A Game-Theoretic Approach to Improve Energy-Related Data

Salem Al-Gharbi, Abdulaziz Al-Majed, Mohamed Mahmoud,* and Abdulazeez Abdulraheem

ABSTRACT: With the increase in the energy demand, the magnitude of energy production operation increased in scale and complexity and went too far in remote areas. To manage such a big fleet, sensors were installed to send real-time data to operation centers, where subject matter experts monitor the operations and provide live support. With the expansion of installed sensors and the number of monitored operations, the operation centers were flooded with a massive amount of data beyond human capability to handle. As a result, it became essential to capitalize on the artificial intelligence (AI) capability. Unfortunately, due to the nature of operations, the data quality is an issue limiting the impact of AI in such operations. Multiple approaches were proposed, but they require lot of time and cannot be upscaled to support active real-time data streaming. This paper presents a method to improve the quality of energy-related (drilling) real-time data, such as hook load (HL), rate of penetration (ROP), revolution per minute (RPM), and others. The method is based on a game-theoretic approach, and when applied on the HL—one of the most challenging drilling parameters—it achieved a root mean square error (RMSE) of 3.3 accuracy level compared to the drilling data quality improvement subject matter expert’s (SME) level. This method took few minutes to improve the drilling data quality compared to weeks in the traditional manual/semiautomated methods. This paper addresses the energy data quality issue, which is one of the biggest bottlenecks toward upscaling AI technology into active operations. To the authors’ knowledge, this paper is the first attempt to employ the game-theoretic approach in the drilling data improvement process, which facilitates greater integration between AI models and the energy live data streaming, also setting the stage for more research in this challenging AI-data domain.

1. INTRODUCTION

With the increase in the energy demand, the magnitude of energy production operation expanded in scale and complexity and went too far in remote areas, leading to challenging energy production operations that could exceed the standard operation crew skills. To address this, the energy industry started equipping their operations with sensors that gauge the operations and send the data in real time to operation centers equipped with groups of subject matter experts. At the beginning, the results were promising. The subject matter experts monitored and analyzed the data and then provided live recommendations to optimize the operations and to address any challenges. Such exercise was very helpful to the operation crew, but with the expansion of the number of installed sensors, the amount of transmitted data became big, overwhelming the operation centers with a massive amount of data beyond the human capability to process—even the subject matter experts! As a result, it became essential to capitalize on the computation power and the artificial intelligence (AI) technologies.

Drilling is one of these energy-related operations that produce a massive amount of data. As a result, multiple drilling-related AI models were developed such as stuck pipe detection,1 BHA monitoring,2 tool wear and flank wear analysis,3,4 rate of penetration (ROP) prediction,5 formation top prediction,6 and much more. But due to the operations’ tough environment, a substantial amount of these data sets was not accurate.7 As a result, developed AI models could not be fully upscaled the real-time operations;8 the bottleneck was the data quality. The low-quality data points negatively affected the AI models’ performance, resulting in bad analysis and poor recommendations. Hence, having a mechanism to improve the data quality in real time is a must to integrate AI power in operations. If done correctly, it will have a direct impact in improving the performance, leading to safer, faster, and cost-effective operations.

The method proposed in this paper addresses—with high accuracy—this data-quality issue and could be applied in other domains as well. Currently, there are different practices to improving the drilling data, which can be classified under four categories: logistic, rig site, administrative, and methods. The logistic practice approaches propose to develop new standards and
improve the operation collaboration. The rig site practices propose to add/develop more sensors and to standardize the transmission and text abbreviation. The administrative approaches propose using key performance indicators (KPIs) to evaluate the quality of data and to develop a data quality dashboard, real-time data quality center, and others. The method approaches apply a manual cleansing process, remove bias, remove outliers, correlate the data together, smooth the data set, and others.

Table 1. Statistics of Raw Data Set 1 and Data Set 2

| data set    | number of data points | bit depth | well depth | hook load |
|-------------|-----------------------|-----------|------------|-----------|
| data set 1  | 153,000               | 0 ft      | 1630 ft    | 20 klb    |
| data set 2  | 124,000               | 0 ft      | 4000 ft    | 20 klb    |

Figure 1. Raw data set. (a) Data set 1 and (b) data set 2.

The first three approaches—logistic, rig site, and administrative—are mainly futuristic solutions; also, they require a lot of time and efforts to be applied. In addition to that, these approaches propose collecting a new data set, neglecting the massive amount of data already collected. The “method” approaches highlighted earlier were the most effective approaches, especially if combining manual and automated processes. Despite their good results, their major limitation was they require a high domain knowledge in the subject data to improve the data set, in addition to a lot of...
time and efforts. Such limitations prevent them from being used in active operations.

The proposed approach in this paper addresses the limitation of the currently used practice. It could improve the data in a much faster manner and has a wider scope covering a bigger number of data sets. Actually, it can be used in active operations, setting the stage for AI models to be integrated directly with energy operations.

2. METHODOLOGY

The work starts by identifying the current practice to improve the real-time energy-related data points, which consists mainly of three major phases. Then a game-theoretic approach is used to mimic these phases, and a mechanism is set to evaluate this game-theoretic approach to ensure that it achieves a good level of matching the existing current practice.

Before moving to the next section, let us shed light on the used data sets. The data sets used in this paper are from actual operations in the Middle East and North Africa (MENA) region covering shallow and deep oil horizontal well drilling operations; both data sets are from the same field. Table 1 sheds light on the statistical characteristics of the used data sets. Figure 1 presents the shape of raw data set 1 and raw data set 2.

3. CURRENT PRACTICES

This section lists the processes conducted by data experts transforming a raw data set to a high-quality data set. The experts’ semi-automated process consists of three major phases:

1. Classify the operation
2. Remove low-density data points
3. Identify the data points’ trend

Figure 2. The drilling-activity data points for data set 1 highlighting the trend and the low-density groups (based on drilling data experts).

Figure 3. Data set 1 after being improved by the data experts.

Table 1
3.1. Classify the Operation. A drilling operation consists of several activities, such as run-in-hole (RIH), pull-out-of-hole (POOH), cementing, and others. The experts start by identifying which activity-related data set will be improved. This paper focuses on the drilling-related data set, which represents activities where the well depth is increased. This represents one of the most challenging activities in the drilling operation. The data set related to this activity can be identified by comparing the bit depth to the well depth. If it is increased, then it belongs to an actual drilling activity. Equations 1 and 2 are a mathematical representation of this classification. It is worth to highlight that this process includes removing invalid data points such as null and out-of-normal-operation values.

\[
\forall x \in \text{drilling data set } \\
\begin{align*}
    f(x_i) &= \begin{cases} 
    x_i \in \text{drilling activity, if } x_i \text{ well depth } < x_i^{(i+1)} \text{ well depth} \\
    x_i \notin \text{drilling activity, if } x_i \text{ well depth } = x_i^{(i+1)} \text{ well depth}
    \end{cases}
\end{align*}
\]

3.2. Remove Low-Density Data Points. The real-time drilling operations generate a massive amount of data in a fast frequency manner; as a result, the trend appears as high dense flow of data points. But due to the harsh environment and the heavy moving equipment during drilling, sensors could record/transmit faulty signals; such signals present out-of-norm data points. Experts could identify them since they appear as a low-density data point and remove them during the data improvement process. Figure 2 identifies groups of low-density data points that need to be removed.

3.3. Identify the Data Points’ Trend. The data points follow specific trends. SMEs use their experience and understanding of the operation to identify the data points’ true trend and then remove data points that deviated from it.

Figure 2 presents the drilling-phase data points, highlighting the data flow trend. It is clear that this process needs subject matter experts and requires a lot of time, which makes the cleansing process a challenging task. Figure 3 presents the improved data set using an expert manual process.

The accuracy result of the current practice is high, but it requires a good amount of time and effort by data experts to be conducted. As a result, this effort is limited to a small portion of the data set and defiantly could not be conducted in a live operation manner, which is one of the major challenges of this approach. The main objective of this paper’s approach is to overcome this challenge.

4. THE PROPOSED METHOD

The objective of this paper is to automate the data quality improvement process based on the game-theoretic approach. This method puts an AI agent in direct competition against a drilling-related data expert; the result is a model that can automatically improve a massive amount of data accurately in a short time. It is worth to highlight that the game-theoretic approach has been applied in the energy sector, for example, in the energy consumption planning,\textsuperscript{19} in trading and pricing,\textsuperscript{20,21} and in pipeline monitoring and security.\textsuperscript{22} Also, the game-theoretic approach was used toward targeting the optimal energy storage.\textsuperscript{23} All of the aforementioned examples capitalized on a big amount of data that required a lot of
time and efforts to reach a sufficient accuracy. This paper—to the authors’ knowledge—is the first to utilize the game-theoretic approach to automatically improve the drilling-data quality, which sets the stage for a new application in this domain.

Since the method is mimicking the experts’ process, it consists of the same three stages:

1. Classify the operation
2. Remove low-density data points
3. Develop the data points’ trend

4.1. Classify the Operation. This stage is based on the rule-of-thumb process, so it will be exactly matching the expert’s manual method described earlier in Section 3.1.

4.2. Remove Low-Density Data Points. The goal of this step is to segregate low- from high-density data points. The heat-map or 3D histogram technique is a good technique fitting this purpose.9

There are two parameters to adjust in this case:

- Block size: the size that will be used to group the data points. Figure 4 presents the data set using a 3D histogram, indicating low- and high-density data points.
- The density threshold: the value that will be used to identify low- from high-density data points.

Note that the higher the block size is, the more data points will be grouped together, and the more the faulty data points will be retained. If the block size is smaller than needed, then a valid data point will be considered as noise and removed. The same is true for the density threshold. If this threshold is set to a low value, it will keep faulty data points, but if it is set higher than the true value, it will eliminate valid data points. As a result, it is essential to select the proper values, and these will be the first parameters the AI agent will manipulate and play with to improve the data quality.

4.3. Develop the Data Points’ Trend. While data experts identify the data points’ trend based on their experience, in this method, the AI agent tries to do the same by experimenting, running a combination of statistical methods (i.e., moving average and running medians) against the data set. The AI agent needs to figure the best sequence of statistical methods leading to the human experts’ result. Note that each statistical

---

Table 2. The Method’s Action Space

| stage | one | two |
|-------|-----|-----|
| scenarios | apply equation | block size | density threshold | statistical methods | average number of internal parameters | number of iterations | # of total space size |
| # of selections | 1 | 15 | 5 | 2 | 100 | 5 | $1 \times 75 \times 100 = 75,000$ |
| # of total scenarios | 1 | $15 \times 5 = 75$ | $2 \times 100 \times 5 = 1000$ |  |  |  |  |

---

Figure 6. Pac-Man game “action space” process.

Figure 7. The method’s “action space” stages.
method has its internal configuration; for example, the moving average has the “backward window length”, and the running medians method has the “median window”. As a result, the AI agent is moving in two dimensions, the statistical method and its internal configuration, to achieve the goal. Figure 5 presents the method process.

One of the basic game-theoretic approaches is the game action space (up, down, left, and right). The agent moves within this space until it achieves the end state. It is worth to highlight that the “Pac-Man” game uses this method (in the game, it eats all the dots or lose if attacked by the ghost). Figure 6 presents this process.

In this paper, the game action space is much bigger. Stage one is straightforward with one equation to apply. Stage two consists of two, block size and the density threshold, and each one has its internal space to select from. Stage three consists of n iterations involving multiple statistical methods to identify and select the internal parameters. While the game action space size of the “Pac-Man” game is 4, the game action space of the method in the paper is of size 75,000. Table 2 sheds light on this method’s game action space, and Figure 7 presents the method’s “action space” stages.

### 5. EVALUATION CRITERIA

The evaluation criteria are based on how close is the method model in improving the data quality compared to the expert’s results. This is done using root mean square error (RMSE), which measures the error between two numeric vectors, which in this paper are the expert’s results and the method model’s results. The smaller the RMSE is, the closer the results are, which is a good indicator.

In each iteration, RMSE will be calculated until the result is lower than 10, which indicates a close match between the results of the method’s AI agent and the human expert.
6. EXPERIMENT RESULTS

After the method’s AI agent played 20,000 times, it achieved an average RMSR of 70. Only 0.30% of the results were below 10. Figure 8 presents sample RMSE results. It is clear that the results were following an arbitrary trend, jumping from below 20 to over 140 in a random manner.

Figure 9 presents the expert’s result (orange) compared to the model’s result (blue). It is clear that the method model’s results are very close to the expert’s results. Most of the data point trends are matching. There are slight deviations in zones that sharply change; despite these deviations, the two trends show the same global picture. The results are very satisfying.

7. APPLICATION TO THE NEW DATA SET

To challenge the method’s model even more, the produced model was evaluated though a new blinded data set from the same field (data set 2). Figure 10 presents the expert’s results compared to the model’s results. Comparing the results of the model and the expert, the model succeeded in eliminating invalid/noisy data points and retaining the valid data points, and the trends were almost matching. Actually, the RMSE value was 3.3, which is very good. Note that our model took less than 2 min to produce this result, which is extremely fast compared to weeks by the SME process.

8. DISCUSSION

The proposed method succeeded to improve data quality using the game-theoretic approach and matched the expert’s results by RMSE of 7.7 and 3.3 for data set 1 and data set 2, respectively. It required 30 h for the model to be developed and less than 2 min to be used on new data sets. This is very fast compared to the current intensive semi-automated work that requires weeks. The proposed method is faster, could cover a wider size of data set, and could be applied in a near real-time manner. Table 3 lists a comparison between the used game-theoretic approach and the currently existing practice.

9. CONCLUSIONS

In conclusion, the proposed approach succeeded in improving the data quality. Its performance significantly exceeded the current practice. The game-theoretic approach achieved good results in a short time, and it could be applied to a wider space of data sets. Actually, it could be used in near-live manner, setting the stage for more AI models to be integrated with active operations in a live manner.

It worth to highlight that this paper’s model mechanism uses the trial-and-error approach to achieve good results. As a result, the accuracy has high fluctuations during the space time. As a future enhancement, this needs to be improved to ensure that the generated accuracy is actually always improving with the space time. Note that this paper is one of the first introducing the game-theoretic approach in automating the process of drilling data quality improvement with great success. This approach could be applied to improve different types of energy data sets as well.

10. AUTHOR INFORMATION

Corresponding Author
Mohamed Mahmoud — Petroleum Engineering Department, College of Petroleum Engineering & Geosciences, King Faisal
The authors would like to thank Dr. Amru Alqurashi for his valuable contribution to this work.

**ACKNOWLEDGMENTS**

The authors declare no competing financial interest.

**REFERENCES**

(1) Alshaikh, Abrar A.; Mohammed K., Albassam; Salem H., Al Gharbi; Abdullah S., Al-Yami. Detection of stuck pipe early signs and the way toward automation. In *Abu Dhabi International Petroleum Exhibition & Conference*. OnePetro, 2018, DOI: 10.2118/192975-MS.

(2) Dasevski, Dubinsky; V., Dubinsky; J. D., Macpherson. “Application of neural networks for predictive control in drilling dynamics.” In *SPE Annual Technical Conference and Exhibition*. OnePetro, 1999, DOI: 10.2118/56442-MS.

(3) Yu, J. Online tool wear prediction in drilling operations using selective artificial neural network ensemble model. *Neural Computing and Applications*. 2011 Jun, 20, 473–485.

(4) Garg, S.; Pal, S. K.; Chakraborty, D. Evaluation of the performance of backpropagation and radial basis function neural networks in predicting the drill flank wear. *Neural Computing and Applications*. 2007 May, 16, 407–417.

(5) Bhutnagar, A.; Khandelwal, M. An intelligent approach to evaluate drilling performance. *Neural Computing and Applications*. 2012, 21, 763–770.

(6) Al-Abduljabbar, Ahmad; Salaheldin, Elkatatony; Mohamed, Mahmoud; Abdulazeed, Abdulrahim “Predicting formation tops while drilling using artificial intelligence.” In *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*. OnePetro, 2018, DOI: 10.2118/192345-MS.

(7) Ashok, Pradeepkumar; Adrian, Ambrus; Dawson, Ramos; John, Lutteringer; Michael, Behounek; Yuhecheng, Li; Yang, Taylor; Thetford; Terrence, Weaver. "A step by step approach to improving data quality in drilling operations: Field trials in north america." In *SPE Intelligent Energy International Conference and Exhibition*. OnePetro, 2016, DOI: 10.2118/187067-MS.

(8) Deshpande, Alisha; Yining, Dong; Gang, Li; Yingying, Zheng; Sizhao, Joe; Qin, Lisa A., Brenskelle. “Advanced Streaming Data Cleansing.” In *SPE Digital Energy Conference and Exhibition*. OnePetro, 2015, DOI: 173415-MS.

(9) Al, Gharbi; Salem, H.; Majed R., Al Zayer. "Automated Real Time Data Cleansing and Summarization; Case Study on Drilling Hook load Real time Data." In *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. OnePetro, 2015, DOI: 10.2118/176755-MS.

(10) Sadlier, Andreas; Duncan, Blue; Thomas, Halland. "Effective operational collaboration delivers substantial real-time service quality improvements." In *IADC/SPE Drilling Conference and Exhibition*. OnePetro, 2016, DOI: 10.2118/178783-MS.

(11) Thorsen, Arve K.; Erland, Saeverhagen; Inge, Svensson. "Enhanced data quality and lowerd risk utilizing remote operations." In *SPE Intelligent Energy Conference & Exhibition*. OnePetro, 2014, DOI: 10.2118/167866-MS.

(12) Arnaout, Arghad; Philipp, Zoellner; N., Johnstone; G., Thonhauser. "Intelligent data quality control of real-time rig data." In *SPE Middle East Intelligent Energy Conference and Exhibition*. OnePetro, 2013, DOI: 10.2118/167437-MS.

(13) Otalvora, William Contreras; Musab, AlKhudirij; Faisal, Alsanie; Binil, Mathew. "A comprehensive approach to measure the realtime data quality using key performance indicators." In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2016, DOI: 10.2118/181315-MS.

(14) Svensson, Inge; Matthew, Wooley; Thomas, Halland. "Improving data quality in witsml data." In *SPE Intelligent Energy International Conference and Exhibition*. OnePetro, 2016, DOI: 10.2118/181038-MS.

(15) Andrianova, Alla; Maxim, Simonov; Dmitry, Perets; Andrey, Margarit; Darya, Serebryakova; Yuriy, Bogdanov; Semen, Budenny; Nikita, Volkov; Artem, Tsenda; Alexander, Bukharev. "Application of machine learning for oilfield data quality improvement." In *SPE Russian Petroleum Technology Conference*. OnePetro, 2018, DOI: 10.2118/191601-18RPTC-MS.

(16) Noshi, Christine; Samuel, Noynaert; Jerome, Schubert. "Data mining approaches for casing failure prediction and prevention." In *International Petroleum Technology Conference*. OnePetro, 2019, DOI: 10.2523/IPTC-19311-MS.

(17) Neufeldt, A. C.; Lai, S. W.; Kristjansson, S. D. An algorithm to automatically zero weight on bit and differential pressure and resulting improvements in data quality. *SPE Drilling & Completion*. 2020 Dec 17, 35, 535–552.

(18) Khudiri, Musab; William, Contreras; Faisal, Sanie; Tony, Jovin; Gary, Hickin; Muhammad, Kashif. "Saudi Aramco Real-Time Drilling Operation Activity Recognition & Analysis Engine." In *SPE Middle East Intelligent Oil and Gas Conference and Exhibition*. OnePetro, 2015, DOI: 10.2118/176799-MS.

(19) Hao, Jun; Ke, Zhang; Xiaoxiao, Dai; Jun Jason, Zhang. "Game-Theoretic Energy Consumption Planning at Distribution Level for Future Smart Grid." In 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), pp. 203–206. IEEE, 2019, DOI: 10.1109/SOLI48380.2019.895054.

(20) Wei, F.; J. Q., Liu; Z. H., Yang; F., Ni. "A game theoretic approach for distributed energy trading in district energy networks." In 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1–6. IEEE, 2017, DOI: 10.1109/EI2.2017.8245568.

(21) Dashi, Zahra Alsadat; Mohammad Mehdi, Lottfi; Mohammadreza, Mazidi. "A Non-Cooperative Game Theoretic Model for Energy Pricing in Smart Distribution Networks Including Multiple Microgrids." In 2019 27th Iranian Conference on Electrical Engineering (ICEE), pp. 808–812. IEEE, 2019, DOI: 10.1109/IranianCEE.2019.8786543.

(22) Islam, Mohammad Saiful; Robert, Nix; Murat, Kantarcigolu. "A game theoretic approach for adversarial pipeline monitoring using Wireless Sensor Networks." In 2012 IEEE 13th International Conference on Information Reuse & Integration (IRI), pp. 37–44. IEEE, 2012, DOI: 10.1109/IRI.2012.6302988.

(23) Scarabaggio, Paolo; Raffaele, Carlì; Mariagrazia, Dotoli. "A game-theoretic control approach for the optimal energy storage under power flow constraints in distribution networks." In 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), pp. 1281–1286. IEEE, 2020, DOI: 10.1109/CASE48305.2020.9216800.