Expert Systems with Applications 36 (2009) 10735–10745

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A portfolio optimization model using Genetic Network Programming with control nodes

Yan Chen [a,*], Etsushi Ohkawa [a], Shingo Mabu [a], Kaoru Shimada [b], Kotaro Hirasawa [a,*]

[a] Graduate school of Information, Production and Systems, Waseda University, 2-7, Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan
[b] Information, Production and Systems Research Center, Waseda University, 2-7, Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, Japan

ARTICLE INFO

Keywords:
Portfolio optimization
Genetic Network Programming
Control node
Reinforcement learning

ABSTRACT

Many evolutionary computation methods applied to the financial field have been reported. A new evolutionary method named "Genetic Network Programming" (GNP) has been developed and applied to the stock market recently. The efficient trading rules created by GNP has been confirmed in our previous research. In this paper a multi-brands portfolio optimization model based on Genetic Network Programming with control nodes is presented. This method makes use of the information from technical indices and candlestick chart. The proposed optimization model, consisting of technical analysis rules, are trained to generate trading advice. The experimental results on the Japanese stock market show that the proposed optimization system using GNP with control nodes method outperforms other traditional models in terms of both accuracy and efficiency. We also compared the experimental results of the proposed model with the conventional GNP based methods, GA and Buy&Hold method to confirm its effectiveness, and it is clarified that the proposed trading model can obtain much higher profits than these methods.

1. Introduction

Nowadays, evolutionary computation has become a subject of general interest with regard to the power to solve complex optimization problems. It has been successfully applied to many fields of science and technology. This paper presents an application of evolutionary computation method named Genetic Network Programming to the problem of multi-brands optimization in the field of financial economics, which is one kind of portfolio optimization. Portfolio optimization in the stock market consists of deciding what brands to include in a portfolio given the investor's objectives and economic conditions. The always difficult selection process includes identifying which brands to purchase, how much, and when. The basic idea is that we want to choose a group of brands from a large number of available issues, in order to maximize the expected return given an acceptable risk rate. A rational investor needs to consider not only maximizing the profit of the investment, but also minimizing the uncertainty or risk resulting from the fluctuations that are expected in the value of the portfolio. The main problem is how to allocate the available capital in order to maximize profit and minimize risk simultaneously.

There have been increased the number of applications of Artificial Intelligence (AI) techniques, mainly artificial neural networks, genetic algorithm and genetic programming, which have been applied to technical financial forecasting (Dempster & Jones, 2001; Goldberg, 1989) as they have the ability to deal with complex non-linear problems and have the self-adaptation for dynamically changing problems. Several applications of Genetic Algorithms (GA) to the financial problems have been done, such as portfolio optimization, bankruptcy prediction, financial forecasting, fraud detection and scheduling (Loraschi, Tettamanzi, Tomassini, & Verda, 1995; Skolpadungket, Dahal, & Harnpornchai, 2007). Genetic Programming (GP) (Koza, 1992) has also been applied to many problems in the time-series prediction. Although these AI approaches possess the properties required for the technical financial forecasting, they have some inherent bottlenecks. For example, Genetic Programming sometimes causes the bloating problem due to its tree structure. Neural networks cannot be used to explain the causal relationships between input and output variables because of their black-box nature.

In the past studies, we have proposed Genetic Network Programming (GNP) and Genetic Network Programming with Reinforcement Learning (GNP-RL) (Eguchi, Hirasawa, Hu, & Ota, 2006; Mabu, Chen, Hirasawa, & Hu, 2007; Mabu, Hirasawa, & Hu, 2007) as an extended method of GA (Goldberg, 1989) and GP (Koza, 1992). Since GNP represents its solutions using graph structures, which contributes to creating quite compact programs and implicitly memorizing past action sequences in the network flows, it has been clarified that GNP is an effective method mainly for complicated problems such as portfolio optimization systems. Moreover,

* Corresponding author.
E-mail addresses: chenyan15153@akane.waseda.jp (Y. Chen), hirasawa@waseda.jp (K. Hirasawa).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.049

in our former research, GNP-RL method was successfully applied to stock trading model (Mabu et al., 2007), and its applicability and efficiency has been confirmed.

Recently, in order to extend the functions of conventional GNP, Genetic Network Programming with control nodes (GNPcn) (Eto, Mabu, Hirasawa, & Hu, 2006) has been proposed. Since GNP has a directed graph structure, the aim of GNPcn is to improve the performance of GNP by extending the evolutionary method of it. In traditional GNP, the current node isn't compulsorily transferred to the start node. However, in the GNPcn method, the number of control nodes are set up and a certain number of processing nodes are executed before returning to one of the control nodes, i.e., we extend the breadth and depth of searching space for GNP. It is clarified from the simulations that the performance of GNP could be improved by the combination with control nodes.

In this paper, we extend our previous research on GNP-RL and propose an algorithm that integrates the GNP-RL and control nodes in order to create an efficient portfolio optimization model for given multi-brands. The features of the proposed method compared with other traditional methods are as follows: The GNPcn method makes a stock trading strategy considering the recommendable information of technical indices and candlestick charts (Izumi, Yamaguchi, Mabu, Hirasawa, & Hu, 2006) for efficient trading decision making. Reinforcement Learning is also used in this paper for taking appropriate actions.

Section 2 presents the literature review. Section 3 describes the proposed GNPcn approach to be studied in this paper. In Section 4, we explain the optimization algorithm in brief. Section 5 presents experimental environments, conditions and results using GNPcn method. The trading profits are presented and compared with both the traditional GNP method and Buy&Hold method. Finally, Section 6 concludes this paper.

2. Literature review

In recent decades, the portfolio problem in financial engineering has received a lot of attention. The classical portfolio problem can be described like the following: Given a finite amount of resources, and the process to which these resources can be allocated, we need to get the allocation which maximizes a given goal function and minimizes the risk simultaneously. The foundation of portfolio optimization was laid by Markowitz (1959), where he proposed a mean–variance optimization model for designing an optimum portfolio based on the idea of minimizing risk and maximizing expected returns. As we know, it is difficult to acquire good estimates for the expected returns, and the calculation of risk becomes more and more complex as the number of available assets grow. In the case of linear constraints, the problem can be solved efficiently by parametric quadratic programming. However, there are many real-world non-linear constraints which limit the number of different assets in a portfolio. As a consequence, evolutionary computation was developed to calculate the optimal portfolio since portfolio problems make the search space become larger. In this paper, how to allocate the given money to a certain number of fixed brands is discussed, which is different from the classical mean–variance optimization model.

Over the last few decades, various approaches have been applied to several financial problems, especially for stock market activities. Generally speaking, these approaches can be separated into two categories: statistical and AI. The statistical methods are widely used to predict the stocks based on the past time-series data. The traditional statistical approaches include ARMA method (Box & Jenkins, 1976), the threshold Autoregressive model (Tong & Lim, 1980), Smooth Transition Autoregressive model (STAR) (Sarantis, 2001), the Autoregressive Conditional Heteroscedastic (ARCH) (Engle, 1982) and multiple Linear Regression model. These methods rely on assumption of linearity among variables and normal distribution. However, with statistical models, problems arise when the variance in the time series increase or when non-linear processes exist in the time series. On the other hand, AI approaches, with the increasing need for more efficient trading models in the stock market, has been confirmed to outperform the conventional statistical models for that it overcomes the limitation of such an assumption (Enke & Thawornwong, 2005).

As a main approach in the AI field, Artificial Neural Networks (ANNs) has been widely used for its ability to forecast financial performance. Dropsy (1996) uses ANNs as a non-linear prediction tool to forecast international equity risk, in which both linear and non-linear forecasting results outperform the random work. Lam (2004) applied the back-propagation algorithm to integrate fundamental and technical analysis for financial performance prediction. The experimental results show that the ANNs outperform the benchmark. Both the forecasting and decision models are significantly outperforming the benchmark market performance. However, despite the wide spread use of ANNs in the financial domain, there are significant problems to be addressed. Since ANNs are data-driven model, the underlying rules in the data are not always apparent, which leads to so-called black-box models, and investors cannot benefit from the knowledge discovery in the analytic process. In addition, the buried noise and complex dimensionality of the stock market data make it difficult to learn or re-estimate the ANNs parameters (Kim & Han, 2000).

Genetic Algorithms (GA), as one of the most popular heuristic optimization techniques, were originally developed by Holland (1975). Subsequently, GA had been applied to many optimization problems in engineering and operations research. GA is a good tool for optimization problems since they make no restrictive assumptions about the solution space. Lin considered the multi-objective genetic algorithm for portfolio selection problem (Lin, Wang, & Yan, 2001). Oh proposed a new portfolio selection algorithm based on portfolio beta by using genetic algorithm (Oh, Kim, Min, & Lee, 2006). Moreover, Lin and Liu (2008) considered Markowitz's model with minimum transaction lots and they presented three other models using GA as their solver. Deck used GA to train a neural network trading system. However, when GA was applied to the portfolio optimization, the problem is that many chromosomes are coded into the same portfolio, or similar chromosomes may be coded into very different portfolios which makes it more difficult for GA to produce better chromosomes from good ones. These problems multiply the GA's search space and makes GA less efficient in finding the optimal portfolio.

Genetic programming (GP), which has been described by Koza (1992), can be considered as an extension of GA. It uses tree-like individuals that can represent mathematical expressions. So far, GP has been applied to wide range financial fields such as stock trading system (Potvin, Soriano, & Vallee, 2004), bankruptcy prediction (Etemadi, Rostamy, & Dehkordi, in press), and etc. In comparison with GA, GP allows the optimization of much more complicated structures and can therefore be applied to a greater diversity of problems (Sette & Boullart, 2001). Recent applications of GP has been done, when the notion of risk is not considered (Marney, Fyfe, Tarbert, & Miller, 2001). Although GP are widely used in the financial field, it occasionally causes some bloating problems for its tree structure.

As another tool based on expert knowledge, fuzzy logic can be used for stock market forecasting domain either independently or hybridized with other methods. For instance, Romahi and Shen (2000) developed an evolving rule based expert system to forecast the financial market activity, where the fuzzy logic and rule induction were combined together to obtain a promising method. However, fuzzy logic still has its limitation when applied to the stock