A Diagnostic Model Based on Support Vector Machine for the Collapse of Horizontal Well Borehole Wall

LIANG HAI-BO1*, HUANG XIAO-QIAN1, SUN YU-QI1, CHENG XIANG-ZHEN2
1 School of Mechanical and Electrical Engineering, Southwest Petroleum University, Chengdu, China
2 science and technology infoshop, China petroleum Huabei oilfield company, Hebei, China

INTRODUCTION

Borehole instability is a serious problem in drilling engineering. The problem of borehole instability in horizontal wells is more complex and more general, especially in developing multi-lateral horizontal wells. Borehole collapse is attributed to two main reasons: one is that the drilling fluid pressure on the formation pressure is less than pressure on the collapsed formation, and the other is the relevant physical–chemical reactions of the drill fluid and formation [1-5]. The following characteristics are observed when borehole collapse occurs:

(1) When a slight collapse occurs, the volume of return bit cuttings increases and the shape of flowback cuttings is mostly angular.

(2) When the collapse layer is the formation being drilled, drilling becomes difficult and pump pressure and torque suddenly increase. The pump pressure also decreases in the common scope after lifting the drill, which cannot be placed at the bottom of the well.

(3) When the collapse occurs above the drilling strata, the pump pressure values rise. After the drill bit is removed from the bottomhole, the pump pressure remains constant and the outlet flow is decreased to less than the flowback. The raising or lowering of the drill bit always results in resistance and the increase in torque.

Borehole collapse also changes the drilling fluid displacement, pump pressure, torque, and drilling time. The current diagnosis methods for the collapse of horizontal well borehole wall apply the ground well logging data directly. However, ground parameters do not always change because of borehole instability, and collapse locating in different places leads to varying ground parameters or the opposite change trend. Changes in surface parameters are not necessarily caused by borehole collapse; they are usually caused by the low precision and poor reliability of using surface parameters to judge borehole collapse. Considering these limitations in borehole collapse characterization, this study proposes a method that uses equivalent circulating density (ECD), massic energy, and friction/torque for the diagnosis of borehole collapse. The three parameters can be precisely obtained by ground survey data. However,
the effect of these parameters on the borehole collapse diagnosis follows different non-linear laws. The fuzzy neural network algorithm cannot minimize the empirical risk and the expected risk simultaneously [6] and thus this method cannot achieve a good classification performance. Thus, this study proposes a diagnostic model based on support vector machine (SVM) for the collapse of horizontal well borehole wall.

2. Calculation model for friction/torque

In this study, a model for calculating the friction/torque of the horizontal well pipe string is established. Considering that the borehole curvature of the deflecting section is relatively large and the deformation of the pipe string is not completely in contact with the shaft wall, the crossbar and flexed beam model is used to calculate these parameters. Given that the vertical section, steady inclined section, and horizontal section of the borehole curvature are relatively small, the pipe string cannot contact with the sidewall completely. Thus, this study adopts the modified soft model by Fan Guangdi [7] to calculate the latter parameters. The calculation model of friction/torque and the tubing string stress analysis diagram of any infinitesimal section are given as follows:

(1) Calculation model of friction/torque in the vertical section and horizontal section:

\[ T_{i+1} = T_i + (Wd \cos \alpha + \mu N_i) \]
\[ M_{i+1} = M_i + \mu N_i r \]
\[ N_i = \sqrt{(T_i \Delta \alpha \sin \alpha)^2 + (T_i \Delta \alpha + Wd \cos \alpha)^2} \]
\[ F = \pm \mu N_i \]

In the formula,
\[ T_{i+1}, T_i \] ——axial force of upper and lower end column in section \( i \);
\[ M_{i+1}, M_i \] ——torque of upper and lower end column in section \( i \);
\[ N_i \] ——conduct normal pressure of tubular element and sidewall in section \( i \);
\[ W \] ——tubular element buoyant weight;
\[ \mu \] ——coefficient of sliding friction;
\[ r \] ——tubular element radius;
\[ F \] ——friction;
\[ \alpha, \Delta \alpha, \Delta \phi \] ——average deviation angle, deviation angle increment, and azimuth increment.

An upward column movement is denoted by “+”, and a downward movement is denoted by “−”.

(2) Calculation model of friction/torque in the building-up section

The method introduced by Fan Guangdi [7] is adopted by projecting well track to the inclination plane (P) and azimuth plane (Q), and then solving the formula for the two planes.

\[ N_{ip} = \frac{M_{i-1} - M_i + T_i (y_{i-1} - y_i)}{L_i} \]
\[ M_{i-1} - M_i + T_i (y_{i+1} - y_i) + q_i L_i + q_{i+1} L_{i+1} \]

The overall support reaction of the touch point is obtained after solving the support reaction of plane (P) and plane (Q), as shown below:

\[ N_i = (N_{ip}^2 + N_{iq}^2)^{\frac{1}{2}} \]  

Pipe string buckling of contact force must also be considered.

\[ \Delta N = \frac{r T^2}{8 E I} \]
\[ \Delta N' = \frac{r' T^2}{4 E I} \]

In the formula,
\[ E \] ——elastic modulus of string material;
\[ I \] ——inertia moment of string.

The axial force of the tubing string, which is caused by string gravity and bit weight, can be calculated from the logging parameters. After determining the string axial force, drilling friction/torque is solved by evaluating whether the string buckling occurred or not on the tubular column, and then calculating the contact force between tubular column and borehole wall.

2. SVM model structure

2.1 Model development

The SVM method is based on statistical learning theory and can be used in various applications, such as pattern recognition, function approximation, and probability density estimation. The diagnosis of
horizontal well borehole wall sloughing belongs to the category of SVM method for function approximation. The support vector regression machine, or support vector regression, is a type of SVM function approximation regression algorithm. The final form of the regression function is [8, 9] as follows:

\[ f(x) = \sum_{s.v.} (a_i - a_i^*) K(x, x_i) + b \quad (6) \]

In the formula, s.v. stands for support vector, s.v. = s, where “s” is the number of the support vector; \( K(x, x_i) \) stands for the kernel function of SVM; \( x_i \) is the sample factor vector of the support vector, \( i = 1, 2, \cdots, s; \) \( x \) is the factor to predict; and \( a_i^*, a_i, b \) represent the undetermined coefficients in the model.

### 2.2 Influencing factors and diagnostic indicators of borehole wall sloughing

According to the above analysis, the parameters that best reflect the collapse of horizontal well borehole wall are specific energy, ECD, and friction/torque. Thus, the three parameters can be used as influencing factors of borehole wall collapse, and the prediction of the occurrence of borehole collapse can be used as an output indicator for the diagnosis of borehole wall collapse. The friction/torque calculation model can determine the friction/torque value at an arbitrary point in the entire well. Accordingly, of the borehole collapse risk can be accurately diagnosed, thereby determining the position of the borehole collapse. Therefore, the collapse position can also be used as an output indicator for the diagnosis of borehole wall collapse.

### 2.3 Establishment of the training set and test set

To test the diagnostic model, the data from 10 horizontal wells of XX oil field are used as the basis data. Then, these data are divided into two parts: the training set (well nos.1–5) and the test set (well nos.6–10). The training set is built for the training (learning) of the unknown rule and is used to diagnose the collapse of horizontal well borehole wall. The test set is established to check the capability of the diagnosis model. Notably, the test set data are not used in the selection process and selection method of model parameters.

The training set and testing set are composed of two parts: the input parameters and output parameters. ECD, specific energy, and friction/torque are the input parameters of the model, while the prediction of collapse occurrence and collapse position are the output parameters of the model. The specific data are shown in Table 1.

<table>
<thead>
<tr>
<th>Sample classification</th>
<th>Serial number</th>
<th>Well number</th>
<th>Drill bit position (m)</th>
<th>ECD (g/cm³)</th>
<th>Massic energy (Mpa)</th>
<th>Friction drag (KN)</th>
<th>Torque (KN.m)</th>
<th>Prediction of collapse occurrence</th>
<th>Collapse position (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training sample</td>
<td>1</td>
<td>X1</td>
<td>2114</td>
<td>1.32</td>
<td>9.74</td>
<td>40.31</td>
<td>14.22</td>
<td>No</td>
<td>\</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>X5</td>
<td>1270</td>
<td>1.20</td>
<td>4.81</td>
<td>27.04</td>
<td>9.41</td>
<td>No</td>
<td>\</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1484</td>
<td>1.23</td>
<td>21.81</td>
<td>42.85</td>
<td>15.63</td>
<td>Yes</td>
<td>1484</td>
</tr>
<tr>
<td>Testing sample</td>
<td>6</td>
<td>X6</td>
<td>1150</td>
<td>1.22</td>
<td>15.02</td>
<td>16.68</td>
<td>6.88</td>
<td>No</td>
<td>\</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2648</td>
<td>1.42</td>
<td>5</td>
<td>14.91</td>
<td>47.65</td>
<td>No</td>
<td>\</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>X10</td>
<td>3439</td>
<td>1.42</td>
<td>2.94</td>
<td>15.54</td>
<td>50.64</td>
<td>No</td>
<td>\</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4151</td>
<td>1.42</td>
<td>1.03</td>
<td>15.55</td>
<td>45.58</td>
<td>No</td>
<td>\</td>
</tr>
</tbody>
</table>

**Table 1** Sample set established from the horizontal well data in XX block
The characteristic parameters have different physical meaning and dimensions such that the original sample data need to be treated first; otherwise, error in the result becomes large. Linear normalization method is used to convert the data in the values between [0,1]. The linear normalized transformation equation is expressed as follows:

\[
\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}
\]  

(7)

In the formula, \(x_i\), \(\bar{x}_i\), \(x_{\min}\), \(x_{\max}\) represent the initial sample data, the data after normalization, as well as the minimum and maximum values in the initial sample data for each sample.

2.4 Model training and diagnostic parameters

At present, the problem of optimizing parameters to build the best model has not yet been thoroughly solved in theory. In practice, the training sample is used to build different SVM diagnosis models by gradually changing the parameter values and utilizing the established diagnosis model to calculate the test samples. Furthermore, the best performing parameters are chosen according to the average relative error of calculation results.

To measure the effect of training and prediction, the following average relative error is used as the evaluation index [10]:

\[
e = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y'_i - y_i}{y_i} \right) \times 100.
\]  

(8)

In the formula, \(y'_i\) is the value of the number \(i\) sample calculated by SVM and \(y_i\) is the sample value of the number \(i\) sample. In the calculation of training and diagnostic accuracy, \(n\) represents the number of training samples and test samples (\(n\) is 5 in this study).

The SVM training algorithm is used to train the normalized learning samples. The model is used after training to conduct the forecast operation on the five groups of test samples shown in Table 1 (sample nos. 6–10). The kernel function types of the model, the size of the kernel function parameters, and the penalty factor are adjusted. The entire modeling process is completed when the result relative to the actual value has the minimum error.

The polynomial kernel function, Gauss radial basis kernel function, and Sigmod kernel function are plugged into the training set and testing set. Then, their resulting values are compared in Table 2.

<table>
<thead>
<tr>
<th>Kernel function type</th>
<th>Learning accuracy (e/%)</th>
<th>Diagnostic accuracy (e/%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynomial kernel</td>
<td>4.2</td>
<td>5.35</td>
</tr>
<tr>
<td>Gauss radial basis kernel function</td>
<td>1.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Sigmod kernel function</td>
<td>2.1</td>
<td>3.3</td>
</tr>
</tbody>
</table>

On the basis of the result of comparative analysis, the Gauss radial basis kernel function is selected.

The concrete form of the Gauss radial basis kernel function is [11]

\[
K(x, x_i) = \exp\left(-\frac{||x - x_i||^2}{\sigma^2}\right)
\]  

(9)

In the formula, \(\sigma\) is the width parameter of the Gauss radial basis kernel function. The generalization capability of the kernel function is weakened as the parameter value increases.

A case study of the SVM model for the collapse of horizontal well borehole wall is conducted. The selection process and selection method of model parameters are discussed. After performing several parameter adjustments, three parameters of the model are found to have control actions: (1) the Gauss radial basic function \(\sigma\), (2) the controlling error \(\varepsilon\), and (3) the penalty factor \(C\) (positive constant; the value of \(a_i, a_i^*\) in Formula (6), i.e., \(a_i, a_i^* \in [0, C]\)). Figure 1 describes the change in learning accuracy and prediction accuracy when selecting parameters \(\varepsilon\) with fixed \(\sigma\) and \(C\) values and variable \(\varepsilon\) value.
The parameter optimization must minimize the learning accuracy and prediction precision simultaneously. Figure 1 shows that, when $\gamma = 0.87$, learning accuracy and prediction precision are the smallest. Thus, $\gamma = 0.87$ is selected as the parameter for the Gauss radial basic function of the model. The penalty factor is $C = 10$ and the controlling error is $\varepsilon = 0.01$.

### 2.5 Application effect

Referring to Table 3, which lists the relative error when applying the model to calculate the test set, and using Formula (10), the computing method for the prediction of the occurrence of borehole collapse is obtained. The results show that the average relative prediction error of the test sample is below 5%; this value meets the engineering requirement.

$$\sigma = \left| \frac{X - X_i}{X} \right| \times 100\%$$  \hspace{1cm} (10)

In the formula, $\sigma$ is the prediction error, $X$ is the sum total of the current sample, and $X_i$ is the point of the correct diagnostic data.

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Collapse occurrence prediction</th>
<th>Calculation results</th>
<th>Collapse position</th>
<th>Fractional error/%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual value</td>
<td>Calculated value</td>
<td>Actual value /m</td>
<td>Calculated /m</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>No</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>No</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>No</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>No</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>No</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
</tbody>
</table>
Figures 2 to 6 show the data for sample nos. 6–10. The diagnosis values of collapse position using the SVM method are compared with the actual values of the samples. As shown in these figures, the degree of coincidence of the calculated value and sample value is acceptable. Therefore, the SVM method has high learning and generalization capability.
3. Comparative analysis of the different input parameters for the SVM method

Two kinds of data are used as input parameters for the SVM method, and their corresponding results are compared and analyzed. The calculation results using ECD, massic energy, and friction/torque as input parameters are compared with the calculation results using direct measurement of the ground parameters(Figure 7 and Table 4).

**Table 4** Comparison of different input parameters for the SVM method-based diagnosis

<table>
<thead>
<tr>
<th>Sample number</th>
<th>Actual value</th>
<th>SVM method (The input parameter is ECD, and so on)</th>
<th>SVM method (The input parameter is surface parameter)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calculated value</td>
<td>Fractional error %</td>
<td>Calculated value</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>
The comparison results show that the diagnostic accuracy using ECD, massic energy, and friction/torque as input parameters is higher than that directly using of the ground parameter data. Although changes in ground parameters can reflect the underground condition, such changes can be attributed to many factors. Thus, the measured change is not necessarily caused by borehole collapse. Therefore, this study demonstrates that the SVM model using ECD, massic energy, and friction/torque as input parameters improves the accuracy of borehole collapse diagnosis.

### 4. Conclusion

1. A diagnostic model for well borehole collapse based on SVM is developed. Unlike the direct use of ground data, the use of ECD, massic energy, and friction/torque as input parameters for the model significantly improved the accuracy by 6.93%.

2. By adopting a microelement method and a soft model, a model for calculating friction/torque is established. An accurate positioning of collapse location is achieved. The friction/torque of the entire well is analyzed to diagnose the collapse risk. The result shows that the error for collapse position is generally within 10 meters.

3. Data from X1–X10 wells in XX block are used to establish the training set and test set for the diagnosis model. Accordingly, the optimal kernel function, kernel parameters, operate miss, and penalty factor are selected. Accordingly, a diagnostic error setoff less than 5% is obtained. The proposed diagnosis model for the collapse of horizontal well borehole wall is feasible. The diagnosis results show that the proposed model has high accuracy and generalization capability and can minimize the empirical risk effectively.

### 5. ACKNOWLEDGEMENTS

The authors acknowledge the mega project of applied basic research in Sichuan province Research on safe drilling data mining and intelligent early warning technology (Grant: JY0049).

### 6. REFERENCES


8. Andreas Fischer, Gerd Langensiepen, Klaus Luig, Rico Strasdat & Thorsten Thies. Efficient optimization of hyper-parameters for

