Drilling Events Detection Using Hybrid Intelligent Segmentation Algorithm

Arghad Arnaout
TDE GmbH
Leoben, Austria
Arghad.Arnaout@tde.at

Bilal Esmael
University of Leoben
Leoben, Austria
Bilal@stud.unileoben.ac.at

Rudolf K. Fruhwirth
TDE GmbH
Leoben, Austria
Rudolf.Fruhwirth@tde.at

Gerhard Thonhauser
University of Leoben
Leoben, Austria
Gerhard.Thonhauser@unileoben.ac.at

Abstract— **Several sensor measurements are collected from drilling rig during oil well drilling process. These measurements carry information not only about the operational states of the drilling rig but also about all higher level operations and activities performed by drilling crew. Automatic detection and classification of such drilling operations and states is considered as a big challenge in drilling industry. Furthermore, the possibility of detecting such events opens the door to detect and analyze hidden lost time of the drilling process. This paper presents a novel algorithm for drilling time series segmentation using Expectation Maximization and Piecewise Linear Approximation algorithms. The suggested algorithm shows that the incorporation of prior-knowledge about the drilling process is a key step to segment drilling time series successfully. The Expectation Maximization algorithm is used to segment drilling time series based on hook-load sensor measurements. In addition, Piecewise Linear Approximation is hired in our approach to slice standpipe pressure, pump flow rate and rotational speed (RPM) and torque of the top drive motor. Merging the results from both, Expectation Maximization and Piecewise Linear Approximation, gives the suggested algorithm the dynamic ability to detect all drilling events and activities.**

Keywords: Drilling Events Detection, Expectation Maximization, Timeseries Segmentation, Piecewise Linear Approximation

I. INTRODUCTION

Automatic detection of drilling events and operations is considered as an urgent need in the drilling industry. Detecting these events gives services of drilling data analysis more aptitude to examine all actions which are done by the drilling crew at the rig site. Furthermore, automatic detection also provides essential mechanisms to judge the performance of the drilling machinery. Moreover, this leads to the possibility to perform sequence mining and analysis on particular drilling process sections.

Usually sensors measurements are collected during the whole drilling process. Such measurements are used by drilling engineers and drilling crews to monitor the drilling process by action/response models. For example, any change in the hydraulic flow rate parameter causes a response in the pump pressure. Likewise, torque measurements are observed through altering the rotational speed of the drill string [2].

The cyclic nature of drilling processes exposes specific patterns for each drilling activity or event in sensor measurements. Furthermore, each sensor data time series has specific statistical distribution. These distributions look very similar in almost all similar-type drilling rigs (offshore or land rigs). This gives the drilling process a similarity property. Here we can find a big possibility to generalize our findings and analyses [1] [2].

Expectation Maximization (EM) is a powerful tool to estimate the parameters of Gaussian distributions in the data. EM has ability to discriminate data into clusters if this data have the nature of mixture models. EM provides the possibility to find and describe main clusters in the data by estimating description parameters of each cluster. Segmenting of data based on a cluster will be a minor task if the parameters are estimated [3]. The Expectation Maximization algorithm is considered with stable performance in data with less amount of noise [4].

Piecewise Linear Approximation (PLA) is another useful tool for time series segmentation. Usually PLA is used to approximate main sections in time series. PLA has no tolerance to data with low of signal to noise ratio, but it can be applied to data with a specific S/N ratio [5].

II. CONTRIBUTION

The contribution of this paper can be outlined as follows:

1. Automatic detection of different drilling events and operations.
2. Incorporation of prior-knowledge on drilling process is a main factor in hybrid intelligent algorithm.
3. Hiring Expectation Maximization algorithm as core algorithm for high-level segmentation.
4. Piecewise Linear Approximation algorithm is applied as low-level time series segmentation.
5. Combination of two algorithms (EM and PLA) to accomplish multi-level drilling time series segmentation.

III. DRILLING PROCESS AND MUD-LOGGING SYSTEMS

Oil well drilling is a process of making a hole in the ground in order to extract oil, gas or any other natural resources from the subsurface; usually performed by a rig. One of the most important parts of such a drilling rig is the drill-string. A drill-string is a chain of connected pipes usually having a length of 10 meters each. The bottom end of drill-string is made of special devices, denoted as bottom hole assembly (BHA). The last part of the BHA is drill-bit [6].

Numerous sensors are mounted at the rig to record different physical measurements during drilling such as block position, hook-load, flow rates, pump and circulation pressures, hole & bit depth and torque, among others [2].

Figure 1 shows a sketch of such sensor data over a period of 20 hours, recorded with a resolution of 0.2 Hz.

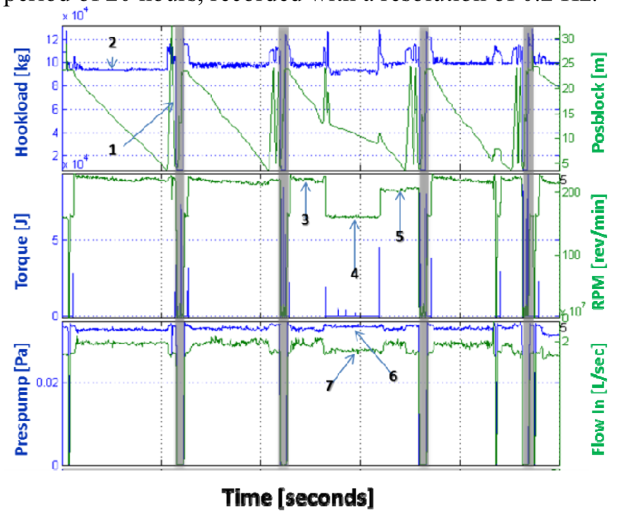

Figure 1: Sketch of drilling time series (20 Hours, 0.2 Hz)

The gray highlighted areas ”1” in Figure 1 refer to a special state in drilling process; drill-string is hanging in the rig floor fixed by slips, thus such a state is denoted as *InSlips*. The non-highlighted areas “2” refer to converse situations denoted as *OutOfSlips*; this means that drill-string is hanging at hook of rig and therefore applies force to the hook-load sensor [6]. Such a hook-load sensor usually measures the weight of the drill-string together with weight of the hook; therefore the hook-load is not zero at *InSlips* state. Two different patterns are formed by hook-load measurements during *InSlips* & *OutOfSlips* states [7]. At *InSlips* state, hook-load is low, the measured value indicates the weight of the hook only. At *OutOfSlips* state the hook-load is higher, the weight of the hook plus the weight of the drill-string hanging at hook is measured.

The separation of *InSlips* from *OutOfSlips* states is one of main steps of an automated drilling operations classification system [8]. Usually, drilling experts set a threshold value manually for the hook-load to separate this states.

Also the situations and states which are tagged in Figure-1 by “3”, “4”, “5”, “6” and “7” are considered as usual and unusual events and states in the time series. “3”, “4” and “5” represent different levels of RPM. The tag ”7” refers to a specific level of the pump flow rate *flowIn*. Tag “6” points to a standpipe pressure as response of the *flowIn* level in tag “6”. From drilling expert viewpoint, no clear reason explains why this level “6” in *flowIn* time series happened.

IV. Drilling Time Series Segmentation Using Expectaion Maximization

In our approach we used the dataset shown in Figure 1 incorporating the knowledge of drilling experts. The approach is mainly based on hook-load sensor measurements.

Figure 2 shows the histogram of the hook-load data over a period of 10 days of drilling. Applying drilling experts’ know-how, knowledge can be explained on this histogram. It shows that hook-load data has the nature of Gaussian Mixture Models, i.e. the data are composed at least out of 4 Gaussian distributions. Each of these distributions reflects information about a specific state of the rig. Obviously, two main distributions are located in the data i.e. *InSlips* and *OutOfSlips*. The left most distribution certainly defines the *InSlips* state.

The statistical parameters of each distribution provide information about the hook-load data for each state. The estimation of the threshold that discriminates *InSlips* states from other states is a significant step in segmentation. The segmentation, which separates *InSlips* state from *OutOfSlips* state, is a high-level segmentation. The advantage of EM is that it can be applied to big data set with acceptable performance.

Arnaout et al. [3] discussed in detail the use of hook-load data to determine automatically the threshold value for separation of *InSlips* and *OutOfSlips* states.

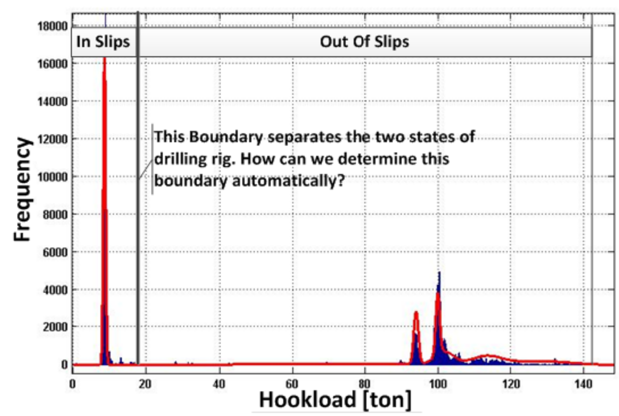

Figure 2: Histogram of hook-load data with indicator to location of threshold between InSlips/OutOfSlips states.

After estimating the parameters of the particular cluster in the hook-load data, the second step is the calculation of the intersection point, which is the threshold, used for separation of *InSlips* and *OutOfSlips* states.

The algorithm below shows how to calculate the intersection point between two clusters based on Bayes' theorem, using the clusters' statistical parameters, mean value and standard deviation.

Intersection Point of two Clusters

Input:

Two univariate clusters C_1 and C_2 assumed to be Gaussian distributed with $\Theta_1=\{\mu_1,\sigma_1\}$ and $\Theta_2=\{\mu_2,\sigma_2\}$.

Output:

The separation threshold x_i of the two clusters.

Do

The probability density $p(x|C_k)$ for the k^{th} cluster of a Gaussian Mixture Model is given by

$$p(x|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \quad (1)$$

According to Bayes' theorem, the separation threshold x_i is located where the posterior probabilities $P(C_k|x)$ of both clusters are identical. Using

$$P(C_1|\,x) = \frac{p(x\,|C_1)\,P(C_1)}{p(x\,|C_1)P(C_1)+p(x\,|C_2)P(C_2)} \quad (2)$$

$$P(C_2|\,x) = \frac{p(x\,|C_2)\,P(C_2)}{p(x\,|C_1)P(C_1)+p(x\,|C_2)P(C_2)} \quad (3)$$

and the prior probabilities $P(C_1)$ and $P(C_2)$ given by

$$P(C_k) = \frac{\text{number of points belonging to cluster } C_k}{\text{total number of points}} \quad (4)$$

the separation threshold x_i can be estimated by solving the equation

$$P(C_1|\,x_i) = P(C_2|\,x_i) \quad (5)$$

End

V. Drilling Time Series Segmentation Using Piecewise Linear Approximation.

The low-level segmentation is applied to each segment obtained from high-level segmentation as discussed above. In our approach Piecewise Linear Approximation is applied to the hydraulic flow rate, denoted as *flowIn*, and the rotational speed of the drill string, denoted as *RPM***.**

The algorithm *Bottom to Up* **[5]** forms the base for the segmentation by Piecewise Linear Approximation based on customized error cost function. The algorithm begins by creating the finest possible approximation of the time series consisting of n samples by using initially n/2 segments. In a subsequent step, the costs of merging pairs of adjacent segments are calculated. The algorithm iteratively merges the pairs with the lowest costs until a stopping criterion is met. Merging pairs of adjacent segments, i and i+1, bookkeeping about the neighborhood merging costs is inevitable. The cost of merging the actual segment with both, right and left neighbors, must be calculated **[5]**.

Figure 3 shows detailed segment of *OutOfSlips* state for sensors: hook-load, flow rate (*flowIn*), and Rotation Speed (*RPM*). In assistance of drilling experts, it is required that each change or event in those sensor measurements should be detected. Applying Piecewise Linear Approximation on each of those time series gives the possibility to detect main and minor changes in these time series. The accuracy of detection depends primarily on customized error cost function of PLA.

VI. The Segmentation Algorithm

In this paragraph, the algorithm of applying EM and PLA on drilling time series is presented. The data showed in Figure 1 is used as sample data.

Segmentation Algorithm

Input

Measurements of sensors as raw data

Output

Segments of time series

Do

1. Estimate of clusters parameters in hookload data using Expectation Maximization algorithm.
2. Calculate the intersection point, which is the threshold as specified earlier in Intersection Point Calculation Algorithm.
3. Create high-level segments (InSlips/OutOfSlips) based on the intersection point (threshold) and hook-load sensor data.
4. Use PLA as specified in previous paragraph to slice each segment from previous step into smaller segments.
5. Merge segments from previous two steps as resulting segments.

End

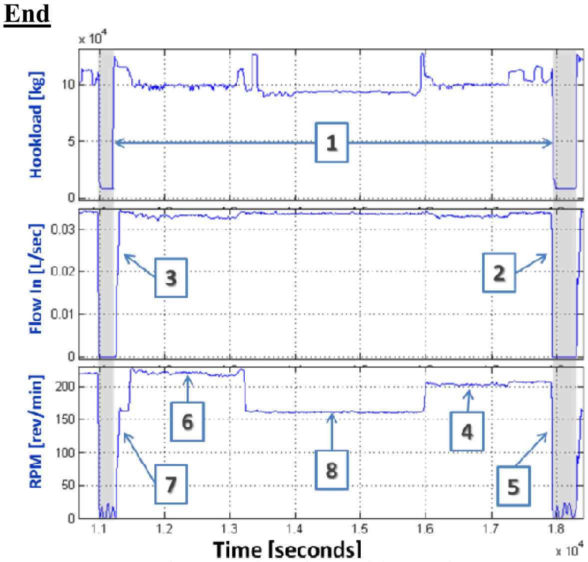

Figure 3: View of *OutOfSlips* section

Figure 3 illustrates detailed view of OutOfSlips section "1". "2" and "3" point to the pump's startup/shutdown procedures."6" represents high level of rotational speed. "7" shows startup procedure of the rotary system over two phases. "8" shows lower level of rotational speed. "4" shows middle level of rotational speed. "5" indicates to procedure of shutting down the rotary system.

VII. Results and Discussion

Figure 4 illustrates the results of high-level segmentation using hook-load sensor measurements as well as low-level segmentation using other sensor data.

The results confirm a high level of accuracy at high-level segmentation. This because the stability of the Expectation Maximization algorithm with noisy data. In addition, prior-knowledge about the fact of locating InSlips distribution as the left most distribution gives more strength to the suggested segmenting algorithm.

The accuracy of low-level segmentation shows sensitivity to the value of predefined error parameter of PLA algorithm. Normalizing the data helps in converging values of error parameters for each sensor data. In most cases, we use same value for all error cost function of PLA.

VIII. Future Work

The suggested algorithm in this paper opens doors to do further analysis and recognition on each detect segment in drilling time series. In addition, more work should be done to reduce/filter the noise in time series before processing. The noise-filtering step will certainly improve the accuracy of segmentation algorithm.

References

[1] Parigot, P., Marine, T., Als, N., Havrevold, L., Offshore, N., & Als, S. (1988). SPE 5789 Intelligent Real-Time Status Analysis for Rig-Site Drilling Engineering, 393–401.

[2] Florence, F., Varco, N. O., Iversen, F., & Systems, D. R. (2010). IADC / SPE 128958 Real-Time Models for Drilling Process Automation: Equations and Applications.

[3] Arnaout, A., Esmael, B., & Fruhwirth, R., Thonhauser, G., Automatic Threshold Tracking of Sensor Data Using Expectation Maximization Algorithm.

[4] A.P. Dempster. (2007). Maximum Likelihood from Incomplete Data via he EM Algorithm, *39*(1), 1–38.

[5] Keogh, E., Chu, S., Hart, D., & Pazzani, M. An online algorithm for segmenting time series. Proceedings 2001 IEEE International Conference on Data Mining, 289–296. doi:10.1109/ICDM.2001.989531

[6] H. Rabia, "Oil Well Drilling Engineering, Principles and Practices", University of Newcastle, 1985

[7] China National Logging Cooperation, "Mud Logging Technology and Services", presentation, 2006

[8] G. Thonhauser, "Using Real-Time Data for Automated Drilling Performance Analysis", OIL GAS European Magazine, 2004

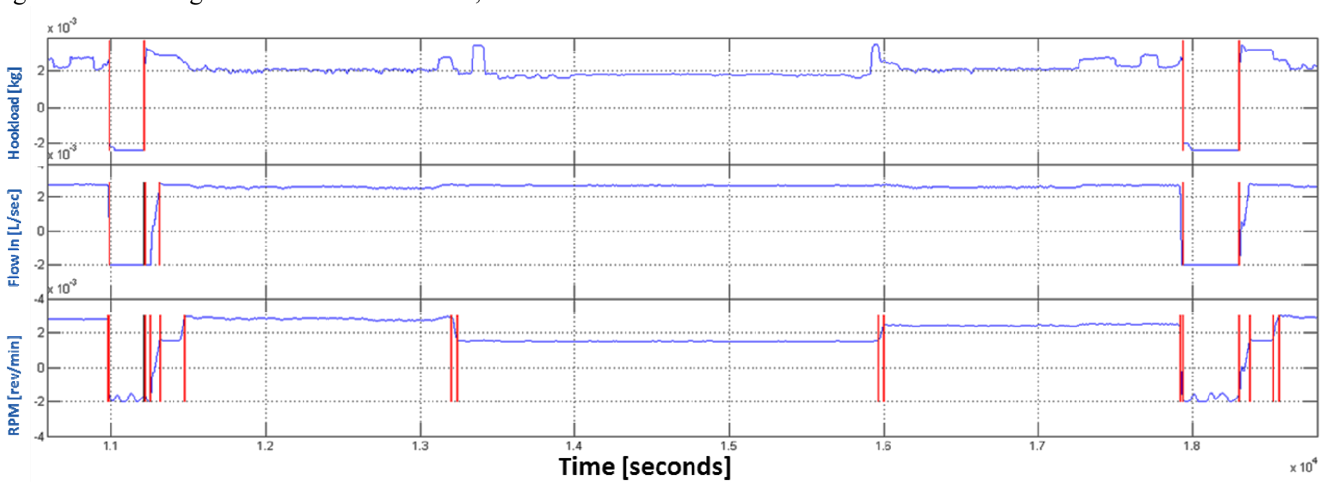

Figure 4: Results of the suggested segmentation algorithm on drilling time series after normalization (Automatic-detected sections in red lines).