Research on Lost Circulation Risk Intelligent Identification and Processing System Based on CBR

Gao Xiaorong, Xu Yingzhuo
Institute of Petroleum Engineering, Xi’an Shiyou University, Xi’an, 710065, China
gaoxr@xsyu.edu.cn

Abstract: Due to the large number of complex and uncertain factors in the drilling process, it is very difficult to identify and deal with the risk of lost circulation by establishing an accurate mathematical model, and the existing intelligent methods are difficult to acquire or require large sample data. In view of the defects, an intelligent decision-making method based on case-based reasoning is proposed, and a software system for intelligent identification and processing of lost circulation risk is developed. In this system, the lost circulation cases of oil fields in the past are represented as case forms in the case base. The design idea of the case retrieval model is to design an improved similarity calculation model for different types of attributes, based on the traditional instance similarity calculation model, which can effectively solve the information uncertainty. The system was tested by field data and the results show that the lost circulation risk identification and processing method can meet the application requirements of the drilling site, and can effectively improve the intelligent level of lost circulation risk management.

1. Introduction
The occurrence of lost circulation in the drilling process has brought great challenges to the oil drilling industry. Once a lost circulation occurs, it will cause the drilling project to be interrupted. It may also cause other accidents such as well collapse and blowout, which seriously affects the safety and progress of the drilling project, and the drilling cost is increased [1]. Therefore, accurately identifying and dealing with the lost circulation risk has become an important problem that drilling engineering technicians need to solve. Since the occurrence and development of the lost circulation is a nonlinear process and the uncertainty and randomness of the information, the traditional identification and processing methods are difficult to meet the requirements. In recent years, with the development and application of artificial intelligence technology, the identification and treatment of the risk of lost circulation are mainly based on intelligent means. At present, most of the common intelligent identification and processing systems adopt the expert system method based on rule reasoning and the neural network method. However, the expert system method has difficulty in acquiring knowledge. The neural network method needs large enough sample data and is prone to the problem of local minimum solutions. In the actual drilling process, the identification and treatment of the lost circulation risk are usually based on experienced experts, through the experience and results of previous similar decision-making to solve the new problems currently faced, and accordingly make the best decisions efficiently. Case-Based Reasoning (CBR) is a kind of reasoning model that uses the previous cases to solve similar problems to obtain the current problem-solving results [2]. It conforms to the cognitive process of drilling experts in decision-making, reflects the characteristics of experts' inference and solving process based on past experience and methods, and is very suitable for decision-
making of lost circulation risk. For this reason, this paper proposes the development of lost circulation risk intelligent identification and processing system based on CBR technology.

2. Overall Structure of the System
The overall structure of the system is shown in Figure 1. It consists of the following components:

- **Case base**: The cases of the lost circulation risk that has occurred in oil field is stored, and is represented by an example.
- **Inference engine**: It is the core of the system. It targets the decision problems and input parameters provided by the user, and retrieves relevant cases from the case base according to the CBR reasoning mechanism to solve the problem until the conclusion is reached.
- **Knowledge acquisition and maintenance module**: including the case base, the acquisition and maintenance of cases and rules in the rule base.
- **Rule base**: Stores the relevant rules for case correction when used for case retrieval.
- **Engineering calculation module**: Responsible for calculating the drilling engineering parameters required for the lost circulation early warning.
- **Reasoning interpreter**: Tells the user the basis for the conclusion of the system to achieve system transparency.
- **Blackboard**: A storage area in which intermediate results are stored, providing a "recording area" for the inference engine.

![Fig.1 The overall structure of the lost circulation risk system](image-url)
3. Main Technology of System Implementation

3.1. Case representation of the lost circulation risk

A case usually consists of the feature attributes and solutions of the problem [3]. For the identification and treatment of the lost circulation risk, through the field investigation and expert consultation, the feature attributes that characterize the problem include:

1. Basic parameters: including the well number, the oil field name, the measure depth when the risk occurs, the formation lithology, the risk occurrence time, and the risk occurrence process.

2. Risk characteristic parameters: including total pool volume, inlet flow rate, outlet flow rate, and standpipe pressure.

3. Geomechanical parameters: including formation fracture pressure, collapse pressure, and pore pressure.

The solution to the problem includes: ① Risk identification type; ② Preventive measures; ③ Treatment plan.

A case of the lost circulation risk is represented by the following two interrelated data tables (Table 1, Table 2). The main table is used to store the basic parameters of the lost circulation case, the risk occurrence process and its solution, and the side table is used to store risk characteristic parameters and geomechanical parameters. Two tables are associated by a "case number", which is the unique identifier of the case.

The “risk occurrence process”, “risk prevention measures” and “risk treatment plan” in Table 1 are described in text form. Due to the large amount of content, it is not expressed in the table.

| Case number: C2001 |
|-------------------|
| **Basic parameters** | |
| Well number: YB21 well |
| Oilfield name: Sichuan Oilfield |
| Measurer depth when risk occurs: 1788m |
| Formation lithology: Siltstone |
Risk occurrence time: 2015-3-14 
Risk occurrence process: Text Description 

Solution 
Risk identification type: Differential Pressure Lost Circulation 
Risk prevention measures: Text Description 
Risk management plan: Text Description 

| Case number: C2001 |
|-------------------|
| Total pool volume: 79.2L |
| Inlet flow rate: 16.27L/s |
| Export flow rate: 13.73 L/s |
| Standpipe pressure: 27.32 MPa |
| Fracture pressure equivalent density: 1.52g/cm³ |
| Collapse pressure equivalent density: 1.05g/cm³ |
| Pore pressure equivalent density: 0.95g/cm³ |

3.2. Design of the case retrieval model
The case retrieval is the core of the system’s inference engine. The effectiveness of the algorithm directly affects the accuracy of the lost circulation identification result. The case retrieval process mainly uses the case retrieval strategy to retrieve the most similar case from the case base. The case retrieval strategy of this system adopts the nearest neighbor method [4], and the similarity of the two cases is measured by the similarity of the feature attributes shared between the cases [4]. Assume the case set in the case base be $SB=\{S_1, S_2, ..., S_k\}$, and the new problem case is $S_0$, which is shared with the case $S_i$ ($i=1, 2, ..., k$) in the case base. There are $n$ feature attributes, set to $F=\{f_1, f_2, ..., f_n\}$. The specific case retrieval algorithm model is as follows.

3.2.1 Similarity calculation model between two case attributes
Due to the complexity of the drilling process and the ambiguity or uncertainty of the information, the description of the identification and treatment of the lost circulation risk includes not only the deterministic (separated numerical and enumerated) feature attributes, but also the uncertain feature attributes, for which the system uses different similarity calculation models for different types of attributes.

1) Similarity calculation model between numerical attributes
The value of the numeric attribute is a continuous value (such as the total pool volume is 85.43L). The similarity between two attributes is defined as:

$$
Sim(f_{0j}, f_{ij}) = 1 - \frac{d(f_{0j}, f_{ij})}{\max\{f_j\} - \min\{f_j\}}
$$

(1)

Where $f_{0j}$ is the value of the j-th feature attribute of the problem case $F_0$, $f_{ij}$ is the value of the j-th feature attribute of the i-th case $F_i$ in the case base, $\{f_j\}$ represents all values of the j-th attribute, similarity $Sim( f_{0j}, f_{ij}) \in [0, 1]$ indicates the degree of similarity between the i-th old case $F_i$ and the j-th feature attribute in the case $F_0$.

2) Similarity calculation model between enumerated attributes
An enumerated attribute is a qualitative description of an attribute. The similarity between two attributes is defined as:

$$
Sim(f_{0j}, f_{ij}) = \begin{cases} 
1 & f_{0j} = f_{ij} \\
0 & f_{0j} \neq f_{ij}
\end{cases}
$$

(2)

Where $f_{0j}$ and $f_{ij}$ have the same meaning as above. Such as "formation lithology" (divided into "shale", "sandstone" or "sandy conglomerate", etc.) belongs to this category.
(3) Similarity calculation model between uncertain attributes

The value of an uncertain attribute is usually represented by an interval number, such as weight on bit is about 20–70kN.\[ \text{Sim}(f_{ij}, f_{ij}) = \text{Sim}([a_1, a_2], [b_1, b_2]) \]

\[ \text{Sim}([a_1, a_2], [b_1, b_2]) = \frac{\int_{a_1}^{b_1} \int_{a_2}^{b_2} \text{Sim}(x, y) dy dx}{(a_2 - a_1)(b_2 - b_1)} \] (3)

Where, \( a_1, a_2, b_1, b_2 \in [a, \beta] \), \( a, \beta \) are the lower and upper bounds of the interval, respectively.

Calculate \( \text{Sim}(x, y) \) according to formula (1), then formula (3) can be transformed into:

\[ \text{Sim}([a_1, a_2], [b_1, b_2]) = \frac{\int_{a_1}^{b_1} \int_{a_2}^{b_2} (1 - |y - x|)/(\beta - \alpha) dy dx}{(a_2 - a_1)(b_2 - b_1)(\beta - \alpha)} = 1 - \frac{\int_{a_1}^{b_1} \int_{a_2}^{b_2} y - x| dy dx}{(a_2 - a_1)(b_2 - b_1)(\beta - \alpha)} \] (4)

3.2.2 Total similarity calculation model for two cases on each feature attribute

During case retrieval, since the weight of each feature attribute of the case has different influence on the similar matching of the case, the system first uses the AHP [1] to determine the weight value of each feature attribute (limited to the length, the process is slightly), then use the weighted average method to calculate the total similarity of the two cases over all of their feature attributes.

Assume the feature attributes of the case be \( \{ f_1, f_2, \ldots, f_n \} \), and determine their influence weights as \( \omega_1, \omega_2, \ldots, \omega_n \) according to the AHP method, meet \( \omega_1 + \omega_2 + \cdots + \omega_n = 1, \omega_i \geq 0 (i = 1, 2, \ldots, n) \). The total similarity between two cases is defined as the weighted sum of the similarity of each feature attribute. The formula is:

\[ \text{Sim}(S_0, S_i) = \sum_{j=1}^{n} \omega_j \times \text{Sim}(f_{0j}, f_{ij}) \] (5)

Where, \( \text{Sim}(S_0, S_i) \) represents the similarity between the i-th old case \( S_i \) and the new problem case \( S_0 \) in the case base, and \( \omega_j (j=1, 2, \ldots, n) \) is the weight of the j-th feature attribute in participating in the similar case retrieval.

Note: In the case retrieval of this system, only the “formation lithology” of the basic parameters of a case participates in the similarity calculation, and the remaining parameters have no meaning for the similarity calculation.

The new problem case \( S_0 \) is compared with the cases in the case base one by one, and their similarities are calculated according to the above algorithm model, and then the case with the largest similarity value is selected. Usually, in the similar case screening, the similarity threshold \( \tau \) should be set first. If the maximum similarity \( \text{Sim}_{\text{max}} \geq \tau \), the old case can be used as a similar case, otherwise it will not be used. The threshold \( \tau \) is usually determined by field expert experience [5] and is set to 0.85 in this system.

4. Application Examples of the System

According to the overall structure of the system and its implementation technology, the lost circulation intelligent early warning system was developed by Microsoft Visual Studio 2015 platform and C# language. The system has been applied in the Yuanba gas field in Sichuan. Following is a case of the lost circulation risk early warning of the gas field “YB21 Well”. When the measure depth is 1858m, the drilling formation is “siltstone”, the formation fracture pressure equivalent density is 1.49g/cm\(^3\), the collapse pressure equivalent density is 1.03 g/cm\(^3\), the pore pressure equivalent density is 0.95 g/cm\(^3\), the inlet flow rate is 16.38 L/s, the outlet flow rate is 13.81 L/s, the total pool volume is 84.4L, and the standpipe pressure is 27.52MPa. Using the system to perform similar case retrieval among the case base, some of the results are shown in Table 3. It can be seen that the case with the highest similarity is "Case 7", the similarity of which exceeds a predetermined threshold \( \tau \), and thus the
current risk is "fractured lost circulation". This conclusion is in line with the actual drilling engineering.

| Case | Formation lithology | Fracture pressure equivalent density (g/cm³) | Collapse pressure equivalent density (g/cm³) | Pore pressure equivalent density (g/cm³) | Inlet flow rate (L/s) | Outlet flow rate (L/s) | Total pool volume (L) | Standpipe pressure (MPa) | Risk type | Case similarity (%) |
|------|---------------------|---------------------------------------------|---------------------------------------------|------------------------------------------|----------------------|-----------------------|-----------------------|------------------------|----------|---------------------|
| 1    | Siltstone           | 1.41                                        | 1.01                                        | 0.93                                     | 16.39                | 14.78                 | 85.23                 | 28.88                  | Differential pressure leak | 85.41    |
| 2    | Siltstone           | 1.45                                        | 1.02                                        | 0.94                                     | 15.39                | 14.23                 | 85.10                 | 29.71                  | Differential pressure leak | 82.12    |
| 3    | Siltstone           | 1.92                                        | 1.08                                        | 1.02                                     | 15.45                | 13.40                 | 82.50                 | 29.01                  | Differential pressure leak | 76.42    |
| 4    | Middle sandstone    | 2.21                                        | 1.04                                        | 1.05                                     | 16.47                | 14.76                 | 82.13                 | 28.52                  | Cracked leak          | 87.37    |
| 5    | Middle sandstone    | 1.52                                        | 1.05                                        | 0.95                                     | 16.27                | 13.73                 | 79.20                 | 27.88                  | Cracked leak          | 88.65    |
| 6    | Silty mudstone      | 1.61                                        | 1.06                                        | 0.98                                     | 16.26                | 13.70                 | 78.57                 | 27.04                  | Differential pressure leak | 78.42    |
| 7    | Mudstone            | 1.49                                        | 1.03                                        | 0.95                                     | 16.34                | 12.84                 | 85.47                 | 27.31                  | Cracked leak          | 95.71    |
| 8    | Siltstone           | 1.77                                        | 1.08                                        | 1.01                                     | 16.40                | 10.84                 | 82.20                 | 27.65                  | Cracked leak          | 89.43    |

5. Conclusion
Because of the large number of complex and uncertain factors in the drilling process, this paper adopts an intelligent decision-making method based on case-based reasoning for the identification and processing of lost circulation risk, and develops a software system for intelligent identification and processing of lost circulation risk. In the case retrieval of the system, the design of the case similarity calculation model is based on the traditional method, and fully considers the characteristics of the identification and processing of the lost circulation risk. Through the preliminary application of the system, the results show that the identification and processing method of lost circulation risk proposed in this paper not only has the consistency with the decision-making thinking of drilling experts, but also overcomes the defects of traditional methods such as difficulty in acquiring knowledge and large sample data. In addition, it can also be used in real-time identification and treatment of lost circulation risk during drilling operations.

Acknowledgments
We are grateful to the National Natural Science Foundation China and the Youth Science and Technology Innovation Fund Project of Xi’an Shiyou University for their financial supported under Grant No.51574194 and No. 2016BS09 for this paper.

References
[1] Zhang Xuehong, Li Qian. Risk pre-warning method of well leakage based on case reasoning[J]. Fault-block Oil & Gas Field, 2017, 24(2): 255-263.
[2] RAJA H S, MORTEN L V. Case-based reasoning: predicting real-time drilling problems and improving drilling performance[R]. SPE141598, 2011.
[3] YAN C F, WANG H B, et al. Fault diagnosis expert system of turbine generator sets based on rule reasoning and case reasoning[J]. Applied Mechanics & Materials, 2014, 513-517.
[4] Gu Shujuan, Gao Xuedong, Sun Ran. An Improved Case Similarity Measure for Case Retrieval in CBR Based on Emergency Decision[J]. China Management Informationization, 2011, 14(9): 50-54.
[5] Sun Qin, Ye Changqing. Fault Diagnosis Algorithm with Case-Based Reasoning[J]. Computer Systems & Applications, 2016, 25(1): 181-186.