International Journal of Hybrid Intelligent Systems 11 (2014) 25–39
DOI 10.3233/HIS-130181
IOS Press

Distributed recognition system for drilling events detection and classification

Arghad Arnaout[a,*], Paul O'Leary[a], Bilal Esmael[b] and Gerhard Thonhauser[a]
[a]*Institute for Automation, University of Loeben, Leoben, Austria*
[b]*Chair of Drilling, University of Leoben, Leoben, Austria*

Abstract. Several sensor measurements collected from drilling rig during oil well drilling process. These measurements carry information not only about operational states of drilling rig but also about all high-level operations and activities performed by drilling crew. The work presented in this paper shed the light on analysis of hidden lost time in drilling process through automatic detection and classification of drilling operations.
This paper develops a novel algorithm for detecting drilling events and operations in sensor data of drilling rig. Expectation Maximization EM and Piecewise Linear Approximation PLA algorithms applied for detecting drilling events. The Expectation Maximization algorithm performs high-level segmentation on hook-load sensor data. In addition, Piecewise Linear Approximation algorithm slices standpipe pressure; pump flow rate; rotational speed and torque of top drive motor into labeled segments (low-level segmentation). Merging results from both Expectation Maximization and Piecewise Linear Approximation gives the suggested algorithm ability to detect all drilling events and activities performed by drilling rig and crew. Moreover, this paper shows the usage of discrete orthonormal basis functions (Gram basis) as a tool to classify drilling operations from detected segments in drilling time series. The classification process performed in cooperation with the concept of Patterns Templates Base. The optimal polynomial degree to represent drilling operations has been concluded through analysis of polynomial spectrum of each drilling operation.

Keywords: Drilling operations, gram polynomials, Expectation Maximization, Piecewise Linear Approximation, orthonormal basis functions

1. Introduction

Improving performance of drilling process is a big challenge in nowadays drilling industry. To improve drilling performance, we need first to measure it [8]. Performance measurement means determining quantitative values or weights that describe each drilling operation and complete drilling process as resultant. For example, duration of each drilling operation considered as a useful measure. In addition, number of drilling operations and distributions of those operations over different well drilling phases are important measures of drilling performance.

Automatic detection of drilling events and operations considered as an urgent need in the drilling industry. Detecting these events gives services of drilling data analysis more aptitude to examine all actions performed by the drilling crew at the rig site [8]. Furthermore, automatic detection also provides essential mechanisms to judge the performance of drilling machinery. Moreover, this leads to the possibility to perform sequence mining and analysis on particular drilling process sections.

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

*Corresponding author: Arghad Arnaout, Institute for Automation, University of Loeben, Leoben, Austria. E-mail: Arghad@stud.unileoben.ac.at.

Through drilling process, a huge amount of data in form of sensors measurements produced over time. This data contains not only readings of sensors but also information about each drilling operation i.e. start, end, and behavior of each equipment. Drilling operations such as formations drilling, making connection for new drillstand, breaking connection, pulling out of hole, running in hole, and cleaning hole are carefully chosen as basic drilling operations performed by drilling crew [6]. Each of those drilling operation has a specific pattern in rig sensors measurements [1].

The cyclic nature of drilling processes exposes specific patterns for each drilling activity or event in sensor measurements. Furthermore, each time series of sensor data 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 [1,2]. Here we can find a big chance to generalize our findings and analyses.

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 Gaussian 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 considered with stable performance in data with less amount of noise [4]. Piecewise Linear Approximation (PLA) is another important tool for time series segmentation. Usually PLA used to approximate main sections in time series. PLA has no tolerance to data with low value of signal to noise ratio, but it applied to data with a limited S/N ratio [5].

Patterns recognition using discrete polynomial moments as features to classify the data is a nowadays trend in patterns recognition domain. Most of the studies in this area perform detection of patterns after calculating polynomial moments from data. The calculated coefficients called features. These features considered as descriptors of pre-defined templates in the data. Those templates extracted during learning phase. Then the templates and calculated features used to classify unknown data and assign classes to it [9].

Detecting drilling operations patterns in sensors data supports rig operators in finding out the state of drilling rig instantly. At the end of the day, it gives detailed information on rig state over any span of time [8]. Therefore, rig's operator can easily observe operating time of drilling rig and how actual drilling performance matches with pre-defined well plan.

2. Contribution

The contribution of this paper outlined as follows:

1. Automatic detection of different drilling events and operations;
2. Show how prior-knowledge on drilling process affects all suggested algorithms;
3. Hiring Expectation Maximization algorithm as core algorithm for high-level segmentation of hookload sensor data;
4. Piecewise Linear Approximation algorithm applied as low-level segmentation of Block Position sensor data;
5. Combine two algorithms (EM and PLA) to accomplish multi-level drilling time series segmentation;
6. Using discrete polynomial moments (Gram basis functions) as descriptors of drilling operations templates;
7. Suggest weighted similarity distance based on Mahalanobis distance and sensors importance matrix;
8. Combining phases of segmentation and classification for drilling operations using distributed component architecture.

3. Drilling process and Rig's sensor measuremtns

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 and 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.

The gray highlighted areas "1" in Fig. 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

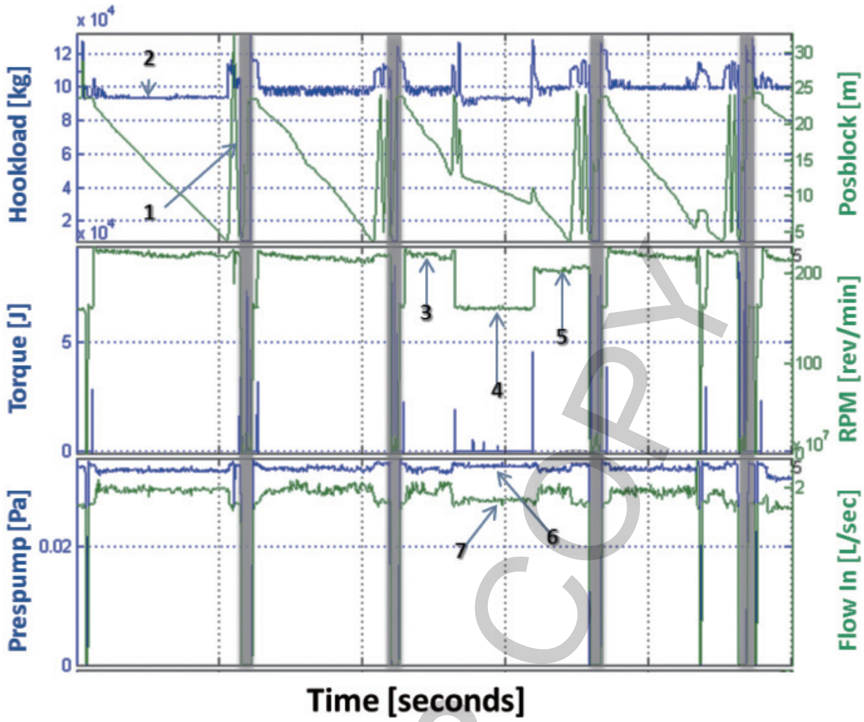

Fig. 1. Sketch of drilling time series (20 Hours, 0.2 Hz). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

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 formed by hook-load measurements during *InSlips* and *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.

In addition, the situations and states, which tagged in Fig. 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.

4. Rig’s sensors systems

Drilling rig performs its functionality in drilling boreholes through collaboration of three main subsystems: Rotary System, Circulation System, and Hoisting System.

Rotary System is the system that turns the drill-string. Top drive as type of rotary system which consists of one or more motors (electric or hydraulic) connected with appropriate gearing to a short section of pipe called a quill, that in turn may be screwed into a saver sub or the drillstring itself. In addition, rotary table another type of rotary system and it consists of revolving or spinning section of the drillfloor that provides power to turn the drillstring in a clockwise direction (as viewed from above). The rotary motion and power are transmitted through the kelly bushing and the kelly to the drillstring. RPM and torque sensors measure the revolution per min and torque of rotation at the surface [2].

Circulation system is defined as the complete, circuitous path that the drilling fluid travels. Starting at the main rig pumps, major components include surface piping, the standpipe, the kelly hose (rotary), the kelly, the drillpipe, drill collars, bit nozzles, the vari-

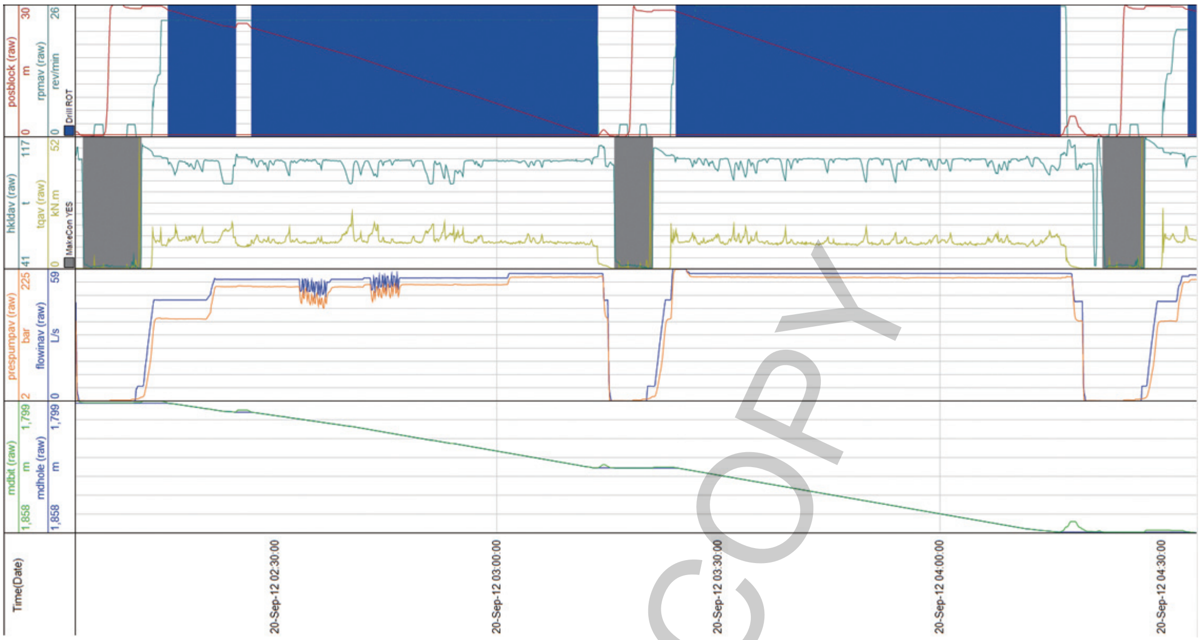

Fig. 2. Drilling operations highlighted on drilling sensors data (Blue color: drilling operation and making hole. Gray color: making connection). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

ous annular geometries of the opened-hole and casing strings, the bell nipple, the flowline, the mud-cleaning equipment, the mud tanks, the centrifugal precharge pumps and, finally, the positive displacement main rig pumps [6]. The first mission of circulation system is keeping opened-hole section in wellbore stable by creating pressure equilibrium on walls of wellbore. The second purpose of circulation systems is cleaning wellbore by removing cuttings and lifting them to surface. The major sensors that are mounted to this system are: FlowIn, FlowOut and Pumps Pressure. FlowIn/Out sensors measure quantity of flow in/out of the mud pump. Pumps Pressure sensor measures the pressure at the standpipe [2].

The main function of hoisting system is getting drilling string or another necessary equipment in/out of borehole safely and efficiently [6]. The main components of hoisting system are: Draw-works, Hoisting tackle including crown and travelling block, hooks and elevators, Deadline anchors, Drilling line and Derrick. The basic sensors measurements related to this system are: Hookload, Position of block. Hookload sensor reads how much weight and load hanged by the hook. While posblock sensor measures the distance between travelling block and floor level of rig [2].

Other surface values provided as sensors readings. In reality, these values not measured but calculated. WeightOnBit calculated by subtracting the string weight from the value of hookload. Rate of penetration calculated as the speed of drilling string movement during drilling operation. Hole depth is estimated through the length of drillstring and the distance between the surface level and the maximum value reached down by drillstring. Bit Depth is the length of drillstring when the drillstring hanged at hook and not stuck in slips at rig floor.

5. Drilling operations

The drilling process is a process of making a hole in the ground until it reaches a pre-specified depth as definite in the well plan. The drilling crew sets up a rig and uses it to generate the hole. Normally the drilling process consists of a sequence of operations. Before start drilling operations, some important main procedures performed and they summarized as following:

1. Setup Bottom Hole Assembly, which is the lower part of drillstring. It consists of drillbit, mud motor, collars, heavy-weight drillpipe, etc.;
2. Attach kelly and rotary table to the drill string;
3. Setup the pumps as part of circulation system to pump the mud through pipes and out of drillbit to carry the cuttings out of hole.

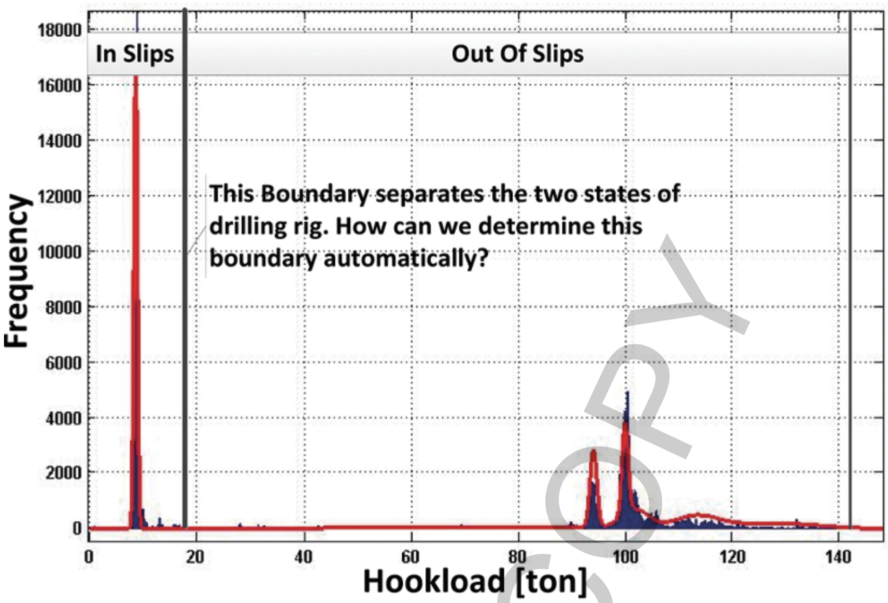

Fig. 3. Histogram of hook-load data with indicator to location of threshold between InSlips/OutOfSlips states. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

After finishing previous phases, the drillstring is ready to make a hole in the ground. Then a sequence of drilling operations starts executed in a recurring manner. These operations defined as follows:

1. Drilling formation operation is the operation of making hole in the ground through applying weight or load on drillbit. Hole generating is mainly achieved through: applying weight on drill bit or BHA; starting rotation of drillstring; and circulating the cuttings. The mud is pumped in drilling pipes through drillbit and then to the annulus (the space between drillstring and walls of wellbore) to the surface. At the surface, all cuttings removed from the mud through specific device called shale shaker.
2. Making a connection is an operation of adding new drillstand to the drillstring to increase string's length and then the hole drilled deeper.
3. Running in hole, in this operation, the drillstring with new BHA is tripping into the hole till the bit touches Hole's bottom.
4. Pulling out of hole is the operation of tripping drillstring out of hole in purpose of changing drillbit and making new Bottom Hole Assembly.
5. Breaking connection is the operation where the drilling crew disconnects drillstand from drillstring to decrease its length and pull BHA out of hole.
6. Cleaning Hole (Circulation): In this operation, the drilling crew pumps the mud continuously to give a chance for the circulated mud to curry the cuttings up to the surface and cleaning the hole from drilling cuttings.

Once the drilling crew reaches the pre-set depth, they must run and cement the casing – place casing-pipe sections into the hole to prevent it from collapsing in on itself. The casing pipe has spacers around the outside to keep it centered in the hole. The casing crew puts the casing pipe in the hole. The cement crew pumps cement down the casing pipe using a bottom plug, a cement slurry, a top plug and drill mud. The pressure from the drill mud causes the cement slurry to move through the casing and fill the space between the outside of the casing and the hole. Finally, the cement allowed to harden and then tested for such properties as hardness, alignment and a proper seal.

In this paper, we try to model and recognize the drilling operations automatically by studying and modeling trends of sensors data (Fig. 2). The drilling operations analyzed to describe each operation using polynomial spectrum of sensors data. The basic operations that we automatically recognize through this paper are Formation Drilling Operations (Making Hole), Making Connections, Breaking Connections, Circulations, Running in Hole, and Pulling out of Hole.

6. High level drilling time series segmentation

In our approach, we used the dataset shown in Fig. 1 incorporating the knowledge of drilling experts. This approach based on hook-load sensor measurements.

Figure 3 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.

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 [14].

Using

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

and

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

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 estimated by solving the equation

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

End

7. Low level drilling time series segmentation

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

The algorithm *BottomToUp* [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 4 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.

7.1. Drilling timeseries segmentation algorithm

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

Figure 4 illustrates detailed view of OutOfSlips section "1". "2" and "3" point to the pump's startup/shut-

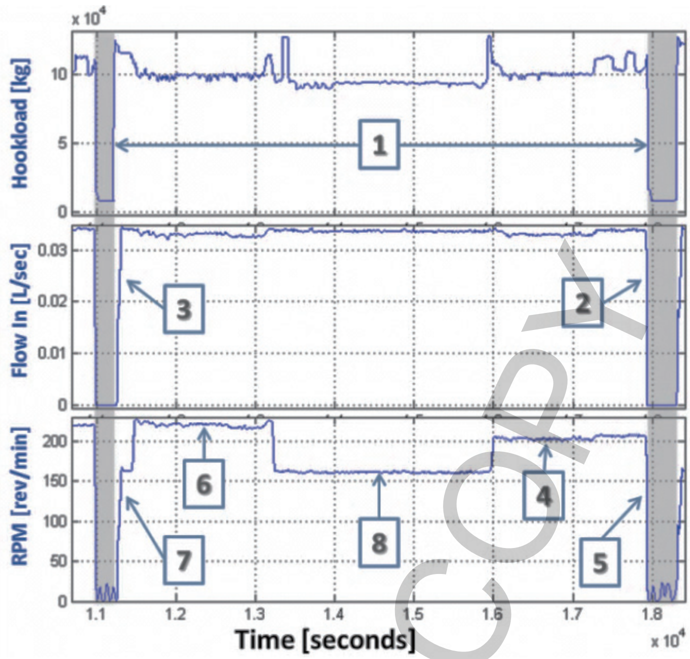

Fig. 4. View of *OutOfSlips* section. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

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

down 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.

8. Results of segmentation algorithm

Figure 5 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.

9. Patterns recognition using basis functions method

In this paragraph, we introduce simple derivation of polynomial discrete moments and how they extracted from drilling sensors data. The main idea is to extract the pseudo inverse B^+ of design matrix B from drilling sensors measurements, then use B^+ to compute the coefficients vector z, which describes the polynomial that fits the measurements of sensors data [11]. The coefficients vector z is the pattern descriptor in our case. For each drilling operation, we derive $z_1 \ldots z_k$ vectors where *k* is the number of sensors that we have.

Consider that we have a drilling dataset D of n data points y_i, where a Gaussian Noise perturbs each data

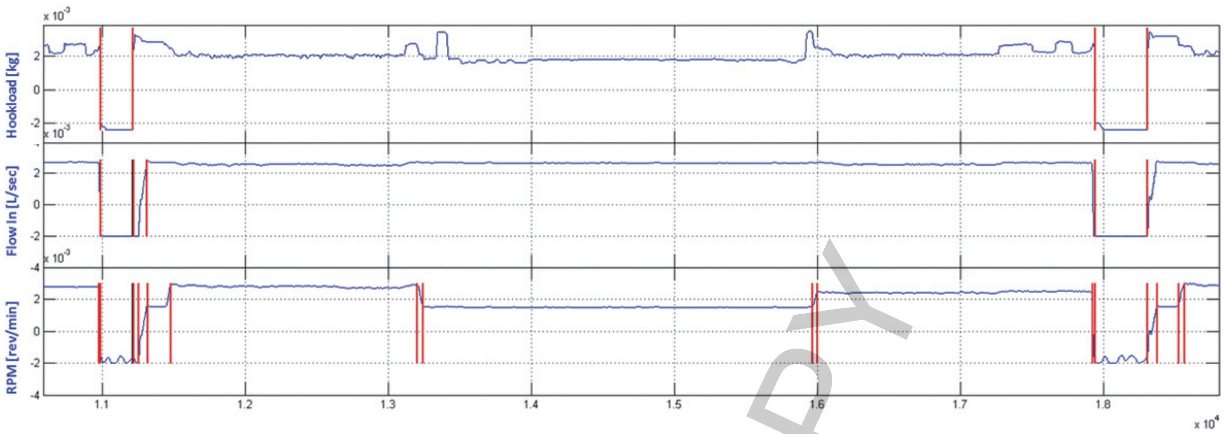

Fig. 5. Results of the suggested segmentation algorithm on drilling time series (Automatic-detected sections in red lines). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

point. The objective of polynomial regression of degree d is to model the data points $\overline{y}_i$ by a sum of monomials, i.e.

$$\overline{y} = a_0 + a_1 x + a_2 x^2 \ldots + a_d x^d = \overset{i=0}{d}\; a_i x^i. \quad (6)$$

Since the data points y_i are perturbed, they in general do not lie exactly on the polynomial. Consequently, there is a residual r_i associated with each point,

$$r_i = y_i - \overline{y}_i = y_i - \overset{i=0}{d}\; a_i x^i. \quad (7)$$

We can rewrite the Eq. (7) in matrix form yields,

$$\begin{bmatrix} r_i \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots x_1^d \\ \vdots & & & \\ 1 & x_n & x_n^2 & \ldots x_n^d \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_d \end{bmatrix} \quad (8)$$

In general, measurements data; the design matrix B and the coefficient vector z are defined as,

$$B = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots x_1^d \\ \vdots & & & \\ 1 & x_n & x_n^2 & \ldots x_n^d \end{bmatrix} \text{ and } z = \begin{bmatrix} a_0 \\ \vdots \\ a_d \end{bmatrix}. \quad (9)$$

Consequently,

$$\overline{y} = Bz. \quad (10)$$

So the error rewritten as,

$$E = \sum_{i=1}^{n} r_i^2 = r^T r. \quad (11)$$

Then, the coefficient vector calculated using,

$$z = \left(B^T B\right)^{-1} B^T y = B^+ y, \quad (12)$$

where B^+ is the pseudo-inverse of B [11].

The column vectors forming the matrix B are the basis functions. The matrix B called Vandermonde matrix. The fact is that the polynomials or the basis functions of Vandermonde matrix are not orthogonal. This causes complexity in computation and calculation of high degrees of polynomials. Moreover, this means that the Vandermode basis functions are not suitable for solving large-scale problems. This fact leads us to choose another set of basis functions that are orthogonal and it is possible to calculate higher degrees of such polynomials. Here are examples on orthogonal polynomials from functional analysis domain: Legendre, Chebyshev, Gram, Gegenbauer, etc.

In our study on drilling sensors data, we will use Gram polynomials to describe drilling time series of different drilling operations. We choose Gram polynomials because not only they are orthogonal but also they have a uniform scaling and this is important to fit some complex shapes in drilling time series with higher performance than what Vandermonde basis can do.

The equation of generating Gram polynomials given by [10]:

$$g_n(x) = 2\alpha_{n-1} x g_{n-1}(x) - \frac{\alpha_{n-1}}{\alpha_{n-2}} g_{n-2}(x). \quad (13)$$

Whereby,

$$\alpha_{n-1} = \frac{m}{n}\left(\frac{n^2 - 1/2}{m^2 - n^2}\right)^{1/2}, \quad (14)$$

and

$$g_0(x) = 1, g_{-1}(x) = 0 \text{ and } \alpha_{-1} = 1. \quad (15)$$

Figure 6 shows us how Gram basis functions look like.

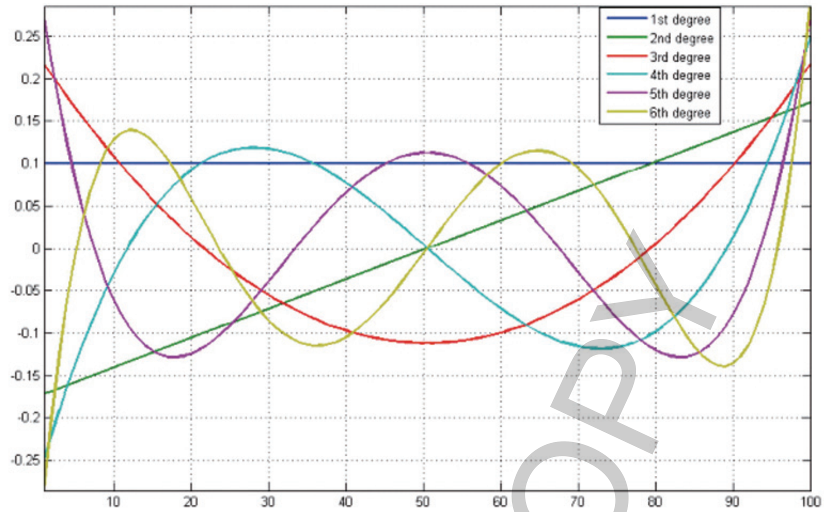

Fig. 6. The first six degrees of Gram basis functions. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

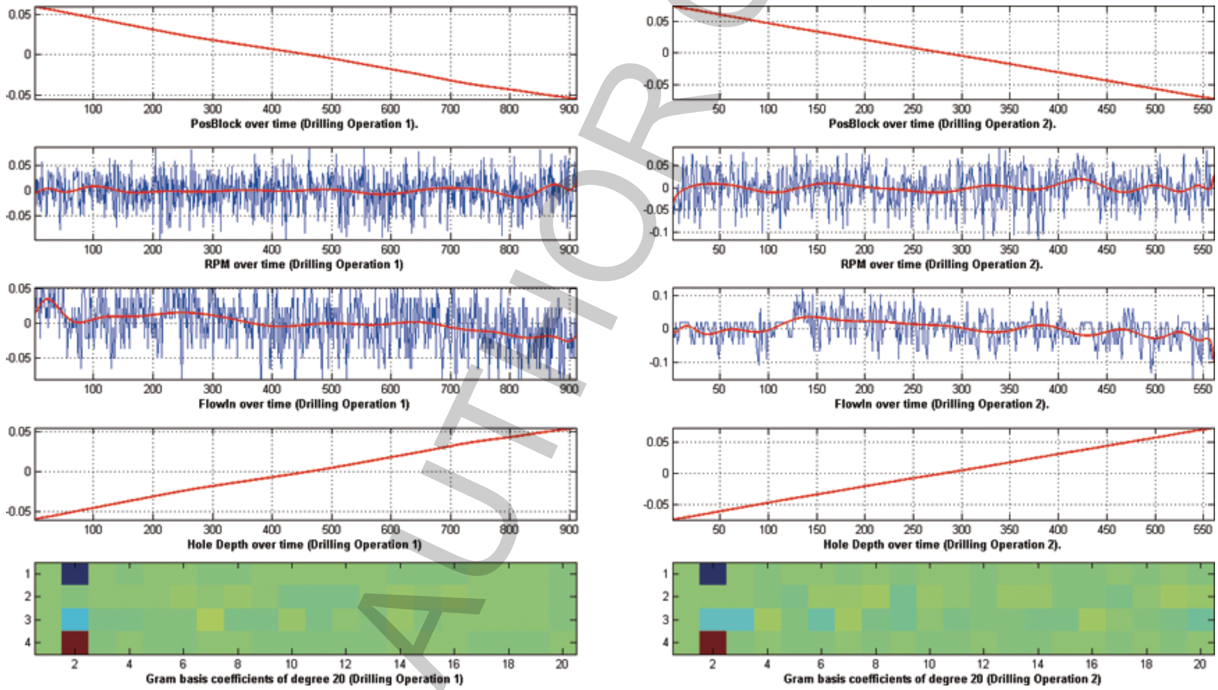

Fig. 7. Two drilling formation operations and their polynomial representations (Patterns). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

10. Analysis of drilling operations using orthonormal basis functions

Figure 7 shows two different formation drilling operations (left and right). The sensors measurements of block position, RPM, flowIn, and hole depth are shown. All the sensors measurements normalized to keep the values of coefficients close to each other and make it comparable. The red lines on upper four sub-plots in Fig. 7 represent the fitted Gram polynomials of each sensor data. While the lowest sub-plot in Fig. 7 shows the values of coefficients vectors z_1, z_2, z_3, and

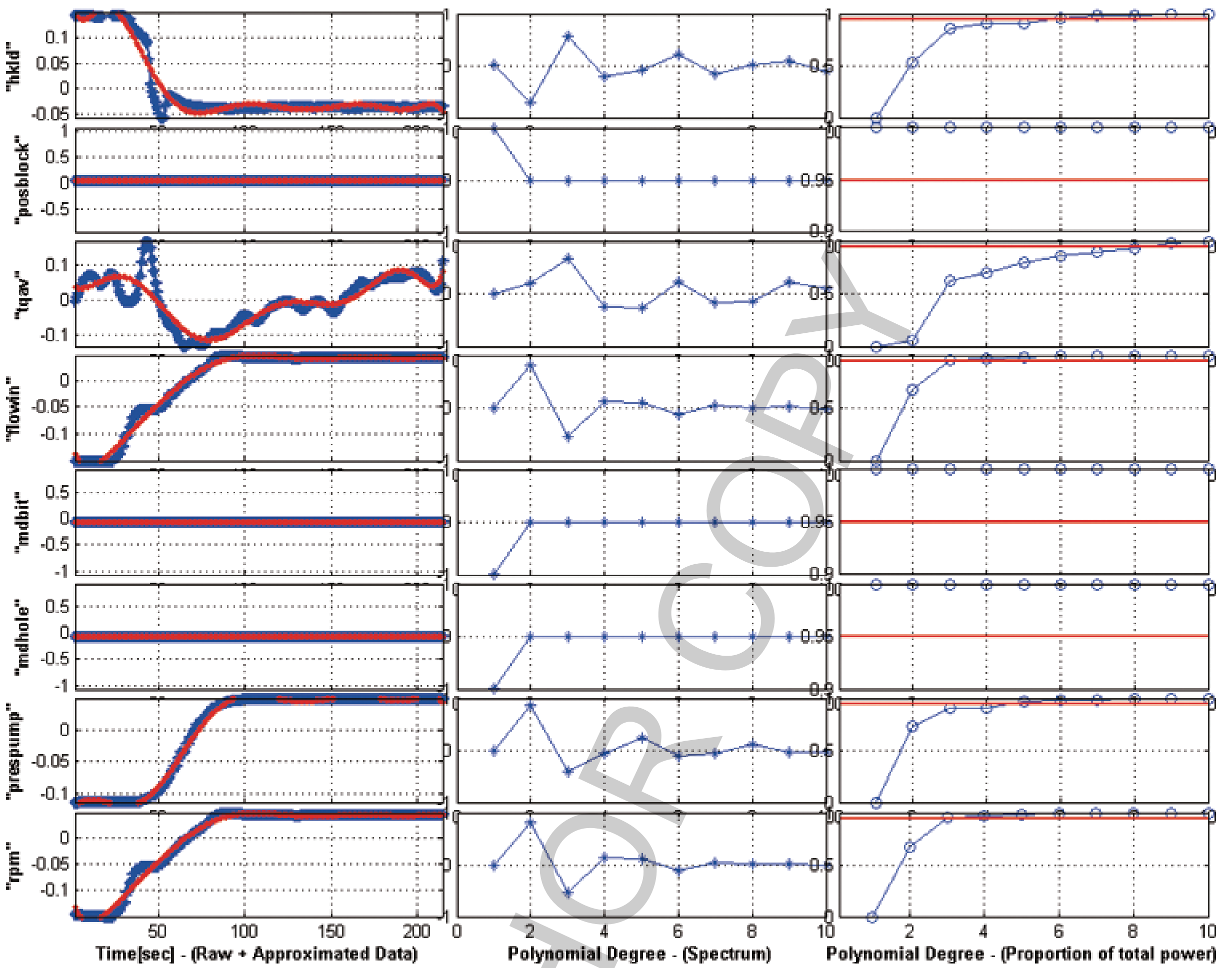

Fig. 8. Applying trend analysis using Gram polynomials on drilling sensors data during cleaning hole (Circulation) operation. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

z_4 as map of colors. We call the coefficient matrix at the lowest sub-plot as pattern descriptor of drilling operation. Gram basis functions of high order used in calculations i.e. each of sensors data in each formation drilling operation in Fig. 7 represented using Gram polynomial of degree 20.

The lowest part of Fig. 7 shows that the two formation drilling operations look similar and their patterns are close to each other. This will be a key indicator in drilling operations recognition process. Furthermore, we notice that the block position data represented as a line and this already shown in the coefficients matrix (lowest sub-plot in Fig. 7, first row in color map). Where from 20 degrees polynomial only the second component has distinct value (blue color). The second value in Gram polynomial spectrum represents line component (see Fig. 6, second degree). The RPM data in drilling operation shows that this data is not exactly at one level, and this is opposite to what was believed from drilling domain where the driller sets and fixes the RPM to a specific value while formation drilling operation. The raw data shows that the values of RPM and flowIn are fluctuated in small range. RPM fluctuated in range [139-142] rev/min, and flowIn fluctuated between 1983 and 1985 L/min at depth of 3070 m. these small fluctuations will be the key indicator to recognize the trends of RPM and flowIn through formation drilling from other states where the pumps are off and the drillstring is not rotating.

Figure 8 presents applying of trend analysis using Gram polynomials on drilling sensors data during cleaning hole (Circulation) operation. The polynomial spectrum shows us how the coefficients fluctuated where each change in values of coefficients gives us information about the importance of the polynomial component in representing and reconstructing the cor-

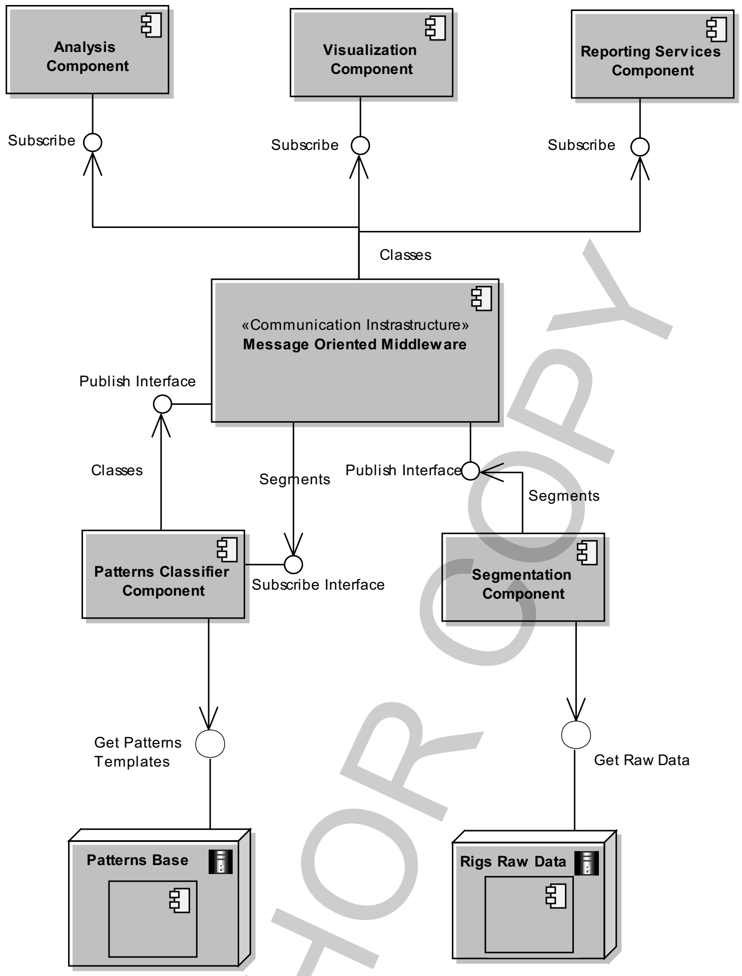

Fig. 9. UML Components diagram of distributed system architecture for drilling operations detection and classification.

responding sensors data. For example, the spectrum of hookload sensor data indicates that the first component in polynomial spectrum carries no information because it has a value of zero, but the components from 2 to 7 are representing most of the information in hookload sensors data during cleaning hole. If we look at proportional of total power of hookload sensor data, we find that the first five components represent around 95% of the information in hookload sensor data and the first component has zero information. Then we can say that we need to keep four components (2 to 5) to be able to represent and reconstruct the hookload sensors data.

If we use same analogy on other sensors data in Fig. 8, we can conclude the following facts about sensors data during cleaning hole (Circulation) operation:

1. To get all information about Block position sensor data, we need to preserve just one component from polynomial spectrum.
2. It is enough to keep the components from 2 to 8 to have 95% information from torque sensor data.
3. We need just two components (2 and 3) to be able to reconstruct the flowIn sensor data.
4. One component for each Hole Depth and Bit Depth is sufficient.
5. Three components (2-3-4) of polynomial spectrum are enough to have 95% of information about pressure of mud pumps.
6. The second and third components of spectrum are adequate for RPM representation.

11. Distributed recognition system

In Fig. 9, we proposed distributed components-based architecture that shows how all proposed algorithms above will work together to perform the intended goal in drilling operations detection and classification.

As the architecture shows, the raw data flows to segmentation component where the segmentation algorithm hosted. The segments have the structure of start index, end index, attributes and data points. These segments published to Message Oriented Middleware MOM in form of messages. Here we consider the phase of detection drilling operations is finished where each segment represents a drilling operation.

After publishing the segments, a classification step for each segment is required. Patterns classifier component receives the segments from MOM through a subscription component. Then the patterns classifier search for the matched pattern class in its patterns base. The matching process between the new received segments and the patterns templates is determined through Patterns Matching Algorithm (as shown in the next paragraph).

Once classification of each segment is finished, the classified segment published again with known class to MOM where all other clients (parties) which are interested in getting such information can subscribe to get these segments with their classes

11.1. Patterns matching algorithm

Patterns Matching Algorithm between data segment and patterns templates

Input:
Data Segment **S**
Output:
Drilling Operation Class of the segment **S**
Start

- Calculate Gram polynomials coefficients of data V in segment S.
- For each patterns template P in Drilling Patterns Base do:
  * Calculate the similarity between V and P (see next paragraph).
  * Record the bigger value in similarity (closest patterns template).
- End For.
- The class of most similar pattern template is the class of **S.**

End

11.2. Drilling patterns base

In cooperation with drilling experts, a drilling patterns base built. This base contains all templates of known drilling operations such as Drilling, Running in Hole, Pulling out of Hole, Making Connections, and Circulation.

Each operation can have many templates. This depends on the rig's type and the drilling process and methods used by the drilling crews.

As a preparation step, we take each possible template data; calculate its Gram polynomial coefficients; and store them in the Patterns base. At the end of the day, we have all the possible patterns templates stored in this base. One of the most useful usages of such base is that we can use it to store the drilling expert's knowledge. The extension of this base by new patterns templates for new drilling operations considered as a trivial task.

11.3. Patterns similarity measure

Given matrix $\boldsymbol{D}$ that represents the sensor data matrix of a segment. We calculate the corresponding gram polynomial coefficients matrix $\boldsymbol{V}$ of matrix $\boldsymbol{D}$. In addition, we consider that all the patterns templates for each drilling operation are stored in patterns base (see Fig. 9). Moreover, we consider that each pattern represented by matrix $\boldsymbol{P}_\mathbf{i}$, where $\boldsymbol{P}_\mathbf{i}$ contains the coefficients of gram polynomials of template raw data that represent drilling operation number $\boldsymbol{i}$ selected by drilling expert.

After the discussion and the analysis of drilling operations in previous paragraph, we find that to increase our accuracy we need to introduce the weighting (importance) matrix $\boldsymbol{W}$ into our similarity measure calculation

$$W = \begin{bmatrix} p1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & pn \end{bmatrix}. \tag{16}$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ are the importance of sensors data.

For example if we want to consider that the weights of sensors as following: Hookload sensor is very important in our detection process then we give it importance or weight as 1.0. Posblock sensor is another important sensors data we assign 1.0 to it. Mdbit does not affect our detection results as much as Posblock, then we give it importance or weight of 0.5 and so on. The weight matrix $\boldsymbol{W}$ is suggested after similar analysis to what we discussed in previous paragraph.

The matrix $\boldsymbol{W}$ should affect both patterns templates $\boldsymbol{P}$,

$$P_w = P\,W, \tag{17}$$

and coefficient matrix of new raw data $\boldsymbol{V}$,

$$V_w = V\,W. \tag{18}$$

Table 1
Learning and test wells information

Well Name	Resolution	Depth	Sensors used in patterns recognition
Learning Well 1	1 data point each second, 1 Hz	4000 m	Hkld, posblock, mdbit, mdhole, flowIn, RPM, torque, prespump
Test Well 1	1 data point each 5 seconds, 0.2 Hz	2550 m	Hkld, posblock, mdbit, mdhole, flowIn, RPM, torque, prespump
Test Well 2	1 data point each 5 seconds, 0.2 Hz	4100 m	Hkld, posblock, mdbit, mdhole, flowIn, RPM, torque, prespump
Test Well 3	1 data point each 5 seconds, 0.2 Hz	1860 m	Hkld, posblock, mdbit, mdhole, flowIn, RPM, torque, prespump

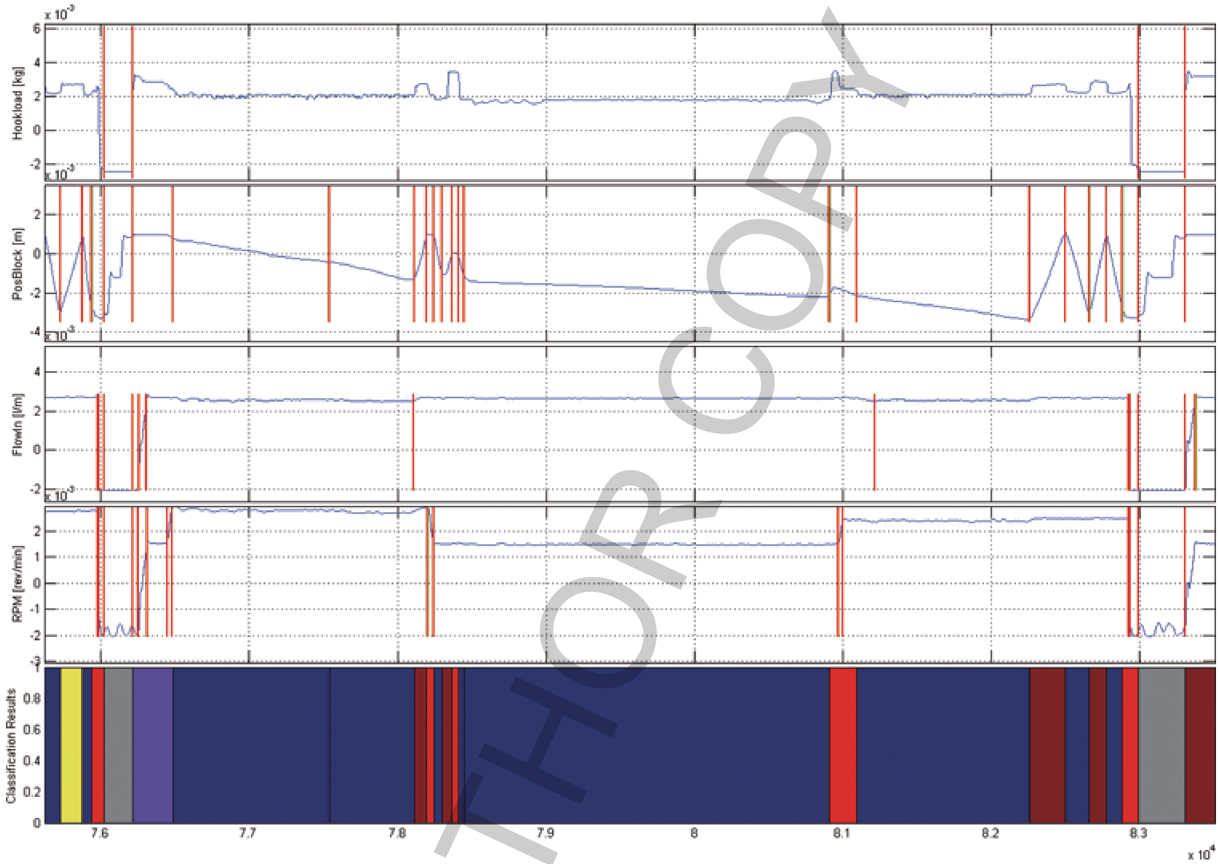

Fig. 10. Results of Segmentation and Classification (≈ 4.5 Hours of Drilling. 0.2 Hz), each color represents different drilling operation. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130181)

The error r represents Mahalanobis distance between $\boldsymbol{P}_\mathbf{w}$ and $\boldsymbol{V}_\mathbf{w}$ given by:

$$r = mahal\,(P_w, V_w). \tag{19}$$

According to definition of Mahalanobis distance then,

$$mahal\,(P_w, V_w) = (P_w - V_w)^T\,\Sigma^{-1}\,(P_w - V_w)\,, \tag{20}$$

where Σ is the covariance matrix of $\boldsymbol{P}_\mathbf{w}$ and $\boldsymbol{V}_\mathbf{w}$.

From Eq. (19), we can write Eq. (20) as

$$d = (PW - VW)^T\,\Sigma^{-1}\,(PW - VW)\,, \tag{21}$$

where $d = mahal(P_w, V_w)$.

Rewrite Eq. (21),

$$d = W^T\,(P - V)\,\Sigma^{-1}\,(P - V)\,W, \tag{22}$$

Noticing that

$$mahal\;(P, V) = (P - V)^T\,\Sigma^{-1}\,(P - V)\,. \tag{23}$$

Then the similarity measurement between patterns templates $\mathbf{P}$ of drilling operations and coefficient matrix $\mathbf{V}$ given by the following equation:

$$d = W^T\,mahal\;(P, V)\,W. \tag{24}$$

In our testing work, we use the similarity measurement suggested in Eq. (24) as a distance function of the proposed patterns classifier.

Test Well 1

Real Operations	Classified Operations (Total Accuracy: 94.05%)					
	Drilling	Run in Hole	Pull out of Hole	Circulation	MakeCon	Accuracy (Operation)
Drilling	**1819**	0	0	1	0	99.95%
Run in Hole	78	**789**	8	9	7	88.55%
Pull out of Hole	81	0	**492**	1	7	84.68%
Circulation	39	16	12	**143**	21	61.90%
MakeCon	15	46	38	23	**3111**	96.23%

Test Well 2

Real Operations	Classified Operations (Total Accuracy: 88.61%)					
	Drilling	Run in Hole	Pull out of Hole	Circulation	MakeCon	Accuracy (Operation)
Drilling	**726**	5	0	1	0	99.18%
Run in Hole	95	**472**	2	3	14	80.55%
Pull out of Hole	39	0	**372**	3	14	86.92%
Circulation	78	36	28	**209**	13	57.42%
MakeCon	35	20	9	15	**1412**	94.70%

Test Well 3

Real Operations	Classified Operations (Total Accuracy: 88.56%)					
	Drilling	Run in Hole	Pull out of Hole	Circulation	MakeCon	Accuracy (Operation)
Drilling	**1095**	0	0	9	0	99.18%
Run in Hole	105	**219**	0	1	0	67.38%
Pull out of Hole	55	0	**340**	1	3	85.21%
Circulation	62	32	12	**312**	30	69.64%
MakeCon	27	5	12	40	**1085**	92.81%

Fig. 11. Confusion matrices of test wells with classification accuracy of each one.

12. Test dataset

The test database, which we worked on, consists of four complete drilled offset wells. The drilling operations highlighted manually by drilling experts on raw data. Table 1 contains information about the testing wells.

13. Results and discussion

Figure 10 illustrates detailed view on how the segmentation and the classification algorithms worked together to accomplish the mission of drilling events detection and classifications. The drilling sensors measurements shown and vertical redlines (segments) plotted over them. The color ranges represent drilling operations that classified using our suggested system. Each color is different class of operations.

Figure 11 demonstrates the results of applying patterns classifier on three offset wells as testing wells. The accuracy of classification process gives a percentage around 90% and we consider that as a high classification rate. It is shown that the confusion of this classifier happen between formation drilling operations and non-drilling operations. This is due to the similarity of flowIn and RPM trends during formation drilling (making hole) and non-drilling situations where both situations show that flowIn and RPM sensors data should be in straight trends. For this reason, the patterns-based classifier is confused between both situations. As a solution for this problem, a threshold level can be defined to employ more information on the trend of flowIn or RPM. Another reason for confusion in the classification results is using all sensors data available. As we discussed in previous paragraph, some sensors data contain same information. For example hole depth, bit depth, and posblock during drilling operation carry same information and they considered as redundant components and this may cause confusion if the quality of those sensors data is bad.

In addition, some sensors data are more important than other sensors data during specific operations. For example during making connection, each of flowIn, pumps pressure, hole depth, and bit depth sensors data are not important for recognizing this operation. The two important sensors data here are Block Position and Hookload.

We believe that the accuracy of results will be more than what it is now, if the previous comments are reflected on the current implementation of the patterns-base classifier. Furthermore, data with no outliers or missing values is expected to have higher accuracy.

14. Future work

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

It is also possible to extend the suggested method of drilling operations classification to recognize any other required operation based on whatever drilling sensors data. The only condition to apply this method is that the trends of sensors data should be obvious.

In addition, the suggested method can be improved by using a rejection classifier to reject specific patterns do not belong to particular operation.

In our work, we used automatic adding for all possible patterns templates in the database and here the error of experts classification is included the results. As improvement, a continuous filtering process should run on the patterns base to remove bad and wrong patterns that cause low classification rate.

Another important improvement is that the results of classification from test wells added to the patterns base, if and only if they reviewed, corrected, and accepted by experts. This supports the idea of extending current patterns base with new knowledge from experts and it shows an example on how to extract the knowledge from experts and teach current running systems to do better in future.

References

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

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

[3] A. Arnaout, B. Esmael, R.K. Fruhwirth and G. Thonhauser, Automatic threshold tracking of sensor data using Expectation Maximization algorithm, Hybrid Intelligent Systems (HIS), 2011 11th International Conference on, vol., no., pp. 551,554, 5–8 Dec. 2011 doi: 10.1109/HIS.2011.6122164.

[4] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society* Series B (Methodological) (1977), 1–38.

[5] E. Keogh, S. Chu, D. Hart and M. Pazzani, 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.

[9] P. O'leary, M. Harker and M.Ã. Bernhard, Discrete polynomial moments and the extraction of 3-D embossed digits for recognition, *Journal of Electronic Imaging* **19**(3) (2010), 031205–031205.

[10] P. O'Leary and M. Harker, Surface Modelling Using Discrete Basis Functions for Real-Time Automatic Inspection, 3-D Surface Geometry and Reconstruction: Developing Concepts and Applications (2012), 216.

[11] P. O'Leary and M. Harker, An Algebraic Framework for Discrete Basis Functions in Computer Vision, 2008 Sixth Indian Conference on Computer Vision, *Graphics and Image Processing* (1) (2008), 150–157.

[12] G. Thonhauser and W. Mathis, SPE 103211 Automated Reporting Using Rig Sensor Data Enables Superior Drilling Project Management, SPE, 2006.

[13] G. Thonhauser, G. Wallnoefer and W. Mathis, SPE 99880 Use of Real-Time Rig Sensor Data to Improve Daily Drilling Reporting, Benchmarking and Planning – A Case Study, SPE, 2006.

[14] G.C. Tiao and A. Zellner, Bayes's theorem and the use of prior knowledge in regression analysis, *Biometrika* (1964), 219–230.