SPE-170680-MS

Predicting Failures from Oilfield Sensor Data using Time Series Shapelets

Om Prasad Patri, Anand V. Panangadan, Charalampos Chelmis, University of Southern California, Randall G. McKee, Chevron U.S.A. Inc., and Viktor K. Prasanna, University of Southern California

This paper was prepared for presentation at the SPE Annual Technical Conference and Exhibition held in Amsterdam, The Netherlands, 27–29 October 2014.

Abstract

Increasing instrumentation of the modern digital oilfield produces streams of data from sensors that monitor the functioning of different components in the field. This data should be converted to actionable information rapidly in order to respond to events as they happen or are predicted. The challenge is therefore to develop technologies that can process these large sensor datasets rapidly and with minimal manual supervision to ensure a data processing system that can scale with the increasing instrumentation.

We consider as a use-case an oilfield with several Electrical Submersible Pumps (ESPs), each instrumented with sensors that continually measure electrical properties of the pump (the streams of sensor data), which are then relayed to a central location. In this paper, we demonstrate how a time-series analysis approach can be applied to failure detection and failure prediction from the streams of sensor data. The method involves identifying "shapelets" – short instances that are particularly distinct – in the streams of sensor data.

The shapelets approach is particularly applicable to large oil and gas enterprise datasets because the algorithm does not need access to the entire historical data. This greatly reduces the amount of data that needs to be stored for data analysis. Moreover, unlike model-based approaches, shapelet-based analysis does not make any assumptions about the underlying nature of the data, making it practical for applications where a detailed physical model of the pump is not available.

We validate our proposed method by analysis on a representative set of instrumented ESPs. We describe the pre-processing steps that were applied in our analysis. We report the results of experiments to study the effects of varying the data processing parameters on the accuracy of fault detection and prediction. These results indicate that shapelet-based approaches are promising for analysis of time-series data in the oil and gas industry.

1. Introduction

The evolving nature of the modern digital oilfield requires large-scale instrumentation and monitoring. Data mining and machine learning approaches have become vital to digital oilfield operations [1, 2] as they move into the age of Big Data. A massive portion of oilfield data today is in the form of sensor streams which necessitates the use of rapid real-time data analysis techniques [3]. Due to the high frequency of sensor data in the oil and gas industry [4], it is inevitable for oil and gas enterprises to leverage efficient data mining techniques, especially those dealing with time series data.

In this paper, we adapt a time-series mining approach that was recently developed in the computer science community, known as *time-series shapelets* [5], for application to sensor measurements typically collected in the oil and gas industry, in particular to component failure detection and prediction. The streaming time series nature of such data is especially suited to this approach. The shapelets method identifies those time segments ("shapelets") within the available sensor data which are most discriminative for differentiating between two classes, for instance those arising from failed pump components in contrast to those that are working normally. Using discovered shapelets from historical data, predictions about future failures or anomalies can be made such that proactive steps can be taken to mitigate their effects. As an example, a shapelet that was discovered from intake pressure measurements from an electrical submersible pump is overlaid over the full time series Figure 1; the short segment shown in red was found to be the most discriminative from this time series and others in the labeled data record for distinguishing failed pumps from normal ones. This shapelet, along with other shapelets, can be used for detecting failures by comparing thsem with real-time sensor data (further details about finding shapelets are provided in Section **4)**.

The shapelets approach is particularly effective for oil and gas enterprise data because it does not need access to the entire historical record of sensor data while making decisions – only the shapelet time segments, identified in an offline step from the historical record, are needed for real-time analysis. This greatly reduces the amount of data needed to be stored for further data mining. Moreover, this approach does not make any assumptions about the nature of the data, making it practical for real world scenarios. The shapelets found are visually interpretable, making deeper root cause analysis possible. Such time series mining methods are directly related to events in the oilfield, and as described in [2], there are several value propositions for efficient event processing [6].

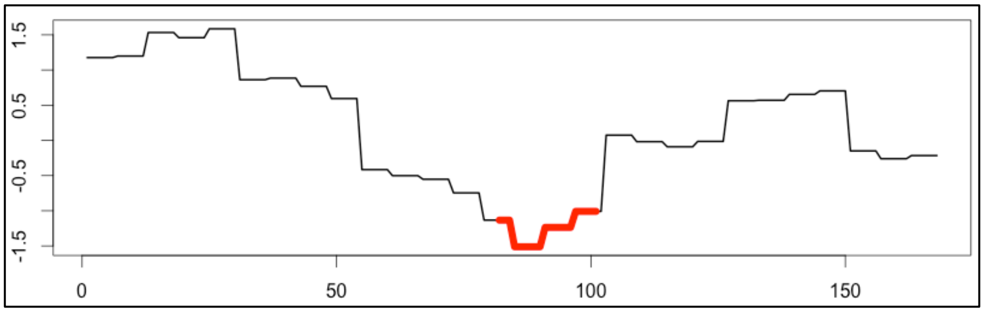

Figure 1: A discriminative segment (shown in red), or *shapelet*, from a time series as computed by the shapelet mining algorithm. The x-axis shows time (in days) while the y-axis represents normalized values of the intake pressure for a specified ESP.

To summarize, the primary purpose of performing time series mining on oilfield sensor data are:

- To detect anomalies in equipment operation automatically
- To identify time segments that are indicators of failures
- Reduce the data storage and computational power requirements for real-time sensor data analysis

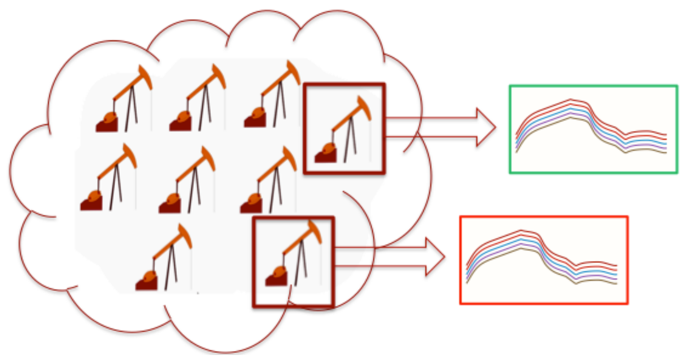

Figure 2: Sensor measurements as time series for failure detection and prediction

1.1 Use-case

Electric submersible pumps (ESPs) are one of the main artificial lift methods for extraction of fluid. The failure of an ESP increases operating costs and affects production. Statistical analysis of ESP failures shows that the failure rates of ESPs vary considerably [7]. There are several factors affecting ESP failures, including reservoir type, sand control, and whether the ESP is a replacement [7]. In an effort to understand the causes of failures, the oil and gas industry has attempted to collect data on all potential factors and exchange this information for better data analysis [8]. However, data aggregation is a difficult task considering the myriad ways of collecting and recording the diverse types of data [9]. Instead of attempting to build a complete model of ESP failures from these disparate data sets, we apply machine learning and time-series data processing techniques to automatically learn failure models from only the sensor measurements (with relatively high temporal frequency) collected directly from the ESP. The goal of our research is therefore to understand the limits of failure detection and prediction using only the most accessible sensor measurements.

It has been estimated that just a 1% improvement in ESP performance world-wide would provide over a half-million additional barrels of oil per day [10]. Considering the high cost incurred by an ESP failure, early detection and prediction of pump failures of even a subset of all operational assets from readily available data can reduce OPEX (Operational expenditure) and can reduce average maintenance cost. The approach proposed in this paper can be generalized for other smart oilfield areas where time series data is plentiful. This use case is illustrated in Figure 2.

The rest of this paper is organized as follows. In Section 2, we describe related work in ESP failure analysis. In Section 3, we provide a brief introduction to the shapelets approach to time series data processing. In Section 4, we describe our approach and Section 5 presents the results from our experiments. We conclude in Section 6.

2. Related Work on ESP failure prediction

Considering the large number of ESPs deployed in an oilfield, the bulk of ESP failure prediction research has focused on computing population-level estimates and causes of failure [11]. Statistical analysis of ESP failures across an oilfield has shown that the failure rates vary considerably [7]. Sawaryn has derived analytical expressions describing failure patterns at a population level [12]. The amount of sand in the extracted fluid is one of the important causal factors affecting failure and Kalu-Ulu et al. have modeled ESP failures in sand producing wells [13]. Furthermore, Liu et al. [14] has adapted data mining classification algorithms and Liu et al. [15] presents a Bayesian network-based machine learning algorithm to predict rod pump failures.

In our work, we explore methods to identify and predict failures of individual ESPs using machine learning techniques. We have also used a technique (similar to this work) based on time series shapelets for electricity disaggregation in [16]. Specifically, in this work, we have used the concept of events as temporal periods of operation (either failure or normal) of ESP pumps as described in Section 4. However, it is possible to instead add more semantics and structure to define and represent events for oil field processes. The Process-oriented Event Model [17] describes the semantics of such an event model and also provides a case study for pump failure event detection.

3. Background on Time Series Shapelets

Time series shapelets, introduced by Ye et al. [5], are discriminative subsequences in time series which can differentiate instances of the positive class (e.g. normal operation of pump) from those of the negative class (e.g. failed or anomalous operation of pump). Time series shapelets are identified based on information gain criteria [5]. Essentially, the subsequence with the highest information gain is identified to be the most discriminative between the classes and thus is the "shapelet". To perform classification of time series using shapelets, a decision tree classifier is built based on distance from the shapelet. When a new time series instance is encountered, the distance to the shapelet is calculated, and depending on which branch of the tree it falls in (distance lesser or greater than distance measure of shapelet), it is assigned a predicted class label. For the case of multiple shapelets from one time series, the same decision tree can be extended to have multiple nodes.

To summarize, our primary motivations for applying the time series shapelets method to oil and gas sensor data is because they have the following advantages over other conventional approaches:

- Shapelet-based methods provide very fast classification times
- They are visually interpretable
- They do not need to use the complete training data for classification (after the shapelet has been discovered from the time series), only the discovered shapelet segments are required
- They make no assumptions about structure of time series (in contrast to conventional autoregressive or ARIMA time series models). The number of parameters involved is also minimal.

Various extensions to time series shapelets have been proposed in the literature, most notable of them being Logical Shapelets [18], Local Shapelets [19], and Fast Shapelets [20]. In this paper we use the Fast Shapelets algorithm, which is currently the fastest supervised shapelet finding method [19]. Though they were originally intended for use with supervised datasets, extending shapelets to include the idea of unsupervised shapelets (U-shapelets) was presented in Zakaria et al. [21]. Instead of maximizing the information gain in the case of (supervised) shapelets, the U-shapelets method maximizes the separation gap, a new metric described in [21]. This method allows unsupervised shapelets to be used for clustering. A stopping criteria based on Rand index is used to decide the number of clusters.

4. Approach

This section describes the dataset used for our experiments and then explains our proposed shapelet mining approach. We developed a pre-processing method to address the irregular nature of real world sensor data and transform the sensor measurements into a format suitable for shapelet-based time series analysis. This method can handle invalid data points, irregularity in sampling intervals and reduce redundant data segments. These are three frequent problems with industrial sensor datasets and are particularly applicable to datasets from the highly instrumented oil and gas industry.

4.1 Dataset description

We used data from electrical submersible pumps (ESPs) from a single onshore oilfield within Chevron North America. We consider a supervised classification scenario. The classification task is to detect and predict the behavior of a new ESP instance over a specified period of time, given sensor readings of its physical attributes (current, voltage and intake pressure). We are provided with labeled data from ESPs defining periods of normal and failure operation. The data ranges from 1 Aug 2011 to 29 July 2013, and measurements are nominally recorded at hourly intervals. We use data from 11 normal operation periods (from 11 ESPs) and 10 failure operation periods (from 8 ESPs, with two ESPs providing two failures each). For these periods, we have sensor data containing time-stamped data points for various attributes or physical properties of the pump. In this paper, we focus on three of these attributes – current, voltage, and intake pressure, which are the most readily available sensor measurements. We also have timestamps for all of the data with separate timestamps for insertion date of the data points as well as the last good scan dates (which we use later for ensuring that the data points are valid). Instances of sensor measurements from selected ESPs during both normal and failed operation periods are shown in Figure 3 (intake pressure), Figure 4 (current), and Figure 5 (voltage). These plots show measurements from two normally functioning ESPs and two failed ESPs. These do not represent all the patterns of normal or failure periods since there is a lot of diversity in the data. Our prediction methods need to incorporate such diverse instances to be able to make decisions even for previously unseen instances without overfitting. Shapelets are suitable for this purpose because they are designed to identify the most discriminative segments in the time series instances, irrespective of the nature of the data.

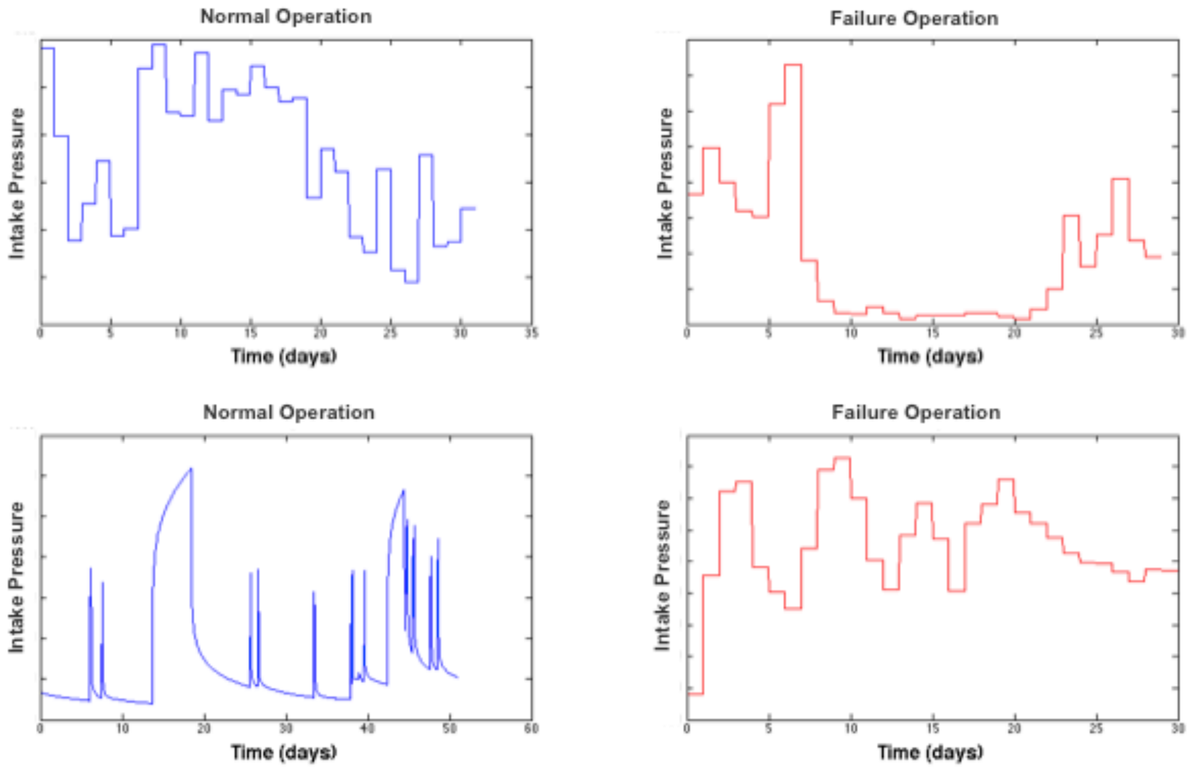

Figure 3: Examples of intake pressure measurements during normal operation (blue) and failed operation (red).

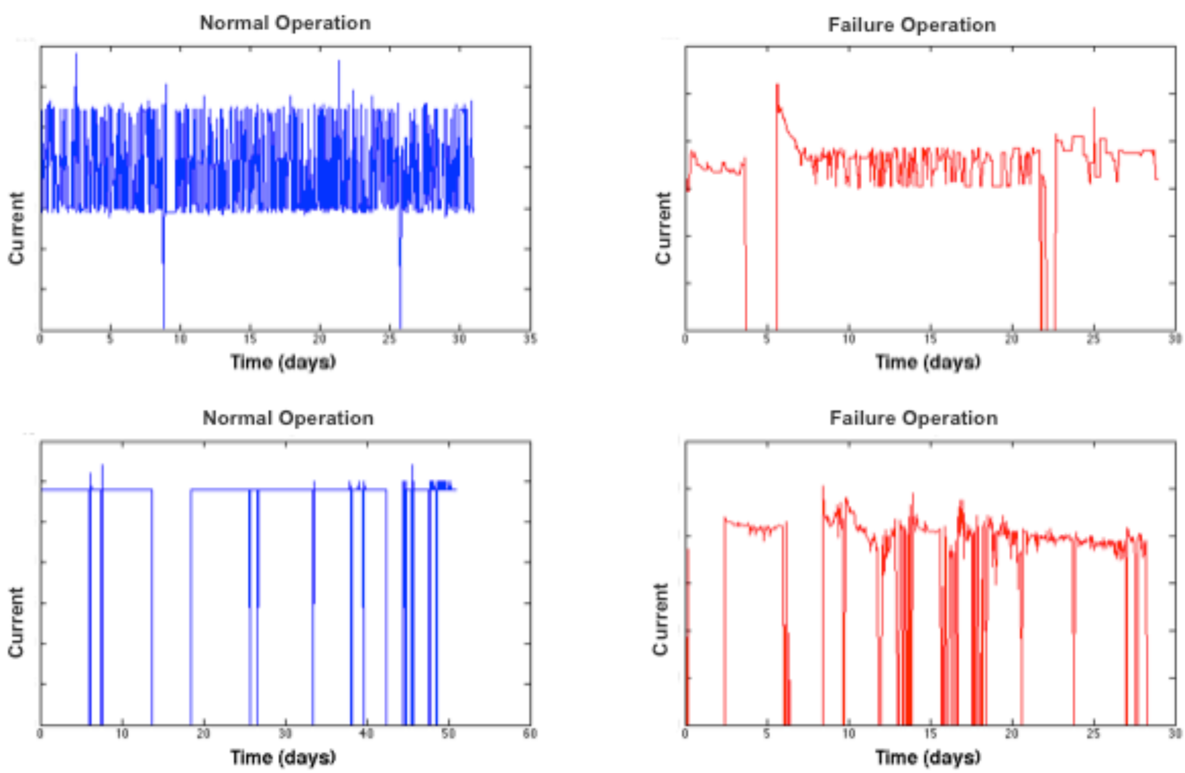

Figure 4: Examples of current measurements during normal operation (blue) and failed operation (red).

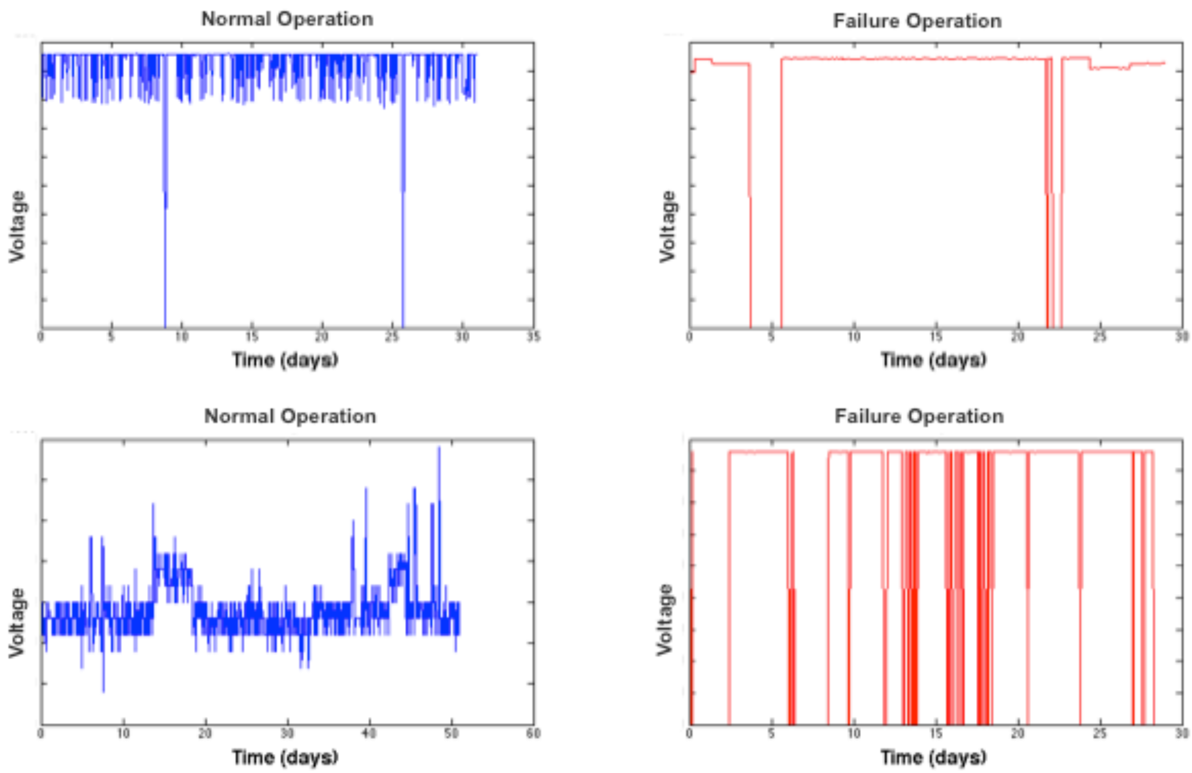

Figure 5: Examples of voltage measurements during normal operation (blue) and failed operation (red).

4.2 Issues with using the raw dataset

Since real oilfield data is irregularly recorded and contains large missing portions, we have developed an appropriate pre-processing method for use before the described shapelets technique can be used for classification. In particular, we have developed a pre-processing method to obtain "clean" data segments. The proposed method divides each sensor data stream into windows of a specified segment size (preferably, the time duration over which predictions need to be made) and ensures that the elements within each segment are regularly sampled. Only a specified degree of overlap is allowed between consecutive segments so as to avoid overly similar time segments. A check is performed on whether the sensor value read was new or not, providing only regularly sampled data segments in the end, suitable for input to shapelet based classification approaches.

Several factors make the raw dataset, collected directly from sensors in an oilfield, unsuitable for direct analysis via the shapelet mining algorithm (or any generic data mining method). We desire a regularly sampled version that can directly be input to a state-of-the-art shapelet finding method (such as Fast Shapelets [20]). To process the dataset, a logical choice would be obtaining one time series per ESP (per failure/normal operation period). However, this is not possible to implement here because of the following three issues.

The first issue is that of validity. Since it is possible for sensors to fail in their functioning, some of the data points recorded may not be valid and need to be eliminated. Keeping track of the "last good scan" timestamp in the data can help us perform this elimination.

The next concern is of regularity in the frequency of sampling the data points. In cases where the sensor was not operational, or for other reasons, we found that the data was sampled for some periods at a 1-day frequency instead of the standard 1-hour frequency imposed on the rest of the data. Figure 6 shows plots of insertion dates and last good scan dates for a single pump. As seen in the plot, it is not regular for the complete duration of the data stream. Regular intervals between data points are essential for the shapelet finding algorithm. One method to handle this is to interpolate for the missing data points, but this can result in localized artifacts, which are then spuriously interpreted as distinctive patterns in the shapelet-based algorithm. Instead, as described later, we find clean segments where the data is regularly sampled for further use.

The third issue is that of redundancy. There may be several data segments in our input training dataset that are very similar to each other with only a few data points being different at the beginning and end of the segments. However, if we do not allow any overlap between the segments at all, then we are left with an extremely limited number of segments and the possibility that the most distinctive patterns are across two segments. To handle this tradeoff, we introduce an overlap parameter, which controls the amount of overlap between consecutive segments.

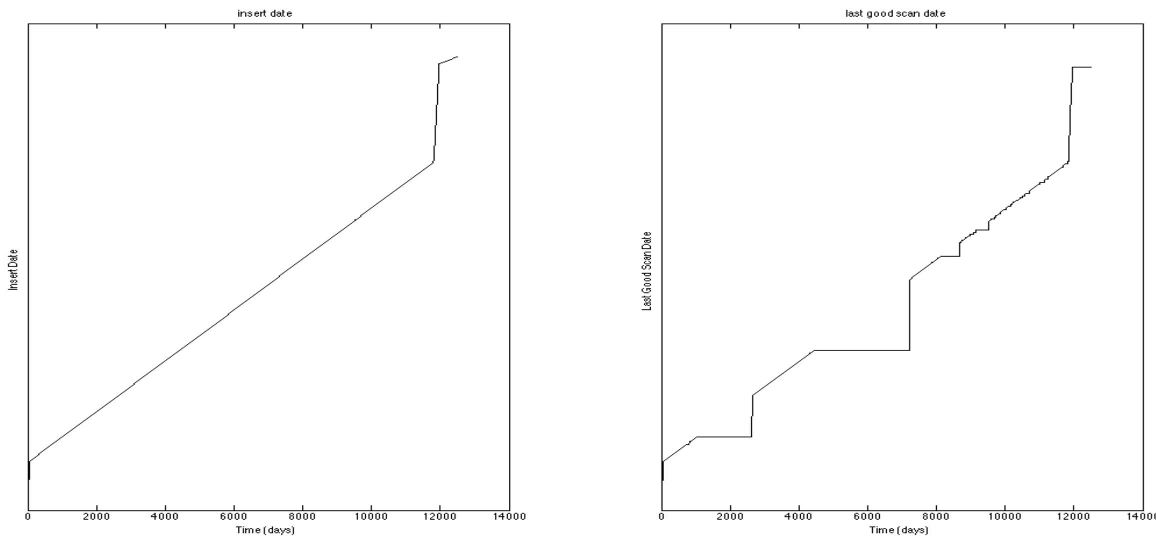

Figure 6: Irregularity in sampling of oilfield sensor data. The period between consecutive data points varies between 1-hour and 1-day. The X-axis shows time (in days) and the Y-axis represents the insertion date of the data point (left) and the last good scan date (right).

4.3 Pre-processing

Since large portions of the data can be missing, we need an effective pre-processing technique to filter useful data. The following filtering step does not modify any of the data or interpolate any values. Our pre-processing method is summarized in Figure 7. We extract segments out of each time series for each pump. The first criterion for keeping a segment is to ensure that it is valid. We ensure this by requiring that the "insertion date" of the data entry is within 1 day of the "last good scan" timestamp. The second condition is to fix a regular time interval for sampling of the values. We only consider those values, which are recorded 1 hour apart. These two pre-processing conditions give us regularly sampled segments of data. However, each segment is very similar to the previous segment with the exception of just one or two data points, which have changed. To eliminate many of these close segments, we add a third condition, which restricts the degree of overlap between adjacent or consecutive segments. This parameter can be changed to reflect various requirements in specific use cases. The more the allowed overlap, the more the number of segments we get after pre-processing. In our experiments, we allow a 75% overlap between adjacent segments, thus forcing at least ¼ of consecutive segments to be different from each other.

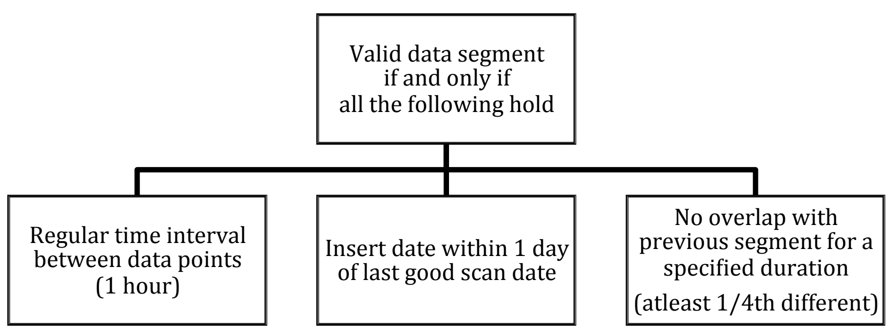

Figure 7: Proposed pre-processing workflow

4.4 Failure event detection and failure event prediction

We describe the two major scenarios of data mining tasks performed via the shapelet mining approach. The first task is failure detection. The failure detection experiment aims to detect whether a given data segment represents a failure or not. It is reactive rather than being proactive. For failure event detection, we label the failure event durations as negative class instances while all other segments are treated to be of the positive class. For example, as described earlier in the pre-processing steps, given a segment length parameter, all possible clean data segments (of the specified segment length), which satisfy all the three processing checks, may be extracted.

Failure event prediction aims to predict failures before they occur, so that maintenance personnel may be notified in advance, taking the proactive anomaly detection approach mentioned in [6, 22]. The failure prediction experiment aims to classify whether a given data segment will eventually lead to a failure. We define a user-customizable lookback parameter, which is to be used for finding precursor segments for building the training and testing data sets in the shapelet-based classification approach. The intuition behind this is that there is a failure event precursor at some point before the failure is detected and we want to make predictions based on including this data point in our training and testing data. We label segments that are precursors to the failure or normal event durations (within the lookback period) as negative or positive class instances respectively. Note that we do not extract or consider any data from the actual failure or normal duration while making a prediction.

Typically, the lookback should be larger than the segment length parameter so as to have multiple data segments for training the system. However, if it is too large, then it would cover a lot of redundant data points which may not have the real precursors for the failure event that is about to happen. Since we do not know the point of time when an exact precursor for a failure event occurs, we fix a moderately long lookback period, typically a few times the segment length. However, the shapelet-based algorithm provides good results (as shown in the next section) for different lookback periods irrespective of the fact that we may not have included the 'real' failure precursor data point or added too many redundant data points. Figure 8 illustrates the extraction of time segments from each sensor dataset for failure detection and prediction.

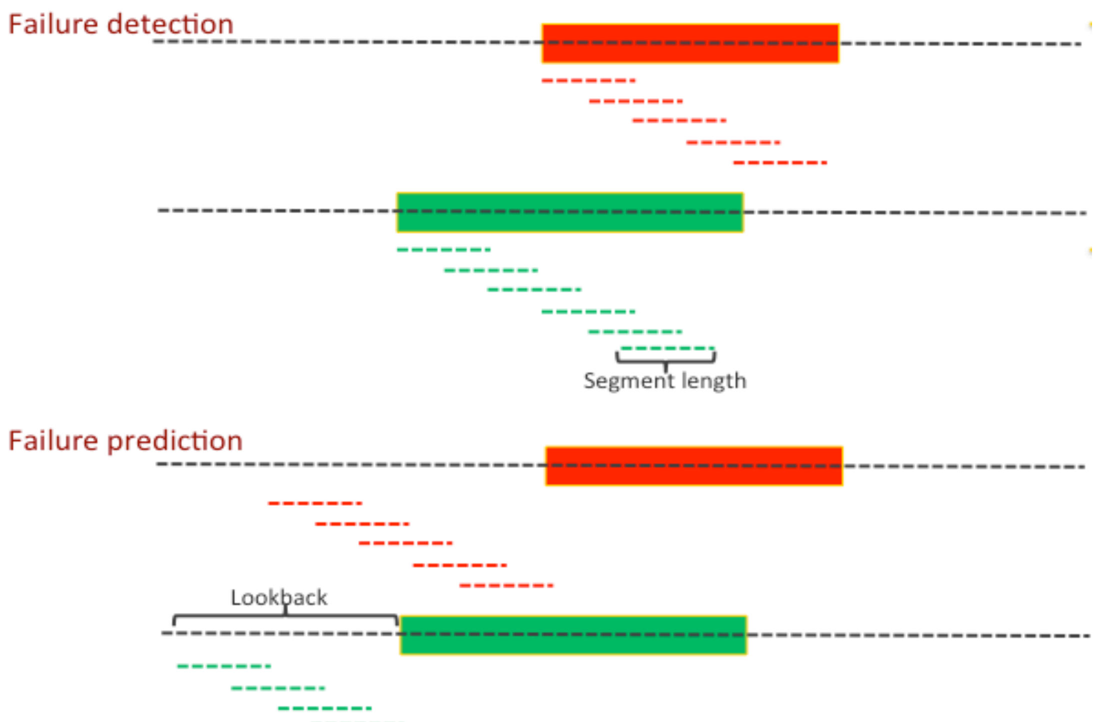

Figure 8: Failure Event Detection and Failure Event Prediction Scenarios. The red shaded portion denotes a failure operation and the green operation period denotes a normal operation period. All segments extracted are of the specified segment length. In the detection scenario, the segments are extracted from the actual failure or normal operation duration, while, for prediction, segments are extracted only from the lookback period without utilizing any data in the actual labeled failure or normal periods.

5. Evaluation

This section describes our experimental framework, experiments performed and evaluation results. We have two major categories of experiments – at the segment-level and at the pump-level, and for each category, there are two scenarios – failure detection and failure prediction, as described earlier. We used the supervised shapelet finding algorithm, Fast Shapelets [20], for finding shapelets in all the experiments. A discussion of the experimental results is given in Section 5.6.

5.1 Segment-level Failure Detection

The first category of experiments is that of segment-level classification – either for failure detection or failure prediction. In this category, we extract data segments from all pumps during pre-processing. All extracted segments (from all pumps) are collected to form the full segment-level dataset. This dataset is split randomly into training and testing sets, which are independent of each other. Note that in this method of evaluation, though a given segment only appears in either the training or test portion, a particular pump can contribute sensor data segments to both the training and test portions. Two shapelets, extracted from one such training experiment, are shown in Figure 9. According to the shapelet framework, these segments are the most discriminative for distinguishing failures from working instances.

We used various parameters in our experiments. We used segment lengths of 1-day, 2-day and 1-week. The allowed overlap was set such that at least one-fourths of the data is different between consecutive segments. We experimented for 25%-75% and 50%-50% random splits for the training and testing sets respectively. We compared our results to a baseline classifier, denoted by ZeroR, and to other common classifiers as well in Section 5.3. The ZeroR classifier simply assigns the label of the majority class in the training data to a new class instance. Thus, if there are more positive class instances than negative in the training data, it will assign the positive class to every test data instance, and vice-versa. This baseline classifier provides the minimum expected accuracy estimates for this dataset. The accuracy results are shown in Table 1.

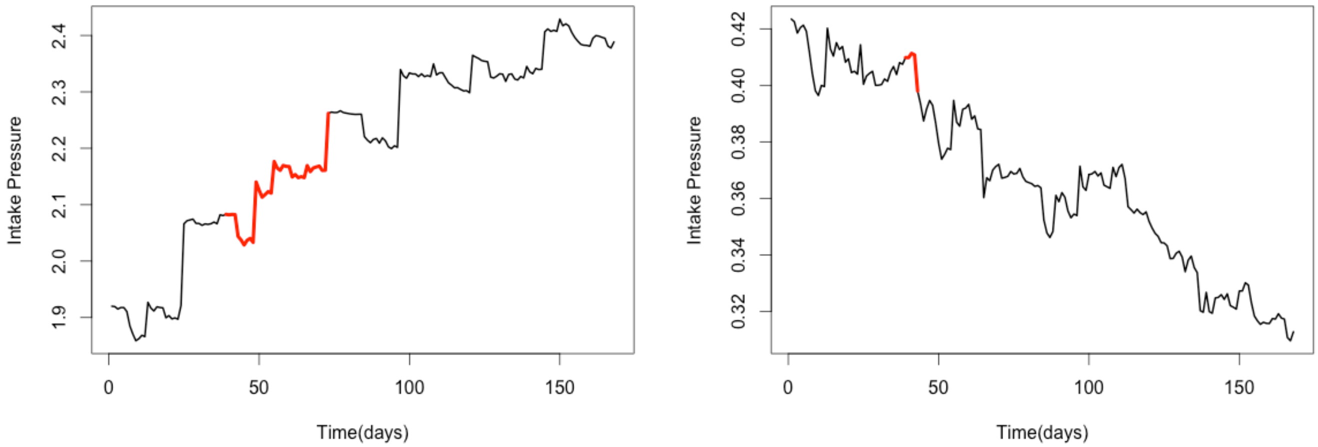

Figure 9: Two shapelets (shown in red) extracted for failure detection based on intake pressure sensor data (normalized). The segment length was 1 week. The first shapelet (left) is of 35 hours duration and the second one (right) is of 5 hours duration. The data was split into a 50%-50% train-test split (half of the data was used for training).

Table 1: Accuracy of detecting failures at the pump level.

Segment length	**1 day**		**2 days**		**1 week**	
% Data in Training	**25%**	**50%**	**25%**	**50%**	**25%**	**50%**
Current	77	77	77	82	72	85
Voltage	81	82	83	85	72	73
Intake Pressure	90	89	90	91	85	96
ZeroR (baseline)	72	72	72	74	73	75

5.2 Segment-level failure prediction

For our failure prediction experiment, we considered data segments that appear before the failure or normal events (i.e., data segments are collected from the lookback period). The lookback periods were set as follows: a lookback of 1-week for 1-day segment length, and a lookback of 4-weeks for both the 2-day and the 1-week segment lengths. The accuracy results are shown in Table 2. The shapelets extracted from training a prediction classifier are shown in **Figure 10**.

Table 2: Accuracy of predicting failures at the pump level.

Segment length	**1 day**		**2 days**		**1 week**	
% Data in Training	**25%**	**50%**	**25%**	**50%**	**25%**	**50%**
Current	60	68	67	73	71	66
Voltage	58	68	64	69	70	67
Intake Pressure	87	83	73	74	67	71
ZeroR (baseline)	56	53	52	52	52	55

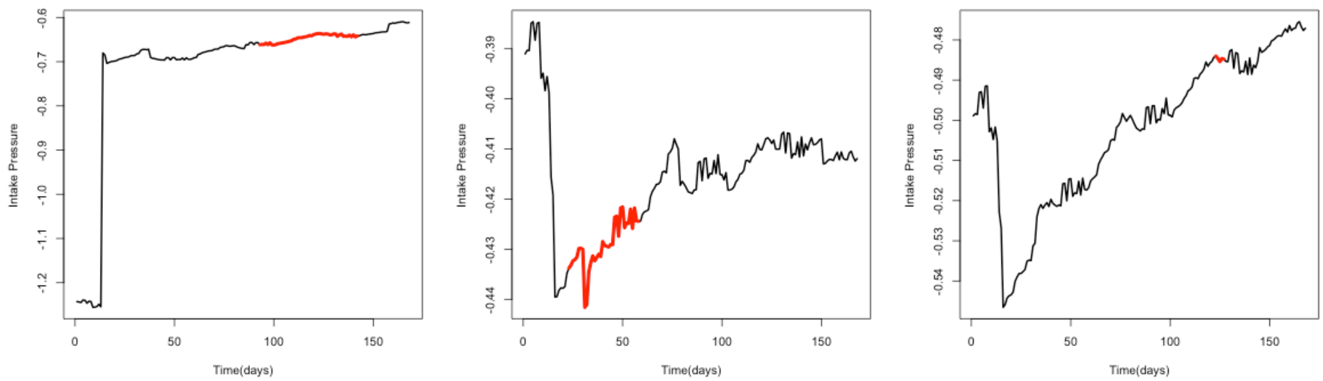

Figure 10: Three shapelets (shown in red) extracted for failure prediction based on intake pressure sensor data (normalized). The segment length was 1 week and lookback period was 4 weeks. The shapelet durations are – 50 hours, 36 hours and 5 hours. The data was divided into a 50-50 train-test split (half of the data was used for training).

5.3 Comparison with other machine learning techniques

We compared our results with those of several state-of-the-art machine learning approaches, widely used for pump failure detection. The results are shown in Table 3. We observe that shapelet-based methods are comparable to other classifiers. In these experiments, the segment length is 1 week and the split for training and testing is 50%. The lookback window for failure prediction is 4 weeks.

Table 3: Comparison of accuracy (%) of shapelet-based segment-level failure detection with other machine learning techniques. The following classifiers are compared to our proposed Fast Shapelets-based method (FS): Logistic Regression, Multi Layer Perceptron (MLP), Support Vector Machines (SVM) using either Gaussian or radial basis function (rbf) kernel, polynomial kernel, or linear kernel, AdaBoost, Decision Tree (J48), Random Forests (RF) and our baseline classifier (ZeroR).

	FS	Logistic	MLP	SVM(rbf)	SVM(poly)	SVM(linear)	AdaBoost	J48	RF	ZeroR
Failure detection	**96**	74	93	**96**	73	73	93	91	95	75
Failure prediction	71	55	69	76	55	55	73	75	**80**	55

5.4 Pump-level failure detection

The segment-level classification experiments attempt to detect and predict failures given a single segment. However, in a typical oilfield monitoring application, it is sufficient to detect and predict failures at a given time using all the sensor measurements available up to that time. We can therefore combine the predictions of all the time segments from a given ESP to get a prediction of the failure for that pump. We call this analysis pump-level failure detection and prediction.

For pump-level analysis, the train-test split is performed differently from the segment-level experiments. We use leave-one-out (LOO) cross validation for evaluating pump-level failure detection and prediction. In LOO, only instances (the data segments extracted after pre-processing) from one of the pumps are set aside for testing the learned classifier while all other segments (along with their failure/normal labels) are used for training using the Fast Shapelets algorithm as described earlier. This evaluation is repeated for each pump and the classification accuracy results are aggregated. Note that unlike segment-level evaluation, none of the data segments from the well set aside for testing are used for training the shapelet-based failure detector or predictor. (In segment-level evaluation, the segments from all the wells were randomly split into training and test instances.)

Table 4 presents the accuracy of pump-level failure detection. There are 8 ESPs which failed and 11 which functioned normally in the available data (column "Failed/Normal?"). The number of data segments extracted after pre-processing is shown (column "#test segments") for each of these 19 cases. The accuracy of applying the shapelet-based classifier for each of the segments from an ESP (the classifier is trained on all the segments from the remaining ESPs) is shown in columns "#detected failure segments" and equivalently as a percentage in column "Segment-level accuracy (%)." In order to combine the segment-level labels into a single label for the pump, we apply a threshold. Setting this threshold enables the precision-recall performance of the classifier to be traded-off against each other. (Precision is defined as the fraction of all positives labeled by the classifier that are true positives; recall is defined as the fraction of all positives in the dataset that are identified as true positives by the classifier.) Increasing the threshold increases the precision (certainty that identified failures are genuine faults) at the cost of recall (missing some genuine faults). In our evaluation, we set the failure detection rule to classify a pump as a failure if 40% or more of the segments were classified as failure segments. This approach is then able to detect ESP failures with a precision of 89% – 8 out of the 9 pumps labeled as failures had actually failed and a recall of 100%

– 8 out of the 8 failed pumps were correctly identified. In Table 4, setting this threshold enables Pump #5, #9, and #16 to be labeled correctly even though their segment-level accuracy is less than 100%.

Table 4: Results of detecting failures at the pump level. Rows in bold indicate pumps that were labeled correctly as failures/normal though their segment-level accuracy is less than 100%

Pump number	Failed/ Normal?	#test segments	#detected failure segments	Segment-level accuracy (%)	Detection label after threshold of 40%
1	F	4	4	100	F
3	F	1	1	100	F
4	F	2	2	100	F
5	**F**	**6**	**4**	**67**	**F**
6	F	28	28	100	F
7	F	2	2	100	F
8	F	10	10	100	F
9	**F**	**10**	**4**	**40**	**F**
10	N	11	0	100	N
11	N	13	13	0	F
12	N	19	0	100	N
13	N	22	0	100	N
14	N	24	0	100	N
15	N	13	0	100	N
16	**N**	**8**	**2**	**75**	**N**
17	N	14	0	100	N
18	N	14	0	100	N
19	N	22	2	91	N
20	N	19	0	100	N

5.5 Pump-level failure prediction

The evaluation described in Section 5.4 is repeated for pump failure prediction. In this case, all the segments for training and testing are taken from the lookback period. The accuracy results are shown in Table 5. In this case, we apply a threshold of 20%. The failure prediction rule is that a pump is predicted to fail if 20% or more of the segments were classified as failure segments. The pump-level shaplets-based predictor is able to predict ESP failures with a precision of 78% (7 out of the 9 pumps labeled as failures will actually fail) and a recall of 78% (7 out of the 9 pumps that would fail were correctly identified).

5.6 Discussion

The computational cost of the Fast Shapelets method that was presented is very different for training and classification. Classification is typically fast since it depends only on the number and length of shapelets identified during training and these are few in number (**Figure 9** and **Figure 10**). The shapelets identified during training can be viewed by domain experts (**Figure 9** and **Figure 10**) and an interpretation with regards to failure causes can be attempted. As expected, prediction of failures is harder than detection (the accuracies of prediction experiments are lower than that of the corresponding detection experiments). Intake pressure is generally a better indicator and predictor of failures compared to current and voltage given the accuracy results in (Table 1 and Table 2). As expected, increasing the amount of data used in training gives higher accuracies (25% versus 50%, Table 1) but the accuracies are comparable in magnitude. Moreover, the accuracy is comparable with other ML techniques even with low amounts of training (Table 3). Our novel threshold-based method of detecting pump-level failures is able to achieve relatively high precision of 89% and recall of 100% when the classifier is trained on all data segments available from other pumps (Table 4). Note that in a fault monitoring application with low failure rates, the operational cost of implementing the monitoring workflow is largely determined by the false alarm rates (100 – precision). Pump-level failure prediction has lower precision and recall (both 78%, Table 5). The pre-processing procedure and the shapelets algorithm have a few parameters that need to be determined before it can be applied for classification. In our case, a segment length of 1 week, and a lookback period of 4 weeks was found to give the best results. The optimal parameters may have to be recomputed for a different dataset.

Table 5: Results of predicting failures at the pump level

Pump number	Will fail/ Normal?	#test segments	# predicted to fail segments	Segment-level accuracy (%)	Prediction label after threshold of 20%
1	F	2	2	100	F
2	F	8	0	0	N
3	F	9	2	22	F
4	F	12	7	58	F
5	F	12	11	92	F
6	F	5	5	100	F
7	F	2	2	100	F
8	F	12	12	100	F
9	F	12	1	8	N
10	N	5	1	80	N
11	N	5	2	60	N
14	N	8	2	75	N
15	N	8	2	75	N
16	N	8	7	13	F
17	N	6	1	83	N
18	N	10	5	50	N
19	N	9	9	0	F
20	N	6	1	83	N

6. Conclusions

Shapelets-based time series classification is attractive for oil and gas applications due to their fast classification time and the possibility of interpreting the underlying shapelets by domain experts. We adapted the shapelets approach for detecting and predicting ESP failures only from readily available sensor data. This required developing an appropriate pre-processing procedure. In our experiments to evaluate the accuracy of this method, we found that the accuracy is comparable to other machine learning-based classifiers even with relatively low amounts of training data. The method is able to detect failures only from intake pressure measurements with a precision of 89% and 100% recall. The accuracy of predicting failures is lower with a precision of 78% at 78% recall. This approach has a few customizable parameters that may have to be adapted to different datasets. In future work, multiple measurement attributes (intake pressure, current, and voltage) can be considered together in the shapelets-based classifier to improve the detection and prediction accuracy.

7. Acknowledgments

This work is supported by Chevron U.S.A. Inc. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

8. References

[1] B. Burda, J. Crompton, H. M. Sardoff, and J. Falconer, "Information architecture strategy for the digital oil field," in *Digital Energy Conference and Exhibition*, 2007.

[2] J. Crompton and C. G. Upstream, "Putting the FOCUS on Data," in *Keynote Address at the 2008 W3C Workshop on Semantic Web in Oil & Gas Industry, Houston, Texas*, 2008.

[3] M. R. Brule, "Big Data in Exploration and Production: Real-Time Adaptive Analytics and Data-Flow Architecture," in *SPE Digital Energy Conference*, 2013.

[4] A. Abou-Sayed, "Data mining applications in the oil and gas industry," *Journal of Petroleum Technology,* vol. 64, pp. 88-95, 2012.

[5] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 947-956.

[6] O. P. Patri, V. Sorathia, and V. K. Prasanna, "Event-driven information integration for the digital oilfield," in *SPE Annual Technical Conference and Exhibition*, 2012.

[7] S. Sawaryn, K. Grames, and O. Whelehan, "The Analysis and Prediction of Electric Submersible Pump Failures in the Milne Point Field Alaska," *SPE production & facilities,* vol. 17, pp. 53-61, 2002.

[8] *Electric submersible pump – Reliability information and failure tracking system Joint Industry Project.* Available: http://jip.esprifts.com

[9] F. Alhanati, S. Solanki, and T. Zahacy, "ESP Failures: Can We Talk the Same Language?," in *Gulf Coast Section ESP Workshop held in Houston, TX*, 2001.

[10] J. Presley. (2013) ESP for ESPs? *Exploration & Production*.

[11] S. Sawaryn and E. Ziegel, "Statistical assessment and management of uncertainty in the number of electrical submersible pump failures in a field," *SPE production & facilities,* vol. 18, pp. 182-191, 2003.

[12] S. Sawaryn, "The dynamics of electrical-submersible-pump populations and the implication for dual-ESP systems," *SPE production & facilities,* vol. 18, pp. 236-246, 2003.

[13] T. Kalu-Ulu, J. Andrawus, and I. George, "Modelling System Failures of Electric Submersible Pumps in Sand Producing Wells," in *Nigeria Annual International Conference and Exhibition*, 2011.

[14] Y. Liu, K.-T. Yao, S. Liu, C. S. Raghavendra, T. L. Lenz, L. Olabinjo, F. B. Seren, S. Seddighrad, and C. Dinesh Babu, "Failure Prediction for Artificial Lift Systems," in *SPE Western Regional Meeting*, 2010.

[15] S. Liu, C. S. Raghavendra, Y. Liu, K.-T. Yao, O. Balogun, L. Olabinjo, R. Soma, J. Ivanhoe, B. Smith, and B. B. Seren, "Automatic Early Fault Detection for Rod Pump Systems," in *SPE Annual Technical Conference and Exhibition*, 2011.

[16] O. P. Patri, A. Panangadan, C. Chelmis, and V. K. Prasanna, "Extracting discriminative features for event-based electricity disaggregation," presented at the IEEE Conference on Technologies for Sustainability, Portland, Oregon, USA, 2014.

[17] O. P. Patri, V. S. Sorathia, A. V. Panangadan, and V. K. Prasanna, "The process-oriented event model (PoEM): a conceptual model for industrial events," in *Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems*, 2014, pp. 154-165.

[18] A. Mueen, E. Keogh, and N. Young, "Logical-shapelets: an expressive primitive for time series classification," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 1154-1162.

[19] Z. Xing, J. Pei, S. Y. Philip, and K. Wang, "Extracting Interpretable Features for Early Classification on Time Series," in *SDM*, 2011, pp. 247-258.

[20] T. Rakthanmanon and E. Keogh, "Fast shapelets: A scalable algorithm for discovering time series shapelets," in *Proceedings of the thirteenth SIAM conference on data mining (SDM)*, 2013.

[21] J. Zakaria, A. Mueen, and E. Keogh, "Clustering Time Series Using Unsupervised-Shapelets," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, 2012, pp. 785-794.

[22] Y. Engel and O. Etzion, "Towards proactive event-driven computing," in *Proceedings of the 5th ACM international conference on Distributed event-based system*, 2011, pp. 125-136.