ConocoPhillips SCHOOL OF GEOLOGY & GEOPHYSICS The University of Oklahoma

Mapping lithologic stacking patterns using Generative Topographic Mapping (GTM)

Atish Roy , Tim J. Kwiatkowski and Kurt J. Marfurt, The University of Oklahoma, Norman*

AASPI **A**ttribute-**A**ssisted **S**eismic **P**rocessing and **I**nterpretation

http://geology.ou.edu/aaspi

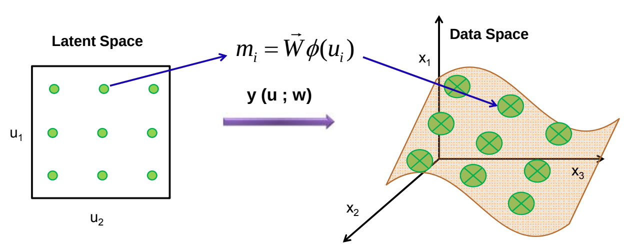
1. Introduction

The most popular seismic attributes fall into three broad categories – those that are sensitive to lateral changes in waveform and structure such as coherence and curvature, those sensitive to thin bed tuning and stratigraphy, such as spectral components, and those sensitive to lithology and fluid properties – such as AVO and impedance inversion. Earlier we presented Kohonen Self Organizing Maps for unsupervised multiattribute clustering which is routinely done by interpreters to differentiate depositional packages characterized by subtle changes in the stratigraphic column as well as lateral changes in texture.

Here we propose a newer clustering algorithm based on the Generative Topographic Mapping (Bishop et al, 1998). Although GTM has its origin from Self-organizing Maps, it overcomes most of the limitations of SOM. GTM is a probability density model which describes the distribution of multi-dimensional data in terms of a smaller number of latent space variables. GTM consists of constrained mixture of Gaussians in which the model parameters are determined by maximum likelihood using Expectation Maximization(EM) algorithm.

2. The Generative Topographic Mapping

GTM is defined by specifying a set of points {$\mathbf{u}_i$} in latent space, together with a set of basis functions {$\phi_j(u)$}. The adaptive parameters **W** and β define a constrained mixture of Gaussians with centers $\mathbf{W}\phi(u_i)$ and a common covariance matrix given by $\beta^{-1}\mathbf{I}$. After initializing **W** and *β* training involves alternating between the E-step in which the posterior probabilities are evaluated and the M-step in which **W** and *β* are re-estimated. Evaluation of the log likelihood using at the end of each cycle can be used to monitor convergence. (Bishop et al, 1998). After re-estimating the W we calculate the posterior mean projection of the data in the latent space and color it by the 2D HSV colorscale.



3. Theory

The aim is to define a non-linear transformation from the latent space to the data space by some linear combination of the basis functions φ so that each point **u** in the latent space is mapped to a corresponding point **y** in the *D*- dimensional data space

$\vec{y} = \vec{W}\phi(\vec{u})$ where **W** is a D x M weight matrix

$m_i = \vec{W}\phi(u_i)$ where m_i are the reference vectors

Define Gaussian functions which becomes the orthogonal basis functions to map the latent space points to the original dimensional space.

$$\phi(u_i) = e^{-\frac{dist}{2*\sigma}}$$

Noise model of the data vector **x**

$$p(\vec{x}\,|\,i) = \left(\frac{\beta}{2\pi}\right)^{D/2} e^{-\frac{\beta}{2}\|m_i-\vec{x}\|^2}$$

The probability distribution function of the GTM model is obtained by summing over all the above equation.

$$p(\vec{x}\,|\,W,\beta) = \sum_{i=1}^{K}\frac{1}{K}\left(\frac{\beta}{2\pi}\right)^{D/2} e^{-\frac{\beta}{2}\|m_i-\vec{x}\|^2}$$

This equation used to monitor convergence.

The initialization of **W** is done by evaluating the data covariance matrix and obtain the first and the second principal eigenvectors. β^{-1} is initialized to be square of half of the grid spacing of the PCA-projected latent points in data space.

The above parametric equation is fitted to the dataset {x_n} by Expectation Maximum (EM) algorithm.
Using Bayes' theorem the posterior probability which every i^{th} component takes for every data point x_n is given by

$$R_{ni} = \frac{e^{-\frac{\beta}{2}\|m_i-x_n\|^2}}{\sum_j e^{-\frac{\beta}{2}\|m_j-x_n\|^2}} \qquad \text{................}\textit{E}\text{ Step}$$

Re-estimate the Weight matrix *W* : $\mathbf{W}_{new}$

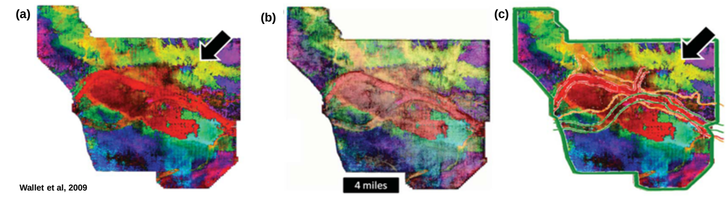
$$(\Phi^T G\Phi + \frac{\alpha}{\beta}I)W_{new}^T = \Phi^T RX \qquad \text{................}\textit{M}\text{ Step}$$

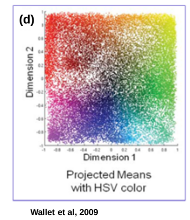
where $G_{ii} = \sum_n R_{ni}$

Re-estimate the *inverse* parameter: $1/\beta_{new}$

$$\frac{1}{\beta_{new}} = \frac{1}{ND}\sum_{n=1}^{N}\sum_{i=1}^{K} R_{ni}\,\|\vec{W}_{new}\phi(\vec{u}_i) - x\|^2$$

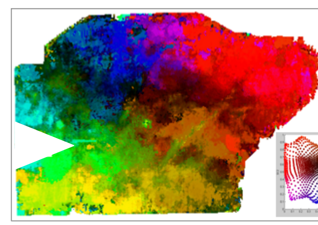
4. Examples of GTM on Red-Fork Formation, Anadarko Basin



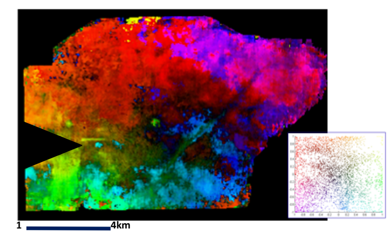


The above GTM analysis was done by (Wallet et al 2009). The input data consisted of 16 sample seismic amplitude as input. (a) is the seismic facies map from the GTM analysis. (b) This GTM image is blended with the coherency image to highlight the channel boundaries better. (c) geologic interpretation map overlaid on the GTM image. The black arrow denotes which may be a crevasse splay. (d) The projected means with HSV colors with each dots represent a cluster with the two dimensional latent space.

5. Lambda-Rho Mu-Rho analysis on Lower Barnett Shale – Comparing SOM with the GTM results



SOM Analysis : The coloring is done by comparing the trained vectors in the latent space with the data.



GTM Analysis : Each x-y data point is colored by its corresponding posterior projected mean in the latent space.

6. Discussions

Both SOM and GTM are unsupervised data driven clustering algorithm. However GTM addresses the shortcomings of SOM clustering algorithm. The GTM model defines a probability density, there is a cost function associated with it and there is a proof of convergence. From the comparison of the SOM and the GTM results for the Lambda-Mu-Rho analysis, the output facies map is similar, mostly the coloring of the facies is different. Different colors in the map in the above analysis should represent the variability of the geomechanical character of the Barnett shale. We are more confident with the GTM results which is a more superior algorithm.

7. References

- Bishop. M. Christopher, Svensen M and Williams K. I, GTM: Generative Topographic Mapping, Neural Computation10, 1998.
- Bishop. M. Christopher, Svensen M and Williams K. I, Developments of the Generative Topographic Mapping, Neural Computation10, 1998.
- Wallet. C. Brad, Matos, M, Kwiatkowski, Tim and Suarez Y., Latent space modeling of seismic data : an overview
- Matlab GTM toolbox.