Research on Server Health State Prediction Model Based on Support Vector Machine

Dan Jin¹, Ce Li², Qiong Wang³ and Yuhang Chen³, *
¹, ²Gansu Tongxing Intelligent Technology Development Co., Ltd., Lanzhou City, Gansu Province, China
³Information System Integration Company, NARI Group Corporation, Nanjing City, Jiangsu Province, China
*Corresponding author

Abstract. With the development of artificial intelligence, machine learning methods have been widely applied to predictive modelling in various fields. Support vector machine is a linear classifier with the largest interval defined on the feature space. In order to improve the current status of server operation and maintenance, and improve the accuracy and efficiency of server health prediction, this paper studies the feature quantity selection based on grey relational analysis, studies the support vector machine applied to data generalization and classification, and establishes a health prediction model. The simulation training of historical data shows that the method has fast learning, high prediction speed, high efficiency, high accuracy and wide application prospects.

1. Introduction
With the steady expansion of the company's business, business data has gradually accumulated, computing resources demand continues to grow, and data center have become an important support for the company's business. As the core computing device of the center, the load and number of servers are also growing. The normal operation of the server is the fundamental guarantee for the normal operation of the business system, so it is especially important for the operation and maintenance of the server [1]. Active operation and maintenance is to predict in advance when a server failure is about to occur, and it can reserve time for the operation and maintenance personnel to deal with the problem. On the one hand, it can ensure the continuous and stable operation of the business system carried by the server, and on the other hand, it can also extend the use of the server run time [2].

This paper establishes a model of server operating parameters and health state through machine learning algorithm, applies gray correlation analysis and support vector machine to build server health prediction model, and trains and optimizes according to historical data to realize server fault warning and provide decision for operation and maintenance personnel. Auxiliary to achieve ‘active’ operation and maintenance of the server.

2. Analysis of Related Technical Principles

2.1. Grey Correlation Analysis
Grey relational analysis is a new method developed on the basis of grey system theory [3]. Correlation analysis is a method to analyse the degree of correlation of various factors in the system, or a method of measuring association. Through the comparison of the correlation degree, the primary and secondary factors can be found in the system, and the main factors affecting a certain variable are obtained. The algorithm flow is as follows:
Step1. First, the dimensionlessness of the parameters is performed.
Step 2. Then calculate the correlation coefficient. Let $X_0 = \{x_0(1), x_0(2), \ldots, x_0(n)\}$ as the data sequence, $X_1 = \{x_1(1), x_1(2), \ldots, x_1(n)\}$ be a comparison sequence, and the correlation coefficient between the two sequences is:

\[
\xi_1(k) = \frac{\min_{i \in k} |x_0(k) - x_1(k)| + \rho \max_{i \in k} |x_0(k) - x_1(k)|}{|x_0(k) - x_1(k)| + \rho \max_{i \in k} |x_0(k) - x_1(k)|}
\]

Where: $|x_0(k) - x_1(k)|$ is the absolute difference between point $x_0(k)$ and $x_1(k)$; $\min_{i \in k} |x_0(k) - x_1(k)|$ is the two-level minimum absolute difference; $\max_{i \in k} |x_0(k) - x_1(k)|$ is the two-level maximum absolute difference; $\rho$ is the resolution coefficient, $0 < \rho < 1$, and generally taken as 0.5. The degree of association between $X_1$ and $X_0$ is:

\[
\gamma(x_0, x_1) = \frac{1}{n} \sum_{k=1}^{n} \xi_1(k) = 1, 2, \ldots, n
\]

It is generally believed that when the degree of association is less than 0.6, the degree of association between the data column and the comparison sequence is small.

2.2. Support Vector Machine Principle

Support vector machines (SVM) are a classic two-class model\cite{4}. The goal is to find a hyperplane to segment the samples. The principle of segmentation is to maximize the interval and eventually transform it into a convex quadratic programming problem. Support vector machines map vectors to a high-dimensional space for association analysis. Two parallel hyperplanes are formed on both sides of the hyperplane separating the data. Establish spaced apart hyperplanes to maximize the distance between two parallel hyperplanes. Assume that the larger the distance or gap between the parallel hyperplanes, the smaller the total error of the classifier, and the higher the accuracy of the prediction.

![Figure 1. Support vector machine classification diagram.](image)

3. Server Health Prediction Model Based on SVM

3.1. Feature Selection

The raw data fields collected by the company server are: [device type, operating system, device name, runtime, IP, CPU usage, memory usage, affiliation, bearer service system]. The gray feature degree is calculated one by one for the nine feature quantities (fields) and the server health state. The gray relation degree of the four feature quantities is less than 0.6, which should be eliminated. They are ‘device name’ and ‘affiliated unit’. Bear the business system, ‘IP’. There are five feature quantities after filtering: [device type, operating system, runtime, CPU usage, memory usage].
3.2. Model Establishment and Training

According to the ‘Next Time Status’ field of the server, the model training set is divided into two categories for training, one type is ‘next time state’ = normal data, and the other type is ‘next time state’ = abnormal data, which helps improve the prediction accuracy of the predictive model.

Figure 2. Principle of server state prediction model modeling based on SVM

Take 80% of the server data processed data as the model training set and 20% of the server data as the model test set. In order to ensure the robustness of the model, multiple training models and test sets are randomly assigned for training. After the support vector machine (SVM)-based model training is completed, the test data is input to the model, and the predicted result is shown in FIG. 3: most of the predicted output and the expected output (the actual value of the test data) obtained by the predictive model are consistent, there are a small number of deviation points. The following analysis of the prediction error, define the error:

\[ \text{error}(i) = \text{pre}(i) - \text{real}(i), i = 1, 2, \ldots, I \]  

(3)

Among them pre(i) is the predicted output of the i-th test data, real(i) is the expected output of the i-th test data. Define the training accuracy rate: the number of accurate predictions / the total number of predictions (the number of test data).

It is calculated that the training accuracy rate is 0.8542, which is equal to 85.42%.

Figure 3. Data center experiment network topology.

3.3. Comparative Analysis

The BP neural network is also more common in the deep learning algorithm. This paper uses the 3-layer BP neural network to establish and train the server state prediction model for the same data. The prediction accuracy based on BP neural network is calculated to be 75.69%, which is lower than the SVM-based prediction model. After analysis, the BP neural network training slow convergence rate is easy to fall into the local minimum and cannot get the global optimal solution, and the server running data type is not rich enough. Therefore, the server state prediction model based on SVM is accurate and efficient.

4. Conclusion

Based on big data thinking, this paper mines the law from a large number of historical operational data, uses BP neural network algorithm in machine learning to establish the model of server operating parameters and
health state, and uses support vector machine algorithm to construct the predictive model of server health state. And training. Compared with BP neural network, SVM has better effects and advantages in analysing server operating parameters. The SVM-based server prediction model can be further optimized to continuously improve the accuracy and realize the intelligent operation and maintenance of the server.

References
1. Doumpos Mz., Zopounidis C., Golfinopoulou V. Additive Support Vector Machines for Pattern Classification, 2007, 3(3).
2. Kuh A., De Wilde P. Comments on “Pruning Error Minimization in Least Squares Support Vector Machines,” 2007, 2(2).
3. Wang Wendong, Yuan Gang. Research on management architecture of the next generation Internet. Journal of Beijing University of Posts and Telecommunications, 2004, 27(Sup): 113-117.
4. Feng Guohe. SVM classification kernel function and parameter selection comparison. Computer Engineering and Applications, 2011, (3).
5. LIN Zhiyong, HAO Zhifeng, YANG Xiaowei. Research status of unbalanced data classification. Computer Application Research, 2008, (2).
6. SHI Hong, ZHOU Xiaodan, LIN Fen, et al. Study on the relationship between the meridian energy health detection system and the twelve meridians. Chinese Medicine Technology, 2009, (4).
7. Sun Kai, Wang Yinglong. Research on the Construction of Mercer Kernel Function in Support Vector Machine. Ordnance Automation, 2008, (11).
8. Wu Tao. The nature and method of kernel function and its application in obstacle detection. 2007-09-01 00:00:00, 2003.
9. Xu Jianhua, Zhang Xuegong, Li Yanda. New Development of Support Vector Machines. Control and Decision, 2004, (5).