Research on Optimization of Crop Planting Schemes Based on Q-Learning and Genetic Algorithm

Jintao Yang[1,a,*,#], Zhuo Liu[1,b,#], Shengkai Wan[1,c,#]

[1]School of Civil Engineering and Transportation, Northeast Forestry University, Harbin, China

[a]2642689088@qq.com, [b]3012008583@qq.com, [c]2727412065@qq.com

[#]These Authors Contributed Equally to This Work

*Corresponding author

Keywords: Linear Programming, Non-Dominated Sorting Genetic Algorithm, Q-Learning

Abstract: This paper focuses on determining the optimal crop planting scheme. After comparing traditional models such as linear programming and stochastic optimization, a modeling approach employing a hybrid optimization algorithm and multi-objective optimization strategy is adopted. Under the assumption of stable crop yields, planting costs, and sales prices, the model utilizes a genetic algorithm (GA) to optimize the allocation of crop planting areas. Additionally, reinforcement learning via Q-learning dynamically adjusts crossover and mutation operations to enhance the global search capability. By introducing two scenarios—unsalable crops and price reductions—and employing penalty functions to manage risks, a multi-objective programming model is established. The model is solved using the non-dominated sorting genetic algorithm (NSGA) to obtain Pareto optimal solutions, leading to the determination of optimal planting schemes for both scenarios. The results demonstrate the model's high profitability and stability across different conditions.

1. Introduction

The primary focus of this inquiry lies in addressing the multi-objective optimization challenge related to rural crop planting, incorporating elements such as uncertainty, crop substitution, and complementarity to devise an optimal planting strategy. By accurately modeling various factors like crop yields, sales prices, and planting costs, we aim to optimize resource allocation and elevate production efficiency.

In related research, Liang Zhou and his colleagues[1]introduced a unique discrete-time data-driven predictive sliding mode control (DDPSMC) approach. They designed a nonlinear integral terminal sliding mode surface to replace the conventional linear sliding mode function, aiming to expedite system error convergence and mitigate chattering. Additionally, S. Iwamoto and team[2]proposed a swift reactive power and voltage control method, formulating the AC load flow equation in a nonlinear manner while considering both lower and upper bounds. By incorporating a desired cost function, they derived a nonlinear programming formulation. Zhe Xu et al.[3]presented a Tuaguchi-ANFIS (Tuaguchi-adaptive neural-fuzzy inference system) methodology, utilizing an orthogonal experiment matrix to minimize training data requirements. They constructed a rapid response mathematical model linking three configuration parameters—fin pitch, fin thickness, and the number of oil channels—with two design indices: heat transfer capacity and working weight.

Furthermore, Tiancai Ma and his co-authors[4]established a comprehensive system model and optimization objective for a novel hybrid renewable energy system (HRES). They combined the hypervolume method with an algorithm framework to propose an enhanced non-dominated sorting genetic algorithm (NSGA-III), addressing the issue of Pareto front degradation due to random selection. Meanwhile, Yang Qi and his team[5]devised a path planning algorithm grounded in bidirectional Q-learning. By refining the bidirectional Q-learning mechanism and bolstering the Q-learning initialization process, the algorithm achieved efficient and effective optimization of

  DOI: 10.25236/icmmct.2025.019

abandoned household appliance recycling routes. This approach facilitated bidirectional updates to the state-action value function, originating from both the starting and goal points.

The determination of optimal crop planting schemes is crucial for maximizing agricultural productivity and economic returns. Traditional methods for addressing this problem include linear programming and stochastic optimization. However, these approaches often struggle with the complexity and dynamic nature of real-world agricultural scenarios, where multiple objectives and constraints need to be considered simultaneously.

Linear programming, for instance, excels in finding optimal solutions under well-defined linear relationships but may fail to capture the nonlinearities and uncertainties inherent in agricultural systems. Stochastic optimization, while better suited for handling uncertainties, can become computationally intensive and may not always guarantee globally optimal solutions.

In response to these limitations, this paper proposes a novel modeling approach that combines a hybrid optimization algorithm with multi-objective optimization strategies. By leveraging the exploratory power of genetic algorithms and the adaptive learning capabilities of Q-learning, the model aims to identify optimal planting schemes that balance various objectives, such as maximizing profit and minimizing risk. In order to verify the feasibility of the algorithm designed in this paper, the solution research is carried out based on the data set provided in a competition.

2. Design of solution method

2.1. Model Comparison

Based on the complexity of crop planting scheme optimization, this paper chooses a hybrid optimization algorithm (genetic algorithm + reinforcement learning) and a multi-objective optimization strategy modeling method. Initially, this paper uses the following three modeling methods for preliminary solution. First, the linear programming model works well when dealing with a single objective, but its performance is limited when facing multiple objectives, and it is difficult to optimize multiple objectives at the same time. Secondly, the stochastic optimization model optimizes by simulating uncertainty, but the model does not grasp the global solution well, is easy to fall into local optimum, and the convergence speed is slow. Finally, due to the lack of adaptive adjustment mechanism, the traditional genetic algorithm has insufficient performance in complex multi-objective optimization problems. It is difficult to dynamically adjust the crossover and mutation operations, which affects the convergence speed and accuracy. The results show that these methods are difficult to take into account the needs of multiple parties when facing multi-objective optimization, and the effect is not ideal.

Therefore, this paper chooses a hybrid optimization algorithm (genetic algorithm + reinforcement learning) and a multi-objective optimization strategy to solve the problem. The innovation is that the crossover and mutation operations are adaptively adjusted by reinforcement learning to improve the global search ability and convergence efficiency of the algorithm, and the multi-objective optimization strategy is used to balance revenue, stability and resource utilization. This makes the model more robust and efficient in complex planting scenarios.

2.2. Solution based on genetic algorithm

Genetic algorithm is an optimization algorithm based on the theory of natural evolution, which simulates the mechanisms of selection, crossover and mutation in the process of biological evolution. The use of genetic algorithm can accurately adjust various production measures, maximize the optimization of input, obtain the highest yield and economic benefits, protect the ecological environment, and realize the sustainable development of agriculture [2]. The main process of the algorithm includes initialization of population, calculation of fitness function, selection, crossover and recombination, mutation and generational update. Firstly, the algorithm randomly generates initial solutions as a population, evaluates the fitness of each solution, and retains good quality solutions into the next generation by selection operation. Then, new solutions are generated by crossover and mutation to ensure population diversity. The process iterates until the algorithm converges or a termination criterion is met. In this paper, reinforcement learning is used to dynamically adjust the crossover and mutation probabilities to further improve the convergence speed and the quality of the solution of the algorithm.

The detailed flowchart of the genetic algorithm is shown in Fig.1

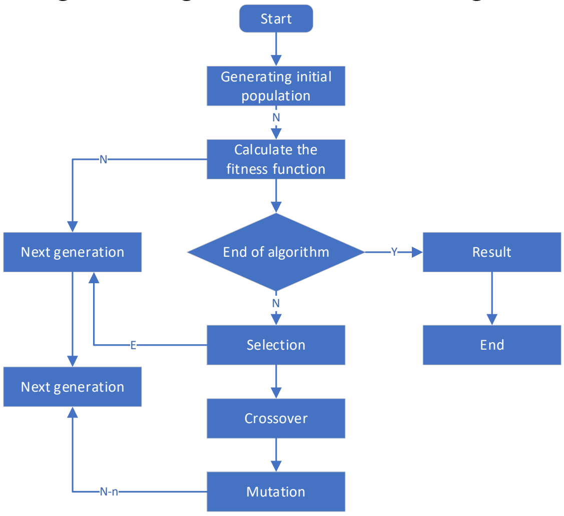

Figure 1 Flowchart of genetic algorithm

As can be seen from Figure 1, the genetic algorithm mainly includes the following five parts:

(1) Generation of initial population: When initializing the population, this paper combines the information of plot area, plot type and crop suitability to randomly generate the planting area allocation scheme of each crop on different plots, while meeting the maximum planting area limit of the plot and the rotation demand of each crop. Each individual represents a cropping scheme for a crop on each plot. In this paper, in order to visually show the distribution of the initialization population, the stacked bar chart is drawn as shown in Fig.2 and Fig.3:

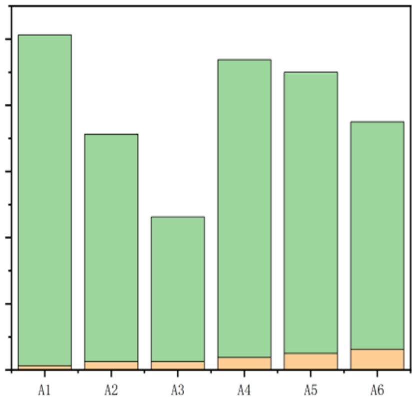

Figure 2 The population is initialized in the flat and dry land

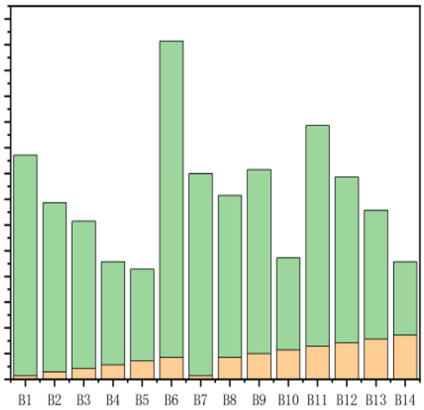

Figure 3 The terrace initializes the population

Taking only flat and dry land and terraces as examples, it can be seen from Fig.2 and Fig.3 that the crop planting area and the type of crop planted in different plots are significantly different. In the flat and dry land (Fig.2), the proportion of crop planting in each block is relatively concentrated. The planting area of A1 and A5 is large, while that of A2 and A3 is small. Crop types are mainly food crops. In the terrace (Fig.3), the planting area is relatively uniform in general, but the distribution of planting crop types in different blocks is different. B6 plot has the largest planting area and has many types.

(2) Fitness function calculation: According to the crop yield, planting cost, selling price and other data in the attachment, calculate the total revenue, planting volatility and resource utilization efficiency of each individual plan, corresponding to the three objectives in the fitness function. The fitness function comprehensively considers these three objectives and is evaluated using the method of weighted summation. In this paper, we use Python to calculate the fitness calculation table and show the specific values of revenue, volatility and efficiency of each scheme through the fitness calculation table. The specific data are shown in Table 1:

Table 1 Fitness calculation table

Crop Name	Revenue	Resource efficiency	Volatility	Crops	Yield	Resource efficiency	Volatility
Soybean	900	1	0.1	sorghum	3380	1.58	0.18
Black Bean	3350	1.25	0.12	Millet	3577.5	1.46	0.17
Red Beans	2950	1.14	0.15	Buckwheat	4050	0.31	0.3
Green Beans	2100	1	0.08	Pumpkin	3500	3	0.07
Crawler Beans	2451.25	1.19	0.13	Sweet potato	5150	1.1	0.14
Wheat	2350	1.78	0.09	Avena sativa	1910	1.05	0.1
Corn	2500	2	0.2	Barley	1487.5	1.5	0.08
Grain	2340	1.11	0.1	Rice	2820	0.74	0.05

(3) Selection: according to the size of the fitness value, the individuals with higher fitness are selected to enter the next generation to ensure that the excellent individuals can continue to reproduce and improve the quality of the overall population solution.

(4) Cross-recombination: by simulating the biological genetic process, the partial genes of two parent individuals are combined to generate a new offspring individual, so as to enhance the diversity of the population and improve the global search ability. In this paper, the simulated binary crossover method is used for the crossover operation. Firstly, two individuals with higher fitness are randomly selected, and the crossover operation is determined according to the predetermined crossover probability. New offspring individuals are then generated based on the weighted combination of the two parent solutions.

(5) Mutation: introduce new solutions by randomly changing some loci in individual genes to prevent the algorithm from falling into local optimum. According to the set mutation probability, a gene in an individual is randomly selected, and it is randomly adjusted in a small range to generate a more diverse population.

2.3. Genetic Algorithm Incorporating Q-learning Reinforcement Learning Approach

In this paper, Q-learning reinforcement learning method is used to dynamically adjust the crossover and mutation operations in the genetic algorithm to further improve the convergence speed and optimization ability of the algorithm. Q-learning continuously adjusts the probability of the operations by learning the impact of the crossover and mutation operations on the algorithm's performance, which enables the algorithm to find a better solution in each generation of optimization. The specific steps are as follows:

(1) Environment definition.

The state is defined as the distribution of fitness value of the current population, population diversity or some kind of statistical measure; the action is defined as the parameter by adjusting the crossover and mutation operation (such as the change of crossover probability and mutation probability; the reward is defined according to the enhancement of the population fitness after the operation: if the average fitness of the population is enhanced, a positive reward will be given, and the opposite will be given a negative reward.

(2) Q-value updating formula.

The optimal action is selected by updating the table, and the updating formula is as follows:

$$Q(s,a) = Q(s,a) + \alpha \times \left(r + \gamma \times \max\left(Q(s',a') - Q(s,a)\right)\right) \tag{1}$$

(3) Selection of actions.

-Greedy strategy : In each step of the greedy strategy, the algorithm randomly selects actions with a certain probability and chooses the action with the highest current Q-value with a residual probability to balance the exploration of new strategies and the utilization of existing knowledge so as to improve the learning efficiency and to avoid local optimums.

(4) Initialization of Q-value table.

The initialization of the Q-value table depends on the dimensions of the state space and action space, and usually all the initial values are set to 0, so that the algorithm can gradually reflect the expected benefits of the state-action pairs through updating during the learning process.

In Q-learning reinforcement learning method, the algorithm gradually learns the optimal action strategy by iteratively updating the Q-value table.The change of Q-value reflects the benefit of taking different actions in different states. In this paper, the Q-value change graph is plotted using Python as shown in Fig. 4:

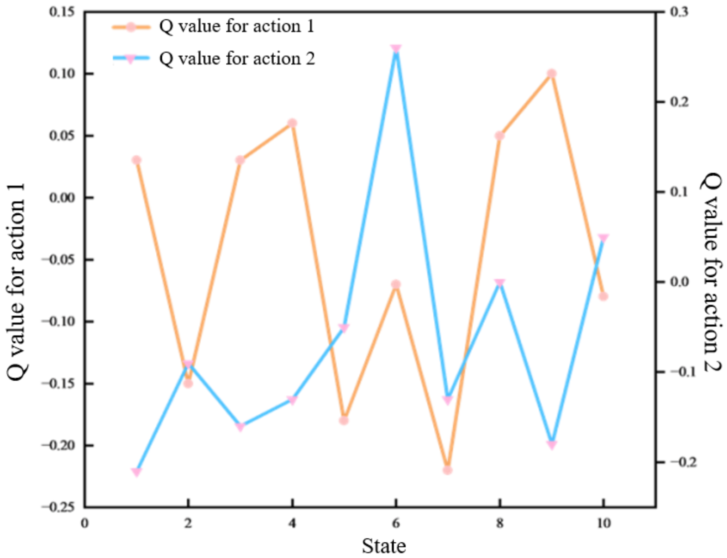

Figure 4 Plot of Q-value change

Fig. 4 shows that the Q value of action 1 gradually improves with the increase of states, showing a steady growth trend, indicating that action 1 is a more optimal choice in most states. However, the Q value of action 2 fluctuates greatly, and although it shows higher payoffs in some states, such as state 6, it is not as stable as action 1 as a whole. Action 2 has negative Q values in multiple states, indicating that the action is not conducive to the optimization objective in these states. From the overall analysis, reinforcement learning gradually guides the algorithm to preferentially choose action 1 for higher long-term payoff.

3. Linear Programming Models for Two Scenarios

3.1. More than part of the crop is stagnant and wasted

In this case, the objective is to maximize the total profit for all crops in each year from 2024 to 2030. For each crop, the sales revenue $R_{i,t}$ and the cost of cultivation $C_{i,t}$ in the $t-th$ year are defined as follows:

$$\max \sum_{t=2024}^{2030} \sum_{i=1}^{N} \left(R_{i,t} - C_{i,t}\right) \tag{2}$$

$R_{i,t}$: Revenue from the sale of the first crop in year i, calculated as:

$$R_{i,t} = P_i \times min\left(D_i, \sum_{j=1}^{2} \sum_{k=1}^{M} x_{i,j,k,t} \times Y_i\right) \tag{3}$$

$C_{i,t}$: The cost of growing crop i in year i, calculated as:

$$C_{i,t} = C_i \sum_{j=1}^{2} \sum_{k=1}^{M} x_{i,j,k,t} \tag{4}$$

(a) Objective function with penalty term:

To penalize the portion of output that exceeds the market demand, an additional penalty term P_{penalty} can be introduced, which is proportional to the portion of output that exceeds the demand. Suppose the penalty coefficient is α and the penalty term is as follows.

$$P_{\text{penalty}} = \alpha\left(Q_{i,t} - D_i\right)_+ \tag{5}$$

Where α_+ denotes that a is taken when a>0 and 0 otherwise, i.e., a penalty is incurred only when production exceeds demand.

(b) Complete objective function.

$$max \sum_{t=2024}^{2030} \sum_{i=1}^{N} \left(R_{i,t} - C_{i,t} - \alpha\left(Q_{i,t} - D_i\right)_+\right) \tag{6}$$

3.2. Excess is sold at a reduced price of 50% of the 2023 sales price

In this case, the objective function is also to maximize the total profit, but for the portion of the crop in excess of the demand to be sold at 50% of the price, the sales revenue is calculated by the formula:

$$R_{i,t} = P_i \times min\left(D_i, Q_{i,t}\right) + 0.5P_i\left(Q_{i,t} - D_i\right)_+ \tag{7}$$

Where $Q_{i,t} = \sum_{j=1}^{2} \sum_{k=1}^{M} x_{i,j,k,t} \times Y_i$: the total production of the first crop in the first year and B denotes $(a)_+ = max(0, a)$ when a > 0 and 0 otherwise.

Constraints.

(1) Variable non-negativity constraint:

In order to ensure that the acreage is reasonable, the acreage of all crops must be non-negative:

$$x_{i,j,k,t} \geqslant 0 \ \forall_{i,j,k,t} \tag{8}$$

(2) Plot planting type restrictions:

Crops may be planted only on plot types that are suitable for their growth, ensuring that plot types and crops are adapted:

$$x_{i,j,k,t} = 0 \text{ if i} \notin L_k \tag{9}$$

(3) Crop rotation constraints:

In order to ensure that the land is rotated into fallow, the same crop cannot be planted consecutively in the same season on the same plot (starting in 2025):

$$x_{i,j,k,t} + x_{i,j,k,t-1} = 0 \forall i, j, k, t \geqslant 2025 \tag{10}$$

(4) Land use constraints:

The total acreage of crops per parcel shall not exceed the total acreage of the parcel to prevent exceeding the carrying capacity of the parcel:

$$\sum_{i=1}^{N} x_{i,j,k,t} \leqslant S_k \forall j, k, t \tag{11}$$

(5) Constraints on the number of plots for crop cultivation:

In order to control over-concentration of crops, the number of plots planted per crop per season cannot exceed the upper limit:

$$\sum_{k=1}^{N} y_{i,j,k,t} \leqslant N_{max} \forall i,j,t \tag{12}$$

(6) Frequency constraints for legume crops:

In order to ensure legume crop rotation, each plot should be planted with legumes at least once in every three years:

$$\sum_{i \in B} \sum_{j=1}^{2} \sum_{k=1}^{M} x_{i,j,k,t} \geqslant e \forall t \tag{13}$$

(7) Crop minimum acreage constraints:

Each crop shall be planted no smaller than the minimum area specified to ensure reasonable planting:

$$x_{i,j,k,t} \geqslant A_{min} \forall i,j,k,t \tag{14}$$

(8) Seasonal planting restrictions:

Restrictions on seasonal planting of certain crops that can only be planted in one season may not be planted again in the second season:

$$x_{i,j,k,t} = 0 \text{ If the plot can only grow one season's worth of crops} \tag{15}$$

(9) Crop diversity constraints:

Each plot can only grow a maximum number of crops in the same cycle to avoid too many monocultures:

$$\sum_{i=1}^{N} x_{i,j,k,t} \leqslant m_{max} \forall j,k,t \tag{16}$$

(10) Minimum planting size constraints:

Each crop shall not be planted on a parcel less than 20 percent of the parcel area to ensure a minimum size for planting:

$$x_{i,j,k,t} \geq 0.2 S_k \forall \mathrm{i}, j, k, t \tag{17}$$

3.3. Multi-objective optimization strategy

When using multi-objective optimization strategy, in addition to maximizing the revenue, it is also necessary to optimize the planting stability and resource utilization efficiency. Therefore, this paper introduces the Pareto frontier multi-objective optimization strategy, and the objective function is designed as follows.

(1) Yield maximization objective.

Maximize the crop yield on each plot, combining the selling price, yield and planting cost:

$$max \sum_{i,j} \left(P_{i,j} \times Y_{i,j} - C_{i,j} \times A_{i,j}\right) \tag{18}$$

(2) Stability Minimization Objective.

Minimize the volatility of crop acreage and ensure that crop acreage remains stable from year to year:

$$min \sum_{i,j} \left|A_{i,j}^{(t)} - A_{i,j}^{(t-1)}\right| \tag{19}$$

(3) Objective of maximizing resource use efficiency.

Maximize resource use efficiency, i.e., the ratio of volume of output to area planted:

$$max \sum_{i,j} \frac{Y_{i,j}}{A_{i,j}} \tag{20}$$

In order to solve the multi-objective optimization problem, this paper adopts the non-dominated sorting genetic algorithm, which gradually approximates the Pareto optimal solution set by initializing the population, calculating the values of the three objective functions of revenue, stability, and resource utilization efficiency, performing the non-dominated sorting, calculating the degree of crowding, selecting the good individuals, and generating a new population through cross mutation. Finally, through many iterations, the non-dominated sorting genetic algorithm can find the optimal balance between the three objectives, and output a set of Pareto frontier solutions for decision makers to choose.

4. Solution result

In order to visually observe the changes in the returns of different processing strategies during the optimization process, and to select the appropriate strategy to maximize the returns and control the fluctuations this paper plots the returns under Case 1 (stagnant and wasteful) and Case 2 (partially sold at reduced prices) as shown in Fig.5 and Fig.6.

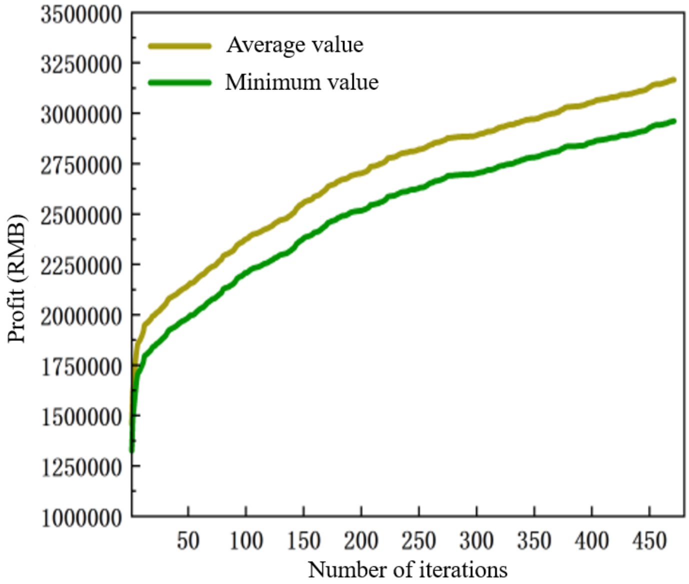

Figure 5 Profit convergence curve for case 1

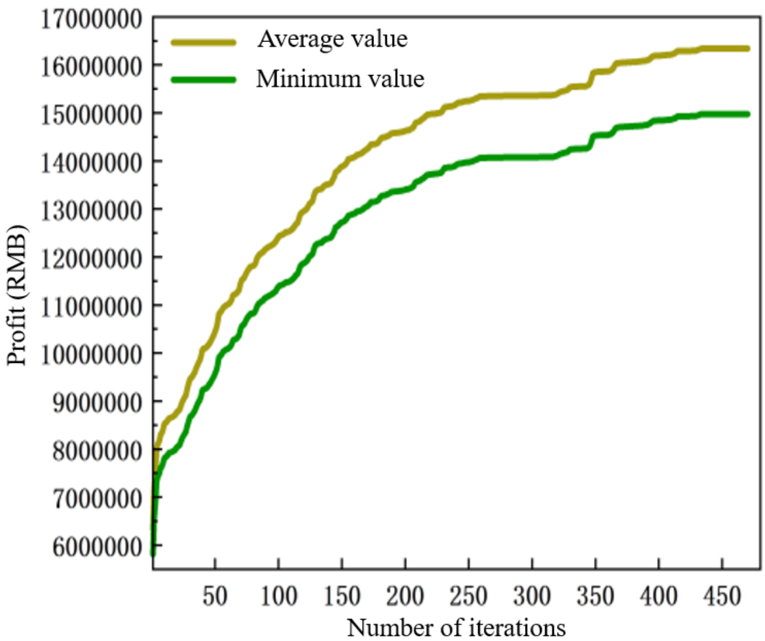

Figure 6 Profit convergence curve for case 2

As can be seen in Fig.5 and Fig.6, in Case 1, with the increase in the number of generation selections, the gains show a trend of steady growth, but there is a large gap between the minimum gain and the average gain, indicating that there are large fluctuations between different scenarios and that individual scenarios outperform others. In Case 2, the price reduction treatment significantly reduces the fluctuation of returns, and the gap between the minimum return and the average return is significantly reduced, indicating a more consistent performance among the schemes. Overall, the return curve in Case 2 flattens out more quickly, suggesting that the treatment is more conducive to maintaining a higher level of return stability over the long term while maximizing total returns.

5. Conclusion

This paper presents a hybrid optimization model that integrates genetic algorithms and Q-learning to address the problem of determining optimal crop planting schemes. By optimizing the allocation of crop planting areas and dynamically adjusting genetic operations through reinforcement learning, the model demonstrates enhanced global search capabilities.

The introduction of unsalable crops and price reduction scenarios, along with penalty functions to control risk, allows the model to account for multiple objectives and constraints, leading to the formulation of a multi-objective programming model. The use of the non-dominated sorting genetic algorithm enables the identification of Pareto optimal solutions, providing a comprehensive set of alternative planting schemes tailored to different scenarios.

The results obtained show that the proposed model consistently achieves high profitability and stability across various conditions. This demonstrates its effectiveness in addressing the complexities and uncertainties associated with real-world agricultural systems. Overall, the research contributes to the advancement of agricultural optimization methods, offering a practical tool for farmers and agricultural planners to make informed decisions regarding crop planting strategies.

References

[1] Zhou, L., Li, Z.-Q., Yang, H. & Tan, C. (2025) Integral predictive sliding mode control for high-speed trains: A dynamic linearization and input constraint-based data-driven scheme. Control Engineering Practice 154, 106139.

[2] Iwamoto, S., Tamura, Y. & Tamura, S. (1985) Fast var and Voltage Control Scheme Combining Fast Second-Order Load-Flow and Recursive Linear Programming. IFAC Proceedings Volumes 18, 399–404.

[3] Xu, Z. et al. (2024) Hybrid of non-dominated sorting genetic algorithm II and invasive weed optimization for fan duct surface heat exchanger configuration optimization design. Case Studies in Thermal Engineering 61, 105149.

[4] Ma, T., Chen, J., Ma, X. & Yang, Y. (2024) Multi-objective optimal configuration of off-grid residential hybrid renewable energy system based on hypervolume-improved Non-dominated Sorting Genetic Algorithm III. International Journal of Hydrogen Energy 87, 277–289.

[5] Qi, Y., Cao, J. & Wu, B. (2024) Bidirectional Q-learning for recycling path planning of used appliances under strong and weak constraints. Communications in Transportation Research 4, 100153.