

To develop Bayes approaches, the researcher can thus use a phantom itself as a training set to ascertain hyperparameters that otherwise would be found by laborious empirical search. Of course, the value λ will not be the same as that found by estimation from noisy projection data, since noise and blur can affect the value of this estimate. However, the two values approach each other in the limit of low noise and blur. For our realistic experiments, our training-set derived hyperparameters perform well even on noisy data, as described later.

The ability to quickly estimate hyperparameters from phantoms assumes greater importance when it is realized that the statistical character of medical objects in ECT is rarely nonstationary; vast improvements can occur if the object is modelled by region-dependent hyperparameters for a given prior model. Physically, the nonstationarity arises because tissues in different anatomical regions take up radiopharmaceutical differently. Thus knowledge of regions via coregistered anatomy can lead to space variant priors that model the “texture” in different regions differently, or encourage boundary formation along region boundaries. Estimation of region dependent hyperparameters from projection data alone requires a simultaneous segmentation (into appropriate regions), reconstruction, and parameter estimation - a difficult task. Phantoms can provide presegmented test sets in which such estimation is fairly easy, and algorithm development using space variant priors can proceed.

A third motivation is that results on a phantom may indeed be generalized in several ways. Hyperparameters calculated for a phantom may be easily related to an intensity-scaled or magnified version of the phantom, and this enables one to quickly scale hyperparameter values with regard to count level and patient size, two parameters of importance in phantom studies.

Likelihood and Prior Models

In ECT, the object to be reconstructed is f_{ij}, identified as the mean of a Poisson gamma-ray emission process from pixel (i,j), radiates gamma rays and creates detector counts $g_{t\theta}$ for a detector positioned at orientation θ and coordinate t. A system matrix H, that accounts for the projection geometry and various physics effects, relates the mean of the detector counts $\bar{g}_{t\theta}$ to f_{ij} via

$$\bar{g}_{t\theta} = \sum_{ij} H_{t\theta,ij} f_{ij}. \tag{1}$$

Though the methods here apply to any form of prior, we consider in this work the well-known membrane prior which has been one of the most popular priors in the emission tomography problem. We use the familiar Gibbs distribution to model the prior distribution:

$$\Pr(\boldsymbol{F}=\boldsymbol{f}) = \frac{1}{Z}\exp[-\lambda E_P(\boldsymbol{f})], \tag{2}$$

where $\boldsymbol{f}$ is a 2-D random field realization for the source intensity, with $\boldsymbol{F}$ the associated random field, and {\bf L,l} are the random field. If we use ij to index the 2-D lattice sites, then F_{ij} is a random variable whose realization is f_{ij}. Note that $f_{ij} \geq 0$ is a real nonegative number whose value corresponds to a mean emission rate. For our case, the phantom $\boldsymbol{f}$ is known. Also, Z is a normalization factor (partition function) to be discussed below, λ is the hyperparameter (the smoothing parameter), and $E_P(\boldsymbol{f})$ is the prior energy. The definition of the prior energy is given by

$$E_P(\boldsymbol{f}) = \sum_{ij}\{f_v^2(i,j) + f_h^2(i,j)\}, \tag{3}$$

where, f_v and f_h are the first order derivatives along the vertical and the horizontal directions, respectively, and are defined here as

$$f_v(i,j) = f_{i+1,j} - f_{i,j}$$
$$f_h(i,j) = f_{i,j+1} - f_{i,j}.$$

With the energy defined as in (3), we may express the partition functions as

$$Z(\lambda) = \int_0^\infty \exp[-\lambda E_P(\boldsymbol{f}|\lambda)]d\boldsymbol{f}. \tag{4}$$

The integration limits in (4) follows from the fact that f_{ij} is a real, nonnegative number.

Caclulating the Smoothing Parameter via Maximum Likelihood

Given the known phantom data $\boldsymbol{f}$, we may use a maximum-likelihood (ML) approach to estimate hyperparameters. For the mem brane prior, the ML solution $\hat{\lambda}$ becomes:

$$\begin{aligned}\hat{\lambda} &= \arg\min_\lambda\{-\log\Pr(\boldsymbol{F}=\boldsymbol{f}|\lambda)\} \\ &= \arg\min_\lambda\{\lambda E_P(\boldsymbol{f};\lambda) + \log Z(\lambda)\}\end{aligned} \tag{5}$$

The difficulty in (5) is the evaluation of $Z(\lambda)$. Despite the inconvenient limits of integration, the evaluation of $Z(\lambda)$ is possible in this case and makes use of a mathematical substitution trick [10]:

$$\begin{aligned} Z(\lambda) &= \int_0^{\infty} \exp[-\lambda E_P(\boldsymbol{f}|\lambda)]d\boldsymbol{f} \\ &= \int_0^{\infty} \exp[-\lambda\{(f_{i+1,j}-f_{i,j})^2+(f_{i,j+1}-f_{i,j})^2\}]df \\ &= \lambda^{-\frac{N}{2}} \int_0^{\infty}\Big[-\sum_{ij}\{(f'_{i+1,j}-f'_{i,j})^2+(f'_{i,j+1}-f'_{i,j})^2\}\Big]df \\ &= \lambda^{-\frac{N}{2}} C, \end{aligned} \tag{6}$$

where $\boldsymbol{f}' = \sqrt{\lambda}\boldsymbol{f}$, N is the number of object pixels, and C is some constant independent of λ. Substituting (6) into (5), using (3) and dropping terms independent of λ, we get

$$\hat{\lambda} = \arg\min_{\lambda}\left\{\lambda E_P(\boldsymbol{f}) - \frac{N}{2}\log\lambda\right\} \tag{7}$$

whose solution is

$$\hat{\lambda} = \frac{N}{2E_P(\boldsymbol{f})}, \tag{8}$$

the correct ML solution for the membrane prior. Note that it is more correct to set N equal to the number of pixels of the object proper, not the zero-valued background pixels.

The derivation of (8) assumed a range of $(0,\infty)$ for f_{ij}. For the range $(-\infty,+\infty)$, the prior model becomes a member of the class of autonormal models; the solution for the hyperparameters for this class was pointed out in [11]. It turns out that the approach above may also be used for the case $(-\infty,+\infty)$, and the solution for $\hat{\lambda}$ is again given by (8). For a general range (a,b), the closed form solution in (8) does not work (C becomes dependent on λ) and approximations are needed.

Numerical Studies

In order to validate our expressions for ML parameter estimates, we conducted "closed-loop" numerical experiments in which images were created by Gibbs sampling from each of the prior probabilities with hyperparamters known. Our estimators were applied to sets of such images, and the sample bias and variance of our estimates were computed. For a phantom that is not necessarily consistent with the underlying prior model, we must resort to evaluating some performance metric, in our case the root mean-squared error (RMSE) as a function of λ, to show that the optimal estimates yield good reconstructions. Here RMSE (λ) is defined as

$$\mathrm{RMSE}(\lambda) = \sqrt{\frac{1}{N}\sum_{ij}(\hat{f}_{ij}(\lambda)-f_{ij})^2}, \tag{9}$$

where $\hat{\boldsymbol{f}}(\lambda)$ is a reconstruction of $\boldsymbol{f}$ obtained using the value λ. For our application, the reconstruction itself was a Bayesian MAP procedure as described in [4].

Table 1. Percent bias and standard deviation of $\hat{\lambda}$

$\hat{\lambda}$	bias %	std %
0.005	16.62	2.01
0.010	8.01	3.26
0.050	5.71	2.30
0.100	4.49	1.58
0.500	6.01	2.69

For our closed-loop experiments, we generated 10 Gibbs samples for each of 5 values λ, and computed the 50 ML estimates according to (8). Each sample was a 64×64 image with 32 discrete grey levels. The results, summarized in Table 1, show a small positive relative bias and small relative standard deviation (expressed as a percentage of the known true value for λ).

In order to test the performance on phantom data (our training set) of our optimal $\hat{\lambda}$ for the membrane prior, we created 40 noisy projection data sets of the phantom in Fig. 1(a). The count level, 732K counts, was equivalent to that of a typical SPECT (single-photon ECT) brain scan on a per-pixel basis. For each projection data set, reconstructions using the membrane prior were performed for nine values $\lambda = \lambda_0 \times 2^n$, where $n = 0, \pm1, \pm2, \pm3, \pm4$ and $\lambda_0 = 0.02$, thus yielding 360 reconstructions. Fig. 2(a) shows a plot of RMSE vs. λ averaged over the 40 noisy projection data sets generated from this phantom, and indicates the position (arrow labeled ML) of the value for $\hat{\lambda}$ obtained by applying (8) on the phantom.

One may also compute an MPL (maximum pseudo-likelihood) estimate [9] for the membrane prior for comparison puroses. The value of the MPL estimate is also displayed (arrow labeled MPL). As seen, the derived value is not far from the one that minimizes RMSE.

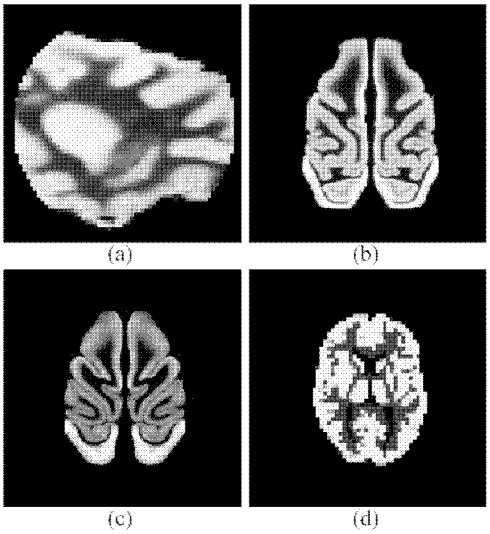

Fig. 1 Phantoms used in the simulations. (a) Phantom A: Primate autoradiograph phantom obtained with the blood flow agent (^{99m}Tc-ECD); (b)(c) Phantoms B and C: Primate autoradiograph phantoms obtained with the benzodiazepine neuroreceptor agent Iomazenil (^{123}I); (d) Phantom D: Hoffman brain phantom.

We have observed empirically that for reconstruction problems with noise levels characteristic of ECT, an ML or MPL estimate for hyperparameters almost always falls within

a factor of 2 or 3 of λ^{rms}, the hyperparameter corresponding to minimum RMSE, and that the RMSE $\hat{\lambda}$ is always within a few percent of RMSE(λ^{rms}). Fig. 2 illustrates a typical case.

Table 2 summarizes trials used to explore anecdotal results regarding issues of generalization (i.e. "similar" phantoms yield approximately equal $\hat{\lambda}$'s). For a given phantom from Fig. 1, each row in Table 2 displays the value $\hat{\lambda}_{ML}$, the ML estimate from training data, $\hat{\lambda}_{RMSE}$, the value of $\hat{\lambda}$ that minimizes the RMSE of a reconstruction with 732K counts, and also $\hat{\lambda}_{MPL}$, a maximum pseudolikelihood estimator [9] listed for comparison. All phantoms were scaled to be consistent with a 732K count level. Phantom A is a distribution of radionuclide density obtained from sagittal autoradiograph primate phantom [9]. Phantoms B and C are very similar transverse slice autoradiographs, from a single primate, spaced at 5 mm, and so we expect the parameter values to be similar for each, and they are. By contrast, we expect the Hoffman brain phantom D, which consists of flat regions with sharp edges, to have fairly different values. This is indeed the case.

Table 2 Hyperparameter estimation using MPL and ML compared with $\hat{\lambda}_{RMSE}$.

Phantoms	$\hat{\lambda}_{MPL}$	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{RMSE}$
A	0.033	0.018	0.020
B	0.039	0.027	0.022
C	0.031	0.022	0.025
D	0.0016	0.0014	0.0016

Summary and Conclusion

We have considered hyperparameter estimation using phantom training data for our medical imaging application, and pointed out several motivations where this approach might be used in place of the difficult problem of hyperparameter estimation from noisy observed projection data. We derived expressions for the ML estimate of λ for the membrane prior. Our initial results indicate that the hyperparameters from training data perform well with regard to bias/variance/rms metrics.

References

1. J.H. Ollinger and J.A. Fessler, "Positron Emission Tomography," *IEEE Signal Processing Mag.*, vol 14, pp. 43-55, Jan. 1997.
2. S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", IEEE Trans. Patt. Anal. Mach. Intell., 6(6), pp. 721-741, Nov. 1984.
3. S.J. Lee, A. Rangarajan, G. Gindi, "Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors", *IEEE Trans. Med. Imaging,* (14)4, pp. 669-680, Dec. 1995.

4. S.J. Lee, I.T. Hsiao, G.R. Gindi, “The Thin Plate as a Regularizer in Bayesian SPECT Reconstruction”, *IEEE Trans. Nuclear Science*, vol. 44, no. 3, pp. 1381-1387, July 1997.
5. S.J. Lee, Y. Choi, G. Gindi, "Validation of New Gibbs Priors for Bayesian Tomographic Reconstruction Using Numerical Studies and Physically Acquired Data", *IEEE Trans. Nuc. Sci.*, 46(6), pp. 2154-2161, Dec. 1999.
6. Z. Zhou, R.M. Leahy, and J. Qi, “Approximate Maximum Likelihood Hyperparameter Estimation for Gibbs Priors”, *IEEE Trans. Med. Imaging*, 6(6), pp. 844-861, Jun. 1997.
7. S.S. Saquib, C.A. Bouman, and K. Sauer, “ML Parameter Estimation for Markov Random Fields with Applications to Bayesian Tomography”, *IEEE Trans. Image Processing*, 7(7), pp. 1029-1044, Jul. 1998.
8. J. Besag, “On the Statistical Analysis of Dirty Pictures”, Journal of the Royal Statistical Society, Series B, 48(3), pp. 259-302, 1986.
9. S.J. Lee, G.R. Gindi, I.G. Zubal, A. Rangarajan, “Using Ground-Truth Data to Design Priors in Bayesian SPECT Reconstruction”, Y. Bizais, C. Barillot, R.D. Paola, eds., Information Processing in Medical Imaging, pp. 27-38, Kluwer Academic Publishers, 1995.
10. G. Chinn and S.C. Huang, “Noise and Resolutio of Bayesian Reconstruction for Multiple Image Configurations”, *IEEE Trans. Nuc. Sci.*, 40. pp. 2059-2063, 1993.
11. J. Besag, “Spatial Interaction and the Statistical Analysis of Lattice Systems”, *Journal of the Royal Statistical Society*, Series B, 36, pp. 192-236, 1974.