A fault diagnosis method based on the Support Vector Machine in rod pumping systems

Juanni Li1*, Jun Shao2
1School of science, Xi’an Shiyou University, Xi’an, Shaanxi, 710065, China
2Mechanical Engineering College, Xi’an Shiyou University, Xi’an, Shaanxi, 710065, China
*Corresponding author’s e-mail: kikilee@xsyu.edu.cn

Abstract. Monitoring the working status of the sucker rod pump is an important part in petroleum engineering. With the development of artificial intelligence technology, more methods have been applied to the fault diagnosis of rod pumping systems. An evolutional fault diagnosis method based on Support Vector Machine (SVM) in sucker rod pumping systems is proposed. Fourier descriptors and Light Field compression algorithm are used in this method to extract the graphic features of the indicator diagram. SVM is used to build fault classification model. This method is verified experimentally through data of indicator diagrams and the results show that it has a shorter training time and higher accuracy.

1. Introduction
Petroleum, as a high-quality energy and a multi-purpose precious chemical raw material, is an ingredient in thousands of everyday items. Sucker-rod pumping is the most widely-used artificial lift method in the oil and gas industry. However, in the process of oil exploitation, it is so difficult to monitor the working status of the rod pumping equipment. In order to improve the reliability of fault recognition, many researchers have done a lot of research work. In 1966, Gibbs, S. G and Nely, A. B. proposed a one-dimensional wave equation for a vertical well under ideal conditions. This represents that the research on sucker rod pump system has entered the stage of theoretical analysis [1]. On this basis, DR Doty and Z Schmidt proposed an improved model for predicting the behaviour of sucker rod pumping systems [2]. Wang Kai et al. considered the influence of the deviation angle, azimuth angle and oil, and selected the micro-element body from the sucker rod string of the multi-stage hybrid rod, and performed a dynamic analysis of the micro-element body [3].

The above-mentioned fault diagnosis methods are mainly to establish a mathematical model of the controlled object within a certain error range, and perform fault diagnosis through analysis and evaluation of relevant parameters [4]. In addition, there are some methods based on data-driven fault diagnosis, such as: the fault diagnosis based on indicator diagram or pump diagram. By measuring the load and the displacement of the polished rod, the closed coordinate curve drawn by a complete stroke is called an indicator diagram. By analysing the indicator diagram, not only the fault of the well pump work can be identified intuitively, but also can assist to analyse the reason for the low oil production efficiency.

With the rapid development of computer intelligence technology, researchers have developed various fault diagnosis systems, and great progress has been made [5]. For example: expert system [6], fuzzy theory [7], support vector machine [8], deep learning [9], artificial neural network [10] methods have
been applied to fault diagnosis in sucker rod pumping systems. However, it is still an open problem in terms of the complexity of feature calculation and the accuracy of fault diagnosis.

In this paper, an evolutional SVM-based fault diagnosis method is proposed. This method comprises the following steps: pre-processing the indicator diagram data to form a standard sample, extracting and merging feature, and optimizing the classification algorithm, establishing the fault diagnosis model. The goal of this method is to reduce the training time and improve the accuracy of the diagnostic model.

2. The Main Idea of the Proposed Method
The indicator diagram can reflect the working conditions of the pumping wells. In the work of the pumping unit, the indicator diagram under ideal conditions without interference from the external environment is called the "ideal indicator diagram". The ideal indicator diagram is shown in Figure 1. The four line segments in the figure correspond to the different working phases of the pumping unit, respectively: AB is the process of load increase, BC is the process of plunger lifting, and CD is the process of load reduction, DA is the process of plunger sinking.

![Figure 1. A theoretical indicator diagram](image)

Under normal working environment, the difference between the generated indicator diagram and the theoretical indicator diagram is very small, and the graph ABCD is approximately a parallelogram. Under the influence of factors such as sand production and delivery valve leakage, the indicator diagram shows different characteristics according to specific influencing factors.

With the development of intelligent computing technology, many artificial intelligence technologies are widely used in oil extraction systems. This paper proposes a method based on the SVM algorithm to classify the indicator diagrams obtained in the actual project and realize the fault diagnosis (As shown as figure 2). The method is divided into the following steps:

1. Obtain indicator diagram data from the actual work process and record relevant condition information;
(2). Analyse and pre-process the original indicator diagram information to form a unified format;
(3). Perform feature extraction on the pre-processed indicator diagram data, which includes two aspects: image data and shape data;
(4). Feature fusion: standardize the extracted feature data;
(5). Optimize classification algorithm, generate fault diagnosis model, and further verify the accuracy and effectiveness of the model.

2.1. Raw data acquisition and pre-processing
The original data comes from the actual oil production process, and indicator diagram data include the data from common fault types and under normal conditions. Generally, these data have two representation forms: one is expressed in the form of numerical values, and the other is expressed in the form of graphics. By analysing the original data, using OpenCV findContours method, the contour coordinates are extracted from the image information, and all the acquired data are expressed in numerical form.

2.2. Feature extraction and feature fusion
Feature extraction is a concept in computer vision and image processing. It refers to the use of computers to extract image information and determine whether each image point belongs to an image feature.

Generally, there are two types of representation methods for shape features, one is contour features, and the other is regional features. The contour feature of the image is mainly aimed at the outer boundary of the object, while the regional feature of the image is related to the entire shape area.

2.2.1 Calculating Fourier descriptors of indicator diagram
There are many methods for shape extraction, such as boundary feature method, set parameter method, shape invariant moment, curvature scale space descriptor, Fourier shape descriptor method and so on. Fourier descriptors (FD) is used in this paper to extract the outline of the indicator diagram information.

Fourier descriptors are a way of encoding the shape of a two-dimensional object by taking the Fourier transform of the boundary, where every point (x, y) on the boundary is mapped to a complex number x+yi. When describing the shape, Fourier descriptors, firstly, find the coordinates of the edge of the shape in a clockwise direction, store these data in a list; secondly, perform complex-valued vector calculations on them, and finally perform Fourier transform on the complex values.

2.2.2 Compression Algorithm
Light Field(LF) Compression is a vector data compression algorithm. The basic idea of LF compress is: for all points on each curve, a fan-shaped region is defined point by point. If the next point of the curve is outside the fan-shaped region, the current point is retained; if the next point of the curve is within the fan-shaped region, the current node is discarded. The vector compression algorithm can greatly reduce the training time by effectively compressing the indicator diagram data.

2.2.3 Multiple Feature Fusion
Integrate information from different sources together, and remove redundant information for the purpose; it is convenient for later analysis and processing. In the method proposed in this paper, the FD is used to describe the outline information of the indicator diagram, and LF algorithm is further used to effectively compress the indicator diagram data. Combine the characteristic information of the indicator diagrams in these two dimensions to form a unified characteristic vector.

2.3. Classification algorithm
According to the number of samples we get and the specific problem to be solved, the SVM method is used in the fault classification of the sucker rod pumping system. SVM is one of the most popular
Supervised Learning algorithms, which is used for Classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.

3. Experimental Results and Analyses
In order to verify the accuracy and effectiveness of the experiment, the indicator diagram data during the actual oil production process was collected. These data include the indicator diagram data under normal conditions and 3 common fault types (Table 1). These data are all labelled data. After pre-processing the data, a uniformly represented labelled sample set is formed.

| Fault type | Indicator diagram (example) | Sample size |
|------------|-----------------------------|-------------|
| Normal     |                            | 43          |
| Gas affected |                           | 26          |
| Stuck pump  |                            | 33          |
| Insufficient liquid supply (ILS) | | 18 |

According to the definition of the indicator diagram, which is described as a closed curve on a 2-dimensional plane. So the outline of the graph is an important feature of the indicator diagram. Through binary image processing, the gray-level image is converted into a binary image. Then through the steps of denoising, boundary connection, boundary tracing, etc., the contour data of the indicator diagram is generated (as shown in Figure 3).

| Serial Number | Label Type | Fourier descriptors |
|---------------|------------|---------------------|
| 1             | N          | 0.3913599, 0.05029733, 0.0756814, 0.0449984, 0.05343696, 0.0311612, 0.0195818, 0.04974325 |
| 2             | N          | 0.16614185, 0.34496466, 0.1525589, 0.07893225, 0.0566788, 0.08718675, 0.04715163 |
| 3             | G          | 0.3520893, 0.24753255, 0.1976806, 0.1189697, 0.1389407, 0.0740735, 0.07534797 |
| 4             | G          | 0.0674999, 0.06158654, 0.0228192, 0.04707838, 0.0551163, 0.03870838, 0.02533862 |
| 5             | G          | 0.22424197, 0.2776464, 0.1664613, 0.04713798, 0.0458576, 0.0489643, 0.03924124 |
| 6             | S          | 0.0394487, 0.13643474, 0.1468117, 0.11381975, 0.09095657, 0.04213997, 0.08117247 |
| 7             | S          | 0.0394487, 0.13643474, 0.1468117, 0.11381975, 0.09095657, 0.04213997, 0.08117247 |

Figure 3. Examples of pre-processing.
The Fourier descriptor is an image feature which used to describe the contour. The basic idea is to use the Fourier transform of the object boundary information as the shape feature, transform the contour feature from the spatial domain to the frequency domain, and extract the frequency domain information as the feature vector of the image. Because the discrete Fourier transform is invertible, all the information about the shape is contained in the Fourier descriptors. Table 2 is the Fourier descriptor examples of the indicator diagram.

By extracting the contour of the indicator diagram, each indicator diagram can get the coordinates of 300 to 400 points on the contour curve. When using these data to calculate graphic features, it takes a lot of time due to the large amount of calculation. In order to reduce the data set while ensuring that the curve shape features do not change, compression algorithms are used in the feature extraction process. As shown in Figure 4, 394 coordinate values are extracted from the original indicator diagram image while 55 coordinate values are extracted after compression by the LF method. By integrating information from different data sources, the characteristic information of each indicator diagram is fused with Fourier descriptors and compressed graphic information.

![Figure 4. Image data compression of indicator diagram.](image)

Because the SVM algorithm suitable for small sample training set has excellent generalization ability, our research mainly uses SVM method to complete indicator diagram fault diagnosis. The most common split ratio: 80:20 is used for splitting the training set and testing set. In order to verify the accuracy and effectiveness of the method, four types of indicator diagram data are used as the research objects, including 43 normal conditions data, 26 gas-affected data, and 33 stuck pump data and 18 data on insufficient liquid supply. Experiments study the training time and accuracy of diagnostic model classification under different feature data. As shown in table 3, The multi-feature fusion method is better in terms of training time and accuracy of the diagnostic model.

![Table 3. Training time and accuracy of different feature extraction methods.](image)

| Methods                       | Training Time(s) | Accuracy(%) |
|-------------------------------|------------------|-------------|
|                               | First  | Second | Third | Average training time | First  | Second | Third | Average accuracy |
| Multiple Feature Fusion       | 4      | 4      | 5     | 4.33 | 91.2 | 90.3 | 90.3 | 90.6 |
| Indicator Diagram Contour     | 13     | 12     | 14    | 13   | 67.0 | 69.2 | 67.8 | 68.0 |

4. Conclusion

In this paper, a classification model for fault diagnosis in rod pumping system based on SVM is proposed. This method is to obtain the characteristic data of the indicator diagram by describing the contour coordinates, further calculating the Fourier descriptor and compressing the coordinate data.
On this basis, the SVM method is applied to establish a fault diagnosis model. Experiment results show that this method is better than traditional method in terms of training time and accuracy.

Acknowledgments
The authors gratefully acknowledge the financial support from the Key R & D Program in Shaanxi Province (No. 2021GY-084).

References
[1] Gibbs, S.G., A.B. Neely. (1966) Computer Diagnosis of Down-Hole Conditions in Sucker Rod Pumping Wells. J Pet Technol 18: 91–98.
[2] Doty D R., Schmidt Z. (1983) An Improved Model for Sucker Rod Pumping. Society of Petroleum Engineers Journal, 23(01):33-41.
[3] Kai Wang. (2010) Modelling method for fault diagnosis to sucker rod pumping system in directional well. Journal of China University of Petroleum, 34(002):130-135.
[4] Xiang Xiao. (2021) Research on Diagnostic Method for Working Conditions of Pumping Wells Based on Deep Learning. University of Science and Technology of China.
[5] Md Nor N., Che Hassan C R., Hussain M A. (2019) A review of data-driven fault detection and diagnosis methods: applications in chemical process systems. Reviews in Chemical Engineering.
[6] Derek, H.J., Jennings, J.W., S.M. Morgan. (1988) Sucker Rod Pumping Unit Diagnostics Using an Expert System. The Permian Basin Oil and Gas Recovery Conference, Midland, Texas.
[7] Li, K., Gao, XW., Zhou, HB. et al. (2015) Fault diagnosis for down-hole conditions of sucker rod pumping systems based on the FBH–SC method. Pet. Sci. 12, 135–147.
[8] Cheng H., Yu H., Zeng P., et al. (2020) Automatic recognition of sucker-rod pumping system working conditions using dynamometer cards with transfer learning and SVM. Sensors (Switzerland), vol. 20, no. 19, 56-59.
[9] Wang, Xiang., He, Yanfeng., Li, Fajun., et al. (2019) A Working Condition Diagnosis Model of Sucker Rod Pumping Wells Based on Big Data Deep Learning. the International Petroleum Technology Conference, Beijing, China.
[10] Souza A M F D, Bezerra M A D, De M, et al. (2009) Using artificial neural networks for pattern recognition of downhole dynamometer card in oil rod pump system. Proceedings of the 8th WSEAS international conference on Artificial intelligence. pp. 230–235.