

Research Article

Complex fuzzy-probabilistic analysis of information on drilling mud losses

GM Efendiyev* and IA Piriverdiyev

Institute of Oil and Gas of the Ministry of Science and Education of Azerbaijan Republic, 9 F.Amirov str., AZ1000, Baku, Azerbaijan

Received: 29 January, 2024
Accepted: 19 February, 2024
Published: 20 February, 2024

***Corresponding author:** GM Efendiyev, Institute of Oil and Gas of the Ministry of Science and Education of Azerbaijan Republic, 9 F.Amirov str., AZ1000, Baku, Azerbaijan, E-mail: galib_2000@yahoo.com

Keywords: Clustering; Classification; Drilling; Lost circulation; Intensity; Cluster analysis; Entropy

Copyright License: © 2024 Efendiyev GM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.mathematicsgroup.com>



Abstract

In recent years, classification and clustering have been widely used for processing and analyzing information for the purpose of structuring, ordering, summarizing, and sorting. Classification and clustering are used when working with information processes both in enterprises (large and medium-sized) and in various fields of scientific activity, which is especially important in the context of the constant growth of processed information.

At the same time, during cluster analysis, an important task is to assess its quality. In this work, cluster analysis was used to identify loss circulation zones when drilling wells and classify them by severity (intensity). To determine the quality of the cluster analysis, the entropy value was calculated, which should tend to a minimum. In our case, it was 0.23, which allows us to judge the fairly high quality of the cluster solution.

Introduction

With the increase in the volume of processed, stored, and received information when working with information processes in large and medium-sized enterprises, as well as scientific activities, its processing in the resulting form becomes difficult. There is a need for primary processing of information for its structuring, identifying features, generalization, and sorting. To achieve this, classification and clustering processes are used. Document classification is the process of arranging or distributing objects (observations) into classes in order to reflect the relationship between them. A class is a collection of documents that have some common feature that distinguishes this collection from others. To classify an object means to indicate the number (or name) of the class to which the object belongs. Classifier training is the process of constructing an algorithm in the case when a finite set of objects is given and it is known which classes they belong to. This set is called a sample. The class affiliation of the remaining objects is unknown [1-3].

Clustering is the process of dividing a given set of objects (observations) into disjoint subsets called clusters, so that each cluster consists of similar objects, and objects in different clusters are significantly different. Clustering is used when data compression is required. If the original sample is too large, it can be reduced by leaving one most characteristic representative from each cluster. Clustering is also used for novelty detection. Atypical objects are selected that cannot be attached to any of the clusters.

Formulation of the problem

As is known, when drilling wells, complications associated with geological conditions arise. One of the most severe types of complications is the loss of drilling fluid. This phenomenon occurs due to a mismatch between the characteristics of the geological section and the drilling fluid (mud). Due to the high porosity and permeability of rocks, the drilling fluid penetrates into their pores, i.e. the rock, as it were, "absorbs" the mud, as a result of which a significant amount of mud is lost and, finally, circulation is lost. The severity of this complication is expressed



by the volume of mud absorbed by the rock (formation). To prevent material losses and loss of time due to the occurrence of this situation, it is necessary to find ways of early forecasting [4,5]. There are various methods that can be used to solve this problem. Due to the influence of various factors, it is not possible to accurately study the well section and establish the influence of these conditions on the severity of losses using traditional methods. As the analysis shows, the most reliable apparatus are methods based on fuzzy logic, in particular, methods that make it possible to classify rocks according to petrophysical characteristics (porosity, permeability) and clarify the correspondence of each homogeneous interval according to these properties with the severity of losses, expressed by the volume of lost drilling fluid. From this point of view, such a correspondence can be established using the clustering method, and the most convenient in this case is the fuzzy cluster analysis algorithm. In this regard, this article, devoted to the early prediction of drilling fluid losses, provides for data collection, statistical data processing, information analysis, including cluster analysis and the establishment of mutual correspondence between the petrophysical properties of rocks and the severity of mud losses in the form of fuzzy rules.

Materials and methods

There are two main classifications of clustering methods. One of them is the division into hierarchical and non-hierarchical (or flat) clustering methods. Hierarchical algorithms (also called taxonomy algorithms) construct multiple partitions of a sample into disjoint clusters and a system of nested partitions. Thus, at the output, we get a tree of clusters, the root of which is the entire sample, and the leaves are the smallest cluster. Top-down hierarchical algorithms work on a top-down principle: first, all objects are placed in one cluster, which is then divided into smaller clusters. More common are bottom-up algorithms, where we start by placing each object in a separate cluster, and then we combine the clusters into larger clusters until all the objects in the sample are contained in one cluster. Thus, a system of nested sections is built. The results of such algorithms are usually presented in the form of a tree – a dendrogram. A classic example of such a tree is the classification of animals and plants. In contrast, non-hierarchical (flat) algorithms build clusters from a single section of objects [1-3,6-8].

In general, clustering can also be classified as soft clustering (overlapping clustering) and hard clustering (or “exclusive” clustering). In the case of hard clustering, each point belongs to only one cluster, whereas in soft clustering, a point belongs to two or more clusters with different degrees of membership. Often soft clustering is more natural because points on class boundaries do not necessarily have to belong entirely to one of them. Rather, they will belong to several classes with varying degrees of membership ranging from 0 to 1. One of the most popular soft clustering methods is Fuzzy C-Means (FCM), and similarly, k-means is one of the most common hard clustering methods [6-8].

For each pair of objects, the “distance” between them is measured – it is the degree of similarity. There are many metrics, and here are the main ones: Euclidean distance, squared Euclidean distance, distance between city blocks (Manhattan distance), Chebyshev distance, and power-law distance.

Results

To assess the influence of geological conditions on the nature of lost circulation in the absence of information, a mutual correspondence was achieved between the indicators of the petrophysical properties of rocks and the degree of lost circulation based on fuzzy cluster analysis. Using a fuzzy clustering algorithm, drilling conditions were classified according to the severity of losses based on data on the petrophysical properties of the rocks being drilled. This is very important for early diagnosis of complications and assessment of their risk [9-11].

One of the important tasks facing us when carrying out cluster analysis is assessing its quality. As noted in the work [12], two measures of the quality of the performed cluster analysis can be distinguished: internal and external. Internal measures are based on assessing the properties of separability and compactness of the resulting data partition, and the function of the sum of squared deviations of objects from the center of the clusters can serve as such a measure. The use of external measures consists of comparing an automatic data partition with a “reference” one obtained from experts or selected on the basis of some theoretical reasoning. In statistical physics, entropy characterizes the “order” in a system that moves from one state to another and is interpreted as a measure of the probability of the system remaining in a given state. The more disorder there is, the greater the entropy. Any system eventually gradually approaches its more probable state. In the process, disorder increases, chaos increases, and therefore entropy increases. As is known, chaos corresponds to the maximum entropy value and means the absence of clusters, while a good cluster solution should tend to minimum entropy values [12-15]. Due to the need to assess the quality of the performed cluster analysis, we performed entropy calculations.

The entropy of one cluster is calculated using the following formula:

$$E(S_r) = -\frac{1}{\log q} \sum_{i=1}^q \frac{n_r^i}{n_r} \log \frac{n_r^i}{n_r} \quad (1)$$

Where n_r is the number of elements in a given cluster;

q is the total number of classes in the entire collection;

n_r^i is the number of elements of the i -th class within cluster r .

Table 1 presents the n_r and n_r^i values we found for each of the five clusters. As can be seen, in the fourth cluster there were no representatives of other classes, and its entropy reaches a minimum value, because $\log(1)=0$, i.e. all elements belong to



the same cluster. According to our calculations, the entropy values for the remaining cases were as follows: 0.23 for the first cluster, 0.29 for the second, 0.22 for the third, and 0.16 for the fifth (Table 2).

- Consequences characterized by a complete disruption of the technological process and lost circulation rates above 40 m³/h are called *catastrophic*; the entropy, in this case, was 0.23, and the cluster included 39 elements.
- *Serious* lost circulation includes those characterized by a significant disruption of the technological process, lost circulation intensity of 32–40 m³/h; the cluster contains 38 objects, and the entropy value is 0.29.
- *Intensive* consequences include those characterized by disruptions in the technological process with lost circulation rates of 23–35 m³/h; the entropy value here is 0.22, and the number of elements in the cluster was 23.
- Consequences are considered *partial* if they are characterized by a relatively minor disruption of the technological process, with lost circulation rates of 14–23 m³/h; in this case, the entropy value is zero, since the cluster did not include elements of other classes, and there are only 6 “own” elements in it.
- *Minor* ones include those that are easily removable and require relatively small financial and time costs with a lost circulation rate of 5–14 m³/h [8–10]. There are 8 elements in the cluster, and the entropy value is 0.16. All this information is reflected in Table 1.

Next, using the formula given in the work [10], the entropy of the cluster solution is found:

$$Entropy = \sum_{r=1}^k \frac{n_r}{n} E(S_r) \tag{2}$$

Table 1: Results of fuzzy cluster analysis.

Cluster	Q		n _r	i n _r ⁱ				
	quant., m ³ /hour	qual.		n _r ¹	n _r ²	n _r ³	n _r ⁴	n _r ⁵
1	5-14	minor	39	39	0	14	0	0
2	14-23	partial	38	0	38	15	1	0
3	23-32	intensive	23	1	2	23	0	0
4	32-40	serious	6	0	0	0	6	0
5	more than 40	catastrophic	8	0	0	0	1	8

Table 2: Entropy values in each selected cluster.

Cluster number	q	n _r	i n _r ⁱ	E(S _r)
1	5	39	14	0,23
2		38	16	0,29
3		23	3	0,22
4		6	0	0
5		8	1	0,16

Our calculations showed that the entropy has a value of 0.23, which indicates a fairly high-quality cluster analysis.

Conclusion

A “hybrid” cluster procedure is proposed, combining fuzzy cluster analysis with a statistical method of information analysis, including the calculation of entropy. Refinement of the results obtained using this method is achieved by applying clustering procedures based on fuzzy logic and estimates based on probabilistic statistical methods.

As a result of the implementation of the FCM algorithm, membership functions were calculated, clusters were identified, the severity of lost circulation was established, entropy values were calculated, and ways to assess the quality of clustering were shown.

The calculated entropy values indicate that the resulting clusters are internally compact and externally noticeably heterogeneous. Integration of probabilistic-statistical methods and methods based on fuzzy clustering makes it possible to predict absorption in conditions of uncertainty expressed by inaccuracy, vagueness of initial data, and lack of information, allowing for early prediction of this severe complication and making timely decisions.

Acknowledgement

This research was carried out with financial support from the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (No AP19674847).

References

1. Cherezov DS, Tyukachev NA. Review of main methods of data classification and clustering. Vestnik VSU, series: system analysis and information technologies. 2009; 2: 25-29.
2. Zadeh LA. Fuzzy sets. Information and Control. 1965; 8: 338-353.
3. Bezdek JC. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York. 1981.
4. Grigorovich AV. Combat complications in real-time. Journal “Neftegaz.RU”, No. 6, June 2020.
5. Panikarovskiy EV, Panikarovskiy VV, Murzaulugov ZA, Panikarovskiy V, Murzaulugov Z. Main types of complications when drilling wells in Achimov deposits. Territory Neftegaz. 2011; 6. <https://cyberleninka.ru/article/n/osnovnye-tipy-oslozhneniy-pri-bureanii-skvazhin-v-achimovskikh-otlozheniyah> [in Russian]
6. Suganya R, Shanthi R. Fuzzy C-Means Algorithm - A Review. International Journal of Scientific and Research Publications. 2012; 2: 11.
7. Jitendrasinh RG. A Review on Fuzzy C-Mean Clustering Algorithm. International Journal of Modern Trends in Engineering and Research. 2015; 2: 751-754.
8. Bora DJ, Gupta AK. A Comparative Study between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. International Journal of Computer Trends and Technology (IJCTT). 2014; 10.
9. Efendiyev GM, Rza-zadeh SA, Kadimov AK, Kouliyev IR. Forecast of drilling mud loss by statistical technique and on the basis of a fuzzy cluster analysis. Seventh International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control. Turkey: Izmir. 2013; 319-322.



10. Efendiyev GM, Mammadov PZ, Piriverdiyev IA, Mammadov VN. Clustering of geological objects using FCM algorithm and evaluation of the rate of lost circulation. 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria. Procedia Computer Science. 2016; 102: 159–162.
11. Efendiyev GM, Mammadov PZ, Piriverdiyev IA, Mammadov VN. Estimation of the lost circulation rate using fuzzy clustering of geological objects by petrophysical properties. Visnyk Taras Shevchenko National University of Kyiv. Geology. 2018; 81: 28-33.
12. Internet Research Laboratory. (n.d.). <https://linis.hse.ru/>
13. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. Journal of intelligent information systems. 2001; 17(2-3): 107-145.
14. Pal NR, Biswas J. Cluster validation using graph theoretic concepts. Pattern Recognition. 1997; 30(6): 847-857.
15. Sivogolovko EV. Methods for assessing the quality of clear clustering, 2011.

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROMEO, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services

<https://www.peertechzpublications.org/submission>

Peertechz journals wishes everlasting success in your every endeavours.