Overview

Thermography uses techniques to study heat radiations in different structures or regions. One of the unique example is its use in detecting tumors in human brain[1][2][3]. It analyses the heat radiated from brain tissues to determine whether it is a benign or a malign tissue. This greatly helps medical professionals to locate the bad tissues in brain. However, the radiations obtained from the surface is highly dynamic and non-stationary with continuous cerebral blood flow changes along with high environmental interference[4]. Therefore, it is of utmost priority to analyze the multidimensional data in such a way that it captures spatial as well as temporal interactions.

The Generalized Linear Models(GLM)[5] is a way of unifying various interactions, such that each outcome $\mathbf{Y}$ of the dependent variables are generated from a particular probability distribution that includes the normal, binomial, poisson, bernoulli and gamma distributions, among others. The mean, μ, of the distribution depends on the independent variables, $\mathbf{X}$, through:

$$\mathrm{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{1}$$

where $\mathbf{E(Y)}$ is the expected value of $\mathbf{Y}$; $\mathbf{X}\beta$ is the linear predictor, β is a linear combination of unknown parameters ; g is the link function which provides the relationship between the linear predictor and the mean of the distribution function. The unknown parameters, β, are typically estimated with maximum likelihood technique which uses iterative re-weighted least squares algorithm[6]. A common shortcoming with this approach is that it is unable to catch random effects in the data which may portray non-linear relationships[7]. Hence, GLM model is extended by non-parametric components which leads to semi-parametric regression or partially linear models[7]. These models combines the deterministic components with non-parametric components such as penalized splines or P-Splines[8]. P-Splines assumes high number of knots to overfit the data and then penalizes it using difference of the adjacent coefficients.

Different regularization methods namely L1 and L2 penalties has been studied[9] for preventing overfitting of data. Focusing on logistic regression, [9] shows that using L1 regularization of the parameters, the sample complexity grows only logarithmically in the number of irrelevant features. This logarithmic rate matches the best known bounds for feature selection, and indicates that L1 regularized logistic regression can be effective even if there are exponentially many irrelevant features as there are training examples.

A Generalized Additive Model(GAM) approach [10] has been developed with a linear predictor depending linearly on unknown smooth functions of some predictor variables. In general , the model has a structure something like:

$$g(\mathrm{E}(Y)) = \boldsymbol{X\beta} + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m). \tag{2}$$

where $\boldsymbol{X\beta}$ is the strictly parametric part of the model, f_i are the functions which may be specified parametrically, or semi-parametrically with smooth functions, Y is the univariate response variable whereas x_i are predictor variables. An exponential family distribution is specified for Y along with a link function g. The P-Splines combined with parametric component gives good approximation of the curve in a generalized additive model (GAM) approach but fails to deal with numerical rank deficiency. Therefore, [11] proposes ridge penalty which is based on the pivoted QR decomposition and the singular value decomposition. The method as shown below deals with rank deficiency even when numerical rank depends on smoothing parameters.

Let $\boldsymbol{X}$ be a $n \times q$ model matrix and $\boldsymbol{S_i}$ is the i^{th} (positive semidefinite) penalty matrix with unknown smoothing parameter θ_i. Wood[17] proposes to form the QR decomposition of $\boldsymbol{X}$,

$$\boldsymbol{X} = \boldsymbol{QR} \tag{3}$$

where $\boldsymbol{Q}$ is made up of columns of an orthogonal matrix and $\boldsymbol{R}$ is upper triangular. Defining $\boldsymbol{S} = \boldsymbol{H} + \sum_{i=1}^{m} \theta_i \boldsymbol{S_i}$ and $\boldsymbol{B}$ as any matrix square root of $\boldsymbol{S}$ such that $\boldsymbol{B^T B} = \boldsymbol{S}$, a singular value decomposition can be formed:

$$\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{B} \end{bmatrix} = \boldsymbol{U D V_T} \tag{4}$$

$\boldsymbol{B}$ can be obtained efficiently by pivoted Choleski decomposition or by eigen vaulue decomposition of the symmetric matrix $\boldsymbol{S}$. The columns of $\boldsymbol{U}$ are columns of an orthogonal matrix, $\boldsymbol{V}$ is an orthogonal matrix, and $\boldsymbol{D}$ is the diagonal matrix of singular values. The examination of these singular values is the most reliable way to detect numerical rank deficiency of the fitting problem. In particular, at this stage any singular values that are too small should be removed along with the corresponding columns of $\boldsymbol{U}$ and $\boldsymbol{V}$. This deletion has the effect of recasting the problem into a reduced space in which the model parameters are identifiable. Therefore, this approach is to be used to find the spline coefficients.

The spline coefficients acts as labels and are assigned to each pixels. They act as a stochastic or random process spatially correlated in one way or the other. A Markov random field[12] (often abbreviated as MRF), is a set of these random variables indexed by spatial positions. In MRFs, each pixel (often termed as a site) of an image act as a random variable having a Markov property described by an undirected graph. In other words, a random field is said to be Markov random field if it satisfies Markov properties[13]. A set of sites is assigned a set of labels. Let S index a discrete set of $n \times m$ sites in a rectangular lattice for a $2D$ image denoted by

$$S = \{(i, j)|1 \leq i \leq n; 1 \leq j \leq m\} \tag{5}$$

The inter-relationship between sites S is maintained by a so-called neighborhood system. Let L be a set of labels. A label set may be categorized as being continuous or discrete. MRFs with discrete labels are known as Discrete Markov Random fields whereas MRFs with continuous labels are known as Gaussian Markov Random fields. A neighborhood system for S is defined as

$$N = \{N_i|\forall_i \varepsilon S\} \tag{6}$$

where N_i is the set of sites neighboring i. In the first order neighborhood system, also called the 4-neighborhood system, every (interior) site has four neighbors as $N_{i,j} = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$, a site at a boundary has three and a site at the corners has two. The pair $(S, N) \equiv G$ constitutes a graph in the usual sense; S contains the sites and N determines the links between the sites according to the neighboring relationship. A clique C for (S, N) is defined as a subset of sites in S. Let $F = \{F_1, ..., F_m\}$ be a family of random variables defined on the set S, in which each random variable F_i takes a value f_i in L. The notation $F_i = f_i$ is to denote the event that F_i takes the value f_i and the notation $(F_1 = f_1, ..., F_m = f_m)$ to denote the joint event. For a discrete label set L, the probability that random variable F_i takes the value f_i is denoted $P(F_i = f_i)$, abbreviated $P(f_i)$ and the joint probability is denoted $P(F = f) = P(F_1 = f_1, ..., F_m = f_m)$ and abbreviated $P(f)$.

F is said to be a Markov random field on S with respect to a neighborhood system N if and only if the following two conditions are satisfied:

$$P(f) > 0, \forall f \varepsilon F \qquad (positivity) \tag{7}$$

$$P(f_i|f_{S-i}) = P(f_i|f_{N_i}), \qquad (Markovianity) \tag{8}$$

Markovianity can be termed as the property of a random variable, if the conditional probability distribution of the random variable depends only upon its neighboring variables N_i instead of all the random variables present in the system $S - i$.

Gibbs Random Field (GRF) is another type of random field that obeys a Gibbs distribution. A Gibbs distribution takes the following form

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{9}$$

where

$$Z = \sum_{f \epsilon F} e^{-\frac{1}{T}U(f)} \tag{10}$$

is a normalizing constant called the partition function, T is a constant called the temperature which shall be assumed to be 1 unless otherwise stated, and $U(f)$ is the energy function. The energy

$$U(f) = \sum_{c \epsilon C} V_c(f) \tag{11}$$

is a sum of clique potentials $V_c(f)$ over all possible cliques C. The value of $V_c(f)$ depends on the local configuration on the clique c. An important special case is when only cliques of size up to two i.e first order neighborhood are considered. In this case, the energy function can also be written as

$$U(f) = \sum_{i \epsilon S} V_1(f_i) + \sum_{i \epsilon S} \sum_{i' \epsilon N_i} V_2(f_i, f_{i'}) \tag{12}$$

The Hammersley-Cliord theorem [14] states that a unique GRF exists for every discrete MRF field and visa-versa as long as Gibbs random field is defined in terms of a neighborhood system. The studies by [16][17][18][15][19] all make use of the Gibbs distributions (GD) for characterizing MRF. [20] and [21] compared Discrete MRFs with Gaussian MRFs in the context of different datasets. All examples featured a two-dimensional regular lattice and the neighborhood structure involved eight immediate neighboring pixels. In both the studies, Discrete Markov Random field has been shown to provide superior performance.

[22] performed a comparative study of Inference Techniques for Discrete Markov Random fields (DMRF). The author discusses Pixel-based-Models where each pixel in a 2D lattice is a variable in the model and has four nearest

neighbors. A large number of different inference methods has been evaluated such as : (i) combinatorial-methods (ii) linear- programming-methods (iii) move-making methods (iv) message passing methods. In all these inference methods, algorithms based on Monte Carlo simulation were not considered. On average, Tree-reweighted Message passing(TRWS) algorithm from Linear programming method afford the best solutions and was the fastest algorithm.

Overall, in order to model spatial-temporal interactions in the thermographic brain imaging data, the semi parametric regression using penalized splines in the generalized additive model approach seems to be a good choice. By discussing two different types of Markov random fields, Discrete Markov Random fields (DMRF) has been to be used for spatial smoothing of spline coefficients with Tree-reweighted message passing algorithm as one the better inference techniques.

References

[1] A. M. Gorbach, J. Heiss, C. Kufta, S. Sato, P. Fedio, W. A. Kammerer, J. Solomon, and E. H. Oldfield. *Intraoperative infrared functional imaging of human brain.* Ann. of Neurol., 54:297–13, 2003

[2] A. M. Gorbach, J. D. Heiss, L. Kopylev, and E. H. Oldfield. *Intraoperative infrared imaging of brain tumors.* J Neurosurg, 101(6):960–9, 2004.

[3] I. A. Shevelev, E. N. Tsicalov, A. M. Gorbach, K. P. Budko, and G. A. Sharaev. Thermoimaging of the brain. J. Neurosci. Methods, 46:49–9, 1993.

[4] Nico Hoffmann, Yordan Radev, Edmund Koch, Gerald Steiner, Matthias Kirsch, and Uwe Petersohn. *Intraoperative identification of the sensory cortex by semiparametric regression of time-resolved thermography.*

[5] Nelder, John; Wedderburn, Robert (1972). *Generalized Linear Models.* Journal of the Royal Statistical Society. Series A (General). Blackwell Publishing. 135 (3): 370–384.

[6] C. Sidney Burrus. *Iterative Reweighted Least Squares.* http://cnx.org/content/m45285/1.12/

[7] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric Regression*, Cambridge University Press, New York (2003)

[8] Eilers, P.H.C. and B.D. Marx, *Flexible Smoothing with B-Splines and Penalties*. Statistical Science, 11(2):89-121,1996

[9] Andrew Y. Ng. Feature selection, L1 vs L2 regularization, and rotational invariance. Computer Science Department, Stanford University, Stanford, CA 94305, USA

[10] Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC. ISBN 978-0-412-34390-2.

[11] Simon N. Wood(2004), *Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models*, Journal of the American Statistical Association, 99:467, 673-686

[12] Abend, K., Harley, T. J. and Kanal, L. N. (1965), *Classification of binary random patterns*. IEEE Trans. Inform. Theory IT-11, 538-544.

[13] Kindermann, R and Snell,J .L. (1980) *Markov Random Fields and their Applications* (Providence, RI, American Mathematical Society).

[14] J.M. Hammersley and P. Clifford. *Markov fields on finite graphs and lattices*. University of California, Berkeley

[15] Derin, H. & Elliot, H. (1987) *Modeling and segmentation of noisy and textured images using Gibbs random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9, pp. 39-55. 8

[16] G. R. Cross and A. K. Jain, *Markov random field texture models*, IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-5, pp. 25-39, 1983.

[17] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, pp. 721-741, Nov. 1984.

[18] F. S. Cohen and D. B. Cooper, *Real time textured image segmentation based on noncausal Markovian random field models*. in Proc. SPIE Conf Intell. Robots, Cambridge, MA, Nov. 1983.

[19] H. Derin and W. S. Cole, *Segmentation of textured images using Gibbs random fields*. Computer Vision, Graphics, Image Processing, vol. 35, pp. 72-98, 1986

[20] Daniel Smith & Michael Smith (2006). *Estimation of Binary Markov Random Fields Using Markov chain Monte Carlo*. Journal of Computational and Graphical Statistics, 15:1, 207-227, DOI: 10.1198/106186006X97817

[21] Peng Wei & Wei Pan (2008). *Network-based genomic discovery: application and comparison of Markov random-field models*. Appl. Statist. (2010) 59, Part 1, pp. 105–125

[22] Jorg H. Kappes and Bj¨orn Andres and Fred A. Hamprecht and Christoph Schn¨orr and Sebastian Nowozin and Dhruv Batra and Sungwoong Kim and Bernhard X. Kausler and Thorben Kr¨oger and Jan Lellmann and Nikos Komodakis and Bogdan Savchynskyy and Carsten Rother. *A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems*. CoRR, abs/1404.0533, 2014, http://arxiv.org/abs/1404.0533