Deep learning based automatic marker separation
Atul Laxman Katole*, Aria Abubakar, Edo Hoekstra, Srikanth Ryali, and Tao Zhao, SLB

Summary

Transitions in the geological layers are referred as markers and they define the key formation boundaries in the regional stratigraphic frameworks. The information pertaining to the markers is gathered from multiple sources as numerous tools prevalent in the Oil and Gas industry, multiple public data bases and vendor supplied datasets. The interpretation of these markers collected from the multiple sources lacks consistency due to the involvement of multiple personnel and different standards. The authors propose a deep learning-based approach to consistently map the markers across the wells in a basin. The proposed automatic marker separation approach is based on dimensionality reduction of the marker waveforms using latent space representation in the transformer models. The latent representation is further transformed to two-dimensional embeddings using Uniform Manifold Approximation and Projection (UMAP). The resultant two dimensional embeddings can be readily segregated using clustering approaches as K-Means and Gaussian Mixture Models (GMM). The probabilistic clustering approach as GMM provides the probability of marker assignment to a specific cluster thereby providing us the confidence in each assignment. The proposed automated clustering approach is evaluated using the six markers available in a well dataset from the Williston basin in the North America. We find that the proposed marker separation approach provides superior accuracy, giving us high precision and recall for all the analyzed markers. The proposed approach paves the way for automated marker separation which leads to consistent mapping of markers across hundreds of wells in a basin.

Introduction

Identification of the formation boundaries is a critical step in the development of the stratigraphic frameworks. Markers in the well logs define these boundaries, therefore marker picking and further correlating them among adjacent wells gains importance. Petrophysicists often spend huge amount of time to pick the markers in hundreds of wells in a basin. Also, the manual marker picking process is highly subjective and prone to human errors. As manual marker picking process is time consuming, it has been mimicked in software using supervised learning approaches which include self-attention convolutional neural networks (Kulkarni et al., 2020). A successful marker picking approach involves simultaneous understanding of the global as well as the local context of the markers. The global context is defined by the measurement logs for the entire well which essentially encodes the order of occurrence of all the markers in the well. The local context is defined by the smaller log window adjacent to the marker (Maniar et al., 2018; Kulkarni et al., 2020). Assessment of the quality of the data for training marker classification is essential as supervised learning approaches demand consistency in training examples. Such quality control mechanisms have been implemented using Siamese Neural Networks (Kulkarni et al., 2018).

The markers when collected from multiple sources need to be scrutinized for consistency as they are picked by different teams of petrophysicists and due to the highly subjective nature of the marker picking process. The inconsistency in the markers also arises as the interpretations are probably performed using different stratigraphic columns. The authors propose a fully automated robust methodology to consistently map the markers across all the wells in the region. In the last couple of years, transformers have successfully modeled the sequential nature of the occurrence of words in the languages and solved complex Natural Language Processing (NLP) problem statements as language translation, question answering, named entity recognition (NER) etc. (e.g., Vaswani et al., 2017; Devlin et al., 2018). The proposed approach models the sequential nature of the well logs using transformers and exploits its capability to extract complex feature hierarchy from well logs for delivering superior automated marker separation solution.

Method

Transformers entirely dispenses with convolutional and recurrence-based approaches, and instead rely on the attention mechanism to model the sequential nature of the symbols occurring in natural languages (Vaswani et al., 2017). We train a transformer-based model utilizing masked language model paradigm, which is widely used to train models in NLP domain. The input to the transformer model includes randomly masked gamma ray (GR), density (DEN) and resistivity (RES) time series measurements from well logs in the dataset. The ground truth for the randomly masked values in the input signal is already known as it is available in the training data. The transformer model can now be trained in a self-supervised manner to predict the randomly masked values in the input signal as shown in the Figure 1. The transformer model trained with self-supervised approach learns the complex hierarchy of features for the GR, DEN & RES measurements. As the transformer model learns to accurately predict the masked input values, it is expected to implicitly learn to distinguish between the subtle differences in the input time series windows.
Deep learning based automatic marker separation

The automatic marker separation workflow is captured in the Figure 2. The time series log window for GR, RES and DEN measurement is cropped at the marker depth and presented as input to the transformer model which is pre-trained with the self-supervised approach.

Instead of using the output of the transformer, the latent space representation of the pre-trained transformer model is used as a reduced dimensional embedding. This reduced dimensional embedding learns distinct representations for distinct markers. In the next step, UMAP based dimensionality reduction is applied on the latent space representation for the markers to obtain a two-dimensional embedding. The two-dimensional embedding for the markers is given to Gaussian Mixture Model (GMM) based clustering approach to probabilistically cluster the markers. GMM based probabilistic clustering automatically clusters the two-dimensional embeddings of same markers together. Also, the clusters formed for distinct markers are well separated thereby achieving automatic marker separation objective. GMM based clustering approach also gives us the probability of assignment of each marker in a cluster thereby providing the confidence score for each assignment.

**Dataset Description**

The proposed automatic marker separation approach is evaluated using a dataset from the Williston Basin. It follows layer cake geology and presents some stratigraphic complexities with pinch outs. There are 452 vertical and deviated wells in this dataset. It is a cleaned up public repository where multiple logs are spliced up and normalized. The dataset includes GR, RES, and DEN triple combination of log measurements. Also, this dataset provides logs for 144 wells with high quality marker picks. There are ten markers UB000, MB000, LB000, LB009, BB000, DP00, TF000, TF100, TF200, TF300, TF330 and TF400 represented in this dataset.

We choose six markers to evaluate automatic marker separation approach. Three out of six markers as DP000, BB000 and UB000 are prominent markers. The rest of the three markers TF000, TF200 and TF330 are subtle markers. We capture five examples per marker in a row, of the GR log pattern window of length 200 samples as shown in Figure 3. We can observe that there is huge variability in the log patterns for these markers. The first three examples
Deep learning based automatic marker separation

constitute consistent log pattern examples and the rest of the two examples capture the inconsistent log patterns that demonstrate the variability aspect for the markers. For example, we can observe that marker UB000 has a pulse like behavior which lasts for 20 samples (10 ft.). The log measurements increase vertically to around 500 GAPI after which it gradually increases to its peak value of around 650 GAPI. This behavior is consistently observed for first three examples whereas in the rest of the examples the pulse is largely deformed or dilated.

Results

We evaluate the proposed automatic marker separation solution using six markers BB000, DP000, TF000, TF330, TF200 and UB000. The results of the marker separation are displayed in the Figure 4. The output of UMAP dimensionality reduction when applied to the transformer latent space representation of the marker is displayed in Figure 4(a). Figure 4(b) represents the output of the K-Means algorithm when applied to the two-dimensional embeddings in Figure 4(a). Figure 4(c) represents the output of the GMM based probabilistic clustering when applied on the two-dimensional embeddings in Figure 4(a). Finally, the Figure 4(d) represents the ground truth. We can observe that the six clusters C1, C2, C3, C4, C5 and C6 in the Figure 4(b) and 4(c) corresponds to the embeddings for the markers UB000, TF200, TF330, TF000, DP000 and BB000 respectively. The precision and recall metric for the evaluation are captured in the Table 1.

Figure 3: GR log pattern windows for markers BB000, DP000, TF000, TF330, TF200 and UB000 are captured with five examples each in a row. The first three examples correspond to the consistent patterns whereas the last two demonstrates the inconsistent patterns.
Deep learning based automatic marker separation

Table 1: Table captures the performance of the marker separation approach using precision and recall values.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB000</td>
<td>1</td>
<td>0.96</td>
</tr>
<tr>
<td>DP000</td>
<td>1</td>
<td>0.95</td>
</tr>
<tr>
<td>TF000</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td>TF330</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>TF200</td>
<td>0.87</td>
<td>0.97</td>
</tr>
<tr>
<td>UB000</td>
<td>1</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Figure 4: Automatic marker separation results for the markers BB000, DP000, TF000, TF330, TF200 and UB000 are captured with four plots. (a) Depicts the marker representations using transformer latent space and UMAP based dimensionality reduction; (b) Depicts the output of K-Means algorithm applied on marker representations in (a); (c) It depicts the output of GMM based probabilistic clustering on marker representations in (a); (d) Presents the ground truth values for marker representations.

Conclusions

Transformer based modeling which is prevalent in the NLP domain is successfully utilized to model the sequential nature of the well logs. The UMAP based two-dimensional projections of the latent representation in transformers can comprehend the distinguishing characteristics of dissimilar markers and similarities in the same marker log pattern. The proposed automatic marker separation approach demonstrates that transformer-based models can successfully model the long-term context in the well logs and can be further utilized to solve many unresolved problems in Petroleum industry.

Acknowledgments

The authors would like to thank SLB for granting the permission to publish this work.