Analysis and Recommendation of Frequent Patterns of Long-Life Pumping Wells Based on Data Mining

Zhong-hui Zhang(✉)

Petroleum Engineering Technology Research Institute of Sinopec Shengli Oilfield Company, Dongying, China
zhangzhonghui.slyt@sinopec.com

Abstract. The theoretical regulation of production parameters in oil production engineering plays a significant role in the management of beam pumps. However, it falls short in identifying the inherent relationships among historical production data, thus failing to address the problem at its core. Valuable information can be extracted from historical well experiences through data mining techniques, offering new insights for adjusting production measures. To achieve this objective, an analysis is conducted to explore the factors and patterns influencing the exemption period of oil wells. Various methods, including expert experience and correlation analysis, are employed to process and selectively identify relevant features. Drawing upon the principles of oil production engineering and leveraging advanced big data processing techniques, these features are encoded to construct a comprehensive sample set that represents long-life wells. Subsequently, association rule mining is applied to uncover frequent patterns exhibited by these long-life wells. By setting a minimum support threshold of 0.01, the mining process encompasses a substantial dataset comprising over 1700 wells, leading to the discovery of more than 100 meaningful association rules. These rules are further prioritized and visualized based on their lift values, providing valuable insights into the experiential knowledge base related to effective measures for long-life well patterns. Consequently, this knowledge base becomes an invaluable asset, offering support for informed decision-making in terms of production parameter control and aiding in the development of scientifically guided production strategies.

J. Lin (Ed.): IFEDC 2023, SSGG, pp. 194–204, 2024.
https://doi.org/10.1007/978-981-97-0272-5_16

Keywords: Association Rules · Data Mining · Beam Pump Wells · Recommendation of Measures

1 Introduction

As the primary production equipment for domestic oil wells, beam pumps play a crucial role in petroleum production. However, issues such as rod parting, pump leakage, rod deviation and wear, and wax deposition significantly reduce the exempt period of beam pumps [1]. Therefore, it is of great significance to explore the intrinsic factors and patterns that affect the exempt period of oil wells, and provide corresponding recommended measures. This can effectively reduce the workload and extend the exempt period, contributing to improved operational efficiency.

The underground structure of oil wells is complex and constantly changing, with strong coupling of production parameters. The maintenance and management of beam pump wells lag behind, and there is an urgent need for production experience-supported information decision-making. Traditional methods for extending the exempt period are often based on empirical knowledge, without extensive utilization of historical real-time data feedback from well production. As a result, it is difficult to track the dynamic information of well production and identify the mixed effects caused by multiple factors. The perspective is limited to single-type problems. For example, in addressing the issue of rod wear, a directional lifting system was designed for a specific well, and no-rod lifting technology was adopted, which fundamentally solved the problem of rod wear [2]. However, this lifting system technology was designed based on the characteristics of a single block, and it has limitations and is difficult to be widely applied.

The advent of the era of big data in the petroleum industry has provided a vast stage for the application of data mining techniques. Accumulating massive historical data during long-term oil well development allows for the use of data mining techniques to process and analyze this data, uncovering valuable insights and experiences embedded within. Association rule mining, as one of the representative techniques in data mining, has made significant advancements in various areas such as disease recognition, drug prediction, and risk prediction of unsafe behaviors [3–5]. In this context, leveraging the mining capabilities of association rule algorithms for intrinsic factors of features promotes the deep integration of petroleum big data mining technology. To achieve this, a long-life well sample library was created based on historical production data from oil wells, combining expert experience with data processing techniques. By applying association rule mining algorithms, a frequent pattern library for long-life wells was constructed, and visual analysis was conducted on these frequent patterns.

The organizational structure of this paper is presented as follows: Sect. 2 introduces the methods for processing and selecting data features, as well as the techniques for association rule mining. Section 3 focuses on the frequent pattern mining of long-life wells in a specific oilfield, analyzing and discussing the frequent patterns and associated rules within these patterns. Section 4 presents the conclusions of the study and provides prospects for future research endeavors.

2 Preparation of the Sample Set for Beam Pump Well Exemption Period

2.1 Construction of Relevant Indicators System for Well Design

In response to the requirements of well design tasks, a comprehensive indicator system tailored for well design was developed by integrating expert knowledge. This indicator system consists of six major categories and includes over 100 parameters. The data sources for each parameter were identified, and a corresponding database was designed. The details of this indicator system are presented in Table 1.

Table 1. Presents the corresponding database for the well design study.

Basic Information	Geological Data	Fluid Data	Mechanical and Production Data	Production Data	Operation Data
Well Number Production Date Oil (Gas) Field Block Unit Unit Name Well Type Lifting Method …	Reservoir Type Exploitation Layer Effective Thickness Saturation Pressure Formation Temperature Layer Porosity Layer Permeability Layer Saturation …	Crude Oil Viscosity Volume Coefficient Water Mineralization Formation Water Type Freezing point Wax Appearance Temperature Wax Content Gas-Oil Ratio …	Structure Data Tubing Data Sucker Rod Data Pump Data Supporting Data Pump Jack Data Wellhead Data Production Parameters …	Daily Water Cut Dynamic Liquid Level Submergence Depth Pump Efficiency Power Consumption System Efficiency Indicator Diagram Oil Pressure …	Pump Testing Period Exemption Period Operation Time Repair Causes Construction Type Failure Point Description …

Combining expert experience, a total of 15 features were selected for the oil well, including reservoir type, sand production, scale deposition, wax deposition, daily fluid production, normal water cut, dynamic liquid level, pump depth, submergence depth, stroke times, pump position wellbore inclination angle, salinity, crude oil viscosity, freezing point, and pump Size.

2.2 Data Integration and Standardization

Based on the operation big data of the pumping unit wells, standardized processing was carried out to address issues such as multiple data sources, varying frequencies, and mixed data types. This included the fusion of multi-source data, integration of data with different frequencies, and digitization of text-based data indicators. As a result, a standardized operation big data set for pumping unit wells was prepared. Refer to Table 2 for details.

Table 2. Operational sample set of beam pumping wells.

Well	Number Reservoir	Category Sand	Depth Pump	Diameter Pump	Efficiency Water	Stroke	Stroke Count
1	Fault Block	No Sanding	2000	44	53	3.05	2.4
2	Medium to High Permeability	Slight Sanding	1999	44	95	3.76	3
3	Medium to High Permeability	Slight Sanding	2003	44	96	3.71	2.2
4	Complex Fault Block	No Sanding	2100	44	95	3.88	2
5	Complex Fault Block	No Sanding	2000	44	68	3.41	2
6	Complex Fault Block	No Sanding	2002	44	21	3.85	2.7
7	Complex Fault Block	Severe Sanding	2007	38	98	2.96	2.5

2.3 Correlation Analysis

When two variables change in some degree as a result of each other's variations, we say they have correlation. Therefore, before data mining, analyzing the correlation between features and removing weakly correlated features can not only reduce workload but also improve model accuracy.

Common methods for correlation analysis include Pearson correlation coefficient, Spearman correlation coefficient, and Kendall correlation coefficient [6–8].

Pearson correlation coefficient formula:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(X_i - \mu_X)E(Y_i - \mu_Y)}{\sigma_X \sigma_Y} \tag{1}$$

where X_i and Y_i represent the values of the i observation, μ_X and μ_Y are the means of variables X and Y respectively, and σ_X and σ_Y are the standard deviations of variables X and Y respectively.

Spearman correlation coefficient formula:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{2}$$

where d_i represents the rank differences between the i variable X_i and Y_i, i.e., X_i - Y_i, and n represents the sample size.

Kendall correlation coefficient formula:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn\left(x_i - x_j\right) sgn\left(y_i - y_j\right) \tag{3}$$

where $sgn(x_i\text{-}x_j)$ and $sgn(y_i\text{-}y_j)$ represent the signs of rank differences between the i and j observations for variables X and Y respectively, and n represents the sample size.

The Pearson correlation coefficient is commonly used for linear correlation analysis, the Kendall correlation coefficient is often used for comparing ordinal correlations, while the Spearman correlation coefficient can reflect both linear and nonlinear relationships between features. Therefore, Spearman correlation coefficient was chosen to analyze the relationship between features and the maintenance period, and the resulting correlation analysis is shown in Fig. 1.

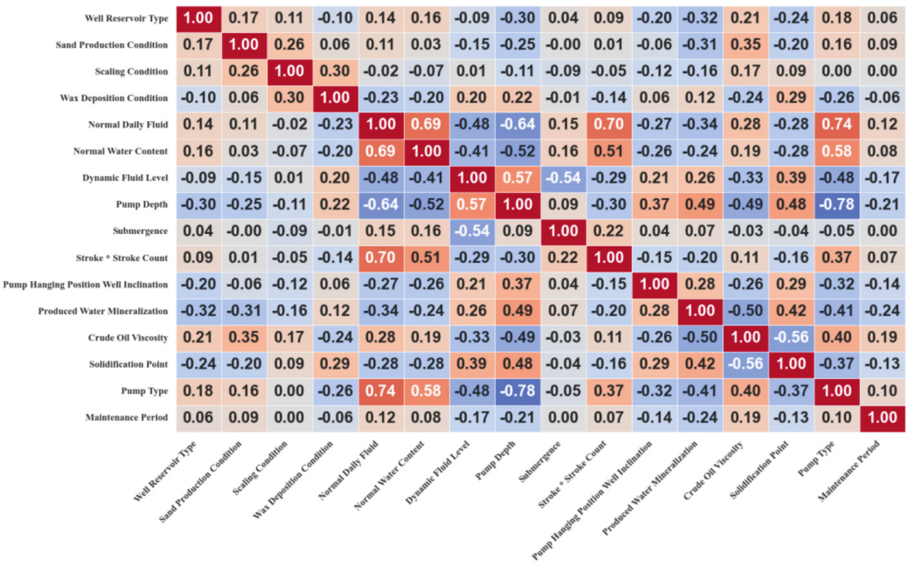

Fig. 1. Spearman correlation analysis graph.

According to the correlation analysis graph in Fig. 1, it can be observed that the features affecting the maintenance period are, in sequence, Salinity, pump depth, and crude oil viscosity. Submergence and scaling condition have insignificant impact on the maintenance period.

3 Measures Recommendation Design for Long-Life Wells Based on Association Rules

3.1 Association Rule Mining Algorithm

Data mining algorithms can be used to discover frequent item-sets and association rules. In order to perform effective clustering, the A-priori algorithm [9–11] is employed for association rule mining. The strength of each rule is evaluated based on indicators such as support, confidence, and lift.

The A-priori algorithm generates frequent item-sets through the process of joining and pruning. The generated frequent item-sets are then used to generate association rules. As shown in Fig. 1, which illustrates the process of association rule mining, the algorithm starts by generating candidate item-sets (C1) based on the item categories in dataset D. Items below a certain threshold are removed, resulting in frequent 1-itemsets (L1). L1 is then combined to form 2-itemsets, generating candidate 2-itemsets (C2). Similarly, items below the threshold are removed to obtain frequent 2-itemsets (L2). This process continues, with L2 being combined to form 3-itemsets and generate candidate 3-itemsets (C3). Again, items below the threshold are removed to obtain frequent 3-itemsets (L3). Finally, the items contained in L3 are permuted and combined to form antecedents and consequents. Support, confidence, and lift are calculated for each rule (A1, A2, A3) based on the relationships between the antecedents and consequents (Fig. 2).

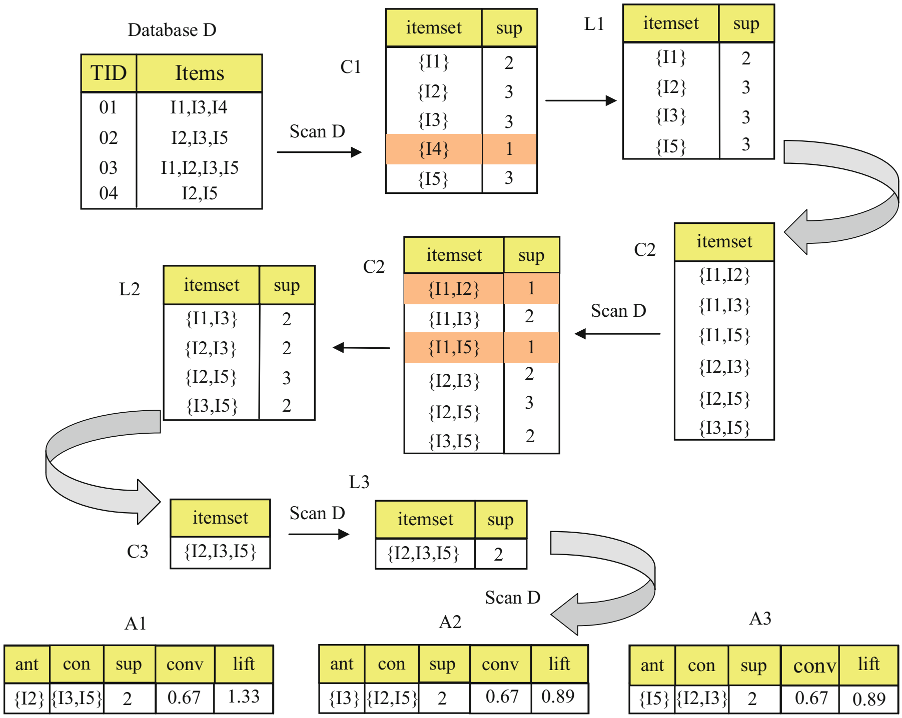

Fig. 2. Illustrates the concept of association rule mining.

3.2 Preparation of Long-Life Well Measures Knowledge Base

The process measures supporting long-life oil wells in each block were statistically analyzed, and an association rule mining algorithm was used to create a process library for long-life wells. By setting the minimum support threshold to 0.01, a total of 104 frequent patterns for long-life wells were obtained, as shown in Table 3.

Table 3. Library of frequent patterns for long-life wells.

rules	frequent item-sets
1	· Stroke * Stroke per Minute [>15] · Wax Deposition Condition [No wax] · Well Reservoir Type [Medium to high permeability] · Water cut [>95] · Pump Size [56 57] · Dynamic Fluid Level [634–889] · Salinity [0–10000] · Angle of Inclination [0–15] · Scaling Condition [Slight scaling] · Crude Oil Viscosity [1000–10000] · Sand Production Condition [Slight sand production] · Submergence [100–300] · Daily Fluid Production [30–80] · Freezing point [null] · Pump Depth [800–1100]
2	· Wax Deposition Condition [No wax] · Stroke * Stroke per Minute [6.5–9.5] · Well Reservoir Type [Medium to high permeability] · Water cut [>95] · Pump Size [56 57] · Daily Fluid Production [10–30] · Dynamic Fluid Level [634–889] · Salinity [0–10000] · Angle of Inclination [0–15] · Scaling Condition [Slight scaling] · Crude Oil Viscosity [1000–10000] · Sand Production Condition [Slight sand production] · Submergence [100–300] · Freezing point [null] · Pump Depth [800–1100]

(*continued*)

Table 3. (*continued*)

rules	frequent item-sets
3	· Wax Deposition Condition [No wax] · Submergence [300–500] · Stroke * Stroke per Minute [6.5–9.5] · Well Reservoir Type [Medium to high permeability] · Water cut [95-] · Pump Size [56 57] · Daily Fluid Production [10–30] · Salinity [0–10000] · Angle of Inclination [0–15] · Scaling Condition [Slight scaling] · Crude Oil Viscosity [1000–10000] · Sand Production Condition [Slight sand production] · Dynamic Fluid Level [306–634] · Freezing point [null] · Pump Depth [800–1100]
…	…

Taking the first frequent pattern as an example, when the frequency of Stroke * Stroke per Minute is greater than 15, Wax Deposition Condition is "no wax deposition", Well Reservoir Type is "medium to high permeability", Water cut is greater than 95%, Pump Size is either 56 or 57, Dynamic Fluid Level is between 634–889 m, Salinity is within the range of 0–10000, Angle of Inclination is between 0–15 degrees, Scaling Condition is "slight scaling", Crude Oil Viscosity is within the range of 1000–10000, Sand Production Condition is "slight sand production", Submergence is between 100–300 m, Daily Fluid Production is within the range of 30–80, Pump Depth is between 800–1100 m, in this pattern, the wells exhibit long lifespan phenomena.

4 Measures Recommendation for Extending the Free Repair Period

Based on the historical data of the wells and expert consultation information, key indicators are determined. Then, the generated solutions from the historical data of the wells and scheduling rules are compared against the determined key indicators to identify expert experiences corresponding to similar well characteristics. Finally, a recommended action plan is formulated in accordance with the expert experiences.

For the target well, long-lived wells in the corresponding block are identified, and measures from these long-lived wells are recommended for the target well, as shown in Table 4.

Table 4. Example of Recommended Measures for Long-lived Wells.

	Target Well	Recommended Wells
Well	Well 1	Well 2
Maintenance Period	1 year	5 years
Well Reservoir Type	Medium-high permeability	Medium-high permeability
Sand Production Condition	Slight sand production	Slight sand production
Scaling Condition	Slight scaling	Slight scaling
Wax Deposition Condition	Slight wax deposition	Slight wax deposition
Water cut	30–90	30–90
Freezing point	0–30	0–30
Pump Size	56/57	56/57
Crude Oil Viscosity	1000–10000	1000–10000
Daily Fluid Production (DFP)	0–10	10–30
Dynamic Fluid Level	889–1260	0–306
Pump Depth	1100–1500	800–1100
Submergence	100–300	>500
Stroke * Stroke per Minute (SSPM)	0–6.5	9.5–12.0
Angle of Inclination	15–30	0–15
Salinity	10000–50000	0–10000

Based on the geological features and process parameters between the target well and the candidate recommended well, the changes in lift values are analyzed based on shared characteristics. This process facilitates the identification of the optimal solution for process selection. As shown in Fig. 3, the lift values are compared for different processes between the target well and the recommended well. The lift values indicate the contribution to the long-term operational performance of the oil wells.

As shown in Fig. 3, when the geological indicators of oil wells, such as reservoir type, sand production, scale deposition, and wax deposition, are consistent, there are significant differences between the pump depth and output water salinity indicators of the target wells and the recommended wells. Therefore, we focus on analyzing the pump depth, submergence, and output water salinity indicators. The pump depth of the target wells ranges from 1100 m to 1500 m, the submergence ranges from 100 m to 300 m, and the output water salinity ranges from 10000 to 50000. In contrast, the recommended wells have a pump depth ranging from 800 m to 1100 m, submergence greater than 500 m, and output water salinity ranging from 1000 to 10000. As the output water salinity indicator is determined by geological conditions and is not easily changed, we can appropriately adjust the pump depth and submergence indicators to prolong the maintenance-free period of oil wells.

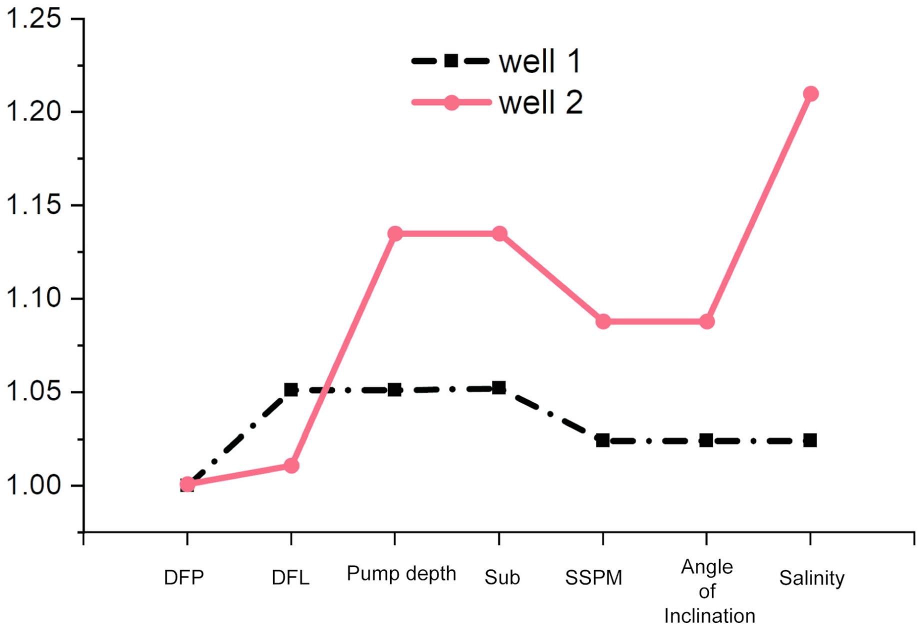

Fig. 3. Comparison of Oil Recovery Measures Indicators

5 Conclusion

Based on the comprehensive achievements of information construction in a certain oilfield, historical production data of 5,789 fully equipped medium-high permeability reservoirs with beam pumping wells were collected. Feature analysis samples and long-lived well samples were designed. Parameters were analyzed from multiple perspectives, including geology, fluid, production, lifting systems, and supporting processes. The focus was on comparing the differences between abnormal wells with extended maintenance intervals and regular wells in various parameters. Based on this analysis, factors and patterns influencing the extended maintenance interval were statistically summarized and identified.

After identifying the factors influencing the extended maintenance interval, the supporting patterns for long-lived wells were explored, resulting in the preparation of 105 frequent patterns for supporting long-lived wells. These patterns can be used to recommend measures and experiences for short-lived wells with similar geological characteristics, providing valuable guidance for the high-value application of oilfield big data.

In the future, efforts will be made to further expand the application of frequent patterns by integrating the recommended supporting patterns with field implementation codes, thereby reducing the difficulty of frontline application.

Acknowledgments. The project is supported by the Sinopec Scientific and Technological Research Project "Research on the Application of Big Data Technology in Oilfield Development" (No. P20071).

References

1. Qun, L., Dingwei, W., Jianhui, L., et al.: Achievements and future work of oil and gas production engineering of CNPC. Pet. Explor. Dev. **46**(1), 145–152 (2019)
2. Qian, K., et al.: Research on fault diagnosis method of electric submersible plunger pump lifting technology. In: Lin, J. (ed.) Proceedings of the International Field Exploration and Development Conference 2021, pp. 5524–5537. Springer Nature, Singapore (2022). https://doi.org/10.1007/978-981-19-2149-0_506
3. Borah, A., Nath, B.: Identifying risk factors for adverse diseases using dynamic rare association rule mining. Expert Syst. Appl. **113**, 233–263 (2018)
4. Vougas, K., Sakellaropoulos, T., Kotsinas, A., et al.: Machine learning and data mining frameworks for predicting drug response in cancer: an overview and a novel in silico screening process based on association rule mining. Pharmacol. Ther. **203**, 107395 (2019)
5. Ruihua, Xu., Luo, F.: Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest. Safety Sci. **135**, 105125 (2021)
6. Porter, T.M.: Karl Pearson. Princeton University Press, Karl Pearson (2010)
7. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. In: Cohen, I., Huang, Y., Chen, J., Benesty, J. (eds.) Noise reduction in speech processing, pp. 1–4. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00296-0_5
8. Myers, L., Sirois, M.J.: Spearman correlation coefficients, differences between. Encycl. Stat. Sci. **12** (2004)
9. Chee, C.-H., Jaafar, J., Aziz, I.A., et al.: Algorithms for frequent itemset mining: a literature review. Artif. Intell. Rev. **52**(4), 2603–2621 (2018)
10. Ping-Hsun, Lu., Keng, J.-L., Kuo, K.-L., Wang, Y.-F., Tai, Y.-C., Kuo, C.-Y.: An apriori algorithm-based association rule analysis to identify herb combinations for treating uremic pruritus using Chinese herbal bath therapy. Evidence-Based Complement. Altern. Med. **2020**, 1–9 (2020)
11. Antomarioni, S., Ciarapica, F.E., Bevilacqua, M.: Association rules and social network analysis for supporting failure mode effects and criticality analysis: framework development and insights from an onshore platform. Safety Sci. **150**, 105711 (2022)