Study on intelligent analysis algorithm for achieving standard of polymer flooding well group

Hanqiang Wang

No.1 Oil Production Plant of Daqing Oilfield Co., LTD, Daqing 163111, Heilongjiang, China

Abstract: In order to master the development effect of polymer flooding well group, it is necessary to accurately analyze the influence of different factors in the whole polymer flooding process on the development index. Combined with the principle of big data analysis, based on the neighborhood rough set theory and Kmeans clustering algorithm, an intelligent analysis algorithm is proposed to determine the achievement of development indicators of polymer flooding well group. Firstly, the neighborhood rough set was used to reduce the attributes of the influencing factors of the Wells with and without the standards. Secondly, Kmeans algorithm was used to cluster the reduced influencing factors to delete the data inconsistent with the actual compliance. Finally, the clustering model is used to judge the standard status of other well groups, and the practical application effect is very good.

Key words: Polyflooding well group; Rough set; Kmeans; Intelligent analysis.

1. Introduction

The geological static factors, production dynamic factors and actual development factors of oilfield development blocks play an important role in the process of oilfield polymer development [1,2]. In order to carry on the long-term reasonable exploitation of the oilfield, it is very necessary to study and analyze the law and development effect of various influencing factors in the development process, and keep abreast of the development effect standard of the polymer flooding well group. Therefore, the prediction method of polymer flooding development index has been paid more and more attention by the oilfield enterprises. Shi Chengfang et al. [3] established a prediction model for fluid production of produced Wells in polymer blocks and production dynamic changes in the initial stage and process of polymer injection by analyzing the corresponding relationship between fluid production of produced Wells and water absorption index of produced reservoirs. Zhao Guozhong et al. [4] conducted index prediction based on the three-layer CBP neural network model by studying the change of water content and its influencing factors in the polymer flooding stage. Qiu Haiyan et al. [5] combined the advantages and disadvantages of HCZ, Weibull and Weng's forecasting models and optimized them, and proposed a weighted combination forecasting model to guide the actual production. Hou Jian et al. [6-7] used numerical simulation technology to analyze the factors affecting the change of polyflooding oil augmenting effect, studied the relationship between characteristic parameters and influencing factors through regression statistical model, and then obtained the prediction model of characteristic parameters to predict the change trend of polyflooding oil augmenting. In this paper, according to the actual dynamic and static data of the development process of polymer flooding in oil field, combined with the principle of big data analysis, based on the neighborhood rough set theory and Kmeans clustering algorithm, an intelligent analysis algorithm is proposed to determine the development index compliance of polymer flooding well group, which has achieved good practical results.

2. Neighborhood rough set theory

2.1 Basic rough set

Rough set [8-10], proposed by Professor Pawlak in 1982, can effectively analyze and process all kinds of incomplete data information, such as data with some characteristics of imprecision, inconsistency and incompleteness. Through rough set theory, the hidden knowledge in data information can be mined, and the hidden law inside data information can be mined. The main principle of rough set theory is to use the method of knowledge reduction to get the classification rules of the problem to be solved without changing the reasoning ability of knowledge classification. The basic idea of rough set theory is to classify things by equivalence relation to recognize knowledge.

1.1.1 Construct knowledge systems. The target of rough set theory is information knowledge system, which is represented by quadruple: $S =< U, A, V, f >$.Among

them， U represents a finite set of objects， $U=\{x_1,x_2,...x_n\}$ ； A represents A finite set of attributes， V represents the set of property values， f Is a function of properties and objects，Attribute set A of information system S is divided into conditional attribute set C and decision attribute set D.

1.1.2 The approximate set. Given a knowledge base $K=(U,R)$ ， And there are arbitrary subsets $X\subseteq U$, Equivalence relation $R\in IND(K)$ ， There are：

$\underline{R}(x)=\cup\{Y\in U/R:Y\subseteq X\}$ ，

$\bar{R}(x)=\cup\{Y\in U/R:Y\subseteq X\}$ ， Among them $\underline{R}(x)$ is X Lower approximation， $\bar{R}(x)$ is X The approximate. $Pos_R(X)=\underline{R}(X)$ Referred to as a subset X R is the domain， $Neg_R(X)=U-\bar{R}(X)$ this is called the R negative domain of the subset X， We can find the boundary of the subset X： $BN_R(X)=\bar{R}(X)-\underline{R}(X)$.

Calculate the information attribute dependence. That is, relative to the set of conditional attributes B ， Calculate the dependence degree of decision attribute set D on it.To determine how important set D is to set B. The calculation method of dependence degree is shown in Formula (1) :

$$\gamma_B(D)=\frac{Pos_B(D)}{|U|} \quad (1)$$

As can be seen from formula (1), the so-called dependence degree of D on B subset is actually the proportion of the positive domain set determined by B subset in the domain U.

1.1.4 Calculate the importance of the attribute. In information knowledge decision system, attribute importance is defined as the influence degree of conditional attribute on decision attribute. Let the information system be denoted by $S=(U,C\cup D,V,f)$ ， $\forall B\subseteq C$ ， If the attributes $\partial\in B$ ， Then the importance degree formula of conditional attribute ∂ to decision attribute D ：

$$Sig(a,B,D)=\gamma_B(D)-\gamma_{B-\{a\}}(D) \quad (2)$$

1.1.5 I'm going to reduce the attribute.Take any subset B of attributes of an information system S， The ambiguous relation $IND(B)$ corresponding to A is defined as:：

$$IND(B)=\{(x,y)\in U^2|\forall a\in B[a(x)=a(y)]\} \quad (3)$$

Let the set of attributes $B\in A$ ， For any property $a\subseteq B$, If you have $IND(B)=IND(B-\{a\})$ ， So call A unnecessary in B， Otherwise, A is said to be necessary in B.If a collection of properties $B\subseteq A$ meet $IND(B)=IND(A)$ ， So B is A reduction of A.

2.2 Neighborhood rough set

Basic rough set theory is aimed at discrete data processing, and so on the continuous data processing, need to first discretization operation, it will introduce error lead to change the original data attributes, which can express the information of the original set of properties, cause the loss of information of information system, and thus the information system of classification performance. Therefore, the neighborhood rough set model [11-13] is adopted in this paper to directly process continuous data to avoid information loss caused by data discretization.

1.2.1 neighborhood. $<U,\Delta>$ A nonempty metric space, $x\in U,\delta\geq 0$ ， According to point set $\delta(x)=\{y|\Delta(x,y)\leq\delta,y\in U\}$ is the δ neighborhood of x .

1.2.2 Neighborhood decision system.Assuming that $U=\{x_1,x_2,...x_n\}$ it's the set of all the samples, D is a set of conditional attributes that describe the sample, $D=\{l_1,l_2,...l_n\}$ is a set of classification decision attributes, $<U,A,D>$ is a neighborhood decision system.

1.2.3 The dependency and importance of the attribute.The dependence degree of decision attribute D on conditional attribute B is：

$$k_D=r_B(D)=\frac{|Pos_B(D)|}{|U|} \quad (4)$$

The importance degree of decision attribute D to condition attribute B is：

$$Sig(a,B,D)=\gamma_B(D)-\gamma_{B-\{a\}}(D) \quad (5)$$

1.2.4 Attribute reduction. When the obtained attribute importance degree is greater than the set lower limit of importance degree， Output the set of reductions red ， red is the set that holds the reduction result.Neighborhood rough set algorithm flow:

Step1: Input decision system $<U,A,D>$ ， Set neighborhood radius calculation parameters and lower importance limit efc；

Step2: Preprocessing, normalizing the original data, and calculating the neighborhood radius δ ；

Step3: Initializes the reduction set $red=\phi$ ；

Step4： For each property a_i in $A-red$ ， Calculate its positive domain： $Pos_{red}(D)=\underline{N}_{red}D$ ；

Step5： Calculate the dependency and importance of each attribute；

Step6： IF is greater than efc; otherwise, output the reduction result red ， else, return to Step4.

3. Attribute data screening based on K-means clustering

After the rough set algorithm is reduced, it is necessary to remove the unqualified data in the classified data. The basic idea is to classify the data based on the reduced data attributes, and delete the data inconsistent with the original standard classification. K-means clustering algorithm is an unsupervised clustering algorithm [14-15], which has the advantages of simple principle, easy implementation and fast clustering speed and is widely used in various fields. However, the algorithm itself is also sensitive to the initial cluster centroid and the comparison of noise and outliers. The clustering principle of K-means algorithm is to continuously divide the data set into different categories according to the centroid through iteration, and verify the clustering effect by evaluating the criterion function, so as to obtain independent inter-class and compact intra-class clustering results.

3.1 Algorithm principle

2.1.1 Select the similarity measurement method between samples. K-means clustering algorithm is not easy to deal with discrete data, but it is very suitable for continuous data. A hypothetical set of data samples $X = \{x_m | m = 1,2,...,total\}$, The sample in X has d properties, Let's call them A1,A2…Ad, And these attributes are continuous data.Then the sample data I and j can be expressed as Xi=(Xi1,Xi2,…Xid), Xj=(Xj1,Xj2,…,Xjd).d(Xi,Xj) is used to represent the similarity between samples Xi and Xj. The smaller the value of d(Xi,Xj), the smaller the distance between samples, and the more similar the two samples are. On the contrary, if the sample distance is larger, it indicates that the two samples are more dissimilar. In the aspect of sample similarity calculation, Euclidean distance and Manhattan distance can be selected according to the specific situation to measure the similarity between data samples. The more commonly used measurement method is Euclidean distance, and the specific calculation method is shown in Formula (6):

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{d} (x_{ik}, x_{jk})^2} \quad (6)$$

2.1.2 Set the criterion function of clustering effect evaluation. The classical K-means clustering algorithm uses the sum of squared error as the criterion function to evaluate the clustering effect. Suppose the data set X is partitioned into k subsets X1,X2... Xk, the number of samples contained in the cluster subset is denoted by N1, N2... , NK, and the cluster centroids are denoted by M1,m2... Mk is denoted by. Then the formula of error sum of squares criterion function is:

$$E = \sum_{k=1}^{k} \sum_{p \in X_i} \|p - m_i\|^2 \quad (7)$$

2.1.3 Calculate the centroid of each cluster subset
1) In the initial state, k centroids are randomly generated, and the sample data are assigned to K clusters according to formula (6);
2) Calculate the average value of the sample data in each cluster and replace the centroid of the cluster with this value;
3) Redistribution was carried out according to the distance between each sample and the centroid of each cluster;
4) Judge whether the evaluation criterion is met, and stop clustering if it is. Otherwise, go to 2) and recalculate k cluster centroids.

3.2 K-means clustering algorithm flow

Input: number of sample clusters K and data set. Output: K clustering results for the dataset. Specific implementation process:
1) Randomly generate K cluster centers;
2) According to the principle of minimum distance, the samples in the data set to be clustered are divided into the nearest clustering set;
3) Calculate the average value of sample data in each cluster set, and replace the center of the original cluster as the cluster center of the next iteration;
4) Repeat Step 2 and Step 3 continuously until the stopping rule is met or the cluster center does not change, then K clusters of the data sample set will be returned.3 Real Data Simulation

3.3 Attribute reduction of influencing factors based on neighborhood rough sets

Was the first to classify a block of polymer flooding well group, and then according to the monthly statistics a standard well group of the well group and not done well group production days, polymer concentration, oil production, to produce liquid, flowing pressure, effective thickness, water cut, oil of nissan, nissan fluid, formation pressure, oil intensity, fluid producing intensity, oil change, water change, flowing pressure changes, and polymer concentration changes, Oil production index, fluid production index, geological reserves, pore volume, injection rate, polymer usage, cumulative injection-production ratio, compliance identifier (1 for compliance, 0 for noncompliance). The statistics of the first well group in January 2018 are shown in Table 1. The theory in Section 2 was used to reduce rough set attributes, delete redundant attributes, and finally determine the factors related to achieving the standard of polymer flooding well group, including water cut change, liquid yield change, flow pressure change, concentration change, polymer dosage change, injection-production ratio, injection rate and liquid extraction index.

Table 1 Influencing factors of Class I well group

NO	Production days (day)	Mining polyethylene concentration (mg/L)	In oil production (t)	Flowing pressure (MPa)	Oil change (%)	Water cut change (%)	Flowing pressure change (%)	Change of concentration of harvesting poly (%)	Fluid productivity index, (t/d·m·MPa)	Injection speed (PV/a)	Amount of polymer (mg/L.PV)	Injection-production ratio	Standard logo
1	25	331	21	5.25	0.95	1.03	0.92	0.98	5.84	0.449	509	1.16	1
2	31	463	365	5.21	0.85	0.99	0.94	0.96	7.18	0.381	584	0.95	1
3	31	71	182	6.79	0.99	1.02	1.16	1.15	6.11	0.304	372	0.9	1
4	31	277	87	2.2	0.95	1.02	1.03	1.07	3.04	0.651	818	1.22	0
5	31	529	241	6.99	0.99	1.00	1.0	1.11	13.64	0.383	549	1.48	0
6	31	42	182	6.19	1.02	1.00	0.91	0.79	8.17	0.486	609	1.1	1

3.4 Actual data screening and well group standard determination based on Kmeans clustering

Kmeans clustering was performed on the 8 attributes after rough set attribute reduction according to the theory in Section 3. Some results are shown in Table 2. It can be seen that the clustering result of well 4 is inconsistent with the standard identification. According to the cluster analysis, there are 142 data of the well group reaching the standard in the first class, among which 15 data of the well group have inconsistent clustering results. There are 120 data in the non-standard well group of the first class, and the clustering results are consistent with the standard identification. Through data analysis and comparison, the correctness and effectiveness of the algorithm are verified. Through the influence factors of 3.1 and 3.2 reduction and data filtering, the intelligent analysis model was established based on clustering algorithm, select the block type of well group in recent 3 months up to standard data to verify the validity of the intelligent analysis algorithm, the block type of well group, 45, which the standard well group 32, not done well group 13, through the method of intelligent analysis, The recognition rate of the well group reaching the standard is 87.5%, and the recognition rate of the well group not reaching the standard is 84.6%.

Table 2 Comparison of actual compliance and clustering results

No	Water cut change (%)	Fluid productivity index, (t/d·m·MPa)	Flowing pressure change (%)	Standard logo	Kmeans Clustering results
1	1.03	5.84	0.92	1	1
2	0.99	7.18	0.94	1	1
3	1.02	6.11	1.16	1	1
4	1.02	3.04	1.03	0	1

4. Conclusion and Understanding

This paper proposes a judging standard of a kind of intelligent polymer flooding well group development analysis algorithm, this algorithm can be combined with data of oil field, has the advantages of adaptive with the increase of oilfield development data, the proposed intelligent analysis method can get a better mark identification model, and improve the accuracy of the identification model, has good popularization value.

References

1. LI Xingrong, TONG Le, Wang Lu, et al. Review of polymer flooding technology [J]. Contemporary Chemical Industry, 2017, 46(6):1228-1230.
2. XU Hao. Optimization of alternate polymer injection parameters in polymer flooding oilfield [J]. Chemical Management, 2017(11).
3. Shi Chengfang, Xiao Wei, Wang Fenglan. Prediction model of polymer flooding development index [J]. Acta petrolei sinica,2005,26(5):78~84.
4. Zhao Guozhong, Meng Shuguang, Jiang Xiangcheng. A neural network method for predicting water content in polymer flooding [J]. Acta petrolei sinica,2004,25(1):70~73.
5. Qiu Haiyan, Ding Xianfeng, Hu Xiaoyun, et al. Prediction of oil production by polymer flooding using combinatorial model [J]. Xinjiang petroleum geology,2010,31(2):194~196.
6. Hou Jian, Zhao Hui, Du Qingjun, et al. Study on sensitive parameters affecting the performance of polymer flooding [J]. Journal of oil and gas technology,2007,29 (1):118~121.
7. Hou Jian, Guo Lanlei, Yuan Fuqing, et al. Quantitative characterization of polymer flooding production performance of different types of reservoirs in shengli oilfield [J]. Acta petrolei sinica,2008,29 (4):577~581.
8. GUO Qing. Research on Uncertain Information System and Decision Making Based on Rough Set Theory [D]. Hefei University of Technology, 2017.
9. Zhang Qinghua, Xue Yubin, Wang Guoyin. Optimal Approximation Sets for Rough Sets [J]. Journal of Software, 2016, 27(2):295-308.
10. Li Jia, Liang Jiye, Pang Tianjie. A multi-attribute decision ranking method based on Dominant Rough Set [J]. Journal of Nanjing University (Natural Sciences), 2016, 52(5):844-852.
11. An Ruoming, So bright. Application of Neighborhood Rough Set in Attribute Reduction and Weight Calculation [J]. Computer Engineering and Applications, 2016, 52(7):160-165.
12. XU Feng, Yao Sheng, Ji Xia, et al. Uncertainty measurement method of information system based on Fuzzy Neighborhood Rough Set [J]. Journal of Nanjing University (Natural Science),2017(5):926-936.

13. Shen Lin, Chen Jianhui. Variable Accuracy Neighborhood Rough Set attribute Reduction Algorithm based on lower approximate distribution [J]. Journal of Guizhou University (Natural Science Edition), 2017, 34(4):53-58.
14. Zhao Wei, Lin Nan, Han Ying, et al. An improved collaborative filtering algorithm based on K-means clustering [J]. Journal of Anhui University (Self-Science Edition), 2016, 40(2):32-36.
15. Ou Hui, Xia Zhuoqun, Wu Zhiwei. Rough Set K-means Clustering Algorithm based on Improved manifold Distance [J]. Computer Engineering and Applications, 2016, 52(14):84-89