

Research on crop planting strategy based on whale optimization algorithm

Leru Sun, Jianyuan Xu *

College of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin, China, 150028

* Corresponding Author Email: xushaoqing528@gmail.com

Abstract. In this paper, the problem of farming strategy of arable land is studied in depth, and the planning model of the optimal planting scheme for decision-making crops. First, multi-objective hybrid planning is constructed by adding constraints. Secondly, the optimal planting strategy that satisfies the multi-objective conditions is found by imitating the law of whale predation and optimized using genetic algorithm; The results showed that the model could significantly improve the planting yield compared with the previous planting structure. Finally, based on this model, the crop planting model of whale optimization algorithm is constructed by considering the uncertainty factor. The results showed that the overall results were more stable after continuous iteration of the offspring, considering the various risks that will be faced in the process of agricultural planting. The research in this paper has significantly increased crop yields and provided a relatively effective planting method, which offers valuable insights for real-world farming strategies.

Keywords: Genetic Algorithm, Whale Optimization, Algorithm, Pearson's coefficient.

1. Introduction

In order to feed the billions of people living on the planet, it will be necessary to increase the production of high-quality food with fewer inputs, but this achievement will be particularly challenging in the face of global environmental change. Plant breeders will need to focus on the traits most likely to increase yields [1]. Similarly, the development of climate change-resistant crops, irrigation water management, adoption of climate-smart agricultural practices, and the promotion of indigenous knowledge can ensure food security by increasing agricultural yields [2]. The findings of Mirzaei A et al suggest that appropriate adaptation strategies and measures can mitigate the impacts of climate change and provide for the restoration of the basin's Jazmulian wetlands with more Water sources [3]. The field of agriculture is one of the most important areas where deep learning applications still need to be explored as it has a direct impact on human well-being. In particular, there is a need to explore how deep learning models can be used as tools for optimal planting, land use, yield improvement, production/disease/pest control and other activities. The large amount of data received from sensors in smart farms makes it possible to use deep learning as a decision-making model in this field. In agriculture, no two environments are identical, which makes testing, validating and successfully implementing such technologies much more complex than in most other industries [4]. Constraints such as plots of each crop not being too spread out and not being too small require the creation of intelligent optimization algorithms to solve for the best planting decisions.

Krishna V used linear regression and random forest algorithms to debug simple parameters such as area, temperature and rainfall, enhancing the algorithms to provide better predictions, showing relatively high accuracy [5]. Maraveas C used an integrated learning approach to integrate various algorithms for agricultural applications: Bioinspired Intelligent Algorithm Ecological Algorithm, Population Based Intelligent Algorithm, Ecological Based Algorithm and Multi-Objective Algorithm [6]. Emphasis is placed on variants of population based intelligent algorithms namely Artificial Bee Colony (ABC), Genetic Algorithm, Flower Pollination Algorithm (FPA), Particle Swarm, Ant Colony, Firefly Algorithm, Artificial Fish Swarm and Krill Swarm algorithms, which have been widely used in the agricultural sector. Accuracy reduces the risk of errors in the application of pesticides, fertilizers, irrigation and crop monitoring, thus increasing yields. Juraev F D utilized the

mechanisms that can improve the model as well as algorithms that activate the mechanism as a way of dealing with the mechanisms that allow the improvement of the model of systematic research is artificial intelligence, the main problem here is the obstacles in the development of algorithms that determine the trajectory of its activity [7]. Wu Y proposed using principal component analysis and neural network methods for selection and detection, which not only improved crop yield but also simplified the complexity of evaluation metrics [8]. Rana proposed that improper selection of planting sites will lead to higher carbon content, and that machine learning methods are needed to recommend planting [9]. Rajasekaran S also proposed that machine learning algorithms are essential to identify biological data in planting [10].

However, due to the previous staff only consider the impact of natural factors such as climate, or only consider the impact of agricultural fertilizers on crop yields, did not take into account the planting structure and crop planting strategy and the distribution of planting structure of the consequences of not considering the full picture. So we propose the optimal planting strategy using the whale optimization algorithm law to satisfy the multi-objective conditions, and in this paper, the use of genetic algorithms for better prediction, the accuracy rate has been improved very much, and it has a far-reaching impact on the future planting strategy of crops.

2. The Establishment of the model

2.1. The basic fundamental of genetic algorithm model

Genetic Algorithm is a planning model for decision making optimal solution with the objective of maximizing profit. By substituting the concept of price index in the genetic algorithm model for cross mutation, the optimal solution to maximize profit is found. This model can be utilized to find the maximum profit of a crop, and calculating the profit involves several key factors: yield, cost, and selling unit price. Of these, yield and cost are known and determined. Yield is the amount of crop that can be harvested from each acre of land, while cost is the total cost per unit area planted. The determination of these two parameters helps us to make basic profit calculations. However, the determination of unit sales price is more complex as it is affected by market conditions, climate change and other external factors.

Linear planning is carried out to maximize the profit of the crop, and a series of conditions such as the amount of human constraints and the influence of external factors are used as constraints as required to build a hybrid planning model of 0-1 planning and linear planning as a whole.

Planning for genetic algorithm models.

Decision variables:

Since the optimal solution needs to be determined and there are only two outcomes for planting and not planting a particular crop in a particular season on a particular plot, the decision variable established is:

$$X_{i,j,k} = 0 \text{ or } 1 \quad (1)$$

where k is the quarter, j-th is the sample plot and i-th is the crop number.

Constraints:

$$\max \sum_{t=2024}^{2030} \sum_{i=1}^{41} x_{i,j,t} \cdot k_n \cdot p_j - y_{i,j} \cdot c_{i,j} \quad (2)$$

$$\begin{cases} x_{i,n,t} \cdot x_{i,n,t} < 1 \\ x_{i,j,t} + x_{i,j,t+1} + x_{i,j,t} \geq 1 \\ k_{n,i} \leq k_n \end{cases} \quad (3)$$

$$1 \leq x_j \leq 15 \quad 0 < j \leq 3 \quad (4)$$

Objective function

Let the benefit be T. From the relationship between benefit and cost, the objective function is.

$$T = \sum_{t=2024}^{2030} \sum_{i=1}^{41} (X_{i,j,k} \cdot L_{j,n} \cdot S_i - X_{i,j,n} \cdot C_{i,j}) \quad (5)$$

where $X_{i,j}$ is crop yield per acre/pound, $C_{i,j}$ is the cost of growing the crop (\$/acre), P_i is the selling price of the crop (yuan/catty), k_n is plot size, $Y_{i,j}$ is the acreage of crop i on plot j (acres) and Z_t^j is expected sales volume (pounds).

Genetic algorithms find optimal solutions in optimization problems by simulating natural selection mechanisms. Our goal is to optimize the cropping solution for a particular piece of land by using genetic algorithm to maximize the profit. To achieve this goal, we need to compute the fitness function which centers on the profit of the land.

In this context, calculating profit involves several key factors: acreage, cost of planting, and selling unit price. Of these, the acre yield and the cost of planting are known and determined. The acre yield is the number of crops that can be harvested from each acre of land, while the cost of cultivation is the total cost required to carry out crop cultivation on each acre of land. The determination of these two parameters helps us in making the basic profit calculation. However, the determination of unit sales price is more complex as it is affected by market conditions, climate change and other external factors.

The unit sales price is the price at which each unit of crop is sold in the market. Since market fluctuations and climate change can lead to uncertainty in the unit sales price, we need to determine its exact value based on actual conditions. This usually means that we need to conduct some market research and forecasting in order to model each possible unit price scenario. The range or probability distribution of unit sales prices can be estimated through historical data analysis, market trend forecasts or expert opinion.

In the genetic algorithm, we will use these estimated unit sales prices to calculate the fitness values for each planting scheme. The fitness function will combine acre yield, planting cost, and unit price of sale to evaluate the performance of each individual through the profit equation. Specifically, profit = (acre yield $\times$ unit price sold) - cost of planting. By doing this calculation, we can obtain the profit for each planting scenario and thus determine its fitness value.

Ultimately, the genetic algorithm will guide the selection, crossover and mutation operations based on these fitness values to gradually optimize the planting scheme to find the best scheme to maximize the profit. In the whole process, the uncertainty of selling unit price is a factor we need to pay special attention to, which directly affects the accuracy of profit calculation and the reliability of the final optimization result. This thesis adopts the price index, i.e., for the p -order VAR model, the following conditions are established:

$$y_t = A_t y_{t-1} + \dots + A_p y_{t-p} + B x_t + \varepsilon_t \quad (6)$$

where y_t is an m -dimensional nonsmooth $I(1)$ sequence, x_t is a d -dimensional deterministic variable, and ε_t is a new interest vector. After deformation, it can be rewritten as:

$$\Delta y_t = \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + B x_t + \varepsilon_t \quad (7)$$

$$\Pi = \sum_{i=1}^p A_i - I_m; \Gamma_i = \sum_{j=i+1}^p A_j \quad (8)$$

Price indices in planning

Since the variables are non-stationary sequences, and each sequence in the endogenous variable vector after the first-order difference is stationary, the result is guaranteed to be flat only if each of the variables constituting $\Pi y_t - 1$ is $I(0)$.

As well as the combination of human factors to solve: for example, after the recovery from the epidemic, the economy has improved, the standard of living of the people to improve and other factors, can be predicted that the unit price of sales: (from crop 1 to crop 41).

In genetic algorithm, we take Darwin's theory of natural selection as the basis, through simulating the selection, crossover and mutation process of organisms, combined with the principle of survival of the fittest, aiming to find the optimal solution of the problem. In specific applications, such as

optimizing crop planting schemes, we first need to define key parameters such as the objective function, constraints, and the number of iterations. These parameters will be used to guide the algorithm's operations and the implementation of constraints. First, define a series of formulas for performing operations and constraints such as the objective function, constraints, and the number of iterations.

Based on the known conditions and constraints, through a randomized algorithm, the initial population is obtained. The selection fitness of the randomly created initial population is obtained through the objective function.

First, based on the known crop cultivation and various types of constraints, we start the algorithm by randomly generating an initial population. Each crop constitutes a set of data representing one possible planting scenario. By means of an objective function, we evaluate the fitness of these initial crops, i.e., how close their effects are to the maximum gain. The level of fitness determines the probability of an individual's retention in subsequent operations. Using the roulette selection algorithm, by comparing the crops corresponding to the maximum gain, individuals with higher fitness have a higher chance of being selected into the next generation. In this way, the next generation of individuals will mainly come from individuals with higher fitness, with a view to gradually approaching the optimal solution.

When entering the crossover phase, we will model the crossover process in biology. In this phase, we define a crossover probability and use this probability to decide whether or not to crossover crops from two lands. If the randomly generated probability is greater than the predefined crossover probability, the crossover operation is performed. The crossover operation randomly reorganizes the selected crops and exchanges their crop planting schemes. The new crossover results are recalculated for their fitness through the objective function and the data are recorded for final comparison.

In addition, the mutation operation is an important step in the genetic algorithm. Even if the probability of mutation is relatively small, we still need to consider its effect. We consider the crop planting scheme of the land as an individual organism and define a mutation probability. By means of a stochastic algorithm, we decide whether or not to mutate. If mutation occurs, new crops are randomly selected for planting as far as the constraints allow, and the value of adaptation after mutation is calculated. The post-variation fitness will be recorded for comparison and evaluation.

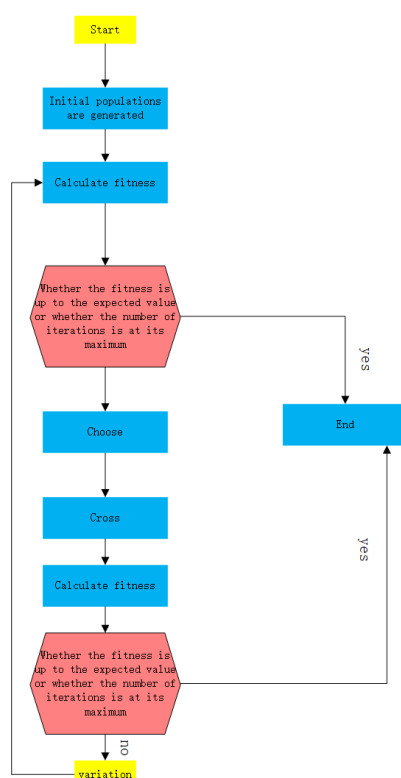


Figure 1. Genetic Algorithm Flowchart

The whole genetic algorithm process is a cyclic iterative process. Each time after selection, crossover and mutation operations, the number of iterations increases. The algorithm continues until a preset number of iterations is reached. During all iterations, we record the fitness after each crossover and mutation and eventually compare these fitness values to determine the optimal fitness and the corresponding crop planting scheme. This systematic approach not only helps to find near-optimal planting schemes, but also deals with complex constraints and optimization problems.

Finally, we compare the adaptations obtained from each crossover and mutation to obtain the best adaptation and the corresponding crop. The Figure 1 shows the flowchart of this modeling.

2.2. The basic fundamental of Whale Optimization Algorithm

The classical whale optimization algorithm originates from the feeding behavior of humpback whales in the deep sea, and the algorithm is known for its simple structure, small number of parameters, powerful search capability, and easy implementation. The whale's predation strategy is known as the bubble net strategy, which can be mainly summarized as follows:

(1) Surrounding prey. After sharing the acquired prey information, approach the whale in the group that is closest to the prey, followed by a gradual narrowing of the encirclement circle. At this point the whales' positions were updated with the formula

$$X(t+1) = X^*(t) - A \cdot |CX^*(t) - X(t)| \quad (9)$$

where D is the distance between the whale and the prey; t is the current number of iterations; $X^*(t)$ is the position vector of the current optimal solution; $X(t)$ is the current position vector of the whale, A and C are the correlation coefficients, which are calculated according to the following formula, i.e.

$$\begin{cases} A = 2a \cdot r - a \\ C = 2 \cdot r \\ a = 2(t_{max} - t)/t_{max} \end{cases} \quad (10)$$

(2) Spiral close to the prey.

$$\begin{cases} D^* = |CX^*(t) - X(t)| \\ X(t+1) = X^*(t) + D^* e^{bl} \cos(2\pi l) \end{cases} \quad (11)$$

(3) Randomly select a prey.

$$X(t+1) = X_{rand}(t) - A \cdot |CX_{rand}(t) - X(t)| \quad (12)$$

The improved whale optimization algorithm achieves the update of the position of the whale optimization algorithm by adding a dynamic weight that changes with the number of iterations, which improves the overall search capability, and the position formula is

$$X(t+1) = \begin{cases} w(t)X^*(t) - A \cdot |CX^*(t) - X(t)|, p < 0.5 \\ w(t)X^*(t) + D^* e^{bl} \cos(2\pi l), p \geq 0.5 \end{cases} \quad (13)$$

$$X(t+1) = w(t)X_{rand}(t) - A \cdot |CX_{rand}(t) - X(t)| \quad (14)$$

In analyzing the relationship between expected sales volume, sales price and expected cost, the Pearson correlation coefficient is an important statistical tool used to measure the degree of linear correlation between two variables. Although there may be complex interactions between these three variables, we can use the Pearson's coefficient to reveal the linear correlation between them. The Pearson's correlation coefficient can help us understand how these variables interact with each other, for example, whether a change in the selling price will have a significant effect on the expected sales volume or whether there is a relationship between the expected cost and the selling price.

Specifically, by calculating the Pearson correlation coefficient between each pair of variables, we are able to obtain the degree of correlation between them. For example, the Pearson coefficient between sales price and expected sales volume can reveal the extent to which changes in price affect sales volume, while the Pearson coefficient between sales price and expected costs helps to understand how price changes are associated with cost fluctuations. Since the Pearson correlation coefficient is specifically designed to deal with situations where there are fewer variables, it can provide a clear and quantitative measure of linear correlation, thus helping us to better analyze and predict the interaction of these variables in real business.

First, the formula and significance of Pearson's coefficient are introduced:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (15)$$

where x_i, y_i are observations of two variables.

From the formula: r is in the range of $[-1, 1]$.

$r = 1$: Perfectly positive correlation, two variables are perfectly linearly correlated and in the same direction.

$r = -1$: Perfectly negative correlation, where two variables are perfectly linearly correlated, but in opposite directions.

$r = 0$: Uncorrelated, two variables have no linear relationship.

The following is the equation for the relationship between sales volume and sales price obtained through correlation.

$$Z_j^t = X_j^{2023} \times (1 + \rho PS \times \Delta P_j^t) \quad (16)$$

where ρ is the correlation coefficient between sales volume and price.

The optimization of the objective function can be performed by combining the whale multi-objective algorithm as well as the relational equation obtained by combining substitution complementarity and correlation.

Specifically, we can use substitutability between crops to adjust the objective function so as to optimize the planting strategy of different crops with limited resources. Substitutability relationships refer to the fact that certain crops are substitutable for each other in terms of resource allocation, and this adjustment helps to maximize the efficiency of resource use in the overall objective function.

Finally, combining the results of correlation analysis, we can further optimize the objective function. By adjusting the correlation equation, we are able to find the best allocation of resources among different crops, further improving the optimal solution of the objective function. By combining these methods, we can more accurately find the optimal planting strategy, thus achieving more efficient resource utilization and yield improvement, and this equation is the adjusted objective function:

$$\max \sum_{t=2024}^{2030} \sum_{i=1}^{34} \sum_{j=1}^M \left(\min(y_{ij}^t \times X_j^t, Y) \times P_j^t - y_{ij}^t \times C_j^t \right) \quad (17)$$

where Eq. and in conjunction with the above, it can be seen that y can be adjusted according to the substitutability between crops, y_{ij}^t can be adjusted according to the substitutability between crops, X_j^t can be updated according to the complementarity between crops, and Z_j^t can be optimized according to the correlation equation, which facilitates the optimal solution.

3. Results

3.1. The establishment of simulation model

Considering that there is a certain degree of substitutability and complementarity between crops as well as a correlation between the expected sales volume and the sales price and the cost of cultivation, there is also a certain degree of correlation between the expected sales volume and the sales price and the cost of cultivation.

Substitution and complementarity between crops: there may be substitution (i.e., one crop can partially substitute for another) and complementarity (i.e., growing one crop promotes the growth of other crops or improves yields) between different crops.

Correlation of sales volume with price and cost: There may be a correlation between the expected sales volume of a crop and the sales price and cost of growing it.

Based on these factors, the optimal planting strategy for 2024-2030 is developed and solved using simulated data, and finally the results are analyzed in comparison with the conclusions in the second point.

Substitutability and Complementarity between Crops: In economics, we usually use the coefficient of cross-elasticity to analyze the relationship between marketed commodities. The cross-elasticity coefficient measures the effect of a change in the price of one commodity on the quantity demanded of another commodity. This analysis is also applicable in the case of crops, where we can use the cross-elasticity coefficient to explore substitutability and complementarity between different crops. Specifically, substitutability and complementarity between crops are very important factors. For example, an increase in sales of one crop may lead to a decrease in sales of another crop. This phenomenon indicates a substitution relationship between the two crops, i.e., a rise in market demand for one crop squeezes demand for the other. Conversely, if an increase in the sales of one crop leads to an increase in the sales of the other, this indicates a complementary relationship between the two crops, i.e., their demand tends to rise and fall together.

When we look at these relationships, the cross-elasticity coefficient can help us quantify this effect. For example, if the cross-elasticity coefficient is positive, it indicates a degree of complementarity between the two crops; if it is negative, it indicates that they are substitutes. In this way, we can better understand the market dynamics between different crops, which in turn provides a scientific basis for developing planting strategies. And as indicated by economics, the cross elasticity coefficient formula is

$$E_c = (\Delta Q_x / Q_x) / (\Delta P_y / P_y) \quad (18)$$

where E_c is the cross elasticity of demand, Q_x is the quantity demanded of commodity x , and ΔQ_x is the change in demand for commodity x , P_y is the price of commodity y , and ΔP_y is the change in price of commodity y .

From this, we can calculate whether there are substitutions and complementarities between various crops based on the data from 2023 to 2027, and we can use the complementarities and substitutions to change our cropping strategies under the influence of factors such as market changes or climate change, for example, using the data and introducing adjustment coefficients and complementarity coefficients, which can give a more optimized objective function, as will be given in the following section of this problem. The adjustment and complementarity coefficients obtained (the objective function and results will be provided below after the analysis is completed).

Adjustment factor.

$$y_{ij}^t = y_{ij}^t + \beta_{AB} \times \Delta P_B^t \quad (19)$$

From this, we can, based on the data from 2023 to 2027.

where is a coefficient based on cross elasticity that represents the impact of price changes for crop B on the planting decision for crop A.

Complementarity factor.

$$X_j^t = X_j^{2023} \times (1 + a_{AB}) \quad (20)$$

where a_{AB} is the complementarity between crop A and crop B.

3.2. Analysis of experimental results

By calculating the Pearson's correlation coefficient, we find that the correlation coefficient between the sales price and the expected sales volume is $r=-0.81$, which indicates that there is a strong negative correlation between these two. This means that when the sales price increases, the expected sales volume usually decreases. This negative correlation may reflect the fact that higher prices inhibit consumers' willingness to buy, which leads to a decrease in sales volume.

On the other hand, the correlation coefficient between selling price and expected cost is $r=0.94$, which is close to a positive correlation.

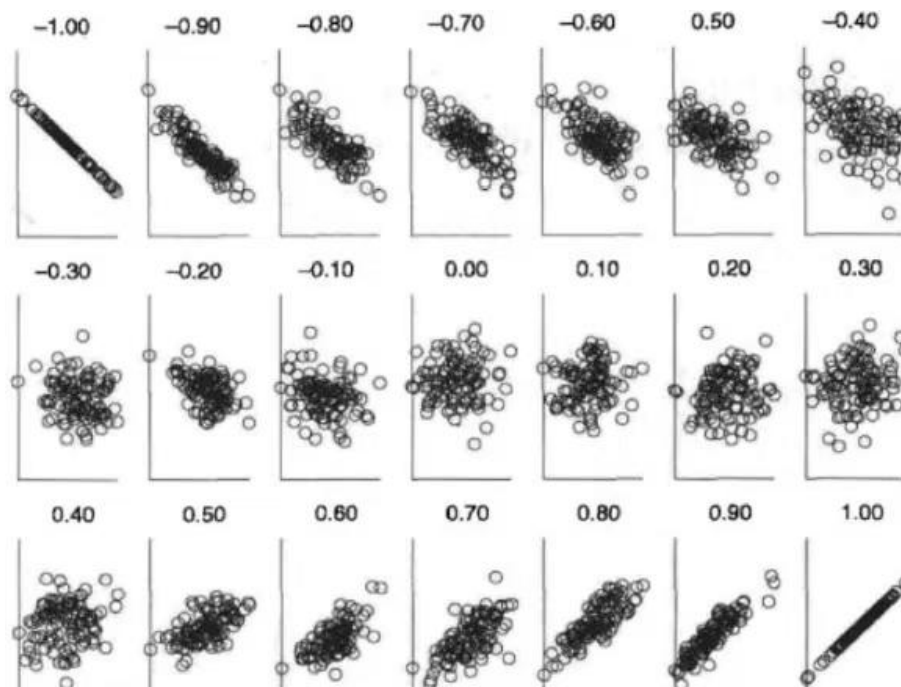


Figure 2. About the Pearson coefficient image

Figure 3 denotes that an increase in costs usually leads to an increase in selling prices. In other words, higher costs drive higher sales prices, and this relationship shows a strong link between costs and prices. However, this does not directly affect the expected change in sales volume, and although prices rise as a result of higher costs, this does not necessarily mean that sales volume will increase as well. This reflects the fact that sales volumes are influenced by a number of factors, not just the relationship between price and cost. The following Figure 3 shows a heat map of sales price plotted against expected costs.

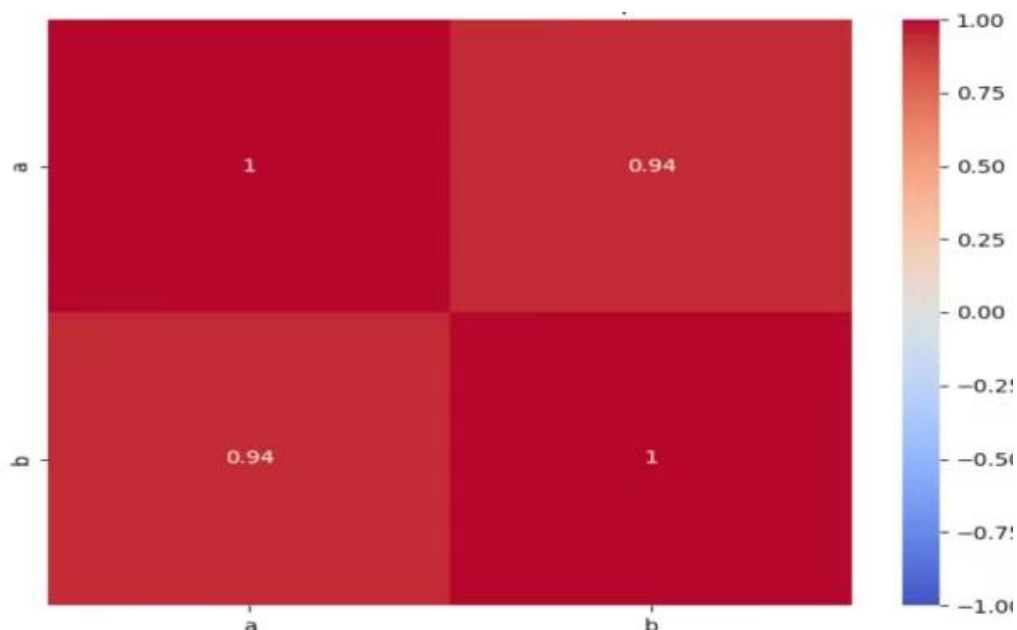


Figure 4. Heat map of sales price vs. expected cost

Figure 5 display $r = 0.94$, indicating a strong positive correlation between the two. There is a strong multicollinearity. When conducting feature engineering, it may be advisable to remove one of the two variables to prevent overfitting caused by multicollinearity.

4. Conclusions

This research method has significant advantages because it can find the global optimal solution, is adaptive, and facilitates robust testing. When estimating the statistical properties of complex systems, it can be combined with genetic algorithms to consider variables more comprehensively and reasonably, providing optimal solutions for multi-objective problems. Meanwhile, the Pearson correlation coefficient simplifies the complex multifactorial crop issues and measures the strength of linear relationships. The application of cross-elasticity coefficients, alongside economic knowledge, not only strengthens the paper but also captures the dynamic relationships in time series data to achieve problem resolution. Due to the relatively large number of crops and land in the farmland that needs solving, the use of genetic algorithms consumes significant computing resources; the rate of convergence may be slower. The Monte Carlo algorithm used in the previous section yields random results that may not be stable given the vast data in this problem. The Pearson coefficients used are applicable only to linear relationships and are more sensitive to outliers, potentially leading to inaccuracies with the large data sets in this issue.

In order to solve the problems of high consumption of genetic algorithm computing resources and slow convergence speed, the algorithm parameters can be optimized in the future, and the population size, crossover rate and mutation rate of the genetic algorithm can be adjusted to balance the computing resources and convergence speed. For Monte Carlo algorithms with low stability, more advanced sampling methods (e.g., importance sampling, Latin hypercube sampling) can be used to improve the stability and efficiency of the estimates. For Pearson coefficients, which are sensitive to outliers and can be inaccurate, the data can be preprocessed before use to remove outliers, normalize them, and improve the fit of the results.

References

- [1] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world[J]. *Science*, 2010, 327(5967):
- [2] Koul B, Yakoob M, Shah M P. Agricultural waste management strategies for environmental sustainability[J]. *Environmental Research*, 2022, 206: 112285.
- [3] Mirzaei A, Zibaei M. Water conflict management between agriculture and wetland under climate change: Application of economic-hydrological-behavioral modelling[J]. *Water Resources Management*, 2021, 35(1): 1-21.
- [4] Alibabaei K, Gaspar P D, Lima T M, et al. A review of the challenges of using deep learning algorithms to support decision-making in agricultural activities[J]. *Remote Sensing*, 2022, 14(3): 638.
- [5] Krishna V, Reddy T, Harsha S, et al. Analysis of crop yield prediction using machine learning algorithms[C]//2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT). IEEE, 2022: 1-4.
- [6] Maraveas C, Asteris P G, Arvanitis K G, et al. Application of bio and nature-inspired algorithms in agricultural engineering[J]. *Archives of Computational Methods in Engineering*, 2023, 30(3): 1979-2012.
- [7] Juraev F D, Mallaev A R, Aralov G M, et al. Algorithms for improving the process of modeling complex systems based on big data: On the example of regional agricultural production[C]//E3S Web of Conferences. EDP Sciences, 2023, 392: 01050.
- [8] Wu Y, Li X, Liu Q, et al. The analysis of credit risks in agricultural supply chain finance assessment model based on genetic algorithm and backpropagation neural network[J]. *Computational Economics*, 2021: 1-24.
- [9] Rana P, Varshney L R. Planting trees at the right places: Recommending suitable sites for growing trees using algorithm fusion[J]. *arxiv preprint arxiv:2009.08002*, 2020.
- [10] Rajasekaran S, Balla S, Huang C H. Exact algorithms for planted motif problems[J]. *Journal of Computational Biology*, 2005, 12(8).