Casing life prediction using Borda and support vector machine methods

Xu Zhiqian[1], Yan Xiangzhen[2]* and Yang Xiujuan[2]

[1] College of Mechanical and Electronic Engineering, China University of Petroleum, Dongying, Shandong 257061, China
[2] College of Storage & Transportation and Architectural Engineering, China University of Petroleum, Dongying, Shandong 257061, China

Abstract: Eight casing failure modes and 32 risk factors in oil and gas wells are given in this paper. According to the quantitative analysis of the influence degree and occurrence probability of risk factors, the Borda counts for failure modes are obtained with the Borda method. The risk indexes of failure modes are derived from the Borda matrix. Based on the support vector machine (SVM), a casing life prediction model is established. In the prediction model, eight risk indexes are defined as input vectors and casing life is defined as the output vector. The ideal model parameters are determined with the training set from 19 wells with casing failure. The casing life prediction software is developed with the SVM model as a predictor. The residual life of 60 wells with casing failure is predicted with the software, and then compared with the actual casing life. The comparison results show that the casing life prediction software with the SVM model has high accuracy.

Key words: Support vector machine method, Borda method, life prediction model, failure modes, risk factors

1 Introduction

Casing failure is a result of a complex interaction among rock strength, pore pressure, casing program, cementing design, production, stimulation treatments, corrosion, earth stress, and engineering practice. Understanding the failure-causing factors can improve well planning, material procurement, field operations and increase the casing life. There are many empirical approaches for predicting casing life. However, they often are not very accurate and suffer from a number of drawbacks. Recently, robust modeling with a support vector machine (SVM) is gaining significant popularity for various advantages including powerful generalization performance and hence a plethora of applications are being considered, e.g. bank performance prediction (Ravi et al, 2008), bankruptcy predictions (Shin et al, 2005), financial forecasting (Kim, 2003; Ding et al, 2008). Similar to the neural network based techniques, the SVM based modeling also involves training and testing of data instances such that the training set is comprised of target outcome variables by mapping several predictor variables. The advantages of SVM include strong inference capacity, generalization ability, fast learning capacity and ability for accurate predictions. The Borda method is characterized as a choice method when the preferences are expressed by means of linear orders (crisp, asymmetric, complete and

transitive relations) (Young, 1974). In this paper, with the comprehensive consideration of casing failure modes and risk factors, the casing life prediction model is completed with SVM and Borda methods.

2 Risk indexes calculation for casing failure modes based on the Borda method

2.1 Casing failure modes and risk factors

The casing failure modes mainly include casing burst, casing collapse, tensile failure, shear failure, slipping and sticking or galling of the threads, thread leakage, and casing instability (lateral deflection, i.e. buckling). The limit state equation based on different failure criteria is expressed as follows:

$$Z = P - P' \begin{cases} Z > 0, & \text{non-failure} \\ Z < 0, & \text{failure} \\ Z = 0, & \text{critical status} \end{cases} \quad (1)$$

where Z, P and P' are failure discriminant value, failure strength, and load stress, respectively.

In this paper, 32 risk factors that cause casing failure are summarized with a comprehensive consideration of geological conditions (Peng et al, 2007), casing program and cementing (Li, 2008), oil and gas production (Yu et al, 2009) and stimulation treatments, earth stress (Chiotis and Vrellis, 1995), and casing corrosion (Carey et al, 2009). There are 16 casing load-influencing factors (01-16) and 16 strength-influencing factors (17-32) listed in Table 1.

Table 1 A list of 32 risk factors that cause casing failure

Casing load-influencing factors	Casing strength-influencing factors
01 Fluid pressure in the casing	17 Low compression strength
02 Pressure caused by high-pressure water injection and fracturing	18 Casing perforations
03 Fluid static pressure in the annulus between the production tubing and casing	19 Cement sheath with uneven thickness
04 Reservoir fluid pressure	20 Manufacture and wear defects
05 Tectonic earth stress	21 Poor cement quality and corrosion
06 Creep of rock salt and swelling of mudstones	22 Low collapse resistance
07 Gravity force acting on the lower part of the casing	23 Uniformly-distributed external load
08 Additional load from cementing operations	24 Poor sealing performance of casing threads
09 Ground settlement	25 Low tensile strength
10 Gravity force acting on the upper part of the casing	26 Ignoring shearing strength
11 Vertical earthquake load	27 Large centralizer spacing or centralizer failure
12 Large dip angle	28 Cavities caused by sand production
13 Faults and earthquakes that cause tangential stress on the casing	29 Low tensile strength of casing threads
14 Horizontal earthquake load	30 Poor-quality threads
15 Additional load from stimulation treatments	31 Low compression strength of casing threads
16 Axial load from formation expansion in the vertical direction caused by water injection in sandstone reservoirs	32 Low strength caused by high temperature and thermal stress

2.2 Failure probability

The probability distribution functions of the casing strength and the load are modeled as normal distributions. This implies that the casing failure probability and reliability can be computed from Eqs. (2) and (3), when the mean value and the standard deviation of the casing strength and the load applied are known. Details of these equations have been previously described (Christensen and Baker, 1982; Milton and Arnold, 1986; Rahman and Chilingarian, 1995).

$$P_f = P(Z < 0) = \Phi\left(\frac{\mu_L - \mu_{RS}}{\sqrt{\sigma_{RS}^2 + \sigma_L^2}}\right) = \Phi(Z_f) \qquad (2)$$

$$R = 1 - P_f \qquad (3)$$

where P_f and R denote the casing failure probability and reliability, respectively; μ_L and μ_{RS} are the mean values of the casing strength and the load applied, respectively; σ_L and σ_{RS} are the standard deviations of the casing strength and the load applied; $\Phi(\cdot)$ represents standard normal distribution function; Z_f is the probability coefficient of the casing failure, $\Phi(Z_f)$ can be directly obtained from a normal distribution table.

2.3 Borda count of each factor for casing failure

The relation between the reliability R and the influence grade are given, as shown in Table 2. The occurrence probability is divided into five levels and the grade scores are listed in Table 3.

Table 2 Corresponding relation between the influence grade and the casing reliability

Influence grade	Grade score	Definition or description	Reliability %
Critical	5	Once the event occurs, it will result in the corresponding mode failure.	0-30
Serious	4-5	Once the event occurs, it will lead to a serious decline in safety performance of the corresponding mode.	30-60
Moderate	3-4	Once the event occurs, the corresponding mode is moderately affected and the casing strength still meets the basic requirements.	60-90
Minor	2-3	After the incident, it has little influence on the corresponding mode and the casing strength still meets the basic requirements.	90-99.999
Negligible	1-2	After the incident, there are no effects on the corresponding mode.	>99.999

Table 3 Grade score interpretation of occurrence probability

Occurrence probability, %	Grade score	Definition or description
0-10	1-2	Absolutely not occur
10-40	2-3	Rarely occur
40-60	3-4	May occur
60-90	4-5	Likely to occur
90-100	5	Very likely to occur

Based on the quantitative grade scores of the influence degree and the occurrence probability, the Borda counts of risk factors for eight failure modes are obtained as follows (Lamboray, 2007):

$$b_{ij} = \prod_{k=1}^{2} r_{ij}^{k} \qquad (4)$$

where subscript i is the ith risk factor that causes casing failure, j is the jth casing failure mode; k denotes the evaluation criterion, $k=1$ represents the influence criterion and $k=2$ represents the probability criterion.

After obtaining the Borda counts of 32 risk factors, the Borda matrix is expressed as follows:

$$\boldsymbol{B} = \left[b_{ij} \right]_{m \times n} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \qquad (5)$$

where m is the number of risk factors, $m=32$; and n is the number of failure modes, $n=8$.

2.4 Risk index of casing failure modes

The risk index represents the risk level priority associated with each failure mode. The Borda matrix elements are weighted according to the column averaging. The risk index for each failure mode is defined as follows:

$$d_j = \frac{a_j}{m} \sum_{i=1}^{m} b_{ij} \qquad (6)$$

where d_j represents the risk index corresponding to the failure mode j, $\boldsymbol{D} = \begin{bmatrix} d_1 & d_2 & \cdots & d_j \end{bmatrix}$ is a row vector of risk index; and a_j represents the weight corresponding to the failure mode j.

The weight set A is determined according to engineering experience as follows:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_8 \end{bmatrix}$$
$$= \begin{bmatrix} 0.118 & 0.106 & 0.162 & 0.172 & 0.156 & 0.148 & 0.042 & 0.096 \end{bmatrix}^{\mathrm{T}}$$

3 Casing life prediction model based on SVM

3.1 Theory of support vector machine

It is very difficult to develop a accurate correlation between the risk factors and the casing life. So, the support vector machine (SVM) method is proposed to predict the casing life in troublesome wells. The fitting function is defined as follows (Vapnik, 1998; Zhang, 2003; Sebakhy, 2007):

$$y(x) = \sum_{i=1}^{L} (\alpha_i'' - \alpha_i') K(x, x_i) + b_0 \qquad (7)$$

$$b_0 = \frac{1}{L} \sum_{i=1}^{L} y_i - \frac{1}{L} \sum_{i,j=1}^{L} (\alpha_i'' - \alpha_i') K(x_i, x_j) \qquad (8)$$

where L is the total number of support vectors, x_i represents a set of input vectors; $y(x)$ refers to target vector, α_i'', α_i'' and b_0 are undetermined coefficients in the SVM model.

In this paper, the radial basis kernel function is considered for SVM modeling.

Radial basis kernel function:

$$K(x, x_i) = \exp(-\gamma |x - x_i|^2) \qquad (9)$$

where γ is taken as constant.

3.2 Basic computing units of casing reliability

Owing to the axial differences along the well casing, it is difficult to calculate the casing reliability as a whole. In this study, the casing is divided into several computing units with the differences among the well sections. The division criteria are summarized as follows:

1) Casing within the same unit has the same geometry and mechanical properties, and load change is linear. There must be one great difference among them within different units at least.

2) Easy to interface with the traditional calculation method and consistent with the actual engineering practice.

3.3 Input vector and output vector determination

Selecting the optimal input and output vectors is an important step in SVM modeling. It is very crucial for accurately predicting the casing life with SVM. The row vector $\boldsymbol{D}$ (risk index vector) obtained from Eq. (6) is defined as input vector, and the life of every unit is defined as output vector.

3.4 Training set selection and model parameter determination

Totally 19 wells with casing failure are selected and their basic data are listed in Table 4.

The 32 risk factors are extracted from the 19 wells with casing failure, and their corresponding Borda counts and the Borda matrix are computed from Eq. (4) and Eq. (5). The row vector $\boldsymbol{D}$ composed of eight risk indexes is derived from Eq. (6). Finally, 19 row vectors $\boldsymbol{D}$ from 19 wells and their corresponding actual lives compose the training set.

By comparing the predicted lives with the actual, the ideal model parameters can be conducted as follows: penalty coefficient $C = 80000$, calculation parameter of the kernel function $\gamma = 104.6$ and control error $\varepsilon = 0.06$.

The lives of the original samples are predicted using the SVM model, and a comparison between the predicted values and the actual is shown in Fig. 1. This figure shows that the tested samples have an absolute deviation less than 1.00 year. So the model parameters given above is optimal.

3.5 Result output of casing life prediction

In this paper, a SVM-based model is proposed to predict the casing life in troublesome wells. The casing life of each computing unit is obtained by inputting the corresponding row vector $\boldsymbol{D}$ into the life prediction model. Obviously, the minimum value of all units' life is the ultimate casing life.

Table 4 Condition data from 19 wells with casing failure

Sample	Damage depth m	Failure mode	Casing structure at the failure location	Formation lithology at the failure location	Actual life, Years	Predicted life, Years
1	952.5	Burst	A single casing + mud outside the casing	Argillaceous sandstone	5.93	5.44
2	1192.0	Collapse	A single casing + external cement sheath	Mudstone	9.02	8.91
3	1171.2-1185.9	Collapse	A single casing + external cement sheath	Rock salt	4.53	4.97
4	1159.0	Shear failure	A single casing + external cement sheath	Arenaceous shale	5.93	6.02
5	820.9-1160.7	Leakage	A single casing + mud outside the casing	Argillaceous sandstone	11.30	11.26
6	3175.9	Collapse	A single casing + external cement sheath	Argillaceous sandstone	8.39	8.25
7	1814.3	Burst	A single casing + mud outside the casing	Argillaceous sandstone	11.54	11.01
8	3025.0-3050.0	Leakage	A single casing + external cement sheath	Argillaceous sandstone	10.50	9.58
9	3324.1-3448.0	Collapse	A single casing + external cement sheath	Argillaceous sandstone	6.20	6.14
10	3206.5	Collapse	A single casing + external cement sheath	Sandstone	4.82	5.26
11	3078.6	Collapse	A single casing + external cement sheath	Rock salt	9.75	8.87
12	3518.0	Collapse	A single casing + external cement sheath	Mudstone	7.84	7.15
13	3099.9	Collapse	A single casing + external cement sheath	Rock salt	6.75	6.25
14	2285.0	Collapse	A single casing + external cement sheath	Mudstone	7.20	6.96
15	1143.4	Burst	A single casing + mud outside the casing	Arenaceous shale	4.93	5.51
16	1024.5	Shear failure	A single casing + external cement sheath	Shale	4.25	4.42
17	925.8-1206.1	Shear failure	A single casing + external cement sheath	Arenaceous shale	5.75	5.56
18	4685.0-4923.1	Leakage	A double casing + external cement sheath	Mudstone/limestone/sandstone	8.62	8.32
19	4673.2-4801.4	Leakage	A double casing + external cement sheath	Argillaceous sandstone	7.56	7.50
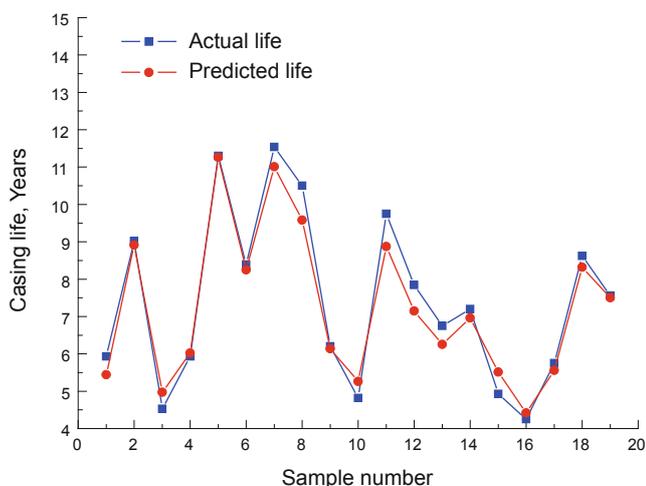
Fig. 1 A comparison between predicted and actual casing life for the training set

The basic structure of the casing life prediction model is shown in Fig. 2.

4 Development and validation of casing life prediction software

4.1 Software development

The SVM model has been trained with actual data. The casing life prediction software is completed with the SVM model as the predictor. The software input parameters mainly include four categories: geological conditions, casing program and cementing, oil and gas production and stimulation treatments, and earth stress and casing corrosion. Thirty-two risk factors are considered in this software, such as rock properties and geologic position, existence of faults and oil reservoirs, casing nominal diameter, wall thickness, steel grade, weight, wear, cementing, perforation, water injection, fracturing, thermal recovery, acidification and other stimulation treatments, earth stress, and casing corrosion.

First, the casing is divided into several computing units according to the axial differences of the casing material, the strength, and the load. Second, 32 risk factors are extracted from the input parameters. Based on the quantitative grade
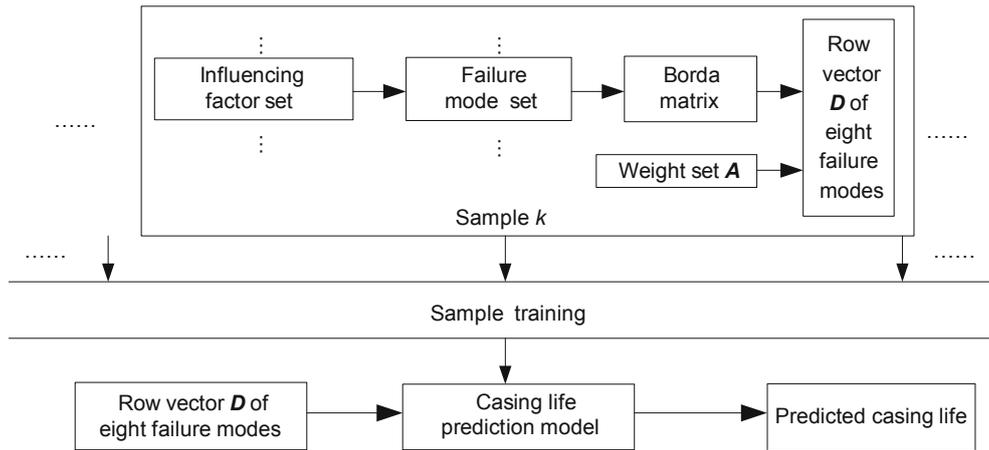
Fig. 2 Basic structure of the casing life prediction model based on SVM

scores of influence degree and occurrence probability, the Borda counts of each risk factor for eight failure modes are obtained. Then the risk indexes of failure modes are computed with the Borda matrix. Finally, the life of each computing unit is predicted with the SVM model as predictor and the eight risk indexes as input vectors. The ultimate casing life is the minimum value of all units' lives.

4.2 Validation

The residual life of 60 wells with casing failure is calculated with the developed program. Fig. 3 shows that 59.2% of the tested samples have an absolute deviation less than 1.0 year and 67.8% of the tested samples have a relative deviation less than 25%. The comparison results indicate that the accuracy of the SVM model is high.
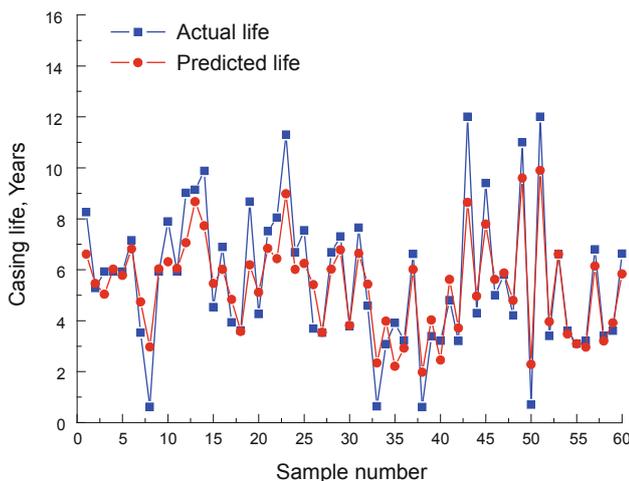
Fig. 3 A comparison between the predicted and actual casing life for non-training sets

5 Conclusions

1) This paper presents 32 risk factors that cause casing failure, and gives a simple calculation method and steps

for calculating the casing failure probability. Based on the quantitative grade scores of the influence degree and the occurrence probability, the risk index for each failure mode is obtained with the Borda method.

2) The SVM method is proposed to establish a relationship between the risk factors and the casing life. For the casing life prediction model, eight risk indexes are defined as input vectors and the casing life is defined as the output vector. The ideal model parameters are obtained with the training set from 19 wells with casing failure.

3) The casing life prediction software is developed with the SVM model as a predictor, and the main life prediction steps are introduced. Comparing the predicted life with actual value from 60 wells with casing failure, the accuracy of the prediction model is high.

Acknowledgments

References

Carey J W, Svec R, Grigg R, et al. Wellbore integrity and CO_2-brine flow along the casing-cement microannulus. Greenhouse Gas Control Technologies. 2009. 1(1): 3609-3615

Chiotis E and Vrellis E. Analysis of casing failures of deep geothermal wells in Greece. Geothermics. 1995. 24(5-6): 695-705

Christensen P T and Baker M J. Structural Reliability Theory and its Applications. Berlin: Heidelberg, New York: Springer-Verlag. 1982. 85-86

Ding Y S, Song X P and Zen Y M. Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Systems with Applications. 2008. 34(4): 3081-3089

Kim K J. Financial time series forecasting using support vector machines. Neurocomputing. 2003. 55(2): 307-319

Lamboray C. A comparison between the prudent order and the ranking obtained with Borda's, Copeland's, Slater's and Kemeny's rules. Mathematical Social Sciences. 2007. 54(1): 1-16

Li Z F. Casing cementing with half warm-up for thermal recovery wells. Journal of Petroleum Science and Engineering. 2008. 61(2-4): 94-98

Milton J S and Arnold J C. Probability and Statistics in the Engineering and Computing Sciences. New York: McGraw-Hill Higher

Education. 1986. 226-237

Peng S P, Fu J T and Zhang J C. Borehole casing failure analysis in unconsolidated formations: A case study. Journal of Petroleum Science and Engineering. 2007. 59(3) 226-238

Rahman S S and Chilingarian G V. Casing Design Theory and Practice. Amsterdam, Netherlands: Elsevier Science B V. 1995. 56-60

Ravi V, Kurniawan H, Kumar P R, et al. Soft computing system for bank performance prediction. Applied Soft Computing. 2008. 8(1): 305-315

Sebakhy E, Sheltami T, Bokhitan S, et al. Support vector machines framework for predicting the PVT properties of crude-oil systems. Paper SPE 105698-MS presented at SPE Middle East Oil and Gas Show and Conference, 11-14 March 2007, Kingdom of Bahrain

Shin K S, Lee T S and Kim H J. An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications. 2005. 28(1): 127-135

Vapnik V. Statistical Learning Theory. John Wiley. New York. 1998. 34-37

Young H P. An axiomatization of Borda's rule. Journal of Economic Theory. 1974. 9: 43-52

Yu Z H, Wang L Y and Liu D W. Investigation on production casing in steam-injection wells and the application in oilfields. Journal of Hydrodynamics. 2009. 21(1): 77-83

Zhang H R, Han Z Z and Li C G. Support vector machine based nonlinear model predictive control. Systems Engineering and Electronics. 2003. 25(3): 330-334

(Edited by Sun Yanhua)