Well logs inversion into lithology classes: comparing machine learning and Bayesian inversion

Miao Tian, Henning Omre, Huaimin Xu

Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway

College of Geoscience, China University of Petroleum Beijing, China

tiamo117@hotmail.com

Introduction

Lithology is a crucial factor in reservoir characterization. The cores from wells are considered as the most accurate categorical information. However, cores are always selectively taken in the favourable depth intervals. Thus, for the remaining depths the quantitative information of well logging data are needed by classification models for lithology recognition and prediction[2]. Spatial characteristics of sediments and convolution properties of well logs data poses challenges in the classification of lithologies. Fortunately, both Bayesian inversion and deep learning frameworks have developed models to process the inputs as sequences while keeping their spatial dependency [1]. In this study we focus on a kernel-based hidden Markov model (HMM) and a Gated recurrent unit (GRU) classifier on the same well logs dataset and compare their results. Also a result from traditional deep neural network (DNN) without any spatial dependency is presented.

Field data and notation

In this case study the sedimentary lithologyes are recognized as medium sandstone (MS), fine sandstone (FS) and siltstone (SS). The five normalized log curves from three wells are acoustic log (AC), density log (DEN), gamma ray (GR), log-deep resistivity (R4) and spontaneous potential (SP) (Fig.1). Pairwise scatterplots of the well logs sorted by lithologies are displayed in Fig.2. Large overlaps exist especially between the FS and the SS. Integrating the categorical information hidden in diverse logs with the spatial dependency is a reasonable choice in the exploration study. The 1D profile along the well path is discretized to $\mathcal{T} = \{1, \ldots, T\}$. At each $t \in \mathcal{T}$, a observation vector $\mathbf{d}_t = (d_{t,1}, \ldots, d_{t,5})$ is provided by five well logs with $\mathbf{d} = \{\mathbf{d}_t; t = 1, \ldots, T\}$. The corresponding categorical attribute of each depth is assigned one of the three lithologies $\kappa_t \in \Omega_\kappa : \{MS, FS, SS\}$. The objective of our study is to assess the full lithology profile represented by the vector $\boldsymbol{\kappa} : \{\kappa_t; t = 1, \ldots, T\}$ given the observations $\mathbf{d}$, i.e. $[\boldsymbol{\kappa}|\mathbf{d}]$.

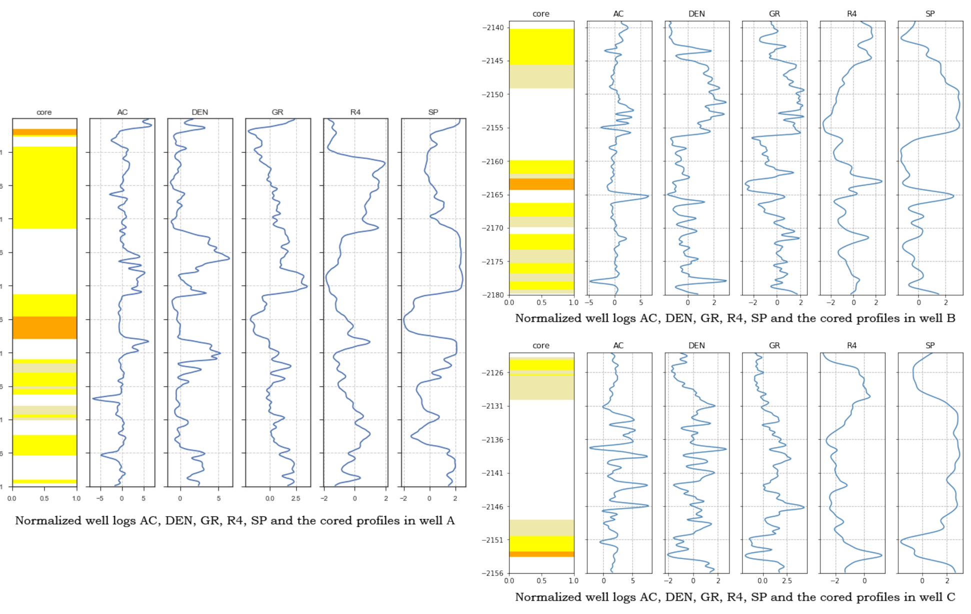

Figure 1: Normalized well logs AC, DEN, GR, R4, SP and the cored profiles in three wells.

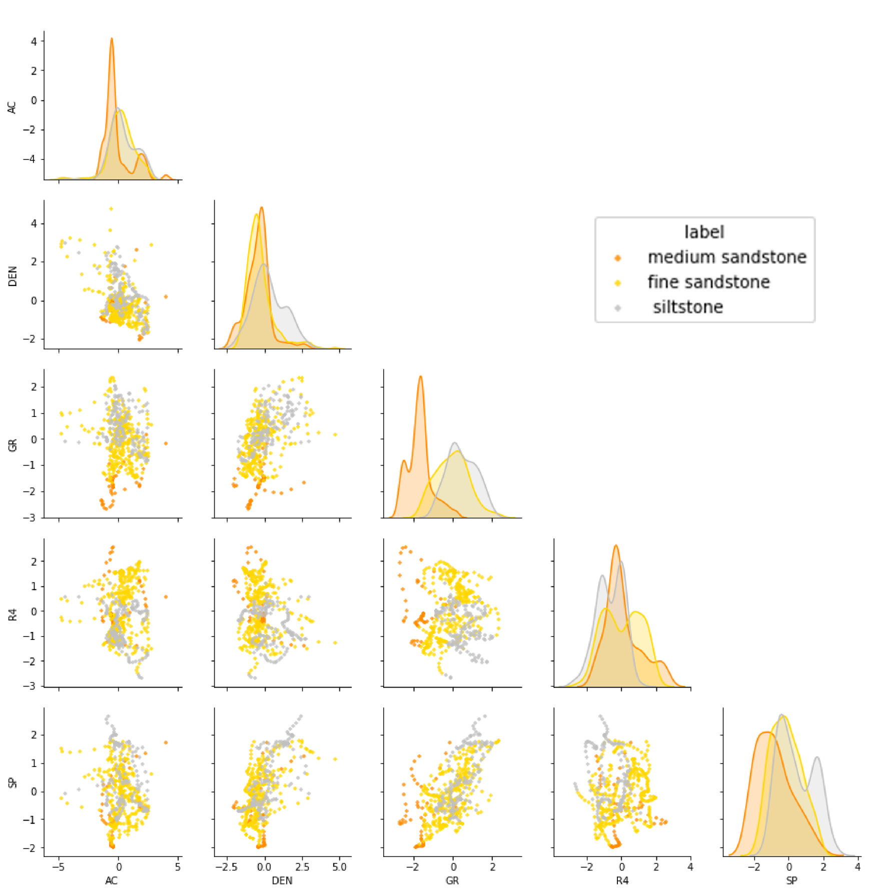

Figure 2: Pairwise scatterplots and histograms of all the well logs sorted by lithologies.

Methods

The HMM and GRU model supply their solutions to the objective in different ways.

Kernel-based hidden Markov model

Bayesian inversion is cast in a probabilistic setting and the solution is the posterior model. The posterior model is provide by Bayes' rule,

$$p(\boldsymbol{\kappa}|\mathbf{d}) = \frac{1}{p(\mathbf{d})} \times p(\mathbf{d}|\boldsymbol{\kappa}) p(\boldsymbol{\kappa})$$

where the likelihood model $p(\mathbf{d}|\boldsymbol{\kappa})$ defines the procedure of well logs data collection, the prior model $p(\boldsymbol{\kappa})$ represents the geological and exploration experience with the interest variable $\boldsymbol{\kappa}$, and $p(\mathbf{d})$ in the equation is a normalizing constant. The prior model represents our original knowledge about the geological setting and sediments. A stationary Markov chain with the first-order Markov property is chosen to guarantee the spatial coupling of the lithologies. It can be given by,

$$p(\boldsymbol{\kappa}) = p_s(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \ldots, \kappa_1) = p_s(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1})$$

We assumed that the likelihood models are conditional independent with single-site response. Hence $p(\mathbf{d}|\boldsymbol{\kappa})$ can be expressed as,

$$p(\mathbf{d}|\boldsymbol{\kappa}) = \prod_t p(\mathbf{d}_t|\boldsymbol{\kappa}) = \prod_t p(\mathbf{d}_t|\kappa_t)$$

Denote the well logs of cores of class κ as $\mathbf{d}^\kappa = (\mathbf{d}_1^\kappa, \ldots, \mathbf{d}_{n_\kappa}^\kappa)$. We can use a kernel estimator to estimate it:

$$\hat{p}_k(\mathbf{d}_t|\kappa_t) = \frac{1}{n_\kappa h_{n_\kappa}} \sum_{i=1}^{n_\kappa} k\left(\frac{\mathbf{d_t} - \mathbf{d_i^\kappa}}{h_{n_\kappa}}\right); \quad \kappa \in \Omega_\kappa$$

where $k(\tau)$; $\tau \in \mathcal{R}^5$ is the kernel function and h_{n_κ} is the band width which defines the smoothness of the density distribution. The kernel function applied here is Gaussian kernel function [3]. Hence the posterior model is then fully defined. It can be assessed by the Forward-Backward algorithm and be expressed as,

$$p(\boldsymbol{\kappa}|\mathbf{d}) = \frac{1}{p(\mathbf{d})} \times \prod_{t \in \mathcal{T}} p(\mathbf{d}_t|\kappa_t) \times p_s(\kappa_1) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1})$$

$$= p(\kappa_1|\mathbf{d}) \prod_{t \in \mathcal{T}_{-1}} p(\kappa_t|\kappa_{t-1}, \mathbf{d})$$

The Viterbi algorithm [3] provides the gobal maximum posterior (MAP) prediction,

$$\hat{\boldsymbol{\kappa}}_{MAP} = \underset{\boldsymbol{\kappa}}{\operatorname{argmax}}\{p(\boldsymbol{\kappa}|\mathbf{d})\}$$

and we also can quantify uncertainty by probability profiles and generate realizations.

Gated recurrent unit neural networks

In GRU, a memory cell with several gates is used to integrate the input data at current depth and the information inherited from deeper depth cells. The memory cells extract and convey the information in the following way [1],

$$\tilde{\mathbf{c}}_t^{[l_R]} = f_{tanh}(\boldsymbol{\omega}_c^{[l_R]}[\boldsymbol{\Gamma}_r^{[l_R]} \times \mathbf{a}_{t-1}^{[l_R]}, \mathbf{a}_t^{[l_R-1]}]^T + \mathbf{b}_c^{[l_R]})$$

$$\mathbf{a}_t^{[l_R]} = \boldsymbol{\Gamma}_u^{[l_R]} \times \tilde{\mathbf{c}}_t^{[l_R]} + (\mathbf{i}_{N_{l_R}} - \boldsymbol{\Gamma}_u^{[l_R]}) \times \mathbf{a}_{t-1}^{[l_R]}$$

$$\boldsymbol{\Gamma}_r^{[l_R]} = f_{sigmoid}(\boldsymbol{\omega}_r^{[l_R]}[\mathbf{a}_{t-1}^{[l_R]}, \mathbf{a}_t^{[l_R-1]}] + \mathbf{b}_r^{[l_R]})$$

$$\boldsymbol{\Gamma}_u^{[l_R]} = f_{sigmoid}(\boldsymbol{\omega}_u^{[l_R]}[\mathbf{a}_{t-1}^{[l_R]}, \mathbf{a}_t^{[l_R-1]}] + \mathbf{b}_u^{[l_R]})$$

where $l_R = \{1, \ldots, L_R\}$ denotes the current GRU layer number with $\mathbf{a}_t^{[0]} = \mathbf{d}_t$, $\tilde{\mathbf{c}}_t^{[l_R]}$ denotes an update candidate N_{l_R} vector at depth t, $\mathbf{a}_t^{[l_R]}$ is a N_{l_R} output vector of the l_Rth memory cell layer at depth t, $\boldsymbol{\omega}_c^{[l_R]}$, $\boldsymbol{\omega}_u^{[l_R]}$ and $\boldsymbol{\omega}_r^{[l_R]}$ are parameter $N_{l_R} \times (N_{l_R} + N_{l_{R-1}})$ matrices, $\mathbf{b}_c^{[l_R]}$, $\mathbf{b}_u^{[l_R]}$ and $\mathbf{b}_r^{[l_R]}$ denote three bias parameter N_{l_R} vectors for corresponding outputs adjustment, $\mathbf{i}_{N_{l_R}}$ is a unit N_{l_R} vector. The functions $f_{tanh}(\cdot)$ and $f_{sigmoid}(\cdot)$ are activation functions which improve the nonlinear regression performance. The final outputs of GRU, $\mathbf{a}_t^{[L_R]}$; at each layer $t \in \mathcal{T}$, is an artificial N_{L_R} vector which capture the spatial information in the well logs. It is processed by an extra classifier to get the $\hat{\kappa}_t$. Here we use a DNN model to do the classification and define the prediction by the depthwise maximum likelihood:

$$\hat{\boldsymbol{\kappa}}_{GRU} : \{\hat{\kappa}_t = \underset{\kappa_t}{\operatorname{argmax}}\{\mathbf{p}_t(\mathbf{a}_t^{[L_R]}|\kappa_t)\}; \quad t = 1, \ldots, T\}$$

Results with discussion

We separately trained a kernel-based HMM, a GRU model with 64-128-64 memory cell layers followed by 32-16-3 general hidden layers and a DNN model with the same structure by using the observations from two training wells. Data from the third well is used as blind test and assigned to these three models to get a locationwise MAP of the full profile from the three models. Furthermore, we calculate the probability profile and generate realizations from HMM. Fig.3 and 4 display the partially cored profile and all the inversion results for one of training well and the blind test well. In the training well (Fig.3), all MS layers are recognized by GRU and HMM. In the core-plugs missing interval at middle of the well HMM prefers to provide a long stable FS layer. However, there appear some thin SS layers according to GRU. At the bottom half of the training well, frequent FS-SS transitions occur. GRU predictions provide excellent results in these locations since it have seen these labels before. The HMM predictions do not capture these transitions since predictions are smoother than reality. The corresponding realizations, which are possible outcomes, do however appear with numerous transitions. In the blind test well (Fig.4), all models fail to predict the thick SS layers at -2126 to -2130m. In the middle unlabeled interval, the predictions from GRU and HMM are very similar to each other. At -2131m, a thin MS layer may occur according to the HMM, while GRU does not capture this effect. HMM is actually quite uncertain about the lithology classes in this interval. The simulations also reflect this uncertainty by provideing different lithologies in different realizations in this segment. At the bottom of the well, all models predict the MS layer around -2153m but none of them get the correct thickness.

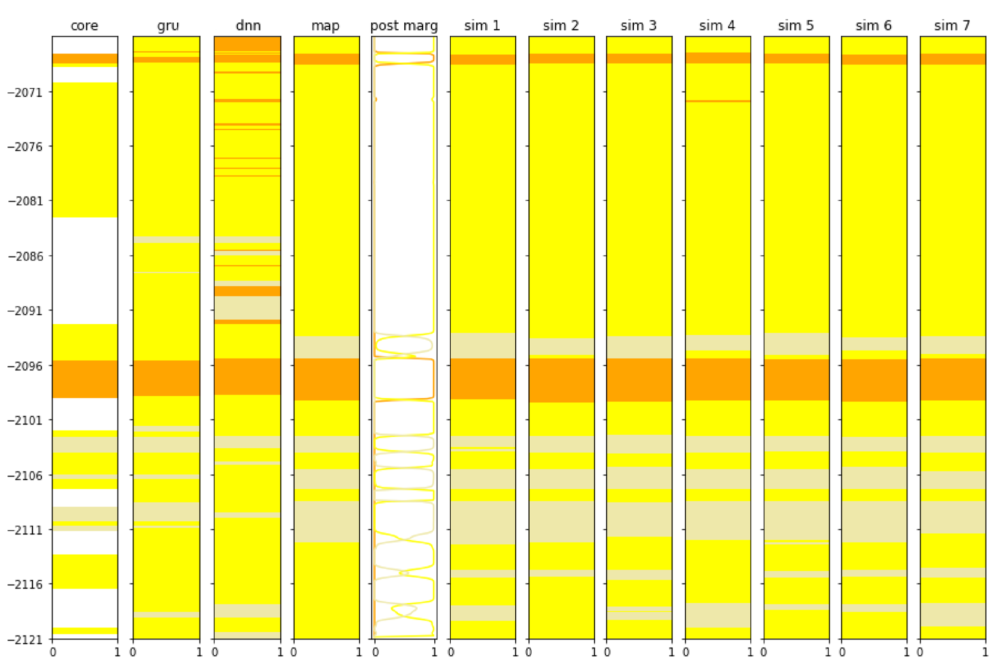

Figure 3: The comparison of the prediction results in the training well.

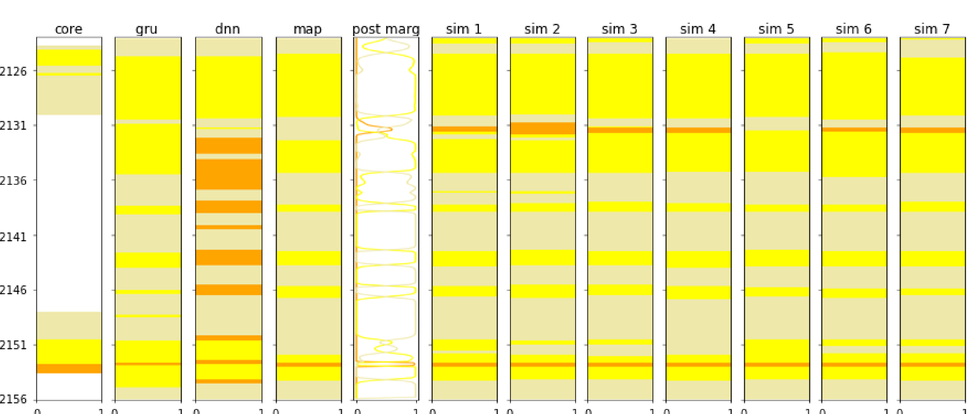

Figure 4: The comparison of the prediction results in the blind test well.

Conclusions

The HMM and GRU models provide very similar predictions on the blind well, they are clearly favorable to the non-spatial DNN predictions. The HMM model is phrased in a consistent probabilistic framework, hence it also provide quantifications of uncertainty as probability profiles and a set of realizations.

References

[1] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014.

[2] Jo Eidsvik, Per Avseth, Henning Omre, Tapan Mukerji, and Gary Mavko. Stochastic reservoir characterization using prestack seismic data. *Geophysics*, 69(4):978–993, 2004.

[3] Selamawit Serka Moja, Zeytu Gashaw Asfaw, and Henning Omre. Bayesian inversion in hidden markov models with varying marginal proportions. *Mathematical Geosciences*, pages 1–22, 2018.