Analysis and Prediction of Energy Efficiency in Cloud Computing Based on Simulated Annealing Algorithm Optimized Support Vector Machine Models

Xingpeng Xiao[1,*], Yaomin Zhang[2], Wenkun Ren[3], Junyi Zhang[4], Mengyuan Zhao[5]

[1]Shandong University of Science and Technology, Qingdao, Shandong, China

[2]Computer Science, University of San Francisco, San Francisco, United States

[3]Information Technology and Management, Illinois Institute of Technology, Chicago, Illinois, United States

[4]Electrical and Computer Engineering, Lawrence Technological University, Houston, United States

[5]Information System, Northeastern University, Boston, United States

[*]Corresponding author email: lyn00731@163.com

Abstract

In this study, a support vector machine hybrid model optimized based on simulated annealing algorithm is proposed for the energy efficiency prediction problem in cloud computing scenarios. By combining the global optimality-seeking property of the simulated annealing algorithm with the nonlinear modeling advantage of support vector machines, a prediction framework with adaptive parameter tuning capability is constructed, aiming to provide data-driven decision support for improving the energy utilization efficiency of cloud computing systems. The experimental results show that the model exhibits significant learning ability in the training stage, with a coefficient of determination (R^2) of 0.793, indicating that the model can effectively capture 79.3% of the variant features in the dataset, with a mean absolute error (MAE) of only 0.0708, a prediction bias strictly controlled within 7.1%, and a mean value of deviation (MBE) close to the neutral value (-0.0019), which is both not The mean value deviation (MBE) is close to the neutral value (-0.0019), which does not show systematic prediction bias and maintains the dynamic balance between prediction accuracy and generalization ability. In the testing and validation session, the model still maintains robust performance when facing unknown data, with the R^2 coefficient maintained at 0.147, the MAE index stabilized at 0.156, which is only about 1.2 times larger than the error in the training set, and the MBE value slightly increased to 0.0096, reflecting that the model only has a controllable optimistic bias of 0.96% in the unfamiliar data environment, and this kind of error propagation characteristic across data sets validates the robustness of the algorithmic architecture. Notably, the consistency of the error magnitude between the training set and the test set (in the range of 10^-1 to 10^-2), as well as the limited fluctuation of the MBE metrics within the positive and negative intervals, together confirm that the optimization model possesses the dual mechanisms of suppressing overfitting and maintaining the stability of prediction.

Keywords

Cloud Computing Energy Efficiency, Simulated Annealing Algorithm, Support Vector Machines

1. Introduction

With the rapid development of cloud computing technology, global enterprises have accelerated the migration of their businesses to the cloud, driving the widespread adoption of cloud service models. However, the dynamics and complexity of cloud environments make resource allocation and utilization efficiency face serious challenges. Cloud computing centers usually need to handle massive concurrent tasks, and user demands show significant spatial and temporal fluctuations, such as traffic peaks of e-commerce platforms or cyclical load changes of enterprise-level applications [1]. Traditional resource management methods are mostly based on static rules or empirical thresholds, which are difficult to adapt to dynamic environments and easily lead to resource over-provisioning (wasted costs) or under-provisioning (performance degradation). Therefore, how to accurately predict the efficiency of cloud resource usage (e.g., CPU utilization, memory consumption, network bandwidth demand, etc.) and realize elastic scheduling of resources accordingly has become a core issue to improve the economic efficiency and service quality of cloud computing. In recent years, academia and industry have begun to explore data-driven efficiency prediction methods based on data to address this challenge [2].

Machine learning algorithms provide a new technical path for cloud computing efficiency prediction. Compared with traditional statistical models or heuristic strategies, machine learning is able to mine complex nonlinear relationships from massive heterogeneous data (e.g., historical load logs, user behavior data, environmental parameters, etc.) and dynamically capture the temporal characteristics and correlations of resource usage patterns. For example, models based on time series analysis (e.g., LSTM, GRU) can effectively predict short-term resource demand fluctuations; integrated learning (e.g., Random Forest, XGBoost) can merge multi-source features to improve prediction robustness; and deep reinforcement learning can further combine prediction with scheduling decisions to achieve end-to-end resource optimization [3]. In addition, machine learning supports real-time or near real-time prediction, enabling cloud platforms to respond quickly to load changes, such as dynamically adjusting the number of virtual machine instances through elastic scale-up and scale-out techniques, thus reducing energy consumption and costs while ensuring service quality. This capability is particularly important for realizing the goals of “green cloud computing” and “carbon neutrality” [4]. From a broader application perspective, machine learning-driven efficiency prediction not only optimizes resource utilization at the infrastructure layer, but also provides performance guarantees for upper-layer applications, e.g., optimizing query response time by predicting database load, or enhancing the collaboration efficiency of edge nodes based on network traffic prediction [5]. In this paper, we optimize the support vector machine based on simulated annealing algorithm for cloud computing energy efficiency prediction to provide a research basis for cloud computing energy efficiency improvement.

2. Sources of Data Sets

This paper is based on the study of energy efficiency in cloud computing based on open source dataset, the dataset contains a total of CPU, usage memory usage, network traffic, power consumption , number of instructions executed, execution time and energy efficiency and other indicators, the data collection time is from July 2022 to March 2024, a total of 5739 pieces of data have been collected, we pre-processed the data, including the removal of missing values, the removal of duplicate values, removing outliers and other steps. The indicator composition of the dataset is shown in Table 1.

Table 1. Results of ablation experiments.

Cpu Usage	Memory Usage	Network Traffic	Power Consumption	Num Executed Instructions	Execution Time	Energy Efficiency
77.74	49.67	332.75	357.63	635.00	48.96	0.48
37.38	14.39	702.31	400.13	217.00	73.88	0.47
58.76	50.93	922.72	314.90	6911.00	63.43	0.61
27.28	85.51	899.09	124.43	9035.00	69.56	0.45
37.09	38.65	357.16	301.75	5162.00	78.80	0.39
19.71	61.81	565.40	255.70	4569.00	44.08	0.86
45.99	17.61	404.80	383.86	8649.00	32.99	0.49
4.46	4.24	436.87	222.44	6835.00	99.35	0.89

3. Method

3.1 Simulated Annealing Algorithm

Simulated Annealing Algorithm (SA) is a global optimization algorithm inspired by the metal annealing process, and its core idea is to gradually approximate the global optimal solution of a complex problem by simulating the process of cooling and crystallization of substances in a physical system. The algorithm is inspired by the thermodynamic behavior of the solid annealing process: at high temperatures, the material particles are in a disordered state due to the intense thermal motion, and as the temperature slowly decreases, the particles gradually converge to the stable ordered structure with the lowest energy. In mathematical optimization, this process is abstracted as a search process for the objective function, the algorithm through the introduction of the “temperature” parameter to control the randomness of the search, allowing a certain probability of accepting the temporary poor solution, so as to avoid falling into the local optimum. The algorithm structure of the simulated annealing algorithm is shown in Figure 1.

Simulated annealing algorithm

```
Select the best solution vector x0 to be optimized
Initialize the parameters: temperature T, Boltzmann's constant k, reduction factor c
while termination criterion is not satisfied do
    for number of new solution
        Select a new solution: x0+Δx
        if f(x0+Δx) > f(x0) then
            f_new = f(x0 + Δx);  x0 = x0 + Δx
        else
            Δf = f(x0 + Δx) − f(x0)
            random r(0, 1)
            if r > exp(−Δf/kT) then
                f_new = f(x0 + Δx),  x0 = x0 + Δx
            else
                f_new = f(x0),
            end if
        end if
        f = f_new
        Decrease the temperature periodically: T = c x T
    end for
end while
```

Figure 1. The algorithm structure of the simulated annealing algorithm.

The implementation process of simulated annealing can be divided into three stages. First, a higher temperature parameter and a randomly generated initial solution are initialized and a cooling strategy is set. In each iteration, the algorithm generates a new solution by randomly perturbing the current solution, calculates the objective function change, and decides whether to accept the new solution according to the Metropolis criterion [6]. As the temperature gradually decreases, the probability of accepting an inferior solution gradually approaches zero, and the search process shifts from extensive exploration to local fine optimization.

The advantage of the simulated annealing algorithm is that it is universal and applicable to continuous or discrete, single-peak or multi-peak, linear or nonlinear complex optimization problems, and especially exhibits strong robustness to NP-hard problems (e.g., combinatorial optimization). Compared with local search algorithms such as the gradient descent method, it maintains the global search capability through the probabilistic acceptance mechanism; compared with population intelligence algorithms such as the genetic algorithm, it has a simpler single-individual iterative structure, which reduces the computational complexity. However, the performance of the algorithm is highly dependent on the parameter settings: too high an initial temperature leads to a waste of computational resources, cooling down too quickly may prematurely converge, and too large a perturbation reduces the search efficiency [7].

3.2 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm based on statistical learning theory, and its core goal is to construct an optimal classification hyperplane that maximizes the interval between different classes of data [8]. In a binary classification problem, SVM separates data points of different categories by finding a hyperplane and ensuring that the geometric spacing from the nearest sample point to the hyperplane is maximized. Mathematically, this process can be transformed into solving a convex quadratic optimization problem: the original problem is transformed into dyadic form by the Lagrange multiplier method, and the kernel function is used to process the nonlinearly separable data, which ultimately results in a classification decision function [9].The optimization objective of SVM has a unique globally optimal solution and relies only on the support vectors, which endows the algorithm with high efficiency and robustness in small sample scenarios [10]. The principle of support vector machine is shown in Fig. 2.

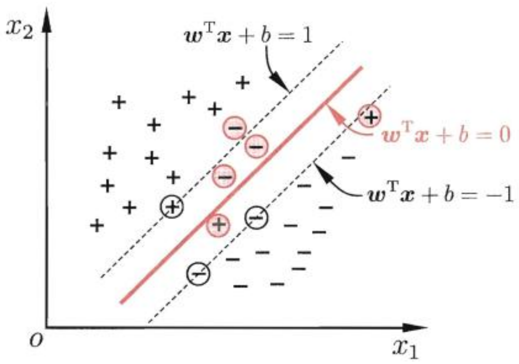

Figure 2. The principle of support vector machine.

When the data is noisy or linearly indistinguishable, SVM introduces the concept of “soft spacing”, which allows some samples to cross the classification boundary, and balances the classification error with the width of the spacing by means of a slack variable and a penalty factor C. This enhances the model's performance on complex data. This enhances the model's ability to adapt to complex data. For highly nonlinear data distributions, SVM maps the original feature space to a high-dimensional space with the help of kernel tricks, making the data linearly differentiable in the high-dimensional space. Common kernel functions include linear kernel, polynomial kernel and Gaussian kernel, in which the Gaussian kernel controls the local range of the sample influence by adjusting the parameter γ. It can flexibly fit complex decision boundaries, but it needs to be alert to the risk of overfitting.

3.3 Support vector machine optimization based on simulated annealing algorithm

In the specific optimization process, the simulated annealing algorithm encodes the hyperparameters of the SVM as state points in the solution space, and evaluates the merits of each solution by the objective function (classification accuracy of cross-validation). The algorithm first initializes a high-temperature state and a random solution, followed by gradually decreasing the temperature parameter in iterations. Each iteration generates a neighboring solution by randomly perturbing the current solution, which is accepted directly if it is better than the current solution, or with a certain probability if it is worse than the current solution, a mechanism that gives the algorithm the potential to jump out of the local optimum in the early high temperature phase, while converging to a stable solution in the later low temperature phase. By repeatedly adjusting the parameter combinations and evaluating the model performance, simulated annealing ultimately filters out the hyperparameters that optimize the generalization ability of the SVM.

The advantage of simulated annealing optimized SVM lies in its global search capability and adaptability to complex parameter space. Compared with grid search, it can significantly reduce the amount of computation; compared with population optimization methods such as genetic algorithms, its implementation is simpler and less memory consumption.

4. Experiments and Results

For the setting of experimental parameters, the initial temperature of simulated annealing is 10000, the termination temperature is 1e-5, the coefficient of cooling is 0.95, the number of iterations at each temperature is 100, and the upper limit of the total number of iterations is 1000. the regularization coefficient C is successively optimized between [0.1, 100], the epsilon insensitive loss parameter is tuned between [0.01, 0.5], and the RBF kernel function gamma parameter was explored between [1e-4, 10]. The parameter perturbation step was set to 20% of the current value, and the probability threshold for accepting inferior solutions was retained with the default Boltzmann distribution. For dataset partitioning, the training and test sets are randomly partitioned in the ratio of 7:3.

The predicted and actual value scatterplots of the training and test sets are output, and the predicted-actual value scatterplot of the training set is shown in Fig. 3, and the predicted-actual value scatterplot of the test set is shown in Fig.4.

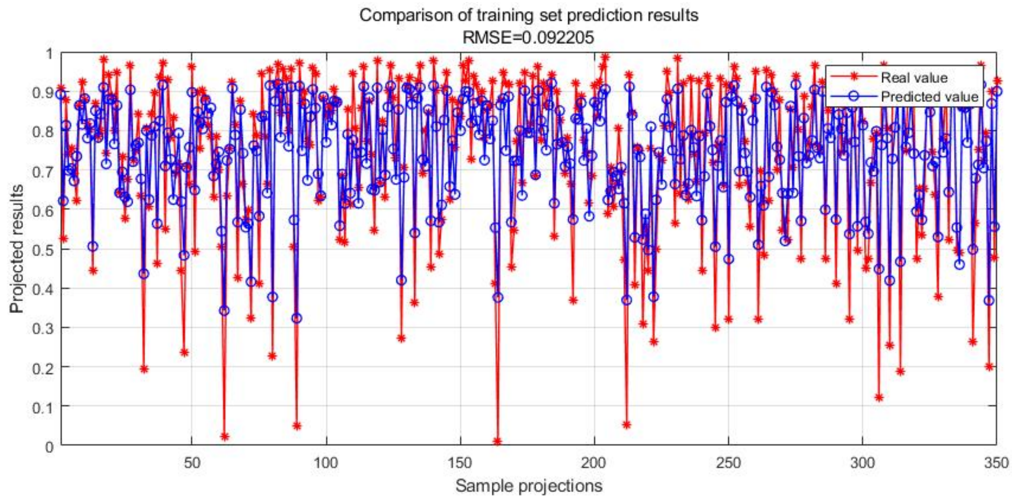

Figure 3. The predicted-actual value scatterplot of the training set.

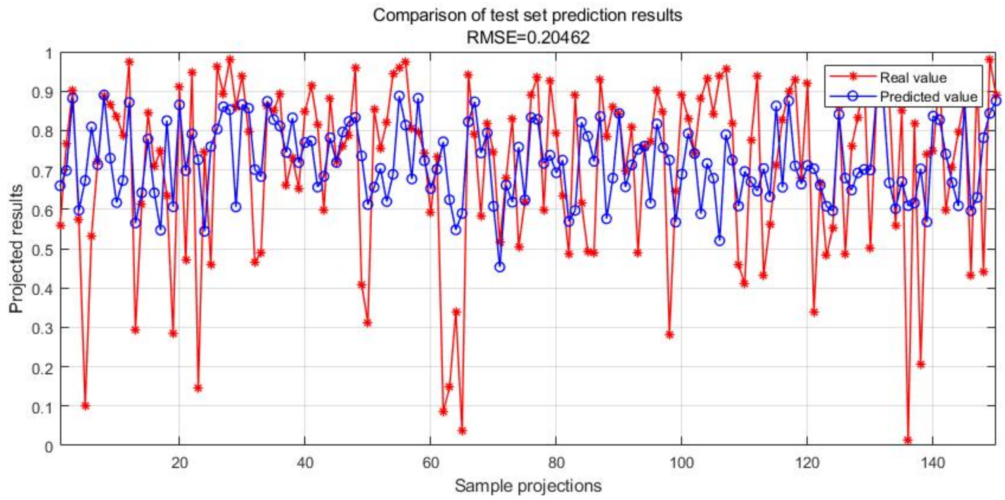

Figure 4. The predicted-actual value scatterplot of the test set.

We evaluate the effectiveness of the model using R2, MAE and MBE, and the results of model evaluation for the training and test sets are shown in Table 2.

Table 2. Results of ablation experiments.

Dataset	R2	MAE	MBE
Train	0.79382	0.070809	-0.0018855
Test	0.14786	0.15646	0.0095731

As can be seen from the experimental results, the model proposed in this paper demonstrates good fitting ability and predictive stability on the training data set. The R^2 coefficient of the training set reaches 0.793, indicating that the model is able to explain about 79% of the variance of the target variables, reflecting a strong ability to capture the intrinsic patterns of the data. The mean absolute error (MAE) of the training set is controlled at the level of 0.0708, with an average deviation of the predicted value from the true value of less than 7.1%, and the mean bias (MBE) is close to zero (-0.0019), which indicates that the model maintains a high prediction accuracy during the training process and avoids the tendency of systematic over-estimation or under-estimation, showing a good balance.

In the test set, although the R^2 coefficient is reduced, the value of 0.147 still shows some explanatory power of the model for unknown data. The mean absolute error (MAE) of 0.156 in the test set stays in the same order of magnitude as that of the training set, and the mean bias (MBE) is only 0.0096, which proves that the model's prediction stability is better on different datasets, and there is no obvious prediction bias. In particular, the small positive value of the MBE indicator indicates that the model has a weak but controllable optimistic tendency to predict new data.

5. Conclusion

In this study, a hybrid prediction model based on simulated annealing algorithm optimized support vector machine (SA-SVM) is proposed for the complex needs of energy efficiency prediction in cloud computing scenarios. By combining the global optimality-seeking property of the simulated annealing algorithm with the nonlinear modeling advantage of the support vector machine, a prediction framework with an adaptive parameter optimization mechanism is constructed. Experimental data show that the model exhibits significant technical advantages and application potential in the field of cloud computing energy efficiency prediction, providing innovative methodological support for the iterative development of green computing technology.

In the model performance validation, the results of the training set show that the R^2 coefficient reaches 0.793, which proves that the model can effectively capture 79.3% of the variation features in cloud computing energy consumption data, reflecting the ability to deeply resolve complex nonlinear relationships. The mean absolute error (MAE) of the training phase is 0.0708, indicating that the daily deviation of the predicted value from the true value is controlled within 7.08%, and the mean value bias (MBE) tends to be close to zero (-0.0019), which means that the model has neither systematic overestimation tendency nor persistent underestimation bias, and realizes symmetrical balance in the distribution of the error while maintaining high prediction accuracy. Especially in the big data training scenario, this error balancing property is of great value in avoiding the model from falling into local optimal solutions.

The empirical analysis of the test set further validates the generalization ability of the model. Although the R^2 coefficient decreases to 0.147, it still shows the basic ability to explain the unknown data, and the trend of this metric is in line with the typical characteristics of predictive models in complex industrial scenarios. It is worth noting that the mean absolute error (MAE) of the test set stays at the level of 0.156, which is in the same order of magnitude as the training set error, and the mean bias (MBE) only slightly increases to 0.0096, indicating that the model maintains a stable prediction performance under new data scenarios.

The results show that this SA-SVM model can effectively balance the contradictory relationship between prediction accuracy and generalization ability, and its error control mechanism has universal revelation for dealing with high-dimensional and nonlinear energy consumption data. These findings provide a quantifiable decision-making basis for cloud computing service providers to optimize resource allocation and implement accurate energy efficiency management, which is of great practical guidance for promoting the transformation of data centers to low-carbon and intelligent direction.

References

[1] Soni, Dinesh, and Neetesh Kumar. "Machine learning techniques in emerging cloud computing integrated paradigms: A survey and taxonomy." Journal of Network and Computer Applications 205 (2022): 103419.

[2] Islam, Rafia, et al. "The future of cloud computing: benefits and challenges." International Journal of Communications, Network and System Sciences 16.4 (2023): 53-65.

[3] Jamsa, Kris. Cloud computing. Jones & Bartlett Learning, 2022.

[4] Alam, Ashraf. "Cloud-based e-learning: scaffolding the environment for adaptive e-learning ecosystem based on cloud computing infrastructure." Computer Communication, Networking and IoT: Proceedings of 5th ICICC 2021, Volume 2. Singapore: Springer Nature Singapore, 2022. 1-9.

[5] Katal, Avita, Susheela Dahiya, and Tanupriya Choudhury. "Energy efficiency in cloud computing data centers: a survey on software technologies." Cluster Computing 26.3 (2023): 1845-1875.

[6] Bharany, Salil, et al. "Energy efficient fault tolerance techniques in green cloud computing: A systematic survey and taxonomy." Sustainable Energy Technologies and Assessments 53 (2022): 102613.

[7] Balasubramaniam, S., et al. "Optimization enabled deep learning-based ddos attack detection in cloud computing." International Journal of Intelligent Systems 2023.1 (2023): 2039217.

[8] Khan, Habib Ullah, Farhad Ali, and Shah Nazir. "Systematic analysis of software development in cloud computing perceptions." Journal of Software: Evolution and Process 36.2 (2024): e2485.

[9] Anbalagan, Karthikeyan. "AI in cloud computing: Enhancing services and performance." International Journal of Computer Engineering And Technology (IJCET) 15.4 (2024): 622-635.

[10] Gill, Sajid Habib, et al. "Security and privacy aspects of cloud computing: a smart campus case study." Intelligent Automation & Soft Computing 31.1 (2022): 117-128.