Short-term Load Prediction of Cloud Computing Based on Fuzzy Information Granulation SVM

Jian Wang¹, Yuanyuan Zhang¹

¹School of information, Central university of finance and economics, Beijing, China
wanderingful@126.com

Abstract. In order to predict the short-term load variation range and trend of cloud computing, this paper proposed a prediction model based on information granulation support vector machine (IGSVM). Taking the historical load value as a sample to do simulation training, through Gravitational Search Algorithm (GSA) to optimize the parameters of SVM, and make regression prediction to three parameters of triangular fuzzy particles, Low, R and Up, to obtain the variation range and trend of short-term load. The result is consistent with the actual situation, which verifies the validity of the model and provides the basis for actual operation and maintenance.

1. Introduction

The goal of cloud computing is to truly configure shared resource pools on demand. In the face of diversified business-oriented, batch data storage and analysis, and complex application fields[1], the users’ requirements for rapid resource allocation and distribution must be taken into account, and effective scheduling among applications, components and virtual machines must be implemented. Cloud computing must have considerable flexibility, so as to ensure efficient processing, to maintain cloud utilization and cloud service level at an acceptable level[2]. Therefore, it is important to actively configure workloads, especially when new virtual machine startup delays, energy minimization and personalized resource allocation requirements arise, the resource needs of cloud computing need to be predicted in advance[3].

Accurate short-term load prediction of cloud computing is one of the most effective measures to deal with the above challenges. In terms of short-term load variation characteristics of cloud computing, it has considerable non-linear, time-varying and uncertain characteristics. There are many accidental factors in the prediction process, so it is difficult to master these special features from the perspective of system description, which often leads to the decline of prediction performance. Traditional prediction methods include pattern matching [4], autoregressive model [5], Bayesian model [6], Hidden Markov model [7] and neural network [8], etc. The short-term resource load variation of cloud computing is affected by many uncertain factors in reality, so it is difficult to be predicted very accurately. Therefore, we can consider enlarging the prediction interval and pay more attention to its variation range and trend. In this paper, support vector machine (SVM) based parameter optimization is combined with fuzzy information granulation model to predict the variation trend and scope of cloud computing short-term load, which has strong self-adaptability and robustness, can guarantee the generalization ability and meet the needs of system fault tolerance at the same time, in addition, it can measure the short-term load variation space of cloud computing.
2. Fuzzy Information Granulation SVM Model

2.1. Fuzzy information granulation

Information Granulation (IG) is the main aspect of granulation and word computing, and study the formation, representation, thickness and semantic interpretation, etc. Essentially, information granule is a set of formation divided by indistinguishability, functional similarity, similarity, function and so on. Granulation computing is a new concept and computing paradigm of information processing, which covers all the theories, methods, techniques and tools of granulation.

Fuzzy information granules are represented by fuzzy sets. Fuzzy granulation of time series by fuzzy set method can be divided into two steps: partitioning window and fuzzification. Partitioning window is to divide the time series into several small subsequences as operation windows; Fuzzification is to fuzzy each window and generate a fuzzy set, i.e. fuzzy information granule. The combination of these two generalized patterns is fuzzy information granulation, namely, f-granulation. In f-granulation, the most important is the process of fuzzification, that is, to create a reasonable fuzzy set on the given window, so that it can replace the original window data, indicating the relevant information concerned by people.

For the given time series, the single window problem is considered, that is, the whole time series \( X \) is regarded as a window for fuzzification. The task of fuzzification is to establish a fuzzy particle \( P \) on \( X \), that is, a fuzzy concept \( G \) which can be reasonably described (a fuzzy set with \( X \) as the universe), \( G \) is determined, and then fuzzy particle \( P \) is determined:

\[
g \triangleq x \quad G
\]

It is assumed that \( A \) is a membership function of fuzzy concept \( G \), that is, \( A = \mu_G \), usually, the granulation is determined the specific membership function \( A \) on the premise of determining the fuzzy concept.

The establishment of fuzzy particles needs to consider how to represent the original data reasonably, at the same time, needs to have certain particularity. There are several basic forms of fuzzy particles commonly used: triangle, trapezoid, Gaussian, parabolic and so on. In this paper, triangular fuzzy particles are used, and the membership functions are as follows:

\[
A(x, a, m, b) = \begin{cases} 
0, & x < \text{low} \\
\frac{x - \text{low}}{R - \text{low}}, & \text{low} \leq x \leq \text{up} - \text{R} \\
\frac{\text{up} - x}{\text{up} - \text{R}}, & \text{up} - \text{R} < x \leq \text{up} \\
0, & x > \text{up} 
\end{cases}
\]  \hspace{1cm} (2)

In which, \( \text{Low}, R \) and \( \text{up} \) are three parameters of fuzzy particles. the parameter \( \text{low} \) describes the minimum value of the corresponding original data variation, parameter \( R \) describes the average quality of the original data variation, and parameter \( \text{up} \) describes the maximum value of the corresponding original data variation.

2.2. SVM algorithm

SVM is a class-two classification model. Its basic model is a linear classifier with the largest spacing defined in the feature space; SVM also includes kernel techniques, which makes it become a non-linear classifier in essence. The learning strategy of SVM is to maximize the interval, which can be formalized as a problem to obtain convex quadratic programming [9].

In the operation of SVM algorithm, the relevant parameters need to be adjusted, and among them, the penalty parameter \( c \) and the kernel function parameter \( g \) have great influence on the performance of the system. The process of parameter selection is equivalent to an optimization process, which may
be the best solution for every point in the search space. Therefore, the best solution can be obtained by finding the point with the least generalization error. In this paper, GSA is used to optimize SVM and then to effectively improve the ability of parameter optimization.

2.3. GSA algorithm
Gravitational search algorithm considers that in nature, every particle will attract all other particles. The algorithm uses the gravitational tropism between particles to achieve intelligent search. The two particles move in opposite direction, the closer they are to each other, the slower the best unknown speed. The best solution [10] will be found by iteration. The algorithm is described as follows:

For a given \( n \)-dimension space with \( k \) particles, the position of particle \( i \) is:

\[
x_i = (x_{i1}, \cdots, x_{id}, \cdots, x_{in}) \quad i = 1, 2, \cdots, k
\]

In which, \( x_{id} \) is the position of particle \( i \) in \( d \)-dimension, and the position of each particle is a potential solution.

The inertia mass of each particle is related to the fitness value of its position, that is, it can be judged that the greater the mass of the particle, the greater the attraction to other objects. In this paper, the inertia mass (mass) of the particles is defined as:

\[
M_i(t) = m_i(t) / \sum_{j=1}^{k} m_j(t)
\]

In which, \( m_i(t) = \frac{\text{fit}_i - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)} \), \( \text{fit}_i(t) \) is the fitness value of particle \( i \) at time \( t \), \( \text{best}(t) \) means the best solution of time \( t \), \( \text{worst}(t) \) means the worst solution of time \( t \). According to the law of universal gravitation, the gravitation calculation formula between particles \( i \) and \( j \) can be obtained as follows:

\[
F_{ij}^d(t) = G(t) \frac{M_i(t) \times M_j(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t))
\]

In which, \( M_i \) is the mass of particle \( i \), \( M_j \) is the mass of particle \( j \), \( G(t) \) is gravitational coefficient, \( \varepsilon \) is minimum constant value, \( R_{ij}(t) \) is the Euclidean distance between particles \( i \) and \( j \), \( G(t) \) means gravitational constant of time \( t \), which can be calculated by \( G(t) = G_0(1 - t / t_{\text{max}}) \).

In which, \( G_0 \) is initial value of gravity, is set to 100, \( t_{\text{max}} \) is maximum number of iterations.

The force of particle \( i \) at \( d \)-dimension can be obtained as:

\[
F_i^d(t) = \sum_{j=1, j \neq i}^{k} \text{rand}_j F_{ij}^d(t)
\]

In which, \( \text{rand}_j \) is a random number in \([0,1]\) and \( F_{ij}^d \) is the gravitational value of particle \( j \) on \( i \).

When a particle is subjected to gravitation, it will generate acceleration. According to the formula, it can be obtained that the acceleration of particle \( i \) at \( d \)-dimension is:

\[
a_i^d(t) = F_i^d(t) / M_i(t)
\]

During each iterative process, the speed and position of particle \( i \) at \( d \)-dimension space can be updated according to the acceleration.

\[
v_i^d(t + 1) = \text{rand}_d \cdot v_i^d(t) + a_i^d(t)
\]

\[
x_i^d(t + 1) = x_i^d(t) + v_i^d(t + 1)
\]

In which, \( v_i^d(t) \) and \( x_i^d(t) \) are the speed and position of particles respectively.
2.4. Load prediction model of fuzzy information granulation SVM

In this paper, GSA algorithm is used to optimize the parameters of SVM model, and the prediction model is generated according to the optimization results. The fitness function value is the mean square error between the training prediction value and the actual value. The optimization process is described as follows:

Step 1: Input cloud computing resource to predict load time series, and construct training set and verification set.

Step 2: Initialize the range of parameters C and σ of SVM, and set relevant parameters of GSA algorithm.

Step 3: Random generation of particle swarm, each particle position vector include C and σ.

Step 4: According to the initial C and σ, the training concentration particles are learned and the fitness value of each particle is calculated.

Step 5: According to the formula (8)–(9) to update the particle position, calculate the corresponding fitness value and find the best fitness value of the swarm.

Step 6: If the number of iterations exceeds the maximum allowable number of iterations, the global optimum shall be output after training, i.e. the values of parameters C and σ.

SVM prediction model will be established by using the parameters found in above process, to predict the time series after fuzzy information granulation, and analyze the prediction results. The algorithm flow is shown in the figure 1.

3. Empirical Analysis

The simulation data is obtained from cloud computing management platform of an IT service company, the data time interval is from 1 August 2016 to 31 August 2016. In which, CPU load is taken as the forecast object. In the management strategy of the company, the dynamic adjustment will be carried out according to the real-time operation status of the cloud platform, and according to actual needs, cloud computing resources will be reconfigured at regular intervals. Therefore, this paper samples and analysis the acquired data every 15 minutes, that is, the data points recorded every day is 96, so as to obtain the corresponding historical resource load time series data. This paper mainly uses the load data of the previous time date to predict the CPU load of the next sampling point, so that the forecast results can play a decision-making role for cloud computing resources to achieve effective real-time configuration and scheduling.

The time series diagram of cloud computing load is shown in figure 2:
3.1. Fuzzy information granulation processing
In this paper, the load data of 15 minutes is taken as a window size, and the original window is divided into 5 to rectify, the result of granulation processing is shown in the figure 3:

3.2. Prediction using fuzzy granulation data
SVM is used to do regression prediction to three fuzzy particles Low, R and Up, and GSA algorithm is used to optimize the parameters of SVM, the results are as follows:
- **Low**: Mean squared error = 0.00843073 (regression), Squared correlation coefficient = 0.956768 (regression)
- **R**: Mean squared error = 0.00635652 (regression), Squared correlation coefficient = 0.957765 (regression)
- **Up**: Mean squared error = 0.00838352 (regression), Squared correlation coefficient = 0.964431 (regression)

| Sample  | Actual load | Predictive load |
|---------|-------------|-----------------|
| Sample1 | 32.45       | [low, R, Up]=[22.06,33.29,39.45] |
| Sample2 | 28.64       |                 |
| Sample3 | 22.49       |                 |
| Sample3 | 34.67       |                 |
| Sample3 | 38.44       |                 |

Optimal parameters are used for fitting and forecasting, we can see from table 1 that the load variation range of cloud computing predicted in this paper is accurate, the accuracy is high and can more effectively predict the short-term range of variation. The prediction method is reliable, can help enterprises to improve work efficiency and has a certain practicality.
4. Conclusion
By the combination of fuzzy information granularity model and SVM, the prediction of cloud computing short-term load based on fuzzy information granulation SVM is proposed. By effective combination of fuzzy information granulation and SVM characteristics, the trend law is described and the variable interval is also captured. On the basis of experimental verification, the variation range and trend of the prediction results are consistent with reality, which proves the validity and practicability of the model. Load variation rate, memory usage rate and other factors, so it still needs to be improved in the future work.

Acknowledgments
This work was supported by National Natural Science Foundation of China (71874215, 71571191, 61671030), Beijing Natural Science Foundation (9182016, 9194031), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (15YJCZH081, 17YJAZH120, 19YJCZH253), Fundamental Research Funds for the Central Universities (SKZZY2015021) and Program for Innovation Research in Central University of Finance and Economics.

References
[1] C. Hochreiner, S. Schulte, and S. Dustdar. Elastic Stream Processing for the Internet of Things. [C]//In Int. Conf. on Cloud Comp. 2016:100–107.
[2] S. Singh and I. Chana. QoS-Aware Autonomic Resource Management in Cloud Computing: A Systematic Review. [J]// Comp. Surveys (CSUR), 48(3): 42:1-42:46.
[3] C. Viswanath and C. Valliyammon. Cpu load prediction using anfis for grid computing. [C]//In Int. Conf. on Advances in Engineering, Science and Management. 2012:343–348.
[4] Caron E, Desprez F, Muresan A. Forecasting for grid and cloud computing on-demand resources based on pattern matching[C]//2nd IEEE International Conference on Cloud Computing Technology and Science.2010:456-463.
[5] Ching C T M, Dusit N, Tham C K. Evolutionary optimal virtual machine placement and demand forecaster for cloud computing[C]//2011 International Conference on Advanced Information Networking and Applications. 2011:348-355.
[6] Di S, Kondo D,Clime W. Host load prediction in a Google compute cloud with a Bayesian model[C]// International Conference on High Performance Computing, Networking, Storage and Analysis. 2012:1-11.
[7] Khan A. Workload characterization and prediction in the cloud:A multiple time series approach[C]// the 3rd IEEE International Workshop on Cloud Management.2012:1287-1294.
[8] Islam S, Keung J, Lee K, et al. Empirical prediction models for adaptive resource provisioning in the cloud[J]// Future Generation Computer Systems. 2011, 28:155-162.
[9] Hsieh,Cho-Jui, Chang Kai-Wei, Lin Chih-Jen, Keerthi, S. Sathiya, Sundararajan S. A. Dual Coordinate Descent Method for Large-scale Linear SVM. [C]// 25th International Conference on Machine Learning, 2008-01-01; 408–415.
[10] H. Askari, S. Zahiri. Decision function estimation using intelligent gravitational search algorithm [J]//International Journal of Machine Learning and Cybernetics. 2012, 3(2): 163-172.