Transactions on Computational and Scientific Methods | Vo. 5, No. 4, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Research on improved zebra optimization algorithm based on Cat mapping and crossover strategy

Di Wu[1], Maoxingyue Zhou[2]
[1]Independent Research, USA
[2]Independent Research, USA
*Corresponding Author: Di Wu; inbox.diwu@gmail.com

Abstract: In order to improve the convergence speed and optimization accuracy of the zebra optimization algorithm (ZOA), this paper proposes an improved zebra optimization algorithm (DCZOA) based on Cat mapping and cross-cutting strategies. Firstly, the Cat mapping method is used to process the initial population to improve the diversity and distribution uniformity of the population; secondly, the cross-cutting strategy is introduced to enhance the global search ability of the algorithm and ensure the optimization update of the zebra position. The simulation experiments on 16 standard test functions show that the improved algorithm has significant improvements in convergence speed and optimization accuracy. The improved algorithm is further applied to the hyperparameter optimization of the random forest regression model, and a combined experiment is carried out using the used car price dataset. The results verify the effectiveness of the algorithm in practical applications. Finally, the advantages of the algorithm are further clarified by combining SHAP analysis.

Keywords: Cat mapping; cross-cutting strategy; SHAP analysis; XGBOOST

1. Introduction

With the continuous advancement of intelligent optimization technology, optimization methods based on biological behavior have shown significant advantages in solving complex global optimization problems. This type of algorithm can find solutions close to the global optimal solution in a complex search space by simulating the cooperation and competition mechanism of biological groups in nature. The zebra optimization algorithm [1] (ZOA) is an emerging group intelligent optimization algorithm. Due to its simple structure and easy implementation, it has been widely used in many fields, such as engineering design, data mining and machine learning [2].

The algorithm is inspired by the behavior pattern of zebra groups, especially the division of labor and cooperation among the three roles of leader, follower and explorer to simulate the optimization process. The leader is responsible for guiding the movement direction of the group, the follower assists the leader in local search, and the explorer conducts random exploration outside the group to maintain the diversity of the population and avoid falling into the local optimum too early. Through information sharing and cooperation strategies, the zebra group gradually narrows the search range and finally approaches the optimal solution.

Although the zebra optimization algorithm performs well in many applications, its performance still has room for improvement. First, the initial population generation method of ZOA is relatively random, which may lead to uneven population distribution and affect search efficiency. Secondly, when dealing with complex multi-peak optimization problems, ZOA is prone to fall into local optimality, especially in high-dimensional problems, and its global exploration ability is insufficient. In addition, as the number of iterations increases, the population diversity gradually decreases, the local development ability weakens, and the convergence speed slows down.

In response to the above problems, this paper proposes an improved zebra optimization algorithm (DCZOA), which improves the global exploration and local development capabilities of the algorithm by introducing a new population initialization method and search strategy. Specifically, DCZOA uses the Cat mapping method [3] to generate the initial population and uses the characteristics of chaotic mapping to ensure the diversity and uniformity of the population. This method effectively avoids the problem of uneven population distribution that may be caused by traditional random initialization, thereby improving the efficiency of early search.

In addition, DCZOA introduces a vertical and horizontal crossover strategy to enhance the global search ability of the algorithm through vertical and horizontal crossover operations to ensure the optimal update of the zebra position. This strategy can effectively prevent the algorithm from falling into local optimality in the early stage and accelerate the convergence speed in the later stage. In order to further improve the local search ability, an adaptive step size adjustment mechanism is adopted to dynamically adjust the search step size according to the quality of the solution, so as to perform more refined development in the search space.

In the experimental part, this paper evaluates the performance of DCZOA through 16 standard test functions and compares it with the classic zebra optimization algorithm and particle swarm optimization algorithm. The experimental results show that DCZOA exhibits faster convergence speed and higher solution accuracy when dealing with complex optimization problems such as high-dimensional, nonlinear and multi-peak. At the same time, DCZOA shows strong stability in multiple independent runs, showing its good robustness [4].

In order to verify the effectiveness of DCZOA in practical applications, this paper applies it to the hyperparameter optimization of the used car price prediction model. The experimental results show that compared with traditional optimization methods, it can significantly improve the prediction accuracy of the model, further proving its practical value in practical problems.

The main contributions of this paper are as follows: (1) An improved zebra optimization algorithm based on Cat mapping and cross-cutting strategy is proposed, which significantly improves the global exploration ability and local development efficiency of the algorithm; (2) The superior performance of DCZOA is verified by standard test functions and practical application cases, showing its wide application potential in complex optimization problems.

2. Method

The Zebra Optimization Algorithm (ZOA)[5] is a heuristic optimization algorithm based on the behavior of zebra groups in nature, proposed by E. Trojovská et al.[1] in 2022. The algorithm solves the optimization problem by simulating the foraging and defense behaviors of zebras in the natural environment, and has the characteristics of strong global search ability and fast convergence speed.

The core idea of ZOA is to achieve position update and optimization search by simulating the foraging and defense strategies of zebra groups. In foraging behavior, zebra members will adjust their positions according to the distribution of pasture, and the zebra with the best position is regarded as a "pioneer zebra" to guide other members to move to a better area. In defensive behavior, zebras will adopt different response strategies depending on the type of predator, such as escaping lions by zigzagging or random movement, or confusing or deterring other predators by gathering in groups.

In the zebra optimization algorithm, the initial population is generated by random method. The specific formula is as follows:

$$x_{i,j} = lb_j + r \cdot (ub_j - lb_j)$$

Among them: $x_{i,j}$ represents the j-th dimension position of the i-th individual; ub_j and lb_j represent the upper and lower bounds of the j-th dimension search respectively; r is a random number in the range [0,1]. This method ensures the uniform distribution of the initial population in the search space, laying a good foundation for the subsequent optimization process.

The zebra optimization algorithm updates the population position by simulating the foraging behavior of zebras. Specifically, the zebra group will move closer to the pioneer zebra (the current optimal individual) to achieve the position update. The position update formula is as follows:

$$x_{r+1} = x_r + r \cdot (PZ - I \cdot x_r)$$

Among them, x_r represents the position of the current individual; PZ represents the position of the current individual; I is a random value in {1, 2}; D is a random number in the range of [0, 1].

Zebras adopt different defense strategies when facing different types of predators. When zebras encounter large predators (such as lions), they will take a zigzag approach to avoid attacks; when facing small predators, zebras will choose to gather in groups to fight back. These behaviors are simulated as the following position update formula:

$$x_{t+1} = \begin{cases} S_1 : x_t + r_1 \cdot (2r_2 - 1) \cdot (1 - \frac{t}{T}) \cdot x_t, r_3 > 0.5 \\ S_2 : x_t + r_4 \cdot (AZ - I \cdot x_t) \qquad else \end{cases}$$

Among them: S_1 represents the defense strategy when facing large predators, which is to escape predators by zigzagging; S_2 represents the defense strategy when facing small predators, which is to fight back by gathering in groups; r_1, r_2, r_3, r_4 is a random number in the range [0,1]; AZ represents the position of the attacked zebra; T is the current number of iterations; t is the maximum number of iterations. This strategy enables the zebra to dynamically adjust its behavior according to different threats, thereby better balancing global exploration and local development during the optimization process.

This paper proposes an improved zebra optimization algorithm (DCZOA) based on chaotic mapping and hybrid crossover strategy. First, the Cat chaotic mapping technology is used to generate the initial population to enhance the diversity of the population and the uniformity of spatial distribution; secondly, combined with the hybrid crossover strategy, the global search ability of the algorithm is improved through vertical and horizontal crossover operations, thereby more effectively optimizing the zebra position update process. This improvement not only improves the convergence speed of the algorithm, but also enhances its robustness in complex optimization problems.

In order to enhance the diversity of the population and ensure the uniform distribution of the initial individuals in the search space, this paper proposes an initialization method that combines chaotic mapping and reverse learning strategy. This method aims to accelerate the convergence speed of the algorithm. Although there are many hybrid optimization algorithms that combine chaotic mapping sequences with traditional optimization algorithms and have achieved certain results, most of these algorithms rely on Logistic mapping. Logistic mapping has problems such as uneven traversal, sensitivity to initial values, and poor traversal and uniformity (for example, the density of mapping points is higher in the edge area and lower in the center area), which limit the performance of chaotic search.

This paper studies the structure and chaotic characteristics of Cat mapping[6]. Cat mapping is a two-dimensional reversible chaotic mapping, and its dynamic equation is as follows:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1,1 \\ 1,2 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \quad \mathrm{mod}\, 1$$

Cat mapping has the characteristics of simple structure, better traversal uniformity and faster iteration speed. In the interval [0,1], the chaotic sequence generated by Cat mapping is evenly distributed.

The population initialization steps based on the chaotic reverse learning strategy are as follows: First, use the Cat chaotic sequence to generate N initial solutions X_i, and then generate the corresponding reverse solution for each initial solution:

$$OP_i = K(X_{\min}^d + X_{\max}^d) - X_i$$

Among them, K is a random number in the interval [0,1], OP_i is the reverse solution of the initial solution $X_{\min}^d$, and $X_{\max}^d$ and d represent the minimum and maximum values of the d-th dimension vector in all initial solutions. Finally, the initial solution and the reverse solution are merged and sorted in ascending order according to the fitness value (for the minimization problem), and the top N solutions with the best fitness value are selected as the initial population.

The vertical and horizontal crossover strategy is a key method that can significantly improve the performance of the dung beetle optimization algorithm. In the search process of the traditional dung beetle optimization algorithm, it is easy for individuals in the population to concentrate in local areas, resulting in the problem of falling into the local optimal solution. In order to overcome this limitation, the vertical and horizontal crossover strategy encourages individuals to explore a wider range in the solution space by crossover operations on individuals in the population. Specifically, the vertical crossover enhances the exploration ability of individuals in the global range, allowing the population to jump out of the local optimal area and find a more promising search direction; the horizontal crossover performs fine-tuning in the local range to improve the algorithm's ability to refine the solution and ensure that the key solution space area will not be missed when approaching the optimal solution. This strategy of combining vertical and horizontal crossover not only increases the diversity of the algorithm, but also effectively accelerates the convergence process. Compared with the traditional dung beetle optimization algorithm, the improved version combined with the vertical and horizontal crossover strategy shows higher optimization accuracy and stability when dealing with complex optimization problems. In the end, the algorithm can find the global optimal solution more quickly and accurately, reduce computational costs, and improve overall performance.

In the zebra optimization algorithm, the horizontal crossover operation is similar to the crossover operation in the genetic algorithm, which mainly performs crossover operations on the same dimension of different individuals. In view of the problem that the global search ability of the zebra optimization algorithm is insufficient, this paper introduces a horizontal crossover strategy to optimize the position of individuals in the population. First, the individuals in the population are randomly arranged, and the crossover operation is performed on the first dimension to update the individual position. This method enhances the global search ability of the algorithm by strengthening the information exchange between individuals, enabling the zebra algorithm to more effectively jump out of the local optimum and find the global optimal solution. The specific formula is as follows:

$$MSx_{i,d}^{t} = r1 \times x_{i,d}^{t} + (1 - r1) \times x_{j,d}^{t} + c1 \times (x_{i,d}^{t} \times x_{j,d}^{t})$$

$$MSx_{j,d}^{t} = r2 \times x_{j,d}^{t} + (1 - r2) \times x_{i,d}^{t} + c1 \times (x_{j,d}^{t} \times x_{i,d}^{t})$$

Among them: $MSx_{i,d}^{t}$ and $MSx_{j,d}^{t}$ represent the d-th dimension individuals generated by the vigilants $x_{i,d}^{t}$ and $x_{j,d}^{t}$ through horizontal crossover, r1 and r2 are random numbers in [0,1], and $c1$ and $c2$ are random numbers in [-1,1].

In the zebra optimization algorithm, after the horizontal crossover operation, individuals can generate offspring individuals with a higher probability in their respective hypercube spaces and outer edges, thereby expanding the algorithm's search space and improving the global search capability. The solution generated by horizontal crossover needs to be compared with the parent generation, and individuals with higher fitness are selected for retention. This means that the number of offspring generated in the outer edge space will gradually decrease linearly as the distance between the parent individuals increases. Under this mechanism, the algorithm can gradually converge to the optimal solution and maintain a high convergence efficiency while ensuring the accuracy of the solution. Through the horizontal crossover operation, the zebra algorithm effectively balances the ability of exploration and development, significantly enhancing the overall performance.

In the later stages of the zebra optimization algorithm, the algorithm is prone to fall into the local optimal solution. This is mainly because some individuals in the population converge to the local optimal value too early in some dimensions, which limits the search range of the entire population and makes it impossible to fully explore the global optimal solution. Through further analysis, it is found that the zebra optimization algorithm lacks an effective diversity maintenance mechanism and cannot effectively adjust individuals that have fallen into the local optimal solution, thus limiting the ability of the algorithm to further approach the global optimal solution. In order to solve this problem, after completing the horizontal crossover operation, it is also necessary to perform a vertical crossover operation on the newly generated individuals to enhance the algorithm's ability to jump out of the local optimal solution. Through vertical crossover, the algorithm can expand the search range of individuals and further improve the global search performance of the population, thereby effectively avoiding the dilemma of falling into the local optimal solution.

The crossover operation is a crossover operation performed on all dimensions of the new individuals. The probability of occurrence is less than that of the horizontal crossover operation, which is similar to the mutation operation of the genetic algorithm.

Assuming that the new individual $x_{i,d}^{t}$ is the offspring individual generated by vertical crossover in the d_1 and d_2 dimensions, the calculation method is as follows:

$$MSx_{i,d}^{t} = r1 \times x_{i,d1}^{t} + (1 - r) \times x_{i,d2}^{t}$$

Among them, $MSx_{i,d}^{t}$ is the offspring individual generated by the vertical crossover of the d1-th dimension and the d2-th dimension of individual $x_{i,d1}^{t}$. Similar to the horizontal crossover operation, the offspring individuals generated by the vertical crossover operation need to compete with the parent generation, and the individuals with higher fitness are selected for preservation. Through this preferential selection mechanism, the dung beetle individuals participating in the crossover will not lose some excellent dimensional information, but will improve the diversity of the population and continuously improve the quality of the solution. After performing the vertical crossover, the individuals that have fallen into the local optimum can make full use of the useful information on each dimension, so that they have the opportunity to jump out of the local optimum.

During the iteration process, if a dimension of an individual jumps out of the local optimum through the vertical crossover operation, it will quickly be combined with the entire population through the horizontal crossover operation to consolidate the dimension of the new solution. This will give other dimensions that are trapped in the local optimum more opportunities to jump out of the local optimum. By combining the horizontal crossover operation with the vertical crossover operation, the convergence efficiency and solution accuracy of the algorithm to get rid of the local optimum can be effectively improved.

3. Experiment

3.1 Dataset

First, this paper selected 16 standard test functions as benchmark data sets to evaluate the performance of the improved zebra optimization algorithm (DCZOA) in terms of convergence speed, optimization accuracy and stability. These test functions cover complex optimization problems such as high-dimensional, multi-peak and nonlinear, and can fully reflect the improvement effect of the algorithm in global search and local development capabilities. By comparing experiments with the original zebra optimization algorithm (ZOA) and particle swarm optimization algorithm (PSO), it is verified that DCZOA has better results and stronger robustness in multiple rounds of independent experiments.

Secondly, in order to further verify the application effect of the algorithm in actual scenarios, the author applied it to the used car price prediction task, using a real used car price regression data set. The data set is divided into training set and test set in a ratio of 7:3. By optimizing the hyperparameters of the XGBOOST model, the prediction performance of the unoptimized model, ZOA optimized model and DCZOA optimized model is compared. The results show that the DCZOA optimized model performs better in indicators such as mean square error and mean absolute error, demonstrating the practical value of the algorithm in actual machine learning tasks.

3.2 Experimental design

In order to verify the optimization performance and application value of the improved algorithm, this paper designed 16 benchmark function experiments to test the effectiveness of the proposed algorithm. In order to effectively compare the difference between ZOA and DCZOA[7] algorithms, in the experiment, the population size was set to 30 and the maximum number of iterations was 500. To ensure the accuracy of the experiment, all experiments were run 30 times, and the optimal value, average value and standard deviation of each experiment were recorded as evaluation indicators of the experimental performance.

3.3 Experimental Results

In order to intuitively see the comparison between the two algorithms, Figure 1 shows the convergence effect diagrams of four benchmark functions: Sphere, Rastrigin, Ackley, and Griewank. It can be seen from the figure that the DCZOA algorithm has a faster convergence speed and convergence effect.

To make the results more convincing, 30 comparative experiments were conducted on 8 benchmark functions, and the optimal value, average value and standard deviation of each experiment were recorded. The average value of the 30 experiments was calculated as the final result for comparison, as shown in Table 1.

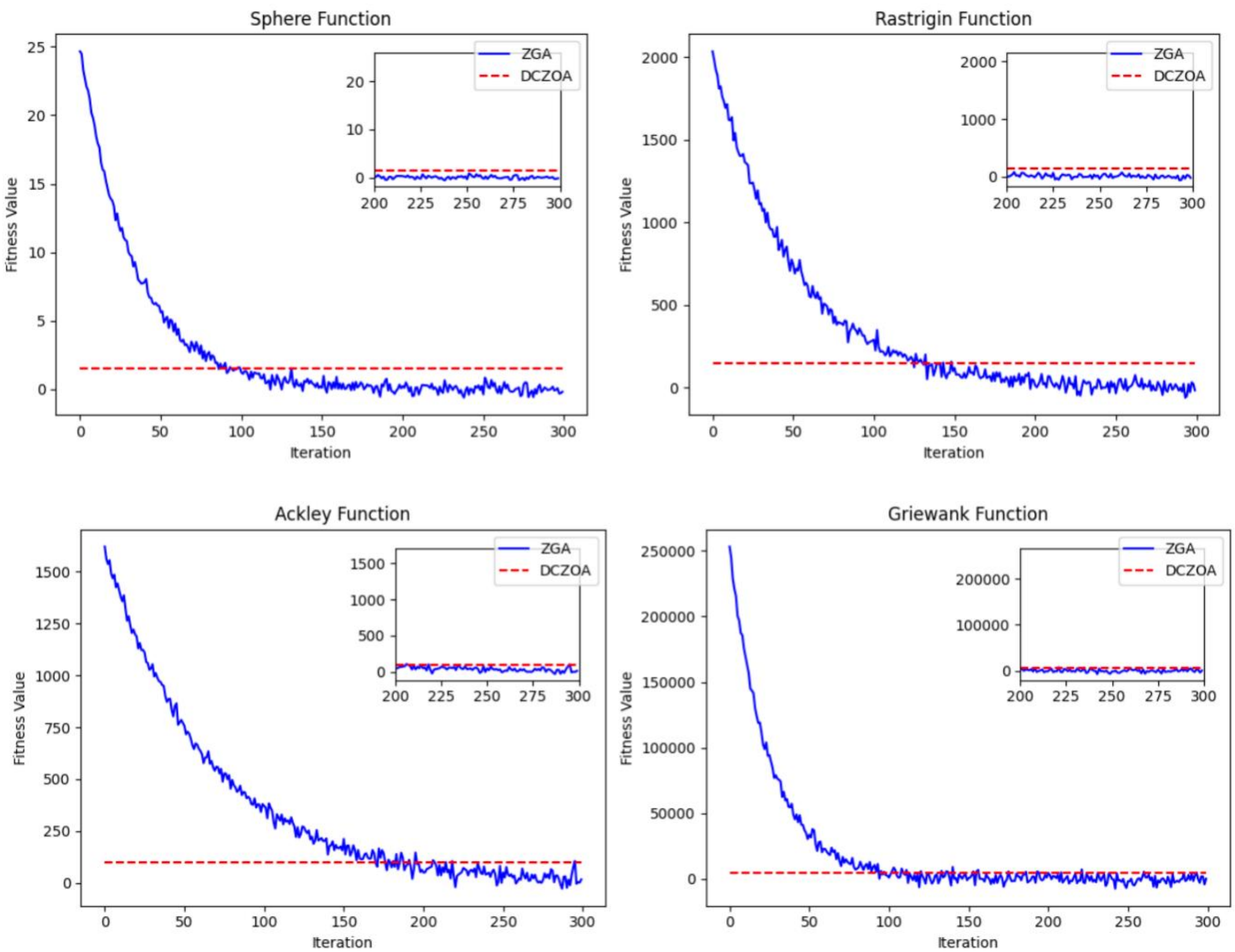

Figure 1. Convergence effect diagram of four benchmark functions

Table 1: Comparative experimental results

Algorithm	**Benchmark Function**	**Optimal value**	**Average value**
ZOA	F1	3.25 x 10^-4	2.17 x 10^-3
DCZOA	F1	1.26 x 10^-5	6.55x 10^-4
ZOA	F2	3.91 x 10^-4	1.25 x 10^-3
DCZOA	F2	3.31 x 10^-6	4.31 x 10^-4
ZOA	F3	2.61 x 10^-5	6.32x 10^-5
DCZOA	F3	1.21 x 10^-6	3.71 x 10^-6
ZOA	F4	6.01 x 10^-5	9.18 x 10^-5
DCZOA	F4	2.01 x 10^-6	5.11 x 10^-6
ZOA	F5	5.02 x 10^-3	7.01 x 10^-1
DCZOA	F5	3.07 x 10^-3	1.08 x 10^-2
ZOA	F6	5.21 x 10^-4	3.21 x 10^-3

DCZOA	F6	3.47 x 10^-5	1.04x 10^-4
ZOA	F7	5.26x 10^-4	1.94 x 10^-3
DCZOA	F7	9.05 x 10^-5	3.21 x 10^-4
ZOA	F8	3.51 x 10^-5	6.11x 10^-5
DCZOA	F8	2.97 x 10^-6	5.15 x 10^-6

After conducting comparative experiments, this paper further verifies the application of the algorithm in downstream tasks. First, a used car price prediction regression dataset is selected, and the dataset is divided into a training set and a test set in a ratio of 7:3. The model is trained on the training set, and the generalization ability of the model is tested on the test set. XGBOOST regression model, ZOA-XGBOOST, and DCZOA-XGBOOST[8] are established respectively, and the models are compared. The performance of the three models is shown in Table 2.

Table 2: Downstream regression experimental results

Method	**MAE**	**MSE**	**RMSE**
XGBOOST	1.1839	14.026	3.943
ZOA-XGBOOST	1.486	11.536	3.369
DCZOA-XGBOOST	1.512	10.438	3.230

As shown in Table 2, XGBOOST optimized with the DCZOA optimization algorithm has better generalization ability.

Through the analysis of experimental results, the DCZOA-XGBOOST model shows excellent prediction performance. In order to further explain the prediction results of the model, this paper adopts the SHAP[9] (SHapley Additive exPlanations) analysis method. The SHAP value is an interpretation method based on the Shapley value theory in game theory, which is used to quantify the contribution of each feature to the model prediction results. Its core idea is to evaluate the impact of features on model output by calculating the marginal contribution of features under different combinations.

First, a feature importance analysis diagram is given, as shown in Figure 2.

A positive SHAP value indicates that the feature has a positive impact on the prediction result, while a negative SHAP value indicates a negative impact. Taking the feature Power as an example, when the Power value is large, the SHAP value is positive, indicating that it has a positive impact on the prediction result, that is, the predicted Price is higher; and when the Power value is small, the SHAP value is negative, indicating that it has a negative impact on the prediction result, that is, the predicted Price is lower. Through this visualization method, we can clearly see the specific impact of each feature on the model prediction result, so as to better understand the decision logic of the model.

Next, a partial dependence diagram is given, as shown in Figure 3.

The partial dependence diagram reflects the relationship between the feature variable and the target variable. Taking Power as an example, when the value is less than 150, the shap value is negative, which indicates a negative impact. When the value is greater than 150, it is positive, which indicates a positive impact, indicating that the predicted Price increases. This means that as Power increases, Price also increases.

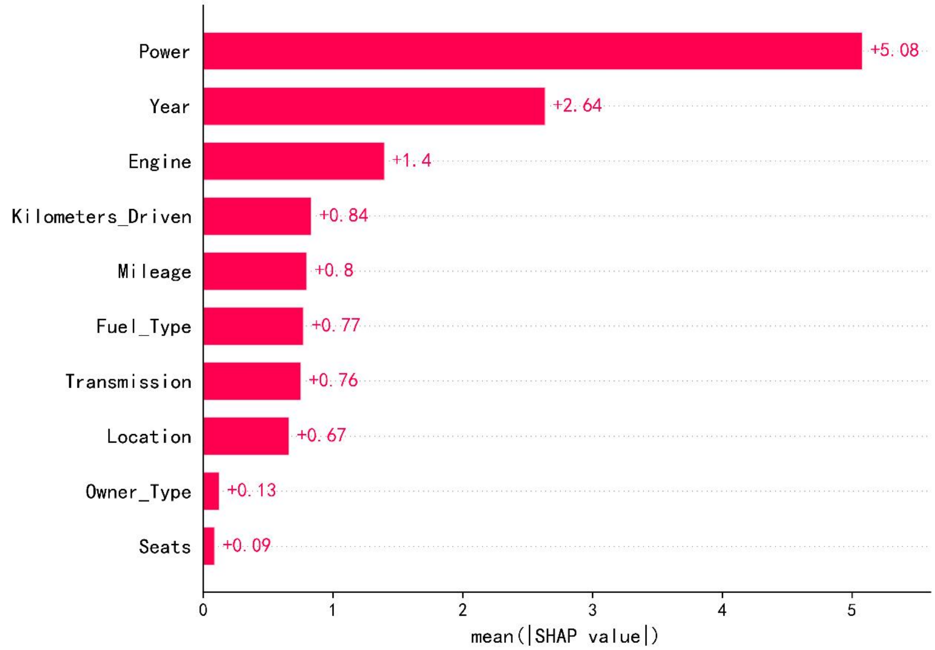

Figure 2. Feature Importance Analysis Chart

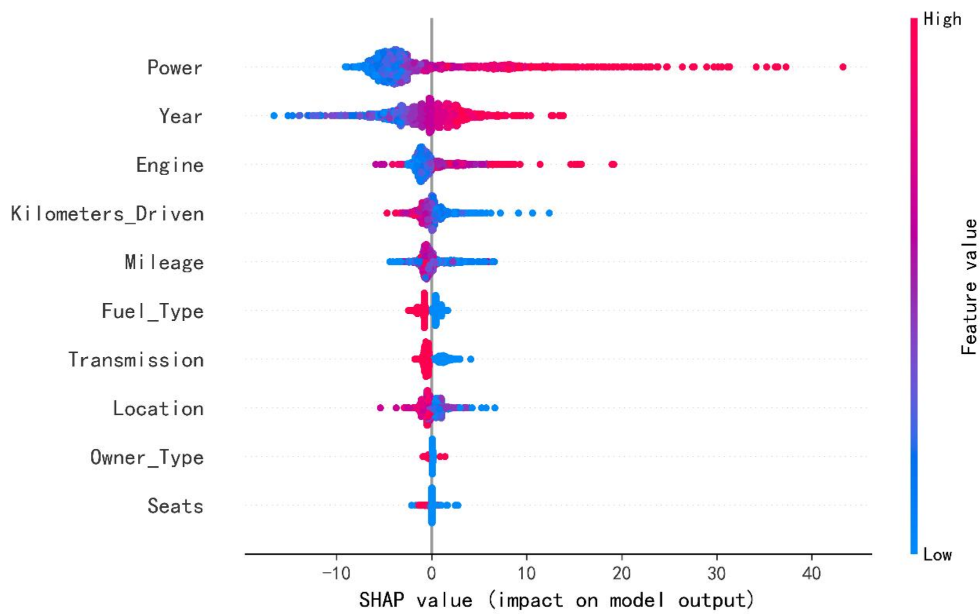

Figure 3. Partial Dependence Plot

4. Conclusion

This paper proposes an improved zebra optimization algorithm (DCZOA) based on Cat mapping and vertical and horizontal crossover strategies, aiming to improve the optimization efficiency and model performance in small sample text classification tasks. By introducing a more chaotic Cat mapping to generate the initial population, the diversity and uniformity of the initial search space are improved;

combined with vertical and horizontal crossover operations, the global search ability and local development ability of the algorithm are effectively enhanced, and the problem of falling into the local optimum is significantly alleviated. The experimental results of standard test functions and actual regression tasks show that the proposed algorithm is superior to traditional ZOA and other comparative algorithms in terms of optimization accuracy, convergence speed and robustness.

Furthermore, in the XGBOOST model hyperparameter tuning task, the DCZOA optimized model shows better generalization performance, and the prediction error indicators are lower than the baseline model, and the SHAP analysis shows the model's potential in interpretability. This study not only verifies the effectiveness of the DCZOA algorithm in theory and practice, but also provides a feasible optimization solution for scenarios that require efficient optimization capabilities such as used car price prediction, showing good application prospects and promotion value.

References

[1] Trojovská E, Dehghani M, Trojovský P. Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm[J]. Ieee Access, 2022, 10: 49445-49473.

[2] Mohapatra S, Mohapatra P. American zebra optimization algorithm for global optimization problems[J]. Scientific Reports, 2023, 13(1): 5211.

[3] Qi Z, Peng S, Wu P, et al. Renewable Energy Distributed Energy System Optimal Configuration and Performance Analysis: Improved Zebra Optimization Algorithm[J]. Sustainability, 2024, 16(12): 5016.

[4] Rana A, Khurana V, Shrivastava A, et al. A ZEBRA optimization algorithm search for improving localization in wireless sensor network[C]//2022 2nd International conference on technological advancements in computational sciences (ICTACS). IEEE, 2022: 817-824.

[5] Amin R, El-Taweel G, Ali A F, et al. Hybrid chaotic zebra optimization algorithm and long short-term memory for cyber threats detection[J]. IEEE Access, 2024.

[6] Anand D, Khalaf O I, Chandra G R. Enhancing the Zebra Optimization Algorithm with Chaotic Sinusoidal Map for Versatile Optimization[J]. Iraqi Journal for Computer Science and Mathematics, 2024, 5(1): 12.

[7] Krithiga G, Senthilkumar S, Alharbi M, et al. Design of modified long short-term memory-based zebra optimization algorithm for limiting the issue of SHEPWM in multi-level inverter[J]. Scientific Reports, 2024, 14(1): 22439.

[8] Rani T J, Kavitha D. Classification of Epileptic Seizures Using LSTM Based Zebra Optimization Algorithm with Hyperparameter Tuning[J]. International Journal of Intelligent Engineering & Systems, 2024, 17(3).

[9] Ghadi Y Y, Neamah N M, Hossam-Eldin A A, et al. State-of-the-art frequency control strategy based on an optimal Fuzzy PI-FOPDF λ for SMES and UPFC integrated smart grids using Zebra Optimization Algorithm[J]. IEEE Access, 2023, 11: 122893-122910.