Abnormal identification of lubricating oil parameters and evaluation of physical and chemical properties based on machine learning

Kun Yang¹,² and Xia Wang¹,²

¹Reliability Engineering Institute, School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China

²National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan 430063, China

E-mail: kunyangwhut@163.com

Abstract. This paper firstly proposes an abnormality identification model for online oil monitoring parameters. The SVR model was established to predict the output power, and the information entropy of the deviation between the predicted value of the regression model and the actual monitoring value was used to identify the abnormal parameters of the oil. The physical and chemical properties of the oil was evaluated and predicted comprehensively. Principal component analysis (PCA) was used to eliminate the correlation between parameters, and then the deviation degree of each principal component is calculated after feature transformation. After adding the deviation degrees of each principal component according to their weights, the performance evaluation index of the oil is obtained. At the same time, a prediction model of oil performance trend based on Gated Recurrent Unit is proposed. The feasibility and applicability of this model are verified by comparing with the results of LSTM neural network and ARIMA model. The evaluation of the physical and chemical properties of the oil can detect the deterioration of the oil condition at the early stage, avoiding or reducing accidents, so as to ensure the safe and reliable operation of the unit.

1. Introduction
The lubricating oil film between the various components of the steam turbine can reduce friction and provide an additional heat transfer path, thereby helping the steam turbine maintain mechanical efficiency, extend the service life and ensure its normal operation. Steam turbines are highly
dependent on lubricating oil for normal operation, so oil products have an important influence on the working state of the unit. At the same time, a lot of wear particles caused by friction of steam turbine parts are also suspended in the lubricating oil. Obviously obtaining real-time oil quality and worn metal element information and taking timely measures can ensure that the steam turbine is in good condition. The extensive application of large-scale machinery and equipment has made oil monitoring an important tool for fault diagnosis. However, most oil monitoring systems currently use parameter threshold over-limit alarms with a single alarm mode and no function to prevent deterioration. Therefore, based on machine learning, this paper proposes an abnormal identification model of lubricating oil parameters and an evaluation model of lubricating oil physical and chemical properties.

2. Anomaly recognition model
Under normal circumstances, the various monitoring parameters of lubricating oil fluctuate in a wide range, it is difficult to visually see its inherent laws and potential fault information. Therefore, based on the oil monitoring data for a period of time before the failure, we propose a model combining support vector regression and information entropy to evaluate the state of the oil parameters and predict potential failures. When the equipment fails or is in an abnormal operating state, its related operating parameters will change abnormally and deviate from the normal operating range. Then the residuals predicted by the model will show dramatic changes, so we propose to use information entropy to measure its degree of change. Information entropy quantifies the degree of ordering of the system or the complexity of the signal, and is of great significance for the identification of parameter anomalies.

2.1. Parameter prediction method based on SVR
Support vector machine (SVM) is a statistical learning method first proposed by Vapnik[1]. In recent years, SVM is used to solve the regression problem, that is, support vector regression (SVR)[2-3]. The basic idea that SVR is used to solve regression prediction is as follows: For the regression problem with the sample set \((x_1, y_1), \cdots (x_l, y_l)\), first find a nonlinear mapping \(\Phi(x)\) from input space to the output space. The data in the sample set is mapped to a high-dimensional space, and linear regression is performed in this space in the feature space using the following linear function, that is:

\[
    f(x) = w \times \Phi(x) + b, w \in R^n, b \in R
\]  

(1)

So, the sum of the empirical risks affecting \(w\) and the \(w^2\) that makes it flat in high-dimensional space, that is,

\[
    R(w) = \frac{1}{2} w^2 + \sum_{i=1}^{l} \epsilon(f(x_i) - y_i)
\]  

(2)

In the formula: \(l\) represents the number of samples; \(\epsilon\) is a loss function, which only calculates sample points other than \([-\epsilon, \epsilon]\), which is defined as follows:

\[
    \epsilon(f(x_i) - y_i) = \begin{cases} 
    0, & |f(x_i) - y_i| < \epsilon \\
    |f(x_i) - y_i| - \epsilon, & |f(x_i) - y_i| > \epsilon 
    \end{cases}
\]  

(3)
There are errors in the model during the fitting process. So we can introduce non-negative relaxation variables $\xi_i$ and $\xi_i^*$:

$$\frac{1}{2} w^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$$

$$\begin{cases} y_i - w \cdot \Phi(x_i) - b \leq \varepsilon + \xi_i \\ w \cdot \Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

(4)

Lagrange multipliers $a, a^*$ are introduced here, and the above problems can be transformed into dual problems:

$$\begin{aligned}
\min & -\frac{1}{2} \sum_{i,j=1}^{l} (a_i^* - a_i)(a_j^* - a_j)(\Phi(x_i), \Phi(x_j)) - \varepsilon \sum_{i=1}^{l} (a_i^* + a_i) + \sum_{i=1}^{l} y_i (a_i^* - a_i) \\
\text{s.t.} & \quad \sum_{i=1}^{l} (a_i^* - a_i) = 0 \\
& \quad 0 \leq a_i \leq C \\
& \quad 0 \leq a_i^* \leq C \\
& \quad i = 1, 2, \ldots, n
\end{aligned}$$

(5)

Therefore, the function regression problem of SVM can be reduced to the quadratic programming problem. Solving the quadratic programming problem, we can get the value of $w$:

$$w = \sum_{i=1}^{l} (a_i - a_i^*) \Phi(x_i)$$

(6)

Where $a_i$ and $a_i^*$ are the solutions that minimize $R(w)$, and the linear regression function can be obtained from:

$$f(x) = \sum_{i=1}^{l} (a_i - a_i^*) (\Phi(x_i), \Phi(x)) + b = \sum_{i=1}^{l} (a_i - a_i^*) K(x, x_i) + b$$

(7)

Where $K(x, x_i) = \Phi(x_i) \cdot \Phi(x)$, $K(x, x_i)$ is called the kernel function. Selecting different forms of kernel functions can generate different SVMs. Commonly used kernel functions include: radial basis functions, polynomial functions, Sigmoid kernel functions, linear functions and so on.

2.2 Anomaly recognition method based on information entropy

When using information entropy to identify the abnormality of monitoring indicators, the first step is to use SVR model to fit the sample. The second step is to calculate the residual of the model. The size of the residual can reflect the state of the parameter. Usually, the smaller the residual, the less severe the anomaly. However, when the trained SVR model is used for the abnormal identification of the
state parameters of the same oil, the residuals and RMSE of the model will shift. Therefore, the method of abnormal identification using residuals is not universal. Because the information entropy proposed by Shannon can be used to measure the degree of ordering of the sequence, we propose an abnormality identification model of oil monitoring parameters based on information entropy[4]. The definition of information entropy is as follows:

$$H_a = -\sum_{i=1}^{N} P_i(x_i)\ln P_i(x_i)$$  \hspace{1cm} (8)$$

Where $H_a$ represents information entropy; $x_i$ refers to a random event that may occur; $P_i$ refers to the probability that an event $x_i$ occurs.

In order to measure the degree of ordering of the residuals, a continuous time series is now divided into time intervals $T_i$ at the same interval, which are divided into a total of $N$; In addition, the number of residuals in each time period is counted and recorded as $n_i$; The information entropy of the model prediction residual within this time period is set to 3 hours in this paper. Therefore, the information entropy $H_e$ formula of the model residual is defined as follow:

$$H_e = -\sum_{i=1}^{N} \left( \frac{n_i}{T_i} \right) \ln \left( \frac{n_i}{T_i} \right)$$  \hspace{1cm} (9)$$

3. Physical and chemical performance evaluation
3.1 Evaluation model of physical and chemical properties of oil
The calculation of the deviation of a single parameter is introduced first. An important parameter in the calculation of the index of the physical and chemical properties of oil is the deviation. First, we analyse the calculation method of the deviation when a single monitoring parameter changes. The optimal value will change with the specific environment of the equipment operation, so the concept of parameter deviation coefficient $s$ is proposed:

$$s = \begin{cases} 
\frac{x_0 - x}{x_0} & x_{F1} \leq x \leq x_{F2} \\
\end{cases}$$  \hspace{1cm} (10)$$

In the formula, $x_{F1}$ is the lower bound of the monitoring parameter failure; $x_{F2}$ is the upper bound of the monitoring parameter failure; $x_0$ is the optimum operation value of the monitoring parameter.

In addition, the degree of deviation $S$ of the parameters needs to be defined again to compare the deviation between different parameters:
\[ S = \frac{S}{S_{\text{max}}} = \begin{cases} \frac{x_0 - x}{x_0 - x_f} & x_{F_1} \leq x \leq x_{F_2} \\ 0 & x_{0_1} \leq x \leq x_{0_2} \end{cases} \] (11)

\( x_f \) represents the failure threshold of the monitoring parameter. If the parameter has the best monitoring range \([x_{0_1}, x_{0_2}]\), then the parameter deviation coefficient is defined as follows:

\[ S = \begin{cases} \frac{x_{0_1} - x}{x_{0_1}} & x_{F_1} \leq x \leq x_{0_1} \\ 0 & x_{0_1} \leq x \leq x_{0_2} \\ \frac{x_{0_2} - x}{x_{0_2}} & x_{0_2} \leq x \leq x_{F_2} \end{cases} \] (12)

The calculation formula of the deviation degree \( S \) is as follow:

\[ S = \begin{cases} \frac{x_{0_1} - x}{x_{0_1} - x_{F_1}} & x_{F_1} \leq x \leq x_{0_1} \\ 0 & x_{0_1} \leq x \leq x_{0_2} \\ \frac{x_{0_2} - x}{x_{0_2} - x_{F_2}} & x_{0_2} \leq x \leq x_{F_2} \end{cases} \] (13)

The physical and chemical performance evaluation indexes of oil products constructed in this section are based on PCA and the calculation of the deviation of single monitoring parameters. For the oil monitoring parameter sample sequence \( x_i (i = 1, 2, \cdots, N) \), \( N \) represents the number of samples. Assume that the original sample sequence has \( t \) features. After the principal component transformation, \( t \) independent principal components are obtained. That is, the parameter at a certain time is \( x_i = [PCA_1^i, PCA_2^i, \cdots, PCA_t^i]^T \). Now, according to the theory of single monitoring parameter, the calculation formula of each principal component deviation degree can be defined as:

\[ S_{\text{pcai}} = \frac{\sum_{k=1}^{n} \mu_{ik} S_k - \sum_{i=1}^{f} \mu_{ik}^{'} \mu_{ik}}{\sum_{k=1}^{n} |\mu_{ik}|} \] (14)

In the formula, \( S_k \) is the deviation degree of the monitoring parameter \( p_k \); \( \mu_{ik} \) is the \( k \)-th element of the feature vector of the \( i \)-th principal component; \( \mu_{ik}^{'} \) is the negative element of the feature vector of the \( i \)-th component; \( f \) is the number of negative feature vectors of the \( i \)-th component.
The results obtained by adding the deviations of the principal components according to weights are recorded as $\Delta R$, then we can use $\Delta R$ to measure the deterioration of the oil. The expression of $\Delta R$ is as follow:

$$\Delta R = \sum_{i=1}^{n} \alpha_i S_{pcai}$$

(15)

Where $S_{pcai}$ is the deviation degree of the $i$-th principal component ($i = 1, 2, \cdots, n$); $\alpha_i$ is the weight occupied by the $i$-th principal component. So far, the deterioration state of the oil over time can be measured by $\Delta R$, and $\Delta R \in (0,1)$.

3.2 Prediction model of physical and chemical properties of oil

The neural network uses multiple neurons and multilayer network settings to significantly improve its ability to characterize nonlinearity. Based on this, the deep learning method RNN uses its own neurons with feedback to introduce loops. The mechanism makes the neural network have the ability of memory, so it can process sequences of any length. Therefore, it is more suitable for sequence prediction than other neural networks. However, traditional RNNs lack the ability to learn for a long time due to problems such as disappearance of gradients, which prevents their prediction result from being improved\cite{5}. Hochreiter and Schmidhuber proposed the Long Short-Term Memory (LSTM) in 1997\cite{6-8}. The "gating unit" can solve long-term problems. The gating unit of long-term and short-term memory networks includes input gates, forget gates, and output gates.

![Figure 1. GRU network structure expansion](image)

In LSTM networks, input gates and forget gates are complementary. The neural network dimension solves the problem that the LSTM using these two gates at the same time is too complicated. The input gate and the forget gate are merged into one gate, that is, the update gate. Therefore, compared to LSTM networks, GRU networks are more concise. In addition, GRU does not introduce additional memory units. Its network structure is shown in Figure 1: where $y_{t-1}$ and $y_t$ represent...
the historical state and current state, \( x_t \) represents the current input, and \( r_t \in [0,1] \). To reset the gate, \( z_t \in [0,1] \) is the update gate\(^{[9,10]} \).

The following describes the working principle of the GRU neural network in simple mathematical language:

\[
\begin{align*}
    z_t &= \alpha(W_z \times [y_{t-1}, x_t]) \\
    r_t &= \alpha(W_r \times [y_{t-1}, x_t]) \\
    \bar{y}_t &= \text{tanh}(W \times [r_t \times y_{t-1}, x_t]) \\
    y_t &= (1 - z_t) \times y_{t-1} + z_t \times \bar{y}_t
\end{align*}
\]

\(\bar{y}_t\) refers to the real-time output of the neuron; \(y_{t-1}\) is the output of the previous neuron, \(\alpha\) is the Sigmoid function, and tanh is the hyperbolic tangent function. \(W_z\) refers to the weight occupied by the update gate. \(W_r\) refers to the weight occupied by the reset gate.

4. Experiment and analysis
4.1 Anomaly identification and analysis of oil monitoring parameters
SVR is first used to establish a regression estimation model for steam turbine output power. Taking a 600MW unit in a power plant as an example, select the density, viscosity, dielectric constant, temperature, 4μm particles, 6μm particles, 14μm particles, capacitance (pF), and moisture content (%) as the input parameters and the output is power. Within the normal operating range of the unit, the operating parameters corresponding to each measurement point are collected through the oil monitoring system. The sampling interval is 1 minute. 90% of the data is used as training samples for training and learning, and the remaining 10% of the data, perform test verification as a test sample. The data was normalized before training, and the model parameters \( C = 28 \) and \( \sigma = 0.53 \) are obtained through the grid search method and cross-validation method search. Figure 2 shows the training and prediction results of the model. It can be seen that the degree of agreement between the predicted value and the actual value is relatively satisfactory, and the error meets the allowable range of the engineering error, as shown in Table 1. At the same time, the BP neural network model is used to predict the output power of the prediction set, and the results are also shown in Figure 2.
Table 1. Evaluation index values of different models

| Model | Evaluation index |
|-------|------------------|
|       | MSE              | $R^2$         |
| SVR   | 0.0646           | 0.9749        |
| BP    | 0.1028           | 0.8956        |

The mean square error (MSE) and the degree of fit $R^2$ of the test set reflect the predictive power of the model. As can be seen from Table 1, the SVR neural network stands out in the prediction of the physical and chemical properties of the oil, and its prediction results also meet the high-precision prediction requirements.

On April 27, 2014, the turbine's oil system failed. Now select the oil monitoring data 10 days before the failure for research and analysis, and use the established SVR model to predict the power to obtain the residual, RMSE and information entropy corresponding to the model prediction, as shown in Figure 3. Figure 3(a) shows the residuals corresponding to the model predictions. It can be seen that the residuals are basically in the range of -20°C to 20°C. In the figure, a small number of points with large residuals occasionally appear, which may be due to the interference of the monitoring data itself or the accuracy of the prediction model. Figure 3(b) is the RMSE change of the model prediction residual. The figure shows that the RMSE value is less than 15MW in most of the time before the time period 60, which is in a normal working state. The amplitude of the RMSE at the time period 60 is almost reaching 23MW, the amplitude has been changing repeatedly until failure. Figure 3(c) shows the change of the residual entropy value during the prediction. The entropy value before the time period 60 is relatively small and fluctuates around 0.75, and it changes back and forth until the time point 60 increases to more than 1, and then decreases. It quickly became larger before the failure. Due to the failure, the output of the prediction model and the actual monitoring value have...
a large difference, and the changes of the residuals start to become chaotic and irregular, so we can measure this disorder by analysing the entropy of the residuals.

![Residual plot](image)

**Figure 3.** (a) Residual, (b) RMSE and (c) information entropy of power prediction in SVR model

### 4.2 Evaluation and prediction of physical and chemical properties of lubricants

In order to measure the degree of deviation of a monitoring parameter at a certain time, it is necessary to determine the optimal value range and failure threshold of the parameters. In this paper, the best monitoring range of each parameter can be obtained by using statistical methods combined with parameter characteristics and each failure mode. At the same time, the mean value of the warning value and the abnormal value obtained by the three-line value method is used as the failure threshold. Table 2 is the optimal value and the failure threshold range.

| Parameter          | Optimal range     | Failure threshold |
|--------------------|-------------------|-------------------|
| Density            | (0.866, 0.867)    | (0.841, 0.892)    |
| Viscosity          | (35.227, 35.293)  | (30.171, 40.352)  |
| Dielectric constant| (2.294, 2.299)    | (1.908, 2.686)    |
Next, a GRU neural network model is established to predict the trend of oil performance changes. The sample size selected for this modelling is 1800, with the first 90% as the training sample and the last 10% as the test sample. This article only studies the case where the neural network contains a hidden layer. Finally, the number of hidden units in the GRU network structure is determined to be 8. In addition, the batch size is set to 1, the learning rate is set to 0.01, and the number of iterations of the model, epoch, is 100.

This article introduces the comparison and analysis between models by introducing LSTM neural network and ARIMA. The prediction results are shown in Figure 4 and the relevant comparison results are shown in Table 3.
It can be seen from the comparison between different models that the GRU neural network stands out in the prediction of the physical and chemical properties of the oil, and its prediction results also meet the high-precision prediction requirements. It is found that when $\Delta R$ is larger than 0.5, the physical and chemical indicators of the oil will be significantly deteriorated, and the oil performance significantly decreased. In addition, when $\Delta R$ is close to 0.8, the oil should be changed in time. We can see that the $\Delta R$ in the second half of the figure shows a continuous upward trend, indicating that the oil of the unit will continue to deteriorate.

5. Conclusion

(1) An SVR-based model is established to predict the output power, and compared with the prediction effect of the BP neural network. The results show that the predicted output power of the SVR is closer to the actual value and the fitting accuracy is higher. Then, based on the combination of RMSE and information entropy, a complete set of state parameter anomaly detection methods is proposed, and the effectiveness of the method is verified.

(2) A comprehensive evaluation model for the physical and chemical properties of oil based on multi-parameter analysis is proposed. The parameters of each index interact with each other, so it is necessary to use PCA to convert each index parameter into an independent comprehensive index, then calculate the deviation degree of each principal component based on the theory of single parameter deviation degree, and finally convert the principal component's The result of adding the degree of deviation according to its weight can be used to evaluate the physical and chemical properties of the oil.

(3) At the same time, a prediction model for the physical and chemical performance of oil based on GRU neural network was further established. By comparing with the prediction results of LSTM neural network and ARIMA model, it was found that the prediction of physical and chemical
performance of GRU neural network showed great superiority. It can accurately predict the performance change trend of the oil in the short term.

References:
[1] Vapnik V N and Lerner A 2008 Pattern recognition using generalized portrait method Automation and Remote Control 24 774-780
[2] Li L J Zhang Z S and He Z J 2004 Research on Condition Trend Prediction of Mechanical Equipment Based on Support Vector Machine Journal of Xi’an Jiaotong University 38 230-233
[3] Akbarzadeh A Naseh V M and NodeFarahani M 2016 Carbon Monoxide Prediction in the Atmosphere of Tehran Using Developed Support Vector Machine Pollution 6 43-57
[4] Yan Y L Li J and Li H 2015 A Wind Turbine Anamaly Detection Method Based on information Entropy and Combination Model Power System Technology 39 737-743
[5] Qiu J L Tian J R and Chen H P 2018 Prediction Method of Parking Space Based on Genetic Algorithm and RNN Advances in Multimedia Information Processing, Pti 452 865-876
[6] Choi J and Lee B 2018 Combining LSTM Network Ensemble via Adaptive Weighting for improved Time Series Forecasting Mathematical problems in engineering 9 1-8
[7] Chen S and Ge L 2019 Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction Quantitative Finance 19 1507-1515
[8] Cao K Kim H and Hwang C Y 2018 CNN-LSTM Coupled Model for Prediction of Waterworks Operation Journal of information processing systems 14 1508-1520
[9] Gao X Li X B and Zhao B 2019 Short-Term Electricity Load Forecasting Model Based on EMD-GRU with Feature Selection ENERGYES 12 1-18
[10] Ungurean L Micea M V and Carstoiu G 2020 Online state of health prediction method for lithium-ion batteries, based on gated recurrent unit neural networks International journal of energy research 14