Framework for automatic globally optimal well log correlation

Oleh Datskiv
odatsk@softserveinc.com
Andrii Struk
astru@softserveinc.com
Mykola Maksymenko
mmaks@softserveinc.com

Mariia Veselovska
mvese@softserveinc.com
Oleh Bondarenko
obond2@softserveinc.com
Volodymyr Karpiv
vkarpi@softserveinc.com

Abstract

Well correlation based on well-logging data is a reliable tool that geological scientists use to interpret and deduce underground sedimentary morphology. Traditional methods are not fully-automated and require additional inputs from the experts to perform well correlation, which complicates the whole process and makes it time-consuming. Well-log data is often noisy and incomplete which significantly reduces performance of well correlation and the accuracy of geological interpretation. To address this issue, we present a framework for the global pattern correlation that is fully automated and does not require additional inputs from the user. Our framework efficiently handles imperfect data with multi-log curve integration. Global optimality in the proposed framework is achieved through adapting Hungarian algorithm to the assignment problem of well log correlation. Finally, we assess performance of the framework on real-world datasets.

1 INTRODUCTION

Attempts to develop automatic systems for correlating horizons in well sections have been made with varying degrees of success since the 1960s. The need for such tools has become particularly urgent in studies of closed areas, where information on sedimentary sequences is obtained mainly by well logging and seismic surveys [1]. Most of the implemented technological solutions give the interpreter a set of tools for analyzing, editing, and visualizing well data. Selecting the correlation variant and proper identification of horizons and boundaries in different wells is the prerogative of the specialist and depends on his installations, experience. The task of boundary identification can be also successfully accomplished with the machine learning approaches. The proposed framework is focused solely on well log correlation and does not consider labelling of stratigraphic layers. In this paper, we propose a solution that prevents the occurrence of discrepancies between scientists, does not depend on their skills and is completely automated.

2 LITERATURE OVERVIEW

Various approaches to automating stratigraphic correlation have been developed in the past. All the algorithms could be divided into two categories according to the number of wells that are considered for the computation, either pairwise-well correlation or multi-well correlation. Pairwise-well correlation is the main topic of this paper.

Rudman et al. [2] proposed an algorithm to estimate the required vertical shift between two well logs that maximizes their cross-correlation. In case the corresponding stratigraphic layers have different

AI for Earth Sciences Workshop at NeurIPS 2020.

thickness, authors propose first to resample the log in a stretched or expanded interval and second to estimate the required vertical shift.

Previously proposed approach was further improved by Mann et al. [3] with the application of Fourier transform to enable computation of correlation in the frequency domain. The calculation of specific scaling factors in the frequency domain helps better solve the problem of differences in layer thickness.

Rule-based approaches to well correlation were studied by Lineman et al. [4] and Startzman et al. [5]. Replicating the logic of experts works well in common scenarios but proved to be challenging for edge cases such as uncorrelated layers or gaps.

One of the first neural network approaches to the problem was proposed by Luthi et al. [6], it extracts the most characteristic geological patterns from each layer to further identify locations of similar markers in the other well logs. Even though the neural network architectures have significantly evolved over the last decade, they still require big volumes of annotated data to perform well.

Previously highlighted correlation methods find locally optimal solutions but may require additional computation to provide globally optimal solutions. Dynamic programming approaches proposed by Nir et al. [7] and Lapkovsky et al. [8] address the problem of global optimality by considering multiple pairs of layers simultaneously. The core principles of stretching and translating the depth are still employed in these works.

3 WELL CORRELATION FRAMEWORK

In this section, we describe our solution, which consists of several steps, which are as follows: data processing, pairwise similarity of layers and, finally, the global correlation algorithm.

3.1 Data pre-processing

Data for each well is represented as L log channels and K depth samples. Each row in this matrix represents multi-log values of one depth and each column represents the values of one log in depth, also referenced as curve in this paper, see Figure 1.

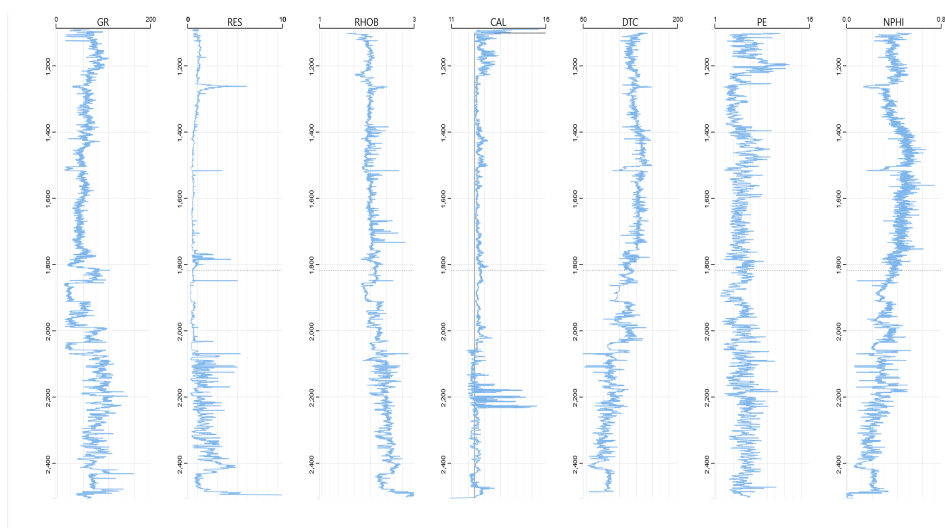

Figure 1: Sample curves data from one well

Well log data is noisy and incomplete with up to 80% of missing information. The proposed framework does not take into account the logs that miss more than 40% of information as these logs will decrease the overall performance of the correlation.

The remaining logs still miss some values that are automatically completed by the framework. Data distribution per each log is typically Gaussian or otherwise distributed within a tight range of values. In the first case the missing values are interpolated, while for the second case missing values are replaced with the previous available one. This approach yield best overall performance.

3.2 Pairwise layer similarity

This chapter describes feature preparation for the pairwise layer similarity measure that will be generalized in the next chapter to compute the globally optimal correlation results.

Feature vectors are assigned to each layer. The framework uses average depth in addition to the averages values of log curves for a given layer. Depth and log values are averaged over the entire depth of a layer. Averaging over the entire layer is robust to errors in boundary localization.

Pairwise similarity measure of two layers is the L2-norm of the respective feature vectors. With this measure, smaller value corresponds to more similar layers.

3.3 Global correlation algorithm

Previously proposed pairwise similarity measure is calculated for all the potential pairs of layers between two well logs. Similarity values are written down as a matrix with number of rows that corresponds to the number of layers in the well 1 and number of columns corresponds to the number of layers in the well 2. The globally optimal solution is then found through solving the assignment problem by an adaptation of the Hungarian matching algorithm.

Hungarian matching algorithm is a $\mathcal{O}(|V|^3)$ algorithm that can be used to find maximum-weight matching in bipartite graphs, which is sometimes called the assignment problem. A bipartite graph can easily be represented by an adjacency matrix, where the weights of edges are the entries.

4 RESULT EVALUATION

The chapter starts with the description of dataset that was used to evaluate results of the correlation framework. Due to a lack of publicly available code for well correlation, we compare performance of our framework with the results of other approaches by specifying the dataset they were originally measured on. Finally, a comparison of feature importance is presented in subsection 4.3.

4.1 Dataset description

Lithofacies data was provided by the FORCE Machine Learning competition with well logs and seismic 2020 [9] was used to assess performance of the framework. Expert geologists additionally labelled correlation marks to enable accuracy measurement of the proposed framework. The dataset contains 98 different wells from the Norwegian region located close to each other. Each well log from the dataset contains 14 different types of curves with a varying rate of missing values. The dataset contains more than 1 million of measurements that are distributed among a large number of groups, formations and lithofacies which enables objective evaluation of the correlation framework performance.

After data pre-processing and filling in the missing values, the following curves were selected as reliable features for the correlation: “CALI”, “DTC”, “GR”, “RDEP”, “RHOB”, “RMED”, “SP”.

4.2 Correlation performance comparison

In this paper the accuracy is measured as a ratio between the number of correctly correlated pairs of layers to the total number of correlated and uncorrelated layers, see Fig. 2. The accuracy of the framework is reported in table Table 1 and compared with the other method that uses the same metrics in the respective paper. This accuracy is reported together with the dataset it was originally measured on.

Table 1: Comparison of approaches

Approach	Accuracy, %	Dataset
Assistive Well-Log Correlation [10]	96.6	Private dataset
Our framework	98.2	FORCE dataset [9]

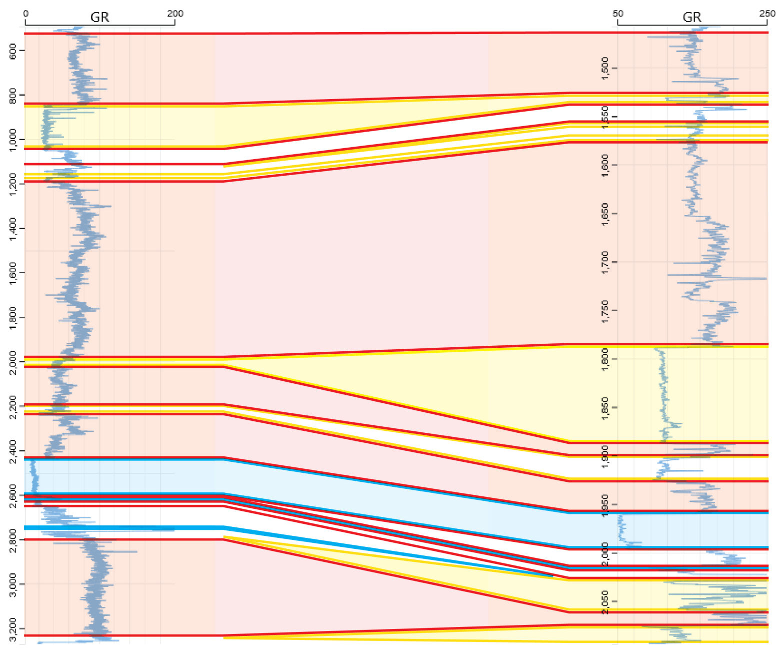

Figure 2: Sample of the pairwise well correlation

4.3 Feature importance

First, we assess the importance of depth feature in the overall performance of the framework. Correlation accuracy drops by more than 19% in case framework does not take depth feature as an input. Depth information induces errors in a couple of edge cases where the layers swap, however it significantly reduces the error for most common scenarios which results in overall accuracy improvements.

Second, we analyze the importance of curve selection for well correlation. Reducing the total number of feature curves from 8 to a minimum set of 2 results in correlation accuracy decrease by 27-61% depending on the curves that remain. The decrease of performance is due to the fact that not all curves have sharp changes on lithoface boundaries. Reducing the number of curves makes correlation theoretically impossible for a certain regions of the dataset. Gamma ray curve is the most important for the precise well correlation, however using more curves leads to even better correlation results.

Curves with a large amount of missing data will worsen the correlation performance because they contain too little new information. Data interpolation for highly incomplete curves will be a poor approximation of the real values. To tackle this issue, the proposed framework does not take into account curves with more than 40% of missing values.

5 CONCLUSIONS

This work introduces a novel framework for the automatic pairwise correlation of well logs. Unlike the others, the proposed approach is not binded to a limited set of curves, some of which may not be available for a given well. Our approach is simple to use, generalizes to different types of wells and does not require additional inputs from the user. Therefore, results of the framework do not depend on the knowledge and skills of the user. Experiments with the framework showcase that correlation results are robust to noise in the inputs and globally optimal.

References

[1] A.I. Kirichkova, V.A. Chizhova, E.K. Stashkova, N.K. Fortunatova, and Boris Shurygin. Stratigraphy in petroleum geology: research methodology and topical problems. *Neftegazovaya Geologiya, Teoriya i Praktika*, 2:1–32, 01 2007.

[2] Temple Smith and Michael Waterman. New stratigraphic correlation techniques. *Journal of Geology*, 88, 07 1980. doi: 10.1086/628528.

[3] C.John Mann and Thomas P.L. Dowell. Quantitative lithostratigraphic correlation of subsurface sequences. *Computers & Geosciences*, 4(3):295 – 306, 1978. ISSN 0098-3004. doi: https://doi.org/10.1016/0098-3004(78)90064-X. URL `http://www.sciencedirect.com/science/article/pii/009830047890064X`.

[4] DJ Lineman, JD Mendelson, M Nafi Toksoz, et al. Well to well log correlation using knowledge-based systems and dynamic depth warping. In *SPWLA 28th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 1987.

[5] Richard A Startzman, Tsai-Bao Kuo, et al. A rule-based system for well log correlation. *SPE Formation Evaluation*, 2(03):311–319, 1987.

[6] Stefan Luthi and Ian Bryant. Well-log correlation using a back-propagation neural network. *Mathematical Geology*, 29:413–425, 01 1997. doi: 10.1007/BF02769643.

[7] I Le Nir, N Van Gysel, D Rossi, et al. Cross-section construction from automated well log correlation: a dynamic programming approach using multiple well logs. In *SPWLA 39th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 1998.

[8] Vladimir Lapkovsky, A.V. Istomin, V.A. Kontorovich, and V.A. Berdov. Correlation of well logs as a multidimensional optimization problem. *Russian Geology and Geophysics*, 56, 03 2015. doi: 10.1016/j.rgg.2015.02.009.

[9] Website. Force: Machine predicted lithology challenge. Xeek, 2020. URL `https://xeek.ai/challenges/force-well-logs/data`.

[10] Seth Brazell, , Alex Bayeh, Michael Ashby, Darrin Burton, , and and. A machine-learning-based approach to assistive well-log correlation. *Petrophysics – The SPWLA Journal of Formation Evaluation and Reservoir Description*, 60(4):469–479, aug 2019. doi: 10.30632/pjv60n4-2019a1. URL `https://doi.org/10.30632%2Fpjv60n4-2019a1`.