Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.32460

Convergence Round Prediction for Deep Learning Training Based on Machine Learning Algorithm

Qixiang Luo[1*†], Zhanrong Liang[1†], Chenxi Li[1], Huixiong Yuan[1]

[1]Dongguan University of Technology, Dongguan, China
**Corresponding Author. Email: 2023414290224@dgut.edu.cn*
[†]These authors contributed equally to this work and should be considered as co-first author.

Abstract. At present, deep learning technology has been implemented on a large scale in computer vision, natural language processing and other fields, and the efficiency of model training and the optimal allocation of resources have become the core research topics focused on by industry and academia. Machine learning algorithm is the core support of deep learning performance prediction task. By mining the intrinsic correlation between hyperparameters, network structure, hardware resources and other features and convergence rounds, the quantitative prediction of training effect can be realized. In view of the shortcomings of traditional algorithms, such as insufficient feature extraction ability and weak timing-dependent capture, this paper proposes HLOA-CNN-BIGRU regression algorithm, which first carries out correlation analysis and violin graph analysis, and then relies on various machine learning algorithms to carry out comparative experiments. The experimental results show that the proposed algorithm is superior to the comparative algorithm in all evaluation indexes, and its MSE, RMSE, MAE and MAPE indexes are the lowest of all models, the highest R^2 value, the smallest prediction error, the best fitting effect and generalization ability; Among the traditional algorithms, integrated algorithms such as GBDT and AdaBoost have similar performance and strong stability, while decision tree and linear regression algorithms have weak prediction performance. This research provides a feasible solution for accurate prediction of deep learning convergence rounds, which can effectively help improve model training efficiency and resource utilization.

Keywords: Deep learning, HLOA, BIGRU, Convergence rounds

1. Introduction

With the widespread implementation of deep learning technology in computer vision, natural language processing and other fields, model training efficiency and resource optimization have become the core issues that industry and academia pay common attention to [1]. The training process of deep learning model is highly dependent on the debugging of hyperparameters and network structure. Different hyperparameters combine model complexity and dataset characteristics, which will directly affect the training rounds required for the model to converge to the target accuracy [2]. The traditional training mode relies on manual trial and error, which not only consumes a lot of computing resources and time cost, but also difficult to accurately match the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.32460

optimal training strategy. Especially in large-scale models and massive data scenarios, blind training will lead to waste of computing power and prolonged training cycle. At present, the demand for automated training and performance prediction of deep learning models is increasingly urgent. Accurate prediction of model convergence rounds can plan the training process in advance, optimize hyperparameter configuration, and realize efficient utilization of resources, which has also become the key technical support for automated deployment of deep learning [3].

Machine learning algorithms play a fundamental supporting role in deep learning performance prediction tasks, and realize quantitative prediction of training effects by mining the potential mapping relationship between features such as hyperparameter network structure hardware resources and convergence rounds [4]. Traditional machine learning regression algorithms, such as random forest gradient boosting tree, can complete basic numerical prediction tasks, but it is difficult to capture the nonlinear correlation and timing dependence characteristics between features in deep learning training process, and the prediction accuracy and generalization ability are limited. The core value of machine learning algorithms lies in getting rid of the traditional empirical trial-and-error mode, building prediction models in a data-driven way, providing quantitative basis for convergence round prediction, laying the foundation for more efficient fusion algorithm design, and promoting the transformation of deep learning training from manual debugging to intelligent prediction [5].

In order to solve the problems of insufficient feature extraction ability and weak timing-dependent capture of traditional algorithms, this paper proposes HLOA-CNN-BIGRU regression algorithm, which combines the advantages of mixed layer optimization algorithm convolutional neural network and bidirectional gated recurrent unit to build a high-precision convergent round prediction model. The algorithm realizes the adaptive optimization of hyperparameters through hybrid layer optimization algorithm, improves the model training efficiency, uses convolutional neural network to extract the deep spatial features of hyperparameters and network structure, and then captures the time series correlation features in the training process through bidirectional gated recurrent unit, and comprehensively mines the internal relationship between input features and convergence rounds. This algorithm makes up for the prediction defects of a single machine learning algorithm, effectively improves the accuracy and stability of convergence round prediction, and provides an innovative technical solution for the automatic and efficient training of deep learning models.

2. Data sources

This dataset contains 10 input feature variables and 1 target regression variable. The input feature covers the core objective factors that affect the convergence of deep learning training, namely, model complexity, batch size, learning rate, optimizer type, dataset size, target convergence accuracy, discard rate, graphics card memory utilization rate, data enhancement strength and initial weighting mode. The target variable is the convergence training rounds required for the model to reach the specified accuracy, and is used as a continuous numerical variable to construct the regression prediction task. First, the correlation analysis is performed, and the correlation heat map is output, as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.32460

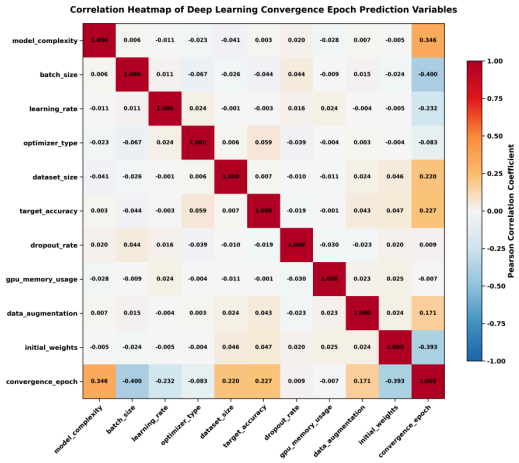


Figure 1. The correlation heat map

Violin plot analysis was performed on each variable and a violin plot was drawn as shown in Figure 2.

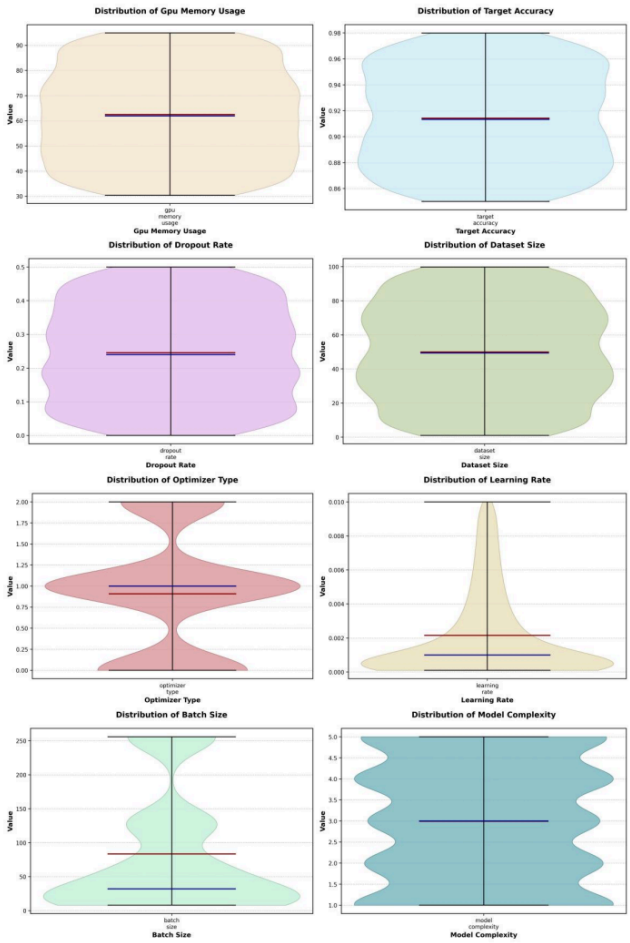


Figure 2. The violin plots

DOI: 10.54254/2755-2721/2026.32460

3. Method

3.1. HLOA

HLOA, or Horned Lizard Optimization Algorithm, is a new meta-heuristic optimization algorithm proposed in 2024. The core is to simulate multiple survival behaviors of horned lizards to build a search mechanism. The algorithm balances global exploration with local development by mathematically modeling behaviors such as skin lightening or darkening, blood spraying, and moving and escaping of horned lizards [6]. The skin color change strategy was used to adjust the search range to suit the needs of different optimization stages. The blood ejection behavior simulates the ejection motion, realizes the jump search of solution space, and enhances the global optimization ability. The move-escape strategy avoids the algorithm from falling into the local optimal solution by moving randomly and rapidly. In the iteration of the algorithm, the individual updates the position according to the behavior rules, and gradually approaches the optimal solution of the objective function.

3.2. CNN

CNN, or convolutional neural network, is the core model for processing grid data in deep learning, which is inspired by the local perception mechanism of biological visual cortex [7]. Its core structure includes convolution layer, pooling layer and fully connected layer. Through three ideas of local connection, weight sharing and hierarchical feature extraction, it solves the problem of parameter explosion and spatial feature loss of traditional fully connected network. The convolution layer uses multiple learnable convolution kernels to slide on the input data, calculate the local area dot product to generate a feature map, and extract basic features such as edges and textures. The pooling layer downsamples the feature map, reducing the data dimension and improving the model robustness. The fully connected layer flattens and weighted the high-level abstract features, and finally maps them to the output space.

3.3. BiGRU

BiGRU, namely Bidirectional Gated Recurrent Unit, is an improved structure of GRU, which is specially designed for bidirectional dependency modeling of sequence data. GRU controls the flow of information through update gate and reset gate, alleviates the gradient disappearance problem of traditional RNN, and captures the long-term dependence of sequence. BiGRU sets two independent GRU layers in parallel, one processing inputs forward in original timing and the other processing sequences in reverse order [8]. The forward GRU learns the impact of historical information on the current moment, and the reverse GRU captures the backtracking association of future information on the current moment. The final hidden state of each time step is formed by splicing forward and reverse outputs, and fuses complete context information. This structure significantly improves the feature expression ability of sequence data [9]. The network structure of BiGRU is shown in Figure 3.

DOI: 10.54254/2755-2721/2026.32460

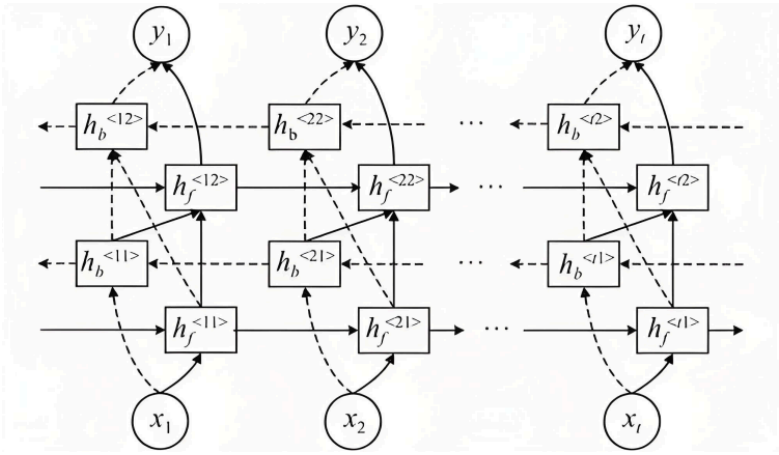


Figure 3. The network structure of BiGRU

3.4. HLOA-CNN-BiGRU

HLOA-CNN-BiGRU regression algorithm is a composite model of fusion optimization, feature extraction and time series modeling, which is used to improve the accuracy of regression prediction. First, the algorithm uses HLOA to optimize the key hyperparameters of CNN and BiGRU, such as the size of convolution kernel and the number of GRU hidden units, so as to avoid the limitation of manual parameter adjustment [10]. The CNN module receives input data, extracts local spatial features through convolution and pooling operations, and converts high-dimensional data into low-dimensional feature sequences. The feature sequence is input to the BiGRU layer, which bidirectionally captures the front and back dependencies of time series data, and learns complex dynamic patterns. Finally, the bidirectional timing features are integrated through the fully connected layer, and the continuous regression prediction value is output.

4. Result

The HLOA-CNN-BIGRU regression project has clear parameter settings, the number of algorithm iterations is set to 100 times, the population size is determined to be 30 search agents, the learning rate is adjusted to 0.001, the discard probability of the spatial dropout layer is set to 0.2, the CNN partial convolution kernel size is 3×3, the pooling layer kernel size is 2×2, the hidden layer dimension of BIGRU layer is set to 128, the batch size is fixed to 32, the number of training rounds is set to 50 rounds, the boundary range of objective function optimization is set to between-10 and 10, the random walk step control is set to 0.1, the weight coefficient of imitation behavior is set to 0.8, the ejection step parameter is set to 0.5, and the Cauchy distribution scale parameter in the initialization stage is adjusted to 1.0.

The comparative experiments of each machine learning algorithm are output, as shown in Table 1. The comparative histogram of each metric of the HLOA-CNN-BIGRU test set is output, as shown in Figure 4.

Table 1. The results of the comparative experiment

Model	MSE	RMSE	MAE	MAPE	R^2
ExtraTrees	250.061	15.813	12.592	10.163	0.801
CatBoost	136.744	11.694	8.5	6.561	0.868

DOI: 10.54254/2755-2721/2026.32460

Table 1. (continued)

Decision tree	328.034	18.112	14.048	11.55	0.682
KNN	734.172	27.096	22.017	17.474	0.336
Random forest	179.763	13.408	10.706	8.552	0.851
AdaBoost	146.732	12.113	9.587	7.696	0.867
GBDT	90.42	9.509	7.51	6.031	0.922
Linear regression	416.89	20.418	16.659	13.601	0.629
HLOA-CNN-BIGRU	80.715	8.984	6.797	5.389	0.933

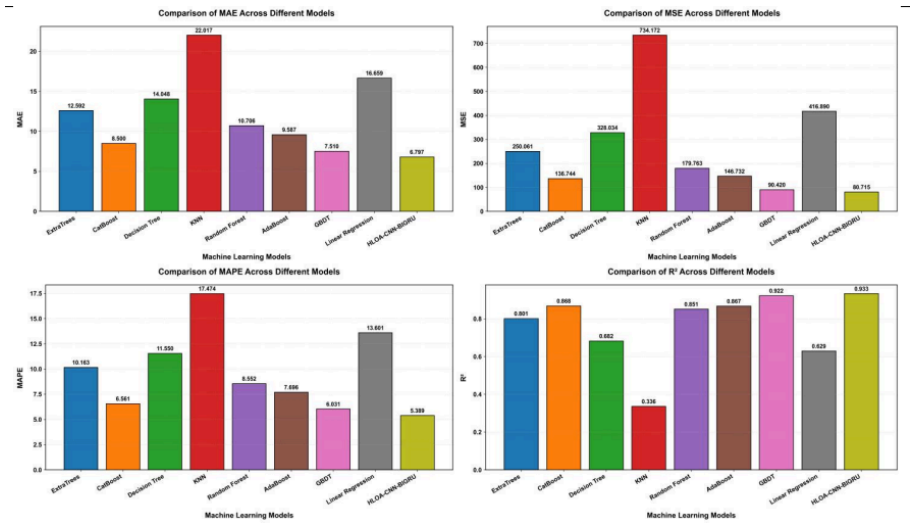


Figure 4. The comparison bar charts of each indicator

Comparing the experimental results of various regression algorithms, it can be clearly seen that the HLOA-CNN-BIGRU algorithm proposed in this paper is superior to other comparison algorithms in various evaluation indexes. The MSE, RMSE, MAE and MAPE of this algorithm are the lowest among all models, and the R^2 value is the highest, indicating that it has the smallest prediction error, the best fitting degree and the strongest generalization ability. In the traditional machine learning algorithms, the overall performance of GBDT, AdaBoost, CatBoost and random forest is relatively close, and they have good stability and accuracy in regression tasks. Among them, the effect of GBDT is relatively outstanding, while the error indexes of extreme random tree, decision tree, K-nearest neighbor and linear regression are obviously higher, the R^2 value is lower, and the overall prediction performance is weak. Especially, the fitting effect of K-nearest neighbor and linear regression is poor in this regression task. Based on various indicators, it can be concluded that the HLOA-CNN-BIGRU algorithm proposed in this paper can effectively reduce the prediction error, improve the fitting accuracy and reliability of the model, and show significant advantages over the traditional machine learning algorithm in convergence round prediction regression task.

5. Conclusion

With the wide application of deep learning technology in computer vision, natural language processing and other fields, model training efficiency and resource optimization have become the core issues of common concern between industry and academia. Machine learning algorithm is the basic support of deep learning performance prediction task, which can realize the quantitative

DOI: 10.54254/2755-2721/2026.32460

prediction of training effect by mining the potential correlation between hyperparameters, network structure, hardware resources and other features and convergence rounds. Aiming at the problems of insufficient feature extraction ability and weak timing-dependent capture of traditional algorithms, this paper proposes HLOA-CNN-BIGRU regression algorithm, and verifies its performance through correlation analysis, violin graph analysis and comparative experiments of various machine learning algorithms. The experimental results show that the algorithm is superior to other comparative algorithms in all evaluation indexes, and its mean square error, root mean square error, mean absolute error and mean absolute percentage error are the lowest, the determination coefficient is the highest, the prediction error is the smallest, the fitting degree is the best and the generalization ability is the strongest. Among the traditional algorithms, gradient boosting decision tree, adaptive boosting, classification boosting tree and random forest have similar performance and good stability. Among them, gradient boosting decision tree has relatively outstanding effect, while extreme random tree, decision tree, k-nearest neighbor and linear regression have weak prediction performance and poor fitting effect. This algorithm effectively improves the accuracy and reliability of convergence round prediction, provides effective technical support for resource optimization and efficiency improvement of deep learning model training, and has important theoretical and practical application value.

References

[1] Lee, Shao-Lun, et al. "Forecasting urban fire severity for enhanced emergency response and resource allocation." Scientific Reports 15.1 (2025): 42165.
[2] Saini, Kanika, Sheetal Kalra, and Sandeep K. Sood. "Disaster emergency response framework for smart buildings." Future Generation Computer Systems 131 (2022): 106-120.
[3] Chaoura, Chaimaa, Hajar Lazar, and Zahi Jarir. "Enhancing emergency response in road accidents: A severity prediction framework using RF-RFE and deep learning model." IEEE Access (2025).
[4] Sahebi, Ali, et al. "Predicting firefighting operation time in urban areas using machine learning: identifying key determinants for improved emergency response." Discover Applied Sciences 7.4 (2025): 250.
[5] Türkan, Yusuf Sait, and Mesut Ulu. "A hybrid approach to traffic incident management: machine learning-based prediction and patrol optimization." IEEE Access 13 (2025): 43455-43472.
[6] Huang, Di, et al. "Ai prediction of power grid faults based on deep learning and improvement of emergency response efficiency in automated repair." Distributed Generation & Alternative Energy Journal (2025): 63-84.
[7] Schubert, Jochen E., et al. "A framework for mechanistic flood inundation forecasting at the metropolitan scale." Water Resources Research 58.10 (2022): e2021WR031279.
[8] Maddireddy, Bhargava Reddy, and Bharat Reddy Maddireddy. "Enhancing Network Security through AI-Powered Automated Incident Response Systems." International Journal of Advanced Engineering Technologies and Innovations 1.02 (2023): 282-304.
[9] Jamshed, Muhammad Ali, et al. "Green uav-enabled internet-of-things network with ai-assisted noma for disaster management." 2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2024.
[10] Van Hoang, Thanh. "Impact of integrated artificial intelligence and internet of things technologies on smart city transformation." Journal of technical education science 19.Special Issue 01 (2024): 64-73.