# Research on Venture Capital Sum Model and loan Credit Evaluation Model based on Ant Colony algorithm

**Rui Chen[1*], Ying Chen[2], Tianyu Zhang[2]**

[1]College of Science, Northwest A&F University, Xianyang, Shaanxi, 712000

[2]College of Economics & Management, Northwest A&F University, Xianyang, Shaanxi, 712000

**Keywords:** Improved genetic algorithm, portfolio, numerical simulation, Analytic hierarchy process

**Abstract:** There is a lack of research on the application of personal credit risk assessment model in online lending and credit information industry, so it is particularly urgent to study the domestic P2P online loan industry which has many chaotic phenomena. Aiming at the problem of risk assessment of investment projects, a risk assessment method based on ant colony-analytic hierarchy process is proposed. under the condition of meeting the consistency requirement, the minimum difference of judgment matrix before and after adjustment is taken as the objective function, and aiming at the problems of weak global searching ability and low convergence accuracy of genetic algorithm (GA), the normal distribution crossover operator is introduced into the crossover operation of genetic algorithm, and is used to measure risk by CVaR. The adjustment problem of judgment matrix is transformed into the traveling salesman problem of ant colony algorithm, and the risk weight vector is calculated. The effectiveness of the method is verified by numerical examples, and the corresponding preventive measures are put forward for the main risks. Firstly, the personal credit evaluation model is established by using support vector machine, and the genetic algorithm is introduced to optimize the parameters of the model, and then the validity analysis and generalization analysis are made for the samples of two P2P network loan platforms. and according to the empirical results to explore the potential risks of credit brushing behavior. The empirical results show that GA-SVM model can effectively solve the problem of personal credit evaluation of P2P network loan platform, and has good robustness and promotion. The results show that when the method is applied to the risk assessment of overseas investment projects, a more reasonable weight vector can be obtained, which is helpful to formulate corresponding risk measures according to the potential main risks.

## 1. Introduction

In recent years, with the continuous development and improvement of the securities market, the enthusiasm of securities investors is increasingly high. But in addition to investment fixed income products (such as government bonds, corporate bonds), investment in stocks, funds are a certain risk. Then how to measure risk becomes the main object of people's research.

Portfolio theory aims at solving the problem of the optimal allocation of wealth among different assets. The mean-variance portfolio model of Harry [1] becomes the basis of the single-stage portfolio selection theory. However, the actual investment activities are usually multi-stage, and investors need to adjust their investment strategies constantly according to the actual situation to achieve their expected goals. Duan and Ng [2] proposed an effective investment strategy for mean-variance model in long-term portfolio problem by using embedding method and stochastic linear control method. Subsequently, a lot of researches on long-term multi-stage asset allocation were carried out , but they were all carried out under complete market conditions. Therefore, portfolio selection in uncertain environment has become a hot topic in recent years. From the perspective of reality, many scholars have considered the impact of various realistic factors on asset allocation, such as transaction costs, inflation, investor views and bankruptcy constraint [3].

In this paper, an ant colony analytic hierarchy process (AHP) is proposed to improve the judgment matrix. Compared with previous studies[4], the problem of adjusting the judgment matrix is

transformed into the travel agent problem of ant colony algorithm, which makes the adjusted judgment matrix satisfy the consistency and at the same time have less difference from the original judgment matrix, and can better reflect the decision maker's subjective judgment intention[5].

## 2. Portfolio model

### 2.1. CVaR risk measurement method

Set $f(\omega, y): R^n \times R^n \to R$ as the loss function of a portfolio, including omega for investment decision-making vector, y is the corresponding portfolio loss random vector. If you choose your fixed portfolio f ($\omega$, y) is a function of y. Assuming that Y is a continuous random variable and its probability density function is P (y), the probability that F ($\omega$, y) does not exceed in the case of allowing the maximum loss of is:

$$\psi(\omega, \alpha) = \int_{f(\omega,y)\le\alpha} p(y)\, dy \tag{1}$$

According to the definition of VaR model and CVaR model, for any confidence level $\in$ (0,1), it can be obtained as follows:

$$VaR_\beta = \min\{\alpha \in R \mid \psi(\omega, \alpha) \ge \beta\} \tag{2}$$

$$\begin{aligned} CVaR_\beta &= VaR_\beta + E[f(\omega, y) - VaR_\beta \mid f(\omega, y) \ge VaR_\beta] \\ &= \frac{1}{1-\beta} \int_{f(\omega,y)\ge VaR_\beta} f(\omega, y) p(y) dy \end{aligned} \tag{3}$$

Due to the complex changes in the market environment, it is difficult to directly obtain the accurate value of P (y), so historical data are used to predict the future distribution of random vectors. For the historical yield data of D trading days of known N securities, relative to $F_\beta(\omega, \alpha)$ the approximate value is:

$$\hat{F}_\beta(\omega, \alpha) = \alpha + \frac{1}{(1-\beta)D} \sum_{d=1}^{D} [f(\omega, y_d) - \alpha]^+ \tag{4}$$

### 2.2. Multistage mean-variance model

Unlike ordinary financial assets, labor income is a non-capital asset that cannot be traded and hedged. Current quantitative methods for uncertainty of labor income can be divided into the following three categories :(1) use proxy indicators such as occupation to measure uncertainty of income, represented by Skinner (2) In the form of variance and standard deviation of income-related data, representative figures are Carroll and Samwick;(3) Questionnaire survey was used to obtain people's perception of the uncertainty of labor income, and the main representative figures were Guiso et al.

Suppose an investor buys n securities, security i (i =1...N) in the capital portfolio $\omega = (\omega_1, \omega_2, \cdots, \omega_n)$ And the proportion in n is $r_i = \frac{1}{D}\sum_{d-1}^{D} y_{id}$, At the same time, in the real securities market, investors cannot avoid various transaction costs in the process of trading. Therefore, for the convenience of calculation, this paper sets the transaction costs of all securities as C =0.00185 according to the regulations of Shanghai and Shenzhen Stock Exchanges. At this time, the expected rate of return of the investment portfolio is $\sum_{i=1}^{n} (r_i - c)\, \omega_i$, At the same time, assuming that the minimum expected return of the investor is no less than, the risk is measured by CVaR, and the mean-CVAR investment portfolio model containing the transaction fee is:

$$\min \hat{F}_\beta(\omega, \alpha) = \alpha + \frac{1}{(1-\beta)D} * \sum_{d=1}^{D} [f(\omega, y_d) - \alpha]^+ \tag{5}$$

### 2.3. Vector machine classification

To find the optimal linear hyperplane first, and then will be used in the optimization algorithm into a convex programming problem and solving, finally using nonlinear mapping from the sample space is mapped to a high-dimensional feature space, so that the linear machine learning method can solve the samples in the feature space cannot be resolved in the original space highly nonlinear classification and regression problems. The so-called support vector refers to the key point of constructing the classification plane, and the optimization of the classification plane is to maximize the interval M between the lines where the support vector is located.

Solving the support vector is transformed to solving a quadratic programming problem with constraints:

$$\begin{aligned} &\min \frac{1}{2} \| w \|^2 \\ &\text{s.t., } y_i(w^T x_i + b) \geq 1, \forall i \end{aligned} \tag{6}$$

Lagrange multiplier method can be used to solve the above quadratic programming problem, and its objective function is:

$$L(w, b, a) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1] \tag{7}$$

## 3. The empirical analysis

### 3.1. The impact of the introduction of income risk on the multi-stage investment strategy

The investor selects 4 stocks from the stock market and obtains the closing price data of these 4 stocks from 2016 to 2020. Taking 1 year as one investment cycle, namely N=4 and T=3, the annual return rate and variance are calculated. The relevant data are shown in Table 1. The risk aversion coefficient of investors is 2, namely =2.The investor invests in the above four stocks and one risk-free asset Treasury bonds. The initial investment amount in the capital market is 1, and the initial labor income is 1. The investor carries out three successive investment activities, and adjusts the investment proportion of each asset at the beginning of each period. The return on risk-free assets is 0.02. In addition, it is assumed that the proportion of labor income invested in the capital market in each period is 0.5, i.e. Lt =0.5.

Tab 1 Annual return rate and variance of stocks

| Time | Assets | Expect | Variance |
|---|---|---|---|
| T=1 | 1 | 0.15 | 0.0235 |
| | 2 | -0.09 | 0.0074 |
| | 3 | 0.33 | 0.0120 |
| | 4 | -0.18 | 0.0277 |
| T=2 | 1 | 0.11 | 0.0027 |
| | 2 | 0.07 | 0.0011 |
| | 3 | 0.09 | 0.0021 |
| | 4 | 0.17 | 0.0058 |
| T=3 | 1 | 0.88 | 0.0041 |
| | 2 | 0.10 | 0.0013 |
| | 3 | 0.35 | 0.0031 |
| | 4 | 0.51 | 0.0170 |

According to the data of China Labor Statistics Yearbook, the average annual growth rate of China's residents' labor income is about 0.15, the variance of annual growth rate of residents' labor income is about 0.0029, and the growth rate of labor income is about the annual stock return is slightly positive with a covariance of 0.006.Considering the other correlation between the growth rate of labor income and risk assets, the covariance is set as -0.006 and 0, that is, the risk of labor income is positively correlated, negatively correlated and uncorrelated with risk assets respectively. In addition,

the growth rate value of labor income is 0.04, 0.1 and 0.2. Three groups of variance data of labor income growth rate are set to represent the change of labor income risk, which are 0.0015, 0.002 and 0.0025 respectively.

Verify the validity of the proposed model. Using the improved genetic algorithm, the risk and return of each phase of the portfolio were considered when the growth rate of labor income was 0.1, the growth rate of labor income was positively correlated with the rate of return on risky assets, and the variance of growth rate of labor income was 0.0015.Figure 1(A ~ c) shows the effective frontier (V-E) of the portfolio with labor income risk in each phase.

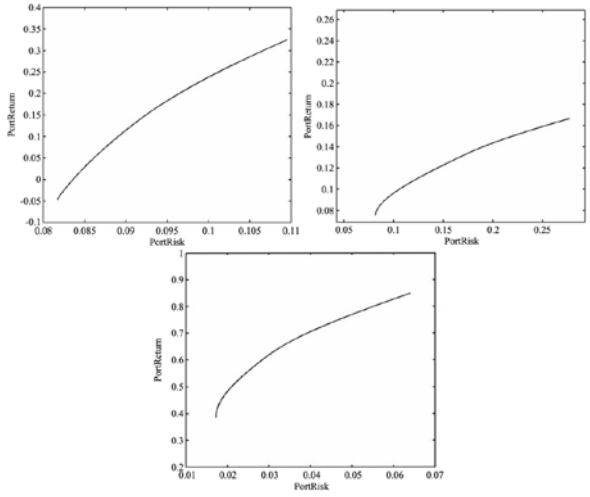


Fig.1 The income risk of the portfolio is effective at each period of the frontier

When the mean value of portfolio returns is the same, the investment risk taking labor income risk into account is greater than the investment risk taking no labor income risk. Therefore, considering labor income risks can better reflect the real investment environment faced by investors. If investors ignore their labor income risks, they will underestimate the investment risks in reality, thus causing investment losses.

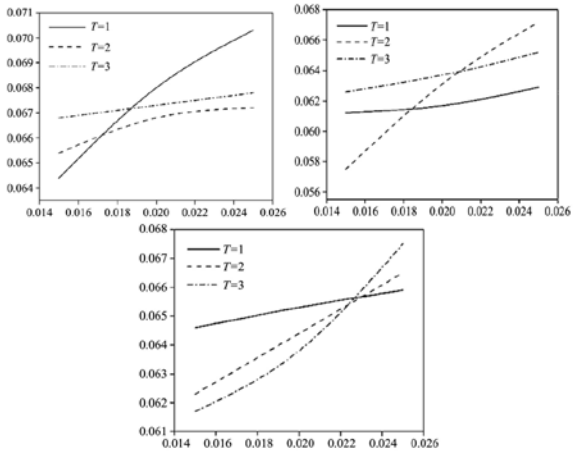


Fig.2 The percentage of risk-free investments in each period when income and risky assets are independent

### 3.2. Genetic algorithm optimization

Genetic algorithm is a kind of theory of nature of biological evolution evolved random search method, and its basic principle is to imitate the nature of biological evolution phenomenon, will become a genetic search space mapping space, possible solutions will be compiled into a single vector - chromosomes, the vector of each element is called a gene.The genetic algorithm selects the best chromosome by continuously calculating the adaptive value of each chromosome, and then obtains the optimal solution .The main advantage of genetic algorithm is that it can operate structure object directly, avoids the limitation of derivation and function continuity, and has excellent global

optimization ability.

The probabilistic optimization method enables it to automatically obtain the search space and carry out relevant optimization, and adjust the search direction adaptively. With the above characteristics, genetic algorithm has been widely used in adaptive control, machine learning, combinatorial optimization and other fields, which is the core technology in intelligent computing. Therefore, this paper uses genetic algorithm to optimize the support vector machine model to determine the value of key parameters.

## 4. Conclusions

By introducing the normal distribution crossover operator into the crossover operation of the conventional genetic algorithm, this paper discusses how to set the average value of the expected return considering the transaction rate. This paper expounds the important position of data cleaning in data application, introduces the distributed storage and distributed computing framework under the background of big data, and describes in detail the experimental process of clustering algorithm based on mahout to detect similar duplicate data. in the actual financial market, investors not only need to deal with the risk of financial assets themselves, but also to resolve the risk of labor income. So that the risk of the investment portfolio can be reasonably dispersed and the expected return can be obtained.

The overseas investment risk assessment method based on ACO-AHP transforms the consistency problem of the adjustment judgment matrix into the traveling salesman problem of the ACO algorithm, in order to solve the problem that the consistency of the judgment matrix in the traditional analytic hierarchy process is difficult to meet the requirements. An example of risk assessment of an overseas investment project is given to illustrate the effectiveness of the method. This paper considers the portfolio selection problem of risk-free investment proportion in each period when there is a positive correlation between labor income and risk assets, establishes a multi-period mean-variance model considering labor income risk, designs an improved genetic algorithm to solve the model, uses MATLAB for numerical simulation, and verifies the effectiveness of the model by using Chinese stock market data. And further analyze the impact of the introduction and changes of labor income risk on multi-period asset selection. The results show that the change of labor income risk has a complex impact on the construction of multi-stage investment portfolio. Specifically, when the labor income and risk assets are independent of each other, the experimental results show that the clustering algorithm can effectively detect similar duplicate records, which provides a solution for a large number of unstructured text data cleaning. it lays an important foundation for follow-up data mining.

## References

[1] Lu Xiaoli, he Guang, Li Gaoxi. Particle swarm optimization algorithm for solving the quantum behavior of self-financing portfolio model [J/OL]. Journal of Wuhan University (Science Edition): 1-7 http://kns.cnki.net/kcms/detail/42.1674.N.2020 1201.0910.001.html.

[2] Zhang Peng, Huang Meiyu. Stochastic fuzzy minimum variance portfolio decision with norm constraints [J]. Fuzzy Systems and Mathematics, 2020 (05): 65-76.

[3] Chen Feng, Xu Shaosong, Li Zhangzheng. Construct the intelligent expansion optimization system of wireless network based on AI algorithm [J].Design Technology of posts and Telecommunications, 2020 (10): 36-Qing.

[4] Wang Fuyu, Tang Tao. Application of two-species Fish Swarm algorithm in distributed portfolio [J/OL]. Journal of system Simulation: 1-12. http://kns.cnki.net/kcms/detail/11.3092.V.20200904.1326.002.html.

[5] Zhao Jian, Huo Jiazhen. Optimization and decision-making of quantitative investment strategy based on genetic algorithm [J]. Shanghai Management Science, 2011 (05): 19-24.