A Review on Anomaly Detection Methods for Optimizing Oil Well Surveillance

Mohd Hilmi Hasan, Azlinda Abdul Malik and Mazuin Jasamai

Universiti Teknologi PETRONAS, Seri Iskandar, Perak, MALAYSIA

Summary
Current economic realities has pushed oil and gas company to produce "more with less". The ever increasing amount of data available to surveillance engineers has caused engineers to spend more time gathering, analyzing them manually which is definitely a daunting exercise and inefficient. Leveraging on data driven surveillance by adopting the principle of management by exception (MBE), the project tries to minimize the manual interaction between data and engineers. The study will focus on monitoring the well production performance through pre-determined parameters with each set of rules. A model (with a certain algorithm) will be built to identify any deviations from the pre-set rules and the model will alert user for deviations that occur. Prediction will be done on when the well be offline if the problem keep on persisting without immediate action from user. The primary benefit of the study is it will allow for proactive measure, faster response time for well intervention, minimize well downtime, safeguard the production as well as contribute to cost saving. Other benefits include better use of practitioner's time (focus on analysis rather than identification), elimination of repetitive data gathering and reformatting tasks, consistency and repeatability of evaluation and better knowledge management. To embark on this study, this paper intends to review previous works related to anomaly detection. The aims are to identify and discuss the characteristics of the available approaches and techniques with respect to adopting MBE in oil well surveillance, and to discover the performance of the techniques so that they can be used to deliver the anomaly detection in surveillance.

Key words:
Well surveillance, anomaly detection, outlier detection, oil production prediction.

1. Introduction

Surveillance has been around the oil and gas industry for many years. The activities involved in surveillance are not only monitoring of behavior, activities or other changing information. Instead, from business point of view it is the aggregation of multiple business practices configured for a particular asset to support and facilitate user decisions needed to manage an asset to its potential in a safe, responsible manner. Whilst surveillance from point of view of tools and processes; the system of observation, visualization, analytics, expert systems and process automation that can provide user with the ability to understand, anticipate and act to improve oil and gas asset performance [1].

At one time, and might still be the current practice in a certain company, production report production target, downtime and operating expenses were all it need for asset surveillance. Data was gathered manually. According to Unneland and Hauser, professionals spend 60-80% of their time finding and preparing data instead of focusing on improving the quality of the decisions [2]. Wells were reviewed sequentially throughout the field on a calendar basis. This was time consuming process with a potential of considerable time lag between problem occurrence and diagnosis. However, engineers do not deal with one or two data sets. One surveillance engineer maybe in-charge to a field which consists of more than hundreds of wells. Monitoring on these wells are done on both surface and subsurface points. With that amount of data available to surveillance engineers, it eventually creating a situation where the engineers spend more time manually to go through the data rather than concentrating on the highest value tasks. In consequence, many opportunities for better performance and risk reduction maybe missed or bypassed [3].

Major improvements in surveillance has been in existence. Some fields have implemented sophisticated monitoring centers that feed data into real-time displays, enabling operation staffs to see the status of key measurements. Moreover, a model-based, integrated workflow has been deployed to provide automation which facilitates the operational excellence [4]. These improvements have changed the landscape of surveillance; as it is no longer a passive, time consuming information delivery system. Instead, by leveraging on real-time information, surveillance has been taken to a new level with the advancement of analytics, expert systems, and process automation because it combines business or operational intelligence with automated technical calculations. These advancements have led to a new generation of hybrid solutions, which integrates elements of data-driven methods including management by exception (MBE), business intelligence (BI), and situational awareness (SA) [1].

Manuscript received November 5, 2017
Manuscript revised November 20, 2017

Our research intends to adopt MBE for monitoring well health status, thus giving heads-up to engineers before it is too late for remedial action. MBE is proposed because it is the most fundamental and widely applied control and optimization technique. MBE identifies significant events, which are normally known as anomaly, that deviate from a standard. By practice, surveillance provides actual measurements or values, which are then compared with a set of target value (performance) that are anticipated a priori. Based on this comparison, an engineer may obtain a list of events that fall outside accepted set-point boundaries. The practice frees up considerable time by removing the burden of gathering data and comparing values manually [5]. Overall, MBE is done to monitor the oil wells based on data-driven well rate estimates [6]. However, MBE does not explain and will not be able to explain the why a problem has occurred or the significance of the problem, which is why technical judgment is still needed by the engineers for them to act towards the problem.

By principal, a production well will have a certain production trend. For example, an oil well with a certain production, water production and gas rates; the trend for another 1-3 months will not deviate much from these figures unless something has happened at downhole or at reservoir level. Sudden increase in water or gas production will hinder the production of oil. Too much deviation, will cause the well to be closed (shut-in) for investigation and remedial action so that it can be back on-line.

Because of multiple workloads and large surveillance data that engineers need to deal with, the production anomalies of the wells are usually overlooked. Current situation is that the anomalies on the well behavior are not detected earlier until the problem peaks up and the well need to be shut-in causing in production loss. This adds up to the time taken for engineers to understand the problem and come out with remedial plan. By adopting the MBE method, any anomalies can be detected earlier, allowing for timely preventive and corrective actions, minimizing the production lost, protecting the hydrocarbon reserves and potential cost saving to the company.

The objective of this paper is to review the previous works related to anomaly detection. The focuses of the review are to identify and discuss the characteristics of the available approaches and techniques with respect to adopting MBE in oil well surveillance, and to discover the performance of the techniques so that they can be used to deliver the anomaly detection in surveillance. This paper is organized as follows; section 2 contains review and discussion on the type of anomaly detection approaches and techniques, section 3 contains review on the selected type of algorithms, and section 4 discusses and concludes the findings and outlines the future works.

2. Categories of Anomaly Detection Technique

One of the main areas that utilizes anomaly detection vastly is network intrusion detection system. Lazarevic et al. generally categorized data mining in intrusion detection techniques into two categories; supervised and unsupervised approaches [7]. In supervised approach, each data point in a data set is labeled, for example as “normal” or “intrusive”. This labeled data is used to train a learning algorithm. A main advantage of supervised detection approach is its high accuracy in detecting known intrusion attacks. Its main drawback is the inability to detect new types of attacks which have not been observed before. In unsupervised approach, anomaly detection is carried out using normal data. This means that it detects intrusions based on deviations from the normal model in data. An advantage of unsupervised approach is that it has the ability to detect intrusion attacks with new signatures that deviate from normal conditions. However, its main drawback is that it suffers from a high degree of false alarm. This is mainly due to the fact that unseen legitimate signatures may be treated as anomalies, and hence categorized as potential intrusions.

Chandola et al. defined anomaly detection as outlier detection, which refers to the problem to find patterns in data that deviate from expected behavior [8]. In the work, they emphasized that the main criteria for denoting whether or not a single data point is normal or anomalous is labels. Obtaining labeled data that is accurate and representing all types of behavior is a daunting task. In normal circumstances, getting anomalous labeled data points which cover all types of anomalous behavior is more difficult that getting labels for normal behaviors. On top of that, anomalous behaviors are normally dynamic, which means that new and unknown types of anomalies may arise, for which there is no labeled data. Overall, labeling requires enormous efforts and is normally done by human experts. In this work, they categorized anomaly detection approach into three categories; supervised, semi-supervised and unsupervised [8,9]. Supervised approach is similar to predictive model which compares normal against anomaly classes. As discussed earlier, supervised approach highly depends on the availability of normal and anomaly training data sets. This has led to a major concern with supervised approach; anomalous data is normally far less than normal data in the training data set. This issue is known as imbalanced class distribution, which can lead to deceivingly high prediction accuracy. Another concern with supervised approach is that obtaining accurate and representative labels, particularly for anomaly class is challenging. The second category of approach, semi-supervised, is an approach where data points are labeled

for only the normal class. Because of this characteristic, semi-supervised approach is more applicable than supervised approach as it is not easy to collect labeled anomalous data. The third type of approach, unsupervised, does not require training data, hence is most widely applicable. The approach makes an assumption that normal data points are far more frequent to be existed in test data than that of anomalous data points. However, this characteristic leads to the drawback if the assumption made is incorrect, hence producing false alarm. One of the ways to solve this drawback is by adapting semi-supervised approach to the unsupervised approach. The unlabeled data set will be used as training data so that the model will learn to become more robust in detecting anomalies.

Each of the three approaches has its own anomaly detection techniques. Patcha and Park categorizes the available techniques into three categories, namely statistical-based, data mining based and machine learning-based techniques [10]. Table 1 lists down some of the methods related to the three techniques [10-12].

One of the advantages of statistical technique is that it does not require prior knowledge of anomalies, hence it has the ability to detect new anomalies. However, the drawbacks of statistical technique are that it is prone to be hacked by skilled attackers, and it is also difficult to determine the threshold of false positive and negative. Furthermore, statistical technique requires accurate distribution to model its methods. However, not all behaviors can accurately be modeled, hence its performance may be reduced. [10,11]. On the other hand, machine learning based approach focuses on developing a model with improved performance based on previous results. It is different from statistical approach that tends to focus on understanding the data generation process. Machine learning approach also offers high flexibility and adaptability [11]. The drawback of machine learning approach is that it is resource expensive. In order to produce better performance in anomaly detection, most machine learning techniques require substantial amount of computational resources. Furthermore, machine learning-based techniques also have high dependency on the assumption on the behaviors of the systems. As for knowledge-based approach, its advantage is robust, flexible and scalable. In some of its techniques, the number of false positive can be kept at low level because it is not based on previous results, instead human experts are used as the inputs to the models. However, the process of getting high quality knowledge is normally time consuming.

Table 1: Anomaly detection techniques

Category	***Method***
Statistical based	Univariate
	Multivariate
	Time series model
Knowledge based/ data mining	Finite state machine
	Description languages
	Expert systems
Machine learning based	Bayesian networks
	Markov models
	Neural networks
	Fuzzy logic
	Genetic algorithm
	Clustering and outlier detection
	Pattern classification
	Single classifiers
	K-nearest neighbor
	Support vector machine
	Self-organizing maps
	Decision trees

2.1 Discussion

Commonly, it is difficult to generate labeled data for oil well surveillance as in supervised approach. The definition of an anomalous state might not be available as it may not have been occurring before since the well is still in operation. Apart from that, as mentioned in earlier, our study focuses on adopting MBE so that it could help engineers in reducing time to spend on monitoring well production. Hence, our study will only be focusing on unsupervised and semi-supervised approaches in detecting anomalies in surveillance data.

With respect to the technique, we will be adopting machine learning-based technique. This is because our intention is to develop predictive model for oil well surveillance based on data. The scope of our study will not cover statistical technique as it requires accurate distribution model in order to generate effective performance. We will also not consider knowledge-based technique as our focus is for the proposed model to carry out anomalies detection works based on data-driven approach.

3. Machine Learning Methods for Anomaly Detection

Machine learning techniques can either be supervised or unsupervised. Supervised machine learning techniques

have the characteristics, advantages and disadvantages of a supervised approach as discussed in the previous section. Among the most commonly used methods of supervised machine learning are supervised neural networks, support vector machine (SVM), k-nearest neighbors, Bayesian networks and decision trees. As discussed in the last section, our study will be focusing on unsupervised machine learning methods. The most commonly used unsupervised machine learning methods are self-organizing maps (SOM), K-Means, C-Means, expectation-maximization meta (EM), and one-class SVM.

Omar et al. presented a review on the pros and cons of unsupervised anomaly detection methods [13]. This is presented in Table 2.

Table 2: Pros and Cons of Unsupervised Machine Learning Methods [13]

Method	***Pros***	***Cons***
Support Vector Machine	• Find the optimal separation hyper plane. • Can deal with very high dimensional data. • Some kernels have infinite Vapnik-Chervonenkis dimension, which means that they can learn very elaborate concepts. • Usually work very well.	• Require both positive and negative examples. • Need to select a good kernel function. • Require lots of memory and CPU time. • There are some numerical stability problems in solving the constraint QP.
Self-organizing maps	• Simple and easy-to-understand algorithm that works. • A topological clustering unsupervised algorithm that works with nonlinear data set. • The excellent capability to visualize high-dimensional data onto 1 or 2 dimensional space makes it unique especially for dimensionality reduction.	• Time consuming algorithm
K-Means	• Low complexity	• Necessity of specifying k. • Sensitive to noise and outlier data points. • Clusters are sensitive to initial assignment of centroids.
C-Means	• Allows a data point to be in multiple clusters. • A more natural representation of the behavior of genes.	• Need to define c, the clusters number. • Need to determine membership cutoff value. • Clusters are sensitive to initial assignment of centroids.
Expectation-Maximization Meta	• Can easily change the model to adapt to a different distribution of data sets. • Parameters number does not increase with the training data increasing.	• Slow convergence in some cases

Singh (2015) in his paper presented the comparison between K-Means and EM methods in network traffic classification problem [14]. The results showed that K-Means performed better than EM in terms of accuracy.

With these comparisons in terms of pros and cons of the machine learning methods, we can conclude that each method has its advantages and drawbacks. Hence, in our study, we will develop the proposed predictive model for oil well surveillance and compare its performance when deployed with different machine learning techniques.

4. Conclusion

This paper presents a review on the approaches and techniques of anomaly detection; with an aim to apply it in MBE of oil well surveillance. It is found that unsupervised or semi-supervised approaches of machine learning technique are appropriate for the said problem. This paper also presents a review on the performance of unsupervised machine learning methods for anomaly detection. There are very limited sources of previous works that we could find in the literature that compared all of the methods. Furthermore, most of the previous works focused on applying machine learning methods for anomaly detection in the area of network intrusion. Finally, it is found that each of the methods has advantages and disadvantages, which requires further experimental works if they are applied in the problem of oil well surveillance prediction. Our future work will be focusing on these identified gaps; applying unsupervised machine learning methods for oil well surveillance prediction model and compare the model's implementation with different kind of methods.

References

[1] Mark J. Lochmann, 2012. The Future of Surveillance. SPE paper 150071 presented at The Intelligent Energy Conference held in Utrecht, The Netherlands, 27-29 March

[2] Schipperijin, P., Thavarajah, R., Simonato, A., 2009. Automated, "By Exception" Well Surveillance: A key to maximizing oil production. SPE paper 123145 presented at SPE Digital Energy International Conference and Exhibition held in Houston, Texas USA, 7-8 April.

[3] Nikolaou, M., Cullick, A.S., and Saputelli, L., 2006. Production Optimization – A Moving-Horizon Approach. Paper SPE 99358 presented at the SPE Intelligent Energy Conference, Amsterdam, The Netherlands, 11-13 April.

[4] Sankaran, S., Olise, M., Meinert, D., and Awasthi, A. 2011. Realizing Value from Implementing i-field TM in Agbami – A Deppwater Greenfield in an Offshore Nigeria Development. SPE Economics and Management 31-44.

[5] Jorge Yero, Thomas A. Moroney. 2010. Exception Based Surveillance in Shell. SPE Paper 128860 presented at Intelligent Energy Conference and Exhibition held in Utrecht, The Netherlands, 23-25 March.

[6] Agrawal, A., Leeuwen, P.V., Hajj, and Demirbas, B. 2016. "Surface Management of Gas Breakthrough in Thin Oil Rim Waxy Reservoir". SPE Paper 181077 presented at the Intelligent Energy International Conference and Exhibition held in Aberdeen, United Kingdom, 6-8 September.

[7] Agrawal, A., Lie, J., Leeuwen, P.V., Adun, D., Briers, J., Pei-Huey, Sniekers, A., and Enns R. 2016, September 6. Leveraging Data-driven Exception Based Surveillance to Maximize Returns in Times of Low Oil Prices. SPE Paper 181080 presented at the Intelligent Energy International Conference and Exhibition held in Aberdeen, United Kingdom, 6-8 September.

[8] A.Lazarevic, L. Ertoz, V. Kumar, A. Ozgur and J. Srivastav, "A comparative study of anomaly detection schemes in network intrusion detection," SIAM Conference on Data Mining, 2003.

[9] R. Kaur and S. Singh, "A survey of data mining and social network analysis based anomaly detection techniques," Egyption Informatics Journal, 17, pp. 199-216, 2016.

[10] A. Patcha and J.M. Park, " An Overview of anomaly detection techniques: existing solutions and latest technological trends," Computer Networks, 51, pp. 3448-3470, 2007.

[11] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maria-Fernandez and E. Vazquez, "Anomaly-based network intrusion detection: techniques, systems and challenges," Computer and Security, 28, pp. 18-28, 2009.

[12] C-F. Tsai, Y-F. Hsu, C-Y. Lin and W-Y. Lin, "Intrusion detection by machine learning: a review," Expert Systems with Applications, 36, pp. 11994-12000, 2009.

[13] S. Omar, A. Ngadi and H.H. Jebur, "Machine learning techniques for anomaly detection: an overview," International Journal of Computer Applications, 79 (2), 2013.

[14] H. Singh, "Performance analysis of unsupervised machine learning techniques for network traffic classification," 5th International Conference on Advanced Computing and Communication technologies, 2015.