
Soft Attention Convolutional Neural Networks for Rare Event Detection in Sequences

Mandar Kulkarni, Aria Abubakar
Schlumberger
(mkulkarni9, aabubakar) @ slb.com

Abstract

Automated event detection in the sequences is an important aspect of temporal data analytics. The events can be in the form of peaks, changes in data distribution, changes of spectral characteristics etc. In this work, we propose a Soft-Attention Convolutional Neural Network (CNN) based approach for rare event detection in sequences. For the purpose of demonstration, we experiment with well logs where we aim to detect events depicting the changes in the geological layers (a.k.a. well tops/markers). Well logs (single or multivariate) are inputted to a soft attention CNN and a model is trained to locate the marker position. Attention mechanism enables the machine to relatively scale the relevant log features for the task. Experimental results show that our approach is able to locate the rare events with high precision.

1 Introduction

Automated event detection/localization in the sequences is an important aspect of temporal data analytics. In this paper, we propose a machine learning based approach for rare event detection in sequences and validate its efficacy for sequential well log data.

Well logs are geological sequences which uniquely characterize sub-surface lithology. Typically, the measurements taken are Gamma Ray (GR), Resistivity (RES), Density (DEN) and the standard resolution of the measurements is 0.5ft. The domain experts mark the depths in the well logs where there is a change in the geological layers. These labels are referred to as well tops or markers. Various markers are picked in all the wells in the basin in order to correlate the layer boundaries across the different wells. This problem is referred to as well log correlation. Fig. 1(a) shows sample well logs with GR measurements labeled with a marker location (indicated in red). Note that each well log has close to or more than 20k samples and only one of the index corresponds to the marker. This renders a marker localization a rare event detection task.

We propose a Soft-Attention CNN framework for localizing these rare marker events in the well logs. For locating a particular marker over the log, all the input log patterns may not be relevant. Soft attention mechanism enables the machine to automatically adjust the scaling of the relevant log features for the marker picking. Experimental results show that the model is able to locate markers with high precision. Moreover, the same model architecture works well across the markers as well as with different log types, thus saving the compute effort on the case-specific hyper-parameter tuning. We observe that, during the manual top picking process, interpreters look at the entire well log to decide the relevant log section to focus on. We intuitively try to emulate the interpreter's workflow through soft attention.

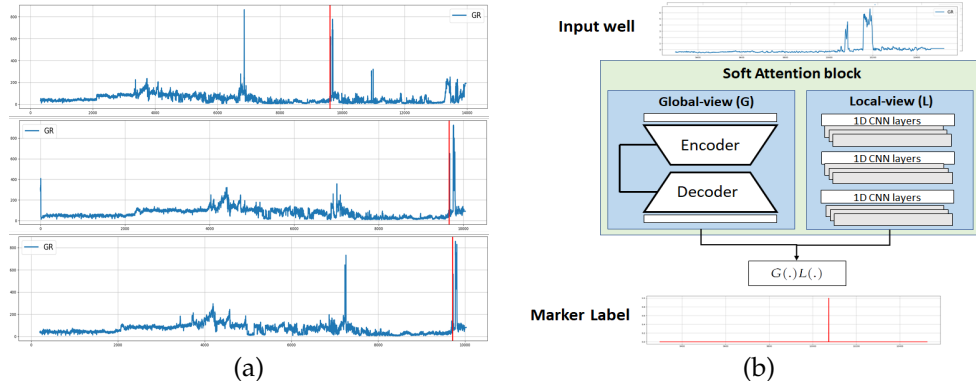


Figure 1: (a) Sample well logs from Bakken dataset labeled with marker location (marked in red). x-axis indicates the depth (b) Proposed soft attention CNN architecture.

2 Methodology

CNNs with soft attention (Wang et al. (2017), Chen et al. (2016)) have shown promising results for image classification/segmentation problems. In this work, we demonstrate its application for the rare event detection in sequences.

Fig. 1(b) depicts the proposed network architecture. Input to the model is the entire sequence (i.e. well log) and the network is trained to predict the marker location. For training the model, the location of an expert marker is converted to a one hot vector and used as the label. Under this setting, we train one model for each marker separately. Since the model is fully convolutional, it can be trained with input well logs of variable lengths.

The well log is simultaneously input to Global-view (G) and Local-view (L) models. The global view model comprise of an encoder-decoder architecture. We resort to the U-Net (Ronneberger et al. (2015)) styled architecture where there are shortcut connections between the corresponding layers of the encoder and the decoder. The encoder model consist of 1D convolution layers followed by average pooling layers while the decoder model consist of 1D convolution layers followed by interpolation (up-sampling) layers. After each convolution layer, we use layer normalization and dropout layers. Due to the successive pooling operations in the encoder, the same convolution filter size provides a larger receptive field in the deeper layers of the encoder and thus helps to capture the log pattern variations over a much longer depth range. Hence, we refer to this model as the global-view model. The local-view model consists of a stack of 1D convolution layers. Since the convolution by nature is a local operation, it helps to model the log pattern variations over a shorter scale.

The output tensor from the Global model is multiplied, element-wise, with the output tensor of the Local model as follows

$$A(x) = G(x)L(x) \quad (1)$$

where A , G and L indicates the outputs of the attention block, the global-view model and the local-view model, respectively and x indicates the well input. For G , we use 'tanh' activation, hence features from L can be scaled positively/negatively as required. The scaled tensor A is then mapped to the output layer using sigmoid activation. Fig. 4 in appendix section 6.3 shows the sample prediction result with attention scores and prediction probabilities for the marker. The convolution layers in global-view as well as local-view model consist of inception blocks (Szegedy et al. (2015)), where convolution is performed with multiple filter sizes and the feature maps are concatenated. Experimentally, it is observed that, for the local-view model, dilated filters provide better results. In ablation study section, we experimentally validate the effectiveness of the global-local view model as compared to the results obtained with either model alone.

In the numerical experiments, the labeled data are divided into training, validation, and test set. The model is trained end-to-end using binary cross entropy loss until the validation loss cease to improve. To mitigate the issue of sparse labels (i.e. label imbalance), we use

Marker	1ft	2ft	5ft	10ft	Recall
UB000	0.82	0.93	0.93	0.93	1.
MB000	0.96	1.	1.	1.	1.
LB000	0.93	0.96	1.	1.	1.
LB009	0.75	0.88	1.	1.	1.
BB000	0.92	1.	1.	1.	0.93
TF180	0.7	1.	1.	1.	0.83
TF200	0.73	1.	1.	1.	0.92
TF300	0.56	0.78	1.	1.	0.82
TF330	0.61	0.83	0.94	1.	1.

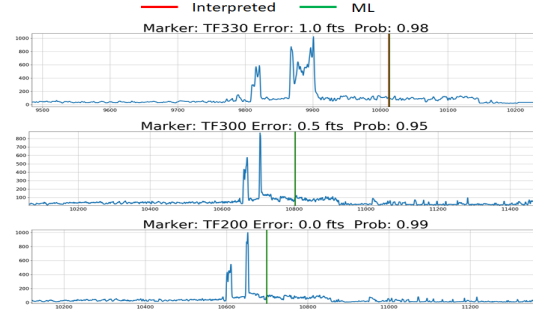
(a) GR

Marker	1ft	2ft	5ft	10ft	Recall
UB000	0.96	1.	1.	1.	1.
MB000	0.83	1.	1.	1.	1.
LB000	0.95	1.	1.	1.	0.91
LB009	0.71	0.86	0.86	1.	1.
BB000	0.82	0.91	1.	1.	1.
TF180	0.5	0.7	1.	1.	0.91
TF200	0.7	0.95	1.	1.	0.91
TF300	0.59	0.76	0.88	1.	0.89
TF330	0.75	0.92	1.	1.	0.71

(b) GR-RES

Marker	1ft	2ft	5ft	10ft	Recall
UB000	0.87	0.96	1.	1.	1.
MB000	0.95	1.	1.	1.	0.95
LB000	1.	1.	1.	1.	0.95
LB009	0.92	0.92	1.	1.	0.93
BB000	1.	1.	1.	1.	1.
TF180	0.5	0.8	1.	1.	0.95
TF200	0.89	1.	1.	1.	0.9
TF300	0.33	0.6	1.	1.	0.83
TF330	0.5	0.86	1.	1.	0.88

(c) GR-RES-DEN



(d) Detection of subtle markers

Figure 2: (a-c) Results with varied number of logs, (d) marker localization plots for subtle markers. Error indicates distance between Interpreted (Expert) and ML marker.

a Gaussian label smoothing where a Gaussian kernel of a fixed scale is convolved with the one hot label. In the ablation study section, we demonstrate the effectiveness of label smoothing strategy.

3 Evaluation metric

To demonstrate the effectiveness of the proposed approach, we devised a following metric. For each marker type, the ML detection depth D_i^{ML} and the error e_i (in fts) between the ML marker and the manual marker is computed as follows

$$D_i^{ML} = \arg \max_j P_{ij} \quad (2a)$$

$$e_i = |D_i^{\text{Expert}} - D_i^{ML}| \quad (2b)$$

P_{ij} indicates the probability prediction over the depths. j ranges over all the depths in the i^{th} target well and $\max_j P_{ij}$ is the probability of the detection.

The precision and recall of the ML predictions is then in calculated as follows

$$Prec_{d_T} = \frac{\sum_i I[e_i \leq d_T]}{M} \quad (3a)$$

$$Recall = \frac{M}{N} \quad (3b)$$

Here, I denotes the indicator function which evaluates to 1 if the condition is satisfied. M denotes the number of ML pick detections. For each marker, we calculate the precision with multiple error tolerances where d_T is set to 1ft, 2ft, 5ft, 10ft. For the recall, N denotes the number of test wells.

To measure the performance on the dataset with all markers, we calculate the F_1 -score as follows

$$F_1 = \frac{2 \text{Mean}(Prec_{d_T}) \text{Mean}(Recall)}{\text{Mean}(Prec_{d_T}) + \text{Mean}(Recall)} \quad (4)$$

where the mean is calculated across all the markers. For all the reported F_1 scores in this paper, d_T is set to 2ft.

3.0.1 Model uncertainty

To estimate the model uncertainty, we use Monte Carlo (MC) Dropout [Gal and Ghahramani (2016)] strategy. For each test well, we make multiple (30) predictions with dropout and the standard deviation of the detection depths is used as the proxy for the model uncertainty. The marker predictions with probability score more than a threshold and model uncertainty less than a threshold are considered valid detections. For all our experiments, we set the uncertainty threshold to 5 fts.

4 Experimental results

We now demonstrate the effectiveness of the proposed architecture under various scenarios. Note that the same network architecture was used for generating all the results.

4.1 Description of the dataset

For the purpose of experimentation, we use a public dataset collected from the Williston basin (Bakken). The wells have logs such as Gamma-Ray (GR), Resistivity (RES), and Density (DEN). We have a dataset with 144 labeled wells. The subset of labeled wells with GR, GR-RES and GR-RES-DEN logs are 144, 122, and 118, respectively. There are two variety of markers present in this dataset. The strong signature markers included in the study are UB000, MB000, LB000 and LB009. For these markers, a prominent signature is visible on the log. The subtle signature markers included in the study are BB000, TF180, TF200, TF300, and TF330. For these markers, subtle signature is observed on the log, which renders the detection challenging.

4.2 Results under various scenarios

Fig. 2(a) shows the marker localization result with GR log where 80% of the wells are used for training and validation while 20% wells are used for testing. It can be seen that the model is able to pick markers with high precision. Majority of the markers are detected within 5ft of error tolerance with relatively good recall. Fig. 2(b) and Fig. 2(c) shows the results with GR-RES and GR-RES-DEN logs with 80% train-val wells. For the majority of markers, better precisions are obtained with multiple logs.

We further investigated the results for the challenging subtle markers. Fig. 2(d) shows the detection result for GR log with subtle markers. Interestingly, even for the markers that are barely visible on the log, our model is able to pick it accurately. We hypothesize that, though a local signature of the marker is not prominent, a global view of the well is guiding the marker localization. For visual validation of the accuracies, we plotted error histograms for all the results. Plots are provided in appendix in section 6.2.

We also performed ablation study to understand the effect of individual components. Results are provided in appendix in section 6.1. Table 1 in section 6.1.1 also shows the F1-score comparison with an existing well correlation approach proposed in Maniar et al. (2018). Note that our approach provides relatively better results.

5 Conclusion

In this paper, we proposed soft attention convolutional neural network based approach for well log correlation. The combination of global-view and local-view model is proposed to locate the markers given a single or multiple input logs. The proposed method achieved high precision on marker picks where the same architecture reliably works well across the markers and variable number of logs. Importance of individual components is verified through Ablation study.

References

- Chen, L.-C., Y. Yang, J. Wang, W. Xu, and A. L. Yuille, 2016, Attention to scale: Scale-aware semantic image segmentation: Proceedings of the IEEE conference on computer vision and pattern recognition, 3640–3649.
- Gal, Y., and Z. Ghahramani, 2016, Dropout as a bayesian approximation: Representing model uncertainty in deep learning: international conference on machine learning, 1050–1059.
- Maniar, H., S. Ryali, M. S. Kulkarni, and A. Abubakar, 2018, Machine-learning methods in geoscience, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 4638–4642.
- Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical image computing and computer-assisted intervention, Springer, 234–241.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, 2015, Going deeper with convolutions: Presented at the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Wang, F., M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, 2017, Residual attention network for image classification: Proceedings of the IEEE conference on computer vision and pattern recognition, 3156–3164.

6 Appendix

6.1 Ablation study

In this section, we performed experiments to understand the effect of individual components on the final result.

6.1.1 Effectiveness of global-view and local-view model

To validate the effectiveness of the combination of global and local models, we performed an experiment, where we trained each model independently and compared their F_1 scores. Table 1 shows the F_1 scores comparison results for Bakken and Permian datasets. It can be seen that, in both the cases, the combined model outperforms the individual global and local model. Table 1 also shows comparison result with an existing well correlation approach.

Dataset	Global (G)	Local (L)	Global + Local	Maniar et al. (2018)
Bakken	0.915	0.929	0.937	0.91
Permian	0.913	0.821	0.939	-

Table 1: Comparison of F_1 scores for individual models.

6.1.2 Effectiveness of label smoothing

To validate the effectiveness of Gaussian label smoothing, we performed an experiment without label smoothing. Table 2 shows the F_1 scores comparison with and without label smoothing. It indicates that the label smoothing gives significantly better results.

Dataset	Without	With
Bakken	0.549	0.937
Permian	0.652	0.939

Table 2: Comparison of F_1 scores for label smoothing.

6.2 Error histograms

Fig. 3 shows the histogram of marker errors. It provides the visual validation of high precision on the marker picks. Note that in all the cases, majority of the markers are picked within 2ft of the error tolerance.

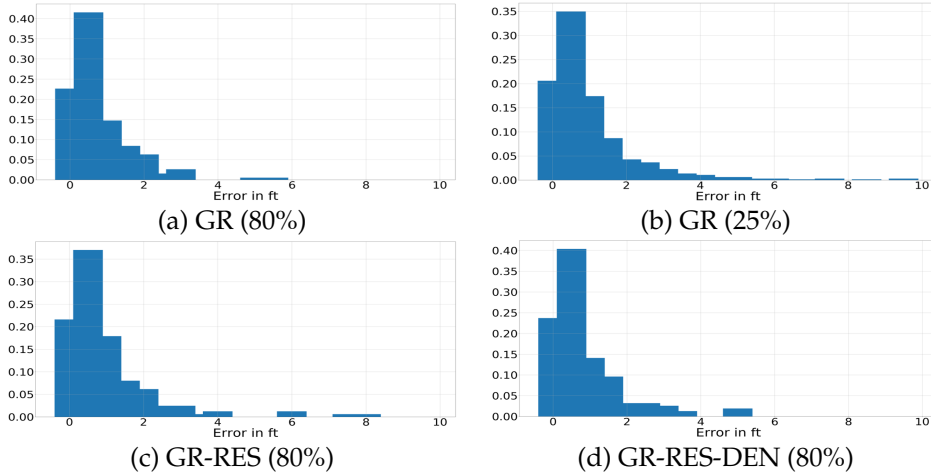


Figure 3: Error histograms. x-axis indicates the error in fts. Percentage of wells used for training is indicated in the brackets.

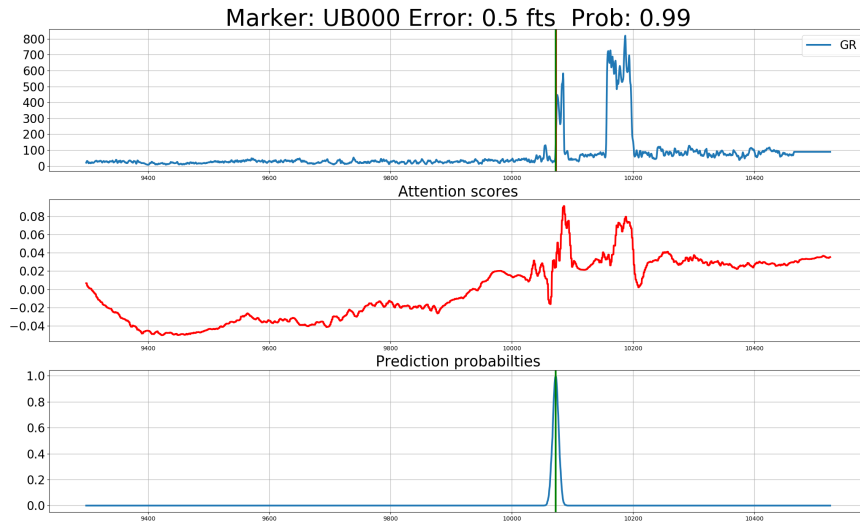


Figure 4: Prediction result with attention scores.

6.3 Prediction result with attention scores

Fig. 4 shows a marker prediction result along with the attention score over depths and prediction probabilities. Attention scores are obtained by averaging tensor G along the feature axis. Note that the different depths have been assigned positive/negative attention score according to their relevance for the marker prediction.