Applied Soft Computing 36 (2015) 534–551

Contents lists available at ScienceDirect

Applied Soft Computing

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

Review article

Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review

Yong Hu[a,*], Kang Liu[b], Xiangzhou Zhang[a,e], Lijun Su[b], E.W.T. Ngai[c], Mei Liu[d]

[a] Big Data Decision Institute, Jinan University, Guangzhou, PR China
[b] School of Business, Guangdong University of Foreign Studies, Higher Education Mega Center, Guangzhou 510006, PR China
[c] Department of Management and Marketing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[d] Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, KS 66160, USA
[e] School of Business, Sun Yat-sen University, China

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form 7 June 2015
Accepted 12 July 2015
Available online 21 July 2015

Keywords:
Literature review
Evolutionary computation
Algorithmic trading
Stock trading rule
Rule discovery
Classification framework

ABSTRACT

Despite the wide application of evolutionary computation (EC) techniques to rule discovery in stock algorithmic trading (AT), a comprehensive literature review on this topic is unavailable. Therefore, this paper aims to provide the first systematic literature review on the state-of-the-art application of EC techniques for rule discovery in stock AT. Out of 650 articles published before 2013 (inclusive), 51 relevant articles from 24 journals were confirmed. These papers were reviewed and grouped into three analytical method categories (fundamental analysis, technical analysis, and blending analysis) and three EC technique categories (evolutionary algorithm, swarm intelligence, and hybrid EC techniques). A significant bias toward the applications of genetic algorithm-based (GA) and genetic programming-based (GP) techniques in technical trading rule discovery is observed. Other EC techniques and fundamental analysis lack sufficient study. Furthermore, we summarize the information on the evaluation scheme of selected papers and particularly analyze the researches which compare their models with buy and hold strategy (B&H). We observe an interesting phenomenon where most of the existing techniques perform effectively in the downtrend and poorly in the uptrend, and considering the distribution of research in the classification framework, we suggest that this phenomenon can be attributed to the inclination of factor selections and problem in transaction cost selections. We also observe the significant influence of the transaction cost change on the margins of excess return. Other influenced factors are also presented in detail. The absence of ways for market trend prediction and the selection of transaction cost are two major limitations of the studies reviewed. In addition, the combination of trading rule discovery techniques and portfolio selection is a major research gap. Our review reveals the research focus and gaps in applying EC techniques for rule discovery in stock AT and suggests a roadmap for future research.

Contents

* Corresponding author. Tel.: +1 973 596 6597; fax: +1 973 596 5777.
E-mail addresses: henryhu200211@163.com (Y. Hu), researcher_lk@foxmail.com (K. Liu), zhxzhou86@foxmail.com (X. Zhang), onnowa@163.com (L. Su), eric.ngai@polyu.edu.hk (E.W.T. Ngai), meiliu@kumc.edu (M. Liu).

http://dx.doi.org/10.1016/j.asoc.2015.07.008
1568-4946/

1. Introduction

Stock investment has attracted a lot of attention. In 2012, the value of share trading-electronic order book trade in the United States stock markets reached USD 23,226,924 million [1]. However, the computerization of stock trading from order book to exchange has been generating large amounts of real-time data [2,3]. At the same time, government, institutions, social media, and listed companies have been releasing an ocean of data on the operating performance of listed companies, such as news, financial statements, and macroeconomic information [2–4]. Discovering useful knowledge from these substantial high-dimensional financial data [5–12] and catching investment opportunities faster than other investors in such a noisy and dynamic market environment [2,10,11,13–15] are significant challenges faced by investors constantly.

Algorithmic trading (AT), an important automatic analysis and trading decision approach for equity investment, gained prominence in the early 1990s and accounted for at least 50% of the total US equity trading volume in 2012 [16]. Generally, AT refers to the use of sophisticated computer algorithms to automatically make certain trading decisions in the trading cycle, including pre-trade analysis (data analysis), trading signal generation (buying and selling recommendations), and trade execution (order management) [17–20]. One of the advantages of AT is the effectiveness and efficiency of machine learning techniques in financial big data analysis [2,21,22]. However, some AT learning models are considered as "black boxes" [23–28] because they involve difficulty in providing easy-to-understand explanations on the interactions between the model inputs and the outputs. Trading with black boxes makes investors uncomfortable and elicits mistrust in the model [2,29,150]. To address this issue, an increasing number of researchers have investigated rule discovery techniques for finding explicit trading rules that can provide explicit knowledge to guide trading.

Rule discovery is an important aspect of data mining because it can generate a set of symbolic rules that describe the relationship among variables in a natural way, and rules can be better understood by the human mind than any other data mining model [30,31]. Numerous studies have demonstrated the necessity of rule discovery in AT, such as the following: (1) increasing investors' approval of the system by improving the comprehensibility of the system decision logic. Thus, investors can justify system decisions using their domain knowledge, and potential investment risks can be reduced [32,33,151]. (2) facilitating the discovery of new knowledge and integration of new and old knowledge [33,34]. (3) reducing errors derived from noise, feature subset selection, or inaccurate parameter settings [35].

Evolutionary computation (EC) has been widely employed in rule discovery. EC is generally defined as a computing tool to solve realistic problems by simulating the evolutionary mechanisms of nature [36–41]. It is mainly based on a population, uses probabilistic transition rules, and directly applies the objectives from the user as "fitness" [36,37,40–42]. Applying EC techniques for rule discovery in stock AT is becoming popular because EC is able to find a sufficiently good solution for a wide range of problems within a relatively short time [14,42]. Previous research suggested that applying EC-based models to trading rule discovery could yield promising results [32,43–45]. Moreover, EC can be employed to optimize the individual trading rule [14,26,44,46,47] and the parameters of the underlying rule discovery algorithm [44,48,49]. This study focuses on the distinctive competence of EC in rule discovery. Over the past few years, several literature reviews have been conducted on stock prediction models. Atsalakis and Valavanis [50] reviewed the application of neural and neural-fuzzy techniques in stock prediction. Guresen et al. [51] presented a comparative survey of different neural network models in NASDAQ Stock exchange index prediction. Bahrammirzaee [52] provided a comparative analysis among artificial neural networks, expert system and hybrid intelligent systems in credit evaluation, portfolio management and financial prediction and planning. However, at present, a systematic and comprehensive review of the EC techniques applied to explicit trading rule discovery remains lacking. Therefore, the objectives of this paper are as follows: (1) develop a classification framework for the application of EC techniques for rule discovery in stock AT; (2) provide a systematic and comprehensive review; and (3) determine research gaps and propose suggestions for directions of future research.