Decision Making Support Mechanism of Portfolio Selection in Multiple Period Consumption and Investment Model

Koji Okuhara*, Junko Shibatab, Nobuyuki Uenoa and Hiroaki Ishiič

*Department of Management and Information Sciences, Hiroshima Prefectural University, 562 Nanatuka, Shobara, Hiroshima 727-0023, Japan.
bDepartment of Artificial Complex Systems Engineering, Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Japan
čGraduate School of Information Science and Technology, Osaka University, 2-1 Yamada-oka, Suita 565-0871, Japan

Accepted in September 2004
Available online

Abstract

In this study, we analyze the portfolio selection in multiple period consumption and investment model that constructs an artificial market. In the model, there exist non-risk assets and risk assets in the market. Each agent determines redistribution of present property for maximizing the expected utility of consumption based on own judgment. Rational actions of investors who take part in the market are virtually realized on a computer by applying reinforcement learning to agents. Moreover, we propose a decision making support mechanism of portfolio selection by using reinforcement learning.

Keywords: Portfolio selection; Artificial market; Reinforcement learning; Multiple period consumption; Investment model

1. Introduction

Finance engineering has developed quickly in the past 30 years. It supports the deregulation and the internationalization in a financial market of recent years from a theoretical viewpoint. Furthermore, with the developments in information technology, application of finance theory becomes possible for the investment strategy incorporating many financial products. The situation surrounding a financial market has also been changing a lot.

Many of conventional mathematical models of the portfolio selection problem in finance theory are aimed at one period. Since an actual employment repeats the inflow, the outflow of funds and the rebalancing, the employment period needs to be multiple periods. However, it is generally difficult to search for the strict optimal portfolio strategy of a multi-period portfolio selection problem. So, the portfolio selection is used to be applied to the approximation models which assume the non-arbitrage condition in a perfect and efficient market (Black and Scholes, 1973).

In the finance theory based on an efficient market hypothesis, dynamics of a financial price is expressed by stochastic process. But an investor cannot always reflect to the information of actual market appropriately and quickly. Therefore, it tends to do many researches of the artificial market by using the multi-agent system that forms a financial price as a result of an investor’s actions (Palmer et al., 1994).

In the situation that is caused by the interaction of the environment of financial market and the action of investment, the investor cannot get to know more about the environment in advance, and decide his action based on it. In such a case, they need the technique of acquiring a conduct code by trial and error based on an interaction with environment. One of such techniques is a reinforcement learning.

In this study, we consider the multiple period consumption and investment model by the multi-agent system. The artificial market is composed of agents who derive a portfolio selection using reinforcement learning (Sutton, and Barto, 1998). In the artificial model, we propose the decision making support mechanism that can analyze the influence of the expected utility to the financial price and consumption.

2. The Outline of Multiple Period Consumption and Investment Model

We explain multiple period consumption and investment model (Konno and Furukawa, 1997). There are non-risk assets \( i = 1 \) and risk assets \( 2 \leq i \leq M \) in a market. The capital market is generally assumed perfect market. That is,

1) There are no transaction fee, dividend and tax.
2) The volume of investment is real number.
3) Action of investment does not affect price and earnings.
4) Investor can profit by short selling.

*Corresponding author: E-mail: okuhara@bus.hiroshima-pu.ac.jp
In order to assure that investor can benefit by investment in financial market, the market assumes the following basic properties of investment income:

\[
\begin{align*}
    r_t & \geq 0 \\
    E[r_t] & \geq \delta + r_t \\
    E[r_t] & \leq K
\end{align*}
\]  

where \( r_t \) is the interest rate at \( t \) term, and \( r_n \) is per-share earnings (random variable) of the investment opportunity \( i \) \((i=2,\ldots,M_t)\). If he invests \( \theta \) an investment opportunity \( i \) at the initial term, he earns \((1+r_n)\theta\) at the term end. The scalar \( M_t(M_1 \leq M) \) denotes the number of available investment opportunity at \( t \) term. \( \delta \) and \( K \) are positive constants, which is defined to assure the money market is favorable game. Assuming that the rate of return \( F_i \) follows Markov chain, and is independent at each term.

\[
F_i(x_2,x_3,\ldots,x_M) = P\{r_2 \leq x_2, r_3 \leq x_3, \ldots, r_M \leq x_M \}
\]

where \( x_i \) denotes arbitrary variable.

The arbitrary \( \Theta \) which contains \( \theta_i \geq 0 \) and such that \( \sum_{i=2}^{M} |\theta_i| = 1 \) satisfies the following condition of the investment opportunity \( i \) which cannot be short.

\[
\Pr\left\{ \sum_{i=2}^{M} (r_i - r_t)\theta_i < \delta_1 \right\} > \delta_2
\]

where \( \delta_1 < 0 \) and \( \delta_2 > 0 \). Equation (3) is equal to the no arbitrage condition, and is necessary and sufficient condition for solving portfolio selection problem.

For maintaining capacity to pay at each term, the invested property \( w_{t-1} \) at the decision making point must satisfy the following equation.

\[
\Pr[w_t \geq 0] = 1 \quad (t = 1,2,\ldots,T-1).
\]

The invested property at \( t-1 \) term is

\[
\sum_{i=1}^{M} z_{it} = w_{t-1} - c_t,
\]

where \( z_{it} \) is loaned money (if \( z_{it} < 0 \), \( z_{it} \) means borrowed money), \( z_{it} \) is the volume of investment in the investment opportunity \( i \) \((i=2,\ldots,M_t)\) at the initial \( t \) term, and \( c_t \) is the volume of consumption.

The value of investment in the dynamic consumption investment model which contains consumption and labor income is defined by

\[
w_e = \sum_{i=2}^{M} (r_i - r_t)z_{it} + (1+r_t)(w_{t-1} - c_t) + y_t \quad (r=1,2,\ldots,T),
\]

where \( y_t \) is labor income.

Here, an investor's purpose can be rewritten as

\[
\max E[U(c_1,\ldots,c_T)]
\]

for which the expected utility from consumption series is maximized under \( c_t \geq 0 \). The utility function \( U \) is a monotonous increase, concave function and reflects an investor's preference.

\[
(a,b,c_1,\ldots,c_T) \succ (b,a,c_1,\ldots,c_T) \quad (a > b),
\]

where \( U_t > 0 \), and \( U_t' < 0 \) for any \( t \), and \( \alpha < 1 \). One of the form of the utility function is given by

\[
U_t(c_t) = \frac{1}{\gamma} c_t^\gamma \quad (0 < \gamma < 1),
\]

where \( \gamma \) represents the degree of the preference of an utility function.

3. The Outline of Artificial Market and the Portfolio Selection by Reinforcement Learning

3.1 The Outline of Artificial Market

We discuss the framework of a more realistic artificial market by easing some conditions of a usual model. There are \( N \) agents who modeled investors. Note that each agent need not behave uniformly.

The amount of investment \( i \) of agent \( n \) in \( t \) term is described by \( z_{it}^n \), and the amount assigned to consumption is set to \( c_t^n \). In an artificial market, a company pays dividend \( d_t^n \) to the agent holding its brand.

Therefore, an investment value in time \( t \) of Eq. (6) is rewritten as

\[
w_t^n = \sum_{i=2}^{M} (1+r_t)z_{it}^n + (1+r_t)z_{it}^n + y_t + \sum_{i=2}^{M} r_d^n d_t^n \frac{z_{it}^n}{p_i(t)} + z_{it}^n, \quad (10)
\]

where \( r_t \) is a positive constant and \( z_{it}^n \) is the repayment in the case of dealings failure (Okuhara, Shibata, and Tanaka, 2003; Shibata, Okuhara, Kato, and Sakawa, 2004). Profit \( r_t \) is given by

\[
r_t = \frac{p_i(t)}{p_i(t-1)} - 1.
\]

where \( p_i(t) \) denotes price at time \( t \).

A dividend is the colored noise of dispersion and it is given by
\[
\log \frac{d^u}{d^u_n} = a^u \log \frac{d^u_{i(t-1)}}{d^u_i} + b^u_i \xi^u_i,
\]

where \( \xi^u_i(t) \) is the gaussian noise with an average 0 and a variance \( \sigma^2_i \) [2]. \( a^u \) and \( b^u_i \) are positive parameters which satisfy \( a^2_i + b^2_i = 1 \). The term \( \log \frac{d^u_i}{d^u} \) has an average 0 and a variance \( \sigma^2_i \), and its autocorrelation function decreases within the correlation time \( r_a = 1/\log a^u \).

The value of the property currently held can be changed or each agent can consume when the next period starts. However, rearrangements are executed under condition that whole property of each agent is fixed, that is,

\[
\sum_{i=1}^{M} z^u_i = w^u_{i-1} - c^u_i.
\]

And the total amount of the risk assets in a market is assumed to be fixed, that is,

\[
\sum_{i=1}^{N} z^u_i = Z_i \quad (2 \leq i \leq M).
\]

Each agent acts so that he makes the following utility maximum by considering a risk and a limited property.

\[
U(c^u_i, \ldots, c^u_n) = u^c_i(c^u_i) + \alpha_i u^c_i(c^u_i) + \cdots + \alpha_i \cdots \alpha_n u^c_n(c^u_n).
\]

By the way, each agent cannot always do a desired dealing. Here, agent \( n \) wants to buy or sell a risk asset \( i \) with an amount \( b^u_i \), \( o^u_i \), respectively. Then the total buying and selling amount at time \( t \) are derived by

\[
B^u_t = \sum_{n=1}^{N} b^u_n, \quad O^u_t = \sum_{n=1}^{N} o^u_n.
\]

Therefore, as desired dealing is possible only for the case of \( B^u_t = O^u_t \), the amount of possession of each agent about risk assets in case of \( B^u_t \neq O^u_t \) shall be set to

\[
z^u_n = z^u_{i(t-1)} + \frac{V^u}{B^u_t} b^u_n - \frac{V^u}{O^u_t} o^u_n,
\]

where \( V^u = \min(B^u_t, O^u_t) \). The refund \( z^u_n \) in the case of dealings failure is set to

\[
\tilde{z}^u_n = \left(1 - \frac{V^u}{B^u_t} \right) b^u_n - \left(1 - \frac{V^u}{O^u_t} \right) o^u_n.
\]

The price \( p^u_{i(t)} \) of a risk asset \( i \) is mainly determined based on dealing of all agents. Under such a situation, the price \( p^u_{i(t)} \) is given as follows.

\[
p^u_{i(t)} = \frac{2p^u_{i(t-1)}}{1 + \exp[-U^u_t/T_t]}, \quad U^u_t = \log \frac{B^u_t}{O^u_t},
\]

where \( T_t \) is a positive constant representing the sensitivity of risk asset \( i \). If it takes small value then it is sensitive to the difference between demand and supply, otherwise it is not sensitive to these values.

### 3.2. The Portfolio Selection by Reinforcement Learning

In an artificial market, agent \( n \) acts by considering the following discount reward

\[
V^u_t = \sum_{k=0}^{\infty} \alpha^u_n u_{t+k}(c^u_{t+k}).
\]

under finite property at time \( t \). In this study, we propose the searching algorithm for the portfolio strategy by a reinforcement learning using neural networks. The neural network (Barto, Sutton, and Anderson, 1983) that realizes an actor-critic model (Doya, 1999) is applied to the reinforcement learning. Note that an actor critic model needs only minimum computational effort. Therefore, when the number of possible action like continuous value action is infinite, this method is effective compared with other methods, such as Q-Learning.

First, the agent \( n \) observes a state \( x_t \) in environment, and actor model generates a control output given by

\[
q^u_n = f \left( \sum_{j=1}^{N_A} w^{A_n}_{ij} g^A_{ij}(x_t) + n^u_t \right),
\]

\[
g^A_{ij}(x_t) = \exp \left( -\frac{1}{2} (x_t - m^A_{ij})^T c^A_{ij} (x_t - m^A_{ij}) \right),
\]

where \( g^A_j \) is the \( j \)-th radial basis function, \( N_A \) is the number of radial basis functions, \( w^{A_n}_{ij} \) is the weight, \( n^u_t \) is standardized gaussian noise, and \( f \) is sigmoid function expressed by

\[
f(x) = \frac{g^u_{i \max}}{1 + \exp[-x/T_n]}.
\]

where \( g^u_{i \max} \) is the maximum of the output \( i \), and \( T_n \) is sensitivity of agent \( n \).

In actor model, agent \( n \) tries trading by

\[
b^u_n = q^u_n w^u_n - z^u_{i(t-1)} (q^u_n w^u_n > z^u_{i(t-1)})
\]

\[
o^u_n = z^u_{i(t-1)} - q^u_n w^u_n (q^u_n w^u_n < z^u_{i(t-1)})
\]

from a control output.

The critic model generates the following evaluation value
\( V_x^n(x_t) = \sum_{j=1}^{N_C} w_j^n g_j^c(x_t), \) (25)

where \( N_C \) is the number of radial basis functions of the critic model. As a result of action, the critic model receives the following reward

\[ R_t = u_t(c^n_t) = \frac{1}{\gamma_\alpha} \left( w^\alpha_{t-1} - \sum_{i=1}^{M^\alpha} c^\alpha_{it} \right), \] (26)

from environment and observes the state \( x_{t+1} \) after transition. The expected utility is given by

\[ E[u(c^n_t)] = V_x^n(x_t) - \alpha V_x^n(x_{t+1}). \] (27)

Then the TD error as a reinforcement signal is defined by the difference between the actual utility and the expected utility as follows:

\[ \delta_t = u_t(c^n_t) - E[u(c^n_t)] = u_t(c^n_t) + \alpha^n_x V_x^n(x_{t+1}) - V_x^n(x_t) \] (28)

After \( \delta_t \) is sent to actor model, the past record of activity is given by

\[ e_{jt}^n = \lambda e_{jt-1}^n + g_j^c(x_t) \] (29)

and the weight is updated by

\[ w_{jt}^C = w_{jt}^C + \eta_C \delta_t e_{jt}. \] (30)

The weight of the actor model is updated by

\[ w_{jt}^A = w_{jt}^A + \eta_A \delta_t g_j^A(x_t) \eta_A, \] (31)

where \( \eta_A, \eta_C \) are the rates of learning and \( \lambda \) is the reduction.

4. A Simulation Result and Consideration

In this section, we show results that the reinforcement learning using neural networks can be applied to the portfolio strategy in case of difficulty apply usual dynamic programming method to it in artificial market. The values acquired in a simulation are the average of 10 trials. Assuming that one trial consists of 2100 steps and the last 100 steps are observed.

First, we consider there are 9 agents, 1 non-risk asset and 3 risk assets in the artificial market. Sensitivities of prices are set to \( T_i = 50, (i = 1, 2, 3) \), and sensitivities of all agents are set to \( T_n = 1, (n = 1, 2, 3, \ldots, 9) \). The initial values of assets are given by \( 10000 + 50\sigma \) and the initial values of prices are given by \( 100 + 5\sigma \), where \( \sigma \) is the uniform random value from -1 to 1. Moreover, Table 1 shows the values of the other parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Simulation’s Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_A )</td>
<td>0.01</td>
</tr>
<tr>
<td>( \eta_C )</td>
<td>0.01</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.7</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.01</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.3</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.5</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Figure 1. Dynamics of Risk Assets.

Figure 2. Dynamics of Value of Utility Function.

Figures 1 and 2 show the dynamics of prices of risk assets and the values of utility functions for agents in one trial, respectively. However, the values of utility functions express 3 agents whose behaviors are typical fluctuation.

Next, we investigate about the influence of \( \alpha_n \) representing the discount rate to future reward, and \( \gamma_n \) representing the degree of the preference of utility function. Hence, assuming that all agents have same combination of the discount rate and the degree of preference, \( (\alpha_n, \gamma_n) = (0.1, 0.1), (0.1, 0.9), (0.5, 0.5), (0.09, 0.1), (0.09, 0.9) \).

Figures 3 and 4 show the prices of risk assets and the values of utility function, respectively. Where each center of line denotes the average, and the top or bottom of category denote the range of \( 3\sigma \). From the results in Figures 3 and 4, we remark that \( \alpha_n \) does not have big influence to
the prices of risk assets and to the values of an utility functions. On the other hand, when $\gamma_n$ becomes large, we find that the prices of risk assets and the values of utility functions become increasing and the variances of utility functions also become large.

5. Conclusion

In this study, we consider the situation with complicated interaction between the environment of financial market and the investors. We further propose the artificial market of the multiple period consumption and investment model by using multi agents, which can take into consideration the case which investors can not determine their action depending on detail information of environment.

Then rational behaviors of investors are realized by applying the reinforcement learning to the agents in virtual computer simulation. Moreover we analyzed the influence of the discount rate for future reward and the degree of preferences for the expected utility to the consumptions and the formed financial prices.

References


