Stock Selection with a Novel Sigmoid-Based Mixed Discrete-Continuous Differential Evolution Algorithm

Lean Yu, *Member, IEEE*, Lunchao Hu, and Ling Tang

Abstract—A stock selection model with both discrete and continuous decision variables is proposed, in which a novel sigmoid-based mixed discrete-continuous differential evolution algorithm is especially developed for model optimization. In particular, a stock scoring mechanism is first designed to evaluate candidate stocks based on their fundamental and technical features, and the top-ranked stocks are selected to formulate an equal-weighted portfolio. Generally, the proposed model makes literature contributions from two main perspectives. First, to determine the optimal solution in terms of feature selections (discrete variables) and the corresponding weights (continuous variables), the original differential evolution algorithm focusing only on continuous problems is extended to a novel mixed discrete-continuous variant based on sigmoid-based conversion for the discrete part. Second, the stock selection model also resolves the gap of the application of differential evolution algorithm to stock selection. Using the Shanghai A share market of China as the study sample, the empirical results show that the novel stock selection model can make a profitable portfolio and significantly outperform its benchmarks (with other model designs and optimization algorithms used in the existing studies) in terms of both investment return and model robustness.

Index Terms—Artificial intelligence, constrained optimization, evolutionary computing, portfolio analysis

1 Introduction

Quantitative asset management involves a set of processes, i.e., ideas proposal, returns forecast, portfolios construction and performance evaluation [1]. Amongst them, stock selection for further portfolio formulation may be one of the most crucial but challenging issues, due to the complexity of financial markets. Generally speaking, a stock selection model includes two main steps, i.e., stock scoring and stock ranking, the former of which may be the core part. According to the existing literature, an abundance of stock evaluation models (or stock scoring mechanisms) have been developed, which can be mainly divided into two categories: traditional statistical regression approaches and computational intelligence (CI) techniques [2]. Both of them have their respective strengths and weaknesses. Traditional statistical regression models are relatively easy to implement and understand due to their simple forms, nevertheless they often appear relatively poor performance [2], [3]. Some important works are as follows. Sharpe [4] first published the capital asset pricing model (CAPM). Ross [5] introduced his arbitrary pricing theory (APT). Fama and French [6] formulated a three factor model.

However, due to the complexity in stock markets, the CI models have fully been shown to be more efficient than the traditional statistical models, though they might be somewhat difficult to understand. In particular, diverse CI models have been applied to stock evaluation, such as artificial neural networks (ANNs), support vector machines (SVMs) and various optimization tools (e.g., differential evolution (DE)). For ANNs, Yu et al. [7] utilized ANNs to select stocks. Wang and Gupta [8] built a stock trading and predicting system based on ANN and wavelet. Despite of the wide application, ANNs often suffer from over-fitting and local optimum problems. To avoid such problems to some degree, SVMs were proposed based on the principle of structural risk minimization and were employed to model stock markets. For example, Yu et al. [9] implemented SVM in stock market analysis and concluded that SVM was a promising alternative in stock evaluation. For model optimization, the DE algorithm, a typical evolutionary algorithm (EA), has widely been applied to financial market analysis. For example, Takahama et al. [10] employed the DE algorithm to optimize the ANN model and improved the prediction accuracy of stock prices. Nizar et al. [11] applied the DE algorithm to the fuzzy set rule exploration for modeling financial market dynamics. These above studies all demonstrated that the CI techniques significantly outperformed the traditional statistical regression approaches in modeling financial markets.

Considering both the interpretation capability of traditional statistical regression models and the prediction accuracy of the CI techniques, this study tends to formulate a CI-based linear stock scoring mechanism. In the proposed model, the stock scores are estimated in terms of a linear combination of various fundamental and technical features.

- L. Yu and L. Tang are with the School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China. E-mail: yulean@mail.buct.edu.cn, tangling_00@126.com.
- L. Hu is with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. E-mail: hulunchao@126.com.

Manuscript received 7 Nov. 2015; revised 21 Feb. 2016; accepted 15 Mar. 2016. Date of publication 23 Mar. 2016; date of current version 1 June 2016. Recommended for acceptance by H. Xiong.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2545660

Accordingly, choosing the appropriate features in a stock market with highly rich information becomes the most key task in a stock scoring mechanism [3]. Furthermore, feature selection can also reduce computational complexity and guarantee model generalization of stock scoring mechanisms [12], [13]. The related methods for feature selection can be mainly divided into three categories: filter, wrapper and embedded approaches [14]. The filter methods remove poorly informative features according to statistical properties. The wrapper methods explore feature spaces to score feature subsets based on prediction power [15]. The embedded methods train models according to preliminary known rules, which otherwise limits their applications due to ignorance of the rules. Generally speaking, the wrapper methods have been much more popularly used, because they focus not only on reducing data dimensions but also on improving prediction accuracy. The wrapper methods usually couple with some powerful CI tools for model optimization, e.g., genetic algorithm (GA) [16] and genetic programming (GP) [3]. All the empirical studies observed the effectiveness of such hybrid wapper-based feature selection models with CI optimization techniques.

Compared with the above CI optimization techniques, the DE algorithm, which searches better solutions through changing the current individuals based on the scaled differences of randomly selected and distinguished members [17], has been shown to be a much simpler but more efficient algorithm [18]. Accordingly, the DE algorithm has been applied to various difficult optimization tasks, such as text mining [19], clustering [20] and engineering [21]. Despite of the successful application, there were few researches introducing DE into stock selection models, to the best of our knowledge. Under such a background, this study especially incorporates the DE algorithm into the proposed stock selection model for model optimization, which fills in the gap of its application to stock selection. Furthermore, as the proposed stock selection model is a mixed discrete-continuous optimization problem with both discrete decision variables (feature selections) and continuous decision variables (the corresponding weights), the typical DE algorithm focusing only on continuous problems is especially extended to a novel mixed discrete-continuous variant with sigmoid-based conversion for the discrete part.

Generally, this paper attempts to propose a stock selection model by introducing and improving the DE algorithm for model optimization. In particular, two main steps are involved in this novel stock selection model: stock scoring and stock ranking. First, a stock scoring mechanism is designed, in which stocks are evaluated based on various fundamental and technical features. Second, the top-ranked stocks are selected to formulate an equal-weighted portfolio as the model output. For choosing appropriate features (discrete decision variables) and optimizing the corresponding weights (continuous decision variables), the powerful CI optimization technique of DE is especially introduced and improved to a novel mixed discrete-continuous variant with sigmoid-based conversion for the discrete part, i.e., the novel sigmoid-based DE algorithm.

The main aim of this study is to propose a stock selection model with a novel sigmoid-based DE algorithm for the mixed discrete-continuous optimization, and to verify its superiority over benchmark models with other model designs (in terms of different decision variables and fitness functions) and other popular optimization techniques. The rest of this paper is organized as follows. Section 2 provides a literature review on the DE algorithm. Section 3 formulates the novel stock selection model and the sigmoid-based mixed discrete-continuous DE algorithm. Section 4 designs the experiment study, in terms of sample data, benchmark models and evaluation criteria. Section 5 reports the empirical results and verifies the effectiveness of the proposed stock selection model and the novel sigmoid-based DE algorithm. Section 6 concludes the paper and notes the main directions for future research.

2 Literature Review on Differential Evolution Algorithm

Since Price and Stone [17] proposed the DE algorithm in 1997, it has become one of the most popular optimization tools due to the simple design and efficient performance [18]. However, the original DE algorithm mainly focuses on continuous problems but finds difficulty in solving discrete problems, which largely limits its application.

To address such a problem, some works have been conducted to extend the traditional DE algorithm to discrete variants. Generally, DE was modified for discrete problems mainly using indirect approaches and direct approaches [22]. On the one hand, the indirect approaches used certain posterior conversion operators to transform real solutions in the original DE into integer, discrete and binary solutions for discrete problems. For example, Lampinen and Zelinka [23] proposed a discrete DE variant, in which real values were rounded to the nearest integers (marked as Round-DE). Angira and Babu [24] added an equality constraint $x(1-x)=0$ in the typical DE to generate the binary solution x. Pampara et al. [25] introduced a trigonometric function into DE to map real spaces into binary spaces and developed angle modulated DE (AMDE) algorithm. However, this indirect searching approach might carry a great computational burden even with a high level of prediction accuracy [25].

On the other hand, the direct DE variants work directly with integer or binary encoded variables without posterior conversion. For instance, Gong and Tuson [26] proposed a new binary encoded DE algorithm (binDE) in which binary solutions were evolved according to the corresponding difference of two randomly chosen solutions. In particular, using the restricted-change DE (Res-DE) mutation, a variable in a solution has the chance to mutate based on the corresponding difference of variables in two randomly selected solutions. Using the any-change DE (Any-DE) mutation, any variable in a solution has the same opportunity to mutate depending on the hamming distance of the two randomly selected solutions. Deng et al. [27] presented a novel binary DE algorithm directly in binary space without scale parameter. Wang et al. [28] proposed a modified binary DE algorithm using a probability estimation operator to manipulate binary solutions. Based on these improvements, the DE algorithm has successfully addressed numerous discrete problems, such as unit commitment [29], load dispatch [30] and knapsack problem [31]. However, the binary DE-variants neglect the mutation operator in the typical continuous DE algorithm, largely reducing the diversity of population [26].

Recently, mixed discrete-continuous problems with both discrete and continuous decision variables have become increasingly popular. For mixed discrete-continuous DE variants, Lampinen and Zelinka [23], Lin et al. [32] and Kim et al. [33] suggested simply rounding continuous values to integers for the discrete parts of solutions. Datta and Figueira [22] developed a binDE directly presenting discrete parts in binary forms to avoid rounding errors. However, these above mixed discrete-continuous DE variants were applied to numerical examples rather than real problems, and their effectiveness should be further verified in real problems.

For the proposed stock selection model with both discrete and continuous decision variables, this study proposes a novel mixed discrete-continuous DE algorithm by introducing an efficient conversion operator, i.e., sigmoid-based conversion, for discrete solutions. In particular, the sigmoid-based conversion, which can generate discrete solutions by following a logistic probability distribution, has widely been considered as one of the most basic statistical approaches in binary classification [34]. Actually, the sigmoid-based conversion has already been employed in various CI optimization algorithms for discrete and discrete-continuous problems, such as PSO [35] and FA [36]. However, to the best of the knowledge, this efficient sigmoid-based conversion has not been introduced into DE so far. Therefore, this paper especially resolves such a literature gap by formulating a sigmoid-based DE algorithm and further verifying its effectiveness in stock selection.

Compared with the existing DE variants, the novel sigmoid-based DE algorithm for mixed discrete-continuous optimization makes contributions from two main perspectives. First, it might be the first try to introduce the efficient sigmoid-based conversion into the traditional DE algorithm for mixed discrete-continuous optimization. Second, this novel sigmoid-based DE algorithm is then incorporated into the proposed stock selection model for feature selection and weight optimization, which finely verifies the effectiveness of the novel mixed discrete-continuous DE variant in stock selection, different from the previous studies based on simple numerical examples.

3 Methodology Formulation

A novel stock selection model is formulated in this section by employing and modifying the DE algorithm for mixed discrete-continuous optimization based on sigmoid conversion, i.e., the sigmoid-based DE algorithm. Section 3.1 first gives an overview of the novel stock selection model. For model design, decision variables (i.e., feature selections and their weights) and objective (i.e., fitness function) are described in Section 3.2. For model optimization, the sigmoid-based DE algorithm is formulated, as discussed in Section 3.3.

3.1 Overall Framework

Generally speaking, a stock selection model mainly includes two key steps, i.e., stock scoring and stock ranking. In the first step, a stock scoring mechanism is proposed based on various features. In the second step, stocks are ranked according to their scores, and the top-ranked ones are selected to formulate an equal-weighted portfolio. In particular, the feature selection and weight optimization in the former step may be the most crucial tasks in the stock selection model. To address such tasks, the simple but efficient CI algorithm of DE is introduced and modified into the novel sigmoid-based DE algorithm to search the optimal mixed discrete-continuous solution for the proposed stock selection model. Fig. 1 illustrates the general framework of the proposed model.

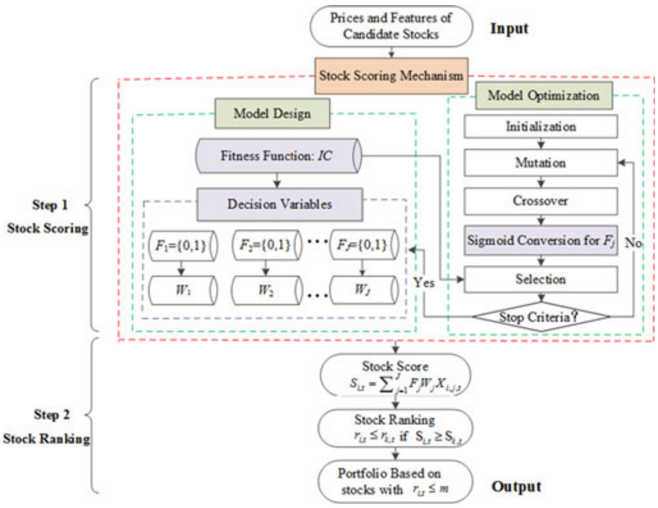


Fig. 1. General framework of the proposed stock selection model with the novel sigmoid-based DE algorithm.

As shown in Fig. 1, the proposed stock selection model is generally composed of the following two main steps:

1) *Stock Scoring*: A stock scoring mechanism is proposed to evaluate all candidate stocks, including two main parts: model design and model optimization. In model design, stocks are scored through various fundamental and/or technical features, and the fitness function of information coefficient (*IC*) helps capture the relationship between features and future returns of stocks, in terms of feature selections (discrete decision variables) and the corresponding weights (continuous decision variables). For this mixed discrete-continuous problem, model optimization is conducted via the novel sigmoid-based DE algorithm.
2) *Stock Ranking*: Candidate stocks are ranked according to their scores generated by the stock scoring mechanism. The m top-ranked stocks, with the highest potentials of increases in future prices, are selected to formulate an equal-weighted portfolio, as the output of the stock selection model.

Obviously, the stock scoring mechanism is the key part in the novel stock selection model, with two main factors: model design (for decision variables and fitness function) and model optimization (for feature selection and weight optimization). The following two sections detail them, respectively.

3.2 Model Design

In general, a higher score of a stock refers to a higher potential of increases in its future prices. To estimate stock scores, various features including profitability, structure, liquidity,

TABLE 1
Candidate Features Used for Stock Scoring

Category	Index	Feature	Descriptions	Indicator	Reference
Price ratio	1	PE	Price-to-earnings ratio = share price / earnings per share	-	[2]
	2	PB	Price-to-book ratio = share price / book value per share	-	[2]
	3	PS	Price-to-sales ratio = share price / sales per share	-	[2]
Profitability	4	ROE	Return on equity (after tax) = net income after tax / shareholders' equity	+	[2]
	5	ROA	Return on asset (after tax) = net income after tax / total assets	+	[2]
	6	ROIC	Return on investment capital = net income after tax / invested capital	+	[40]
	7	EPS	Earnings per share = profit preferred dividends / weighted average common shares	+	[40]
	8	NPM	Net profit margin = net income after tax / net sales	+	[40]
	9	EBIT/IC	Earnings before interest and tax to invested capital = earnings before interest and tax / invested capital	+	[40]
Structure	10	DE ratio	Debt-to-equity ratio = total liabilities / shareholders' equity	-	[41]
	11	AE ratio	Asset-to-equity ratio = total assets / shareholders' equity	+	[41]
	12	D/IC	Debt to invested capital = total debt/ invested capital	-	[40]
	13	LD/OC	Long-term debt to operating capital = total long-term debt / operating capital	-	[40]
Liquidity	14	CR	Current ratio = current assets / current liabilities	+	[41]
	15	QR	Quick ratio = quick assets / current liabilities	+	[41]
	16	SC	Sales cash ratio = operating cash flow / sales	+	[40]
	17	CFPS	Cash flow per share = cash flow / number of shares	+	[40]
	18	FCFPS	Free cash flow per share = free cash flow / number of shares	+	[40]
Efficiency	19	ITR	Inventory turnover rate = cost of goods sold / average inventory	+	[41]
	20	RTR	Receivables turnover rate = net credit sales / average accounts receivable	+	[41]
	21	AT	Asset turnover rate = cost of goods sold / total asset	+	[41]
Growth	22	OPG	Operating profit growth = (quarterly operating income / the corresponding quarterly operating income last year)-1	+	[40]
	23	NPG	Net profit growth = quarterly net income after tax / the corresponding quarterly net ncome after tax last year	+	[40]
Momentum	24	RSI	Relative strength index = 100-100 / (1+RS), RS = average gain / average loss	-	[40]
	25	RSV	Raw stochastic value = (current price - 52 week low) / (52 week high - 52 week low)	-	[40]

efficiency, growth, momentum and others can be used, as listed in Table 1. Based on these candidate features, the performance of a given stock can be then quantitatively evaluated in terms of a score, under the assumption that a higher value implies a better performance in the near future.

Let $Y_{i,j,t}$ denote the score of stock i assigned by feature j at time t, i.e., the Z-score normalization of $V_{i,j,t}$ [37] where $V_{i,j,t}$ is the actual score by feature j. Especially, if feature j is return on asset (ROA), a larger value implies that the assets of the corresponding corporate might be more profitable in generating revenues. Accordingly, $Y_{i,j,t}$ can be defined by:

$$Y_{i,j,t} = \frac{V_{i,j,t} - \overline{V_{j,t}}}{D_{j,t}}, \quad (1)$$

where $\overline{V_{j,t}} = \frac{1}{N}\sum_{i=1}^{N} V_{i,j,t}$ is the average value of feature j across all N stocks at time t, and $D_{j,t} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(V_{i,j,t} - \overline{V_{j,t}})^2}$ is the standard deviation. If feature j is a price-to-book ratio (P/B ratio), a small value indicates that the market value of the related corporate is relatively undervalued to its book value [38]. Therefore, $Y_{i,j,t}$ can be calculated according to the following form:

$$Y_{i,j,t} = \frac{\overline{V_{j,t}} - V_{i,j,t}}{D_{j,t}}. \quad (2)$$

The score $Y_{i,j,t}$ is assumed to follow a normal distribution with mean zero and deviation one. It is worth noticing that the directional indicator on each feature, i.e., ROA (+) or P/B ratio (-) as labeled in Table 1, is determined according to the existing studies.

As mentioned above, feature selection (i.e., whether a feature is selected as a main feature) and the corresponding weight optimization (i.e., the assignment of importance to the selected feature) are the most crucial tasks in the stock selection model. First, binary variable $F_j = \{0, 1\}$ is utilized to represent whether feature j is used in stock evaluation ($F_j = 1$) or not ($F_j = 0$). Moreover, let W_j denote the weight on the jth feature, for capturing the relationship between feature j and stock scores. Accordingly, the final score $S_{i,t}$ of stock i at time t can be estimated in terms of a linear combination of various features $j = \{1, 2, \ldots, J\}$ [39], [40]

$$S_{i,t} = \sum_{j=1}^{J} F_j W_j Y_{i,j,t}. \quad (3)$$

As a higher score reflects a higher potential of increases in stock prices, the candidate stocks can be ranked according to their scores. Let $r_{i,t} = \{1, 2, \ldots, N\}$ denote the ranking of stock i at time t, i.e., $r_{i,t} \leq r_{k,t}$ if $S_{i,t} \geq S_{k,t}$, where $i,k \in \{1, 2, \ldots, N\}$ represent any two different stocks. Thus, a more highly ranked stock has a higher potential of price increase, and an equal-weighted portfolio can be constructed for the next period by selecting the stocks with the top m rankings $r_{i,t} = \{1, 2, \ldots, m\}$ at the end of each period. Accordingly, the performance of the formulated portfolio for the next period can be evaluated as the average return of all selected stocks

$$R^p_{t+1} = \frac{1}{m}\sum_{r_{i,t}=1}^{m} R_{t+1}(r_{i,t}), \tag{4}$$

where $R_{t+1}(r_{i,t})$ is the next period return of the stock with current ranking $r_{i,t}$ at time t, and R^p_{t+1} is the next period return of the portfolio constructed by the proposed model.

For effectively capturing the relationship between features and future returns, *IC* is especially chosen as the objective function (i.e., fitness function), due to its powerful capability in judging stock rankings [1]

$$\begin{aligned} \min F &= -\frac{1}{T}\sum_{t=1}^{T} IC_t, \\ IC_t &= \frac{\mathrm{cov}(r_{i,t}, r'_{i,t+1})}{\sqrt{\mathrm{var}(r_{i,t})\mathrm{var}(r'_{i,t+1})}}, \end{aligned} \tag{5}$$

where $r_{i,t}$ is the score ranking of stock i by the proposed model at time t, $r'_{i,t+1}$ is the actual return ranking in the next period, and T is the total number of training periods. The functions cov($*$) and var($*$) are the covariance and variance estimations, respectively. Obviously, IC_t is actually the Spearman correlation between the currently predicted rankings of stocks $r_{i,t}$ and their actual return rankings $r'_{i,t+1}$ in the next period.

3.3 Model Optimization

To determine the optimal solution in terms of feature selections F_j (discrete decision variables) and the corresponding weights W_j (continuous decision variables), this study introduces the powerful CI algorithm of DE for model optimization. Since the original DE form focuses on solutions with continuous values but otherwise finds difficulty in solving discrete optimization problem, this paper improves it into a mixed discrete-continuous variant for the proposed stock selection model. This section first gives a description of the original DE algorithm, and then formulates the novel sigmoid-based DE algorithm.

3.3.1 Typical DE

The DE algorithm proposed by Storn and Price [17] has been considered as a simple but efficient CI tool for global optimization. As a population-based heuristic algorithm, DE produces optimal solutions iteratively through a series of stages, i.e., initialization, mutation, crossover and selection, until the stop criteria are met.

a) *Initialization*: For a D-dimension optimization problem, the DE algorithm first randomly generates a population of P initial feasible solutions in terms of chromosomes $\mathbf{X}_{p,g=0} \in \mathbf{R}^D, p = \{1, 2, \ldots, P\}$,

$$\mathbf{X}_{p,g=0} = \mathbf{X}_{LB} + \mathbf{Z}\cdot(\mathbf{X}_{UB} - \mathbf{X}_{LB}), \tag{6}$$

where $g \in \{0, 1, \ldots, G\}$ is the step of iteration, $\mathbf{X}_{LB}$ and $\mathbf{X}_{UB} \in \mathbf{R}^D$ are the lower boundary and upper boundary of feasible domain, respectively, and $\mathbf{Z} \in \mathbf{R}^D$ is a random vector with random components following a uniform distribution between 0 and 1.

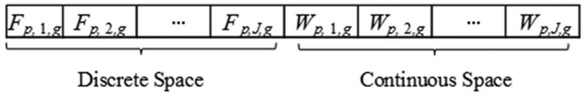


Fig. 2. Encoding for decision variables in the proposed stock selection model.

b) *Mutation*: The mutation stage aims to create donor individuals from the current ones. In the DE algorithm literature, the parent chromosome at the current generation is marked as a target vector, and a mutated one obtained through several mutations is a donor vector. The offspring individual, i.e., a trial vector, can be obtained by recombining the target vector and its donor vector [18]. In this study, a simple but efficient approach for mutation is employed to create the donor vector $\mathbf{V}_{p,g} \in \mathbf{R}^D$ for individual p,

$$\mathbf{V}_{p,g} = \mathbf{X}_{r^p_1,g} + \beta\cdot(\mathbf{X}_{r^p_2,g} - \mathbf{X}_{r^p_3,g}), \tag{7}$$

where $\mathbf{X}_{r^p_1,g}$, $\mathbf{X}_{r^p_2,g}$ and $\mathbf{X}_{r^p_3,g} \in \mathbf{R}^D$ are randomly sampled vectors form the current population excluding the individual p at generation g. The random indices r^p_1, r^p_2 and $r^p_3 \in \{1, 2, \ldots, p-1, p+1, \ldots, P\}$ are mutually exclusive integers. β is a scaled factor typically set on the range of $(0, 1)$.

c) *Crossover*: The crossover operator is further to generate the trial vector $\mathbf{U}_{p,g} = \{u_{p,d,g}\}$ based on the target vector $\mathbf{X}_{p,g} = \{x_{p,d,g}\}$ and its donor vector $\mathbf{V}_{p,g} = \{v_{p,d,g}\}$, to enhance the diversity of population:

$$u_{p,d,g} = \begin{cases} v_{p,d,g} & \text{if } r_{p,d} \le C_r \text{ or } r_d = d \\ x_{p,d,g} & \text{otherwise,} \end{cases} \tag{8}$$

where $r_{p,d}$ is a random term on the range of $(0, 1)$, $r_d \in \{1, 2, \ldots, D\}$ is a randomly chosen index, and C_r is the crossover rate predefined between 0 and 1.

d) *Selection*: Selection is to finally determine whether the target vector $\mathbf{X}_{p,g}$ or its trail vector $\mathbf{U}_{p,g}$ can survive as the solution in the next generation $\mathbf{X}_{p,g+1}$,

$$\mathbf{X}_{p,g+1} = \begin{cases} \mathbf{U}_{p,g} & \text{if } f(\mathbf{U}_{p,g}) \le f(\mathbf{X}_{p,g}) \\ \mathbf{X}_{p,g} & \text{otherwise,} \end{cases} \tag{9}$$

where $f(*)$ is the fitness function of optimization to be minimized in general, e.g., *IC* in the proposed stock selection model. Accordingly, if the trial vector $\mathbf{U}_{p,g}$ can yield better utility in terms of fitness function, the current solution, i.e., the target vector $\mathbf{X}_{p,g}$, will be replaced, or otherwise the target vector will remain to work.

3.3.2 Sigmoid-Based Mixed Discrete-Continuous DE

The proposed stock selection model is obviously a mixed discrete-continuous problem, in which the decision variables of feature selections $F_j = \{0, 1\}, j = \{1, 2, \ldots, J\}$ are binary variables whereas the weights W_j are continuous ones. Accordingly, the decision variables can be divided into two parts: discrete space for feature selection variables $F_j = \{0, 1\}$ and continuous space for their respective weights $W_j \in [0, 1]$, as shown in Fig. 2. In DE, the discrete term $F_{p,j,g} = \{0, 1\}$ represents the selection decision on feature j in the pth solution at iteration g, and the continuous term $W_{p,j,g} \in [0, 1]$ is the corresponding weight. It is worth noticing that if feature j is not selected in solution p at iteration g, i.e., $F_{p,j,g} = 0$, the corresponding weight $W_{p,j,g}$ is accordingly set to 0.

However, the typical DE form focuses on the solutions with continuous values but otherwise finds difficulty in solving discrete or mixed discrete-continuous optimization problems. Therefore, the original DE algorithm should be improved to a mixed discrete-continuous variant for the novel model here, by adding a conversion operator after the crossover step.

In conversion, the most popular binary classification model, i.e., sigmoid function, is introduced to identify a candidate feature as a key factor for stock scoring ($F_{p,j,g} = 1$) or a poorly informative one ($F_{p,j,g} = 0$). Based on the sigmoid function, the conversion from the continuous variable $x_{p,d,g}$ in the original DE to the binary form $F_{p,d,g}$ is conducted according to the probability $P(x_{p,d,g})$ which follows a logistic distribution [34]

$$F_{p,d,g} = \begin{cases} 1 & \text{if } r_{p,d,g} \leq P(x_{p,d,g}) \\ 0 & \text{otherwise,} \end{cases} \qquad P(x_{p,d,g}) = \frac{1}{1+e^{-x_{p,d,g}}}, \tag{10}$$

where $r_{p,d,g}$ is a random term following a uniform distribution on the range of $(0, 1)$. After adding such a conversion operator for the discrete part after the crossover operator, the traditional DE algorithm can be extended to the sigmoid-based DE for mixed discrete-continuous problems.

4 Experiment Design

For illustration and verification purposes, this study utilizes the novel model to select stocks in the Shanghai A share market of China. Section 4.1 gives a brief description to the sample data, Section 4.2 formulates benchmark models for comparison, and Section 4.3 defines evaluation criteria.

4.1 Data Descriptions

The Shanghai A share market is selected as the study sample in this study due to two main reasons: the increasingly important role of China's stock market in global financial market and the dominant role of the Shanghai A market in China's financial market. First, amongst international stock markets, China has been the second largest market in terms of trading volume, and the third largest market in terms of capitalization (about 3.7 trillion in 2013) just behind the U.S. and Japan [42]. Second, the Shanghai A share market is the biggest stock market in China, in terms of both trading volume and capitalization [43].

In particular, a total of 483 stocks are considered, excluding the stocks in financial industry due to the different balance sheet structure [1], [3] and those ever labeled as Special Treatment (ST) because they are unsteady and do not deserve long-term investment [43]. The quarterly sample data of stock prices and financial features are obtained from Wind Database (http://www.wind.com.cn), covering the period from the first quarter in the year 2005 to the fourth quarter in 2012. Stock returns are calculated in terms of the natural logarithm of price ratio (i.e., $R_{i,t} = \ln(P_{i,t}/P_{i,t-1})$, where $R_{i,t}$ and $P_{i,t}$ represent the return and price of stock i at time t, respectively. The yearly Chinese demand deposit rates are used as the risk free rates, due to data availability [44].

TABLE 2
Data Splitting Strategy: Training Periods (in Gray) and Testing Periods (in Black)

Case	05Q1	05Q2	05Q3	05Q4	...	12Q3	12Q4
1							
2							
...				...			
29							
30							

Both fundamental and technical variables are chosen as candidate features for stock evaluation, according to the previous studies (e.g., [2], [39], [40], [41], [45], [46]). In particular, a total of 23 fundamental features and two technical features are considered in this study, which can be grouped into seven categories, as shown in Table 1.

All these time series data are divided into training sets and testing sets, as presented in Table 2, with a total of 30 cases throughout the whole sample periods. Taking the first sample (in the first row) for example, the grey area (the first two quarters of 2005) indicates the training period for model training, and the black area (the third quarter of 2005) is the testing period for performance evaluation. Similarly, for the second row, the training period is from the first quarter of 2005 to the third quarter of 2005, and the testing period is the fourth quarter of 2005. Notice that this data splitting strategy, with a varying size of training set across different periods, is somewhat different from the regular sliding window procedure with a fixed training sample size. This strategy used in this paper attempts to utilize all available information that can be obtained in the current period to train the model [47]. For example, in the first case ($t = 3$), all historical observations, i.e., the data in the first two quarters ($t = 1, 2$), are used as the training dataset.

4.2 Benchmark Models

The performance of stock selection models highly depends on stock scoring mechanisms, with the key factors of decision variables, fitness functions and optimization methods. Therefore, to verify the effectiveness of the novel stock selection model, a set of benchmark models are formulated for comparison, by introducing different decision variables, fitness functions and optimization methods.

4.2.1 *Benchmarks with Other Decision Variable Designs*

In the proposed model, feature selection variables and their weights are utilized as the decision variables. However, some existing studies omitted one of them (e.g., [6], [40]), and/or introduced some other variables such as directional indicators for determining a feature as a ROA or P/B ratio (e.g., [39]). Accordingly, Benchmarks A1–A4 with different decision variables are designed, as listed in Table 3. Notably, when the decision variables of feature selections are not considered (e.g., in Benchmark A1), all candidate features

TABLE 3
Benchmark Models with Different Decision Variables

Model	Feature Selections	Directional Indicators	Weights	Reference
A0	√	–	√	This Study
A1	–	–	–	[40]
A2	√	–	–	
A3	√	√	–	
A4	√	√	√	[39]

Note: The symbol "√" means that the corresponding decision variables are considered in the model, whereas the symbol "–" means that the corresponding variables are not selected.

are used for stock scoring. If the variables of directional indicators are ignored (e.g., in the proposed model A0 and Benchmarks A1 and A2), features are labeled as ROA (+) or P/B (-) ratios according to the existing studies without optimization, as presented in Table 1. When the weights are not optimized (e.g., in Benchmarks A1–A3), an equal weight is assigned on each feature to calculate stock scores.

4.2.2 Benchmarks with Other Fitness Functions

A proper fitness function is also important to guarantee the efficiency of stock selection models [48]. To verify the effectiveness of the *IC* fitness function in the proposed model, benchmarks with other fitness functions popularly used in the previous studies are formulated for comparison.

According to Kuhn and Johnson [49], a model using fewer inputs is more efficient, under a given level of prediction accuracy. Therefore, a positive penalty is introduced on the number of selected features [49]:

$$F1 = (1-\alpha)F + \alpha\frac{M}{25}, \tag{11}$$

where M is the total number of selected features, F is the *IC* fitness function as designed in (5), and α is a trade-off between the *IC* fitness function and the control on the size of features. In this study, by changing α from 0.05 to 0.50 with a step length of 0.05, a total of 10 benchmark models with different α in the form of $F1$ can be formulated, to compare with the proposed model with $\alpha = 0.00$.

Besides, various other popular fitness functions can be also introduced for comparison, such as intra-fractile hit rate (*IFHR*) [3] as designed in (12), spread (*SPREAD*) [3] in (13) and cumulative return (*CR*) [39] in (14) and mean-to-variance (*MV*) ratio [50] in [15].

$$\begin{aligned} F2 &= -\frac{1}{T}\sum_{t=1}^{T} IFHR_t \\ IFHR_t &= \frac{\sum_{r_{i,t}=1}^{m} \operatorname{sgn}(R_{t+1}(r_{i,t}) - M_{t+1})}{2m} \\ &+ \frac{\sum_{r_{i,t}=N-m+1}^{N} \operatorname{sgn}(M_{t+1} - R_{t+1}(r_{i,t}))}{2m}, \end{aligned} \tag{12}$$

$$\begin{aligned} F3 &= -\frac{1}{T}\sum_{t=1}^{T} Spread_t \\ Spread_t &= \frac{\sum_{r_{i,t}=1}^{m} R_{t+1}(r_{i,t}) - \sum_{r_{i,t}=N-m+1}^{N} R_{t+1}(r_{i,t})}{m}, \end{aligned} \tag{13}$$

$$F4 = -CR = -\prod_{t=1}^{T} R_{t+1}^{p}, R_{t+1}^{p} = \frac{1}{m}\sum_{r_{i,t}=1}^{m} R_{t+1}(r_{i,t}), \tag{14}$$

$$\begin{aligned} F5 &= -MV = -\frac{1}{T}\sum_{t=1}^{T} MV_{t+1} \\ MV_{t+1} &= \frac{R_{t+1}^{p}}{\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(R_{t+1}(r_{i,t}) - R_{t+1}^{p}\right)^2}} \end{aligned} \tag{15}$$

where M_t denotes the average return of all candidate stocks at time t, T is the total number of training periods, N is the number of candidate stocks, m is the number of selected stocks, and the indicator function sgn(x) is equal to 1 when $x \geq 0$ or 0 when $x < 0$. In particular, the fitness function *MV* considers risks (i.e., the standard deviations of return rates) of selected stocks.

4.2.3 Benchmarks with Other Optimization Algorithms

To test the performance of the novel sigmoid-based DE algorithm, two categories of optimization algorithms are introduced. First, as the proposed optimization algorithm aims to select features for evaluating stocks, some typical feature selection models should be considered. Second, other powerful CI-based optimization algorithms should be also performed for comparison.

Typical Feature Selection Algorithms: The proposed algorithm is to compute the optimal feature set and the optimal feature weights. Therefore it should be first compared with baselines and state-of-the-art feature selection algorithms. In this study, the Pearson correlation filter method [51] is introduced as a basic feature selection algorithm, which ranks the input features according to the Pearson correlation coefficients. The least absolute shrinkage and selection operator (LASSO), a state-of-the-art feature selection method [52], is employed.

CI-based benchmark algorithms: Amongst CI-based algorithms, other popular population-based optimization algorithms are introduced, including four existing DE variants and three other popular algorithms. The DE variants refer to the improved DE forms for discrete problems in the previous studies, i.e., the Round-DE [23], AMDE [25] and two binDEs (i.e., Res-DE and Any-DE) [26]. Moreover, three popular population-based tools are utilized, i.e., PSO, FA and GA. It is worth noticing that for consistency, all these benchmark optimization tools use similar discrete-continuous encoding to the proposed sigmoid-based DE algorithm. In particular, the chromosomes of all CI-based optimization tools are designed the same, including the discrete part (for feature selection) and the continuous part (for weight optimization). The principles and designs of PSO, FA and GA can be found in [53], [54] and [55], respectively.

4.3 Evaluation Criteria

The main aim of stock selection models is to select promising stocks to formulate profitable portfolios with high investment returns. Therefore, this study evaluates the model performance in terms of the average return of the formulated portfolios over all testing periods:

TABLE 4
Parameters Settings of Model Optimization Algorithms

Algorithm	Parameter	Value	Algorithm	Parameter	Value
DE variants	P	30	GA	P	50
	C_r	0.5		cr	0.6
	β	0.6		mr	0.5
FA	P	30	PSO	P	40
	η	0.5		c_1	1.495
	λ	0.2		c_2	1.495
	γ	1		w	$[0.4, 0.9]$

Note: For the DE family, C_r and β respectively refer to the crossover rate and the scale factor. For GA, cr is the crossover rate and mr is the mutation rate. For FA, η refers to the weight of randomness, λ is the attractiveness factor, and γ is the absorption coefficient. For PSO, c_1 refers to the local exploration coefficient, c_2 represents the global exploration coefficient, and w is the intra weight declining from 0.9 to 0.4.

$$MR = \frac{1}{T'} \sum_{t=1}^{T'} R_{t+1}^{p}, \tag{16}$$

where R_{t+1}^{p} is the next period return of the portfolio formulated at time t, and T' is the total number of testing periods.

Moreover, investment risk should also be considered in stock selection, and a risk adjusted return, i.e., Sharpe ratio, is adopted to assess the performance of a portfolio. In particular, Sharpe ratio measures the excess return per unit deviation of a given portfolio [56]:

$$SharpeRatio = \frac{E[R_t^p - R_t^f]}{\sigma^p}, \tag{17}$$

where R_t^f is the return of risk free asset at time t, $E[R_t^p - R_t^f]$ is the expected portfolio return beyond the risk free return, and σ^p is the standard deviation of the excess return.

5 Empirical Results

The proposed model is performed to select stocks in the Shanghai A share stock market of China, and this section discusses the corresponding empirical results. The parameter settings of the proposed model and diverse benchmark models are presented in Section 5.1. For a clear discussion, the empirical results are analysed from two perspectives. First, the effectiveness of the proposed model is discussed to check whether it can obtain a significantly higher investment return than market average performance and Shanghai A share index, as presented in Section 5.2. Second, Section 5.3 further compares the proposed model with various benchmark models with other decision variable designs, fitness functions and model optimization tools. Finally, Section 5.4 summarizes the empirical study.

5.1 Parameters Settings

The parameters of the proposed method and diverse benchmark models mainly lie in model optimization algorithms. For consistency, the DE family, including the proposed sigmoid-based DE algorithm and four existing DE variants, follow a similar path in parameter settings according to [17]. As for the other three CI benchmarks, FA, GA, and PSO, the population size P is selected via the grid search method on the range of [10,60], and other parameters are specified according to the related studies (e.g.,[57], [58], [59]).

TABLE 5
Comparison Results Between the Proposed Model and the Market Performance in Terms of Portfolio Returns

Case	R_t^p	$Std.$	$R1$	$R2$	Case	R_t^p	$Std.$	$R1$	$R2$
1	10.87	0.58	9.24	6.72	16	20.22	0.84	15.56	22.10
2	−1.13	0.29	−2.40	0.57	17	4.15	0.50	−0.90	−6.31
3	15.50	1.04	13.03	10.88	18	30.04	0.63	26.35	16.43
4	45.99	1.18	36.16	25.58	19	2.93	0.62	5.62	−5.30
5	5.08	0.52	2.79	4.57	20	−25.95	1.49	−27.18	−25.98
6	13.91	0.30	10.57	42.51	21	29.01	1.43	25.18	10.12
7	50.89	1.34	50.47	17.29	22	7.71	0.99	4.94	5.53
8	26.85	1.70	18.51	18.09	23	9.55	0.89	4.94	4.19
9	45.33	1.76	33.65	37.38	24	−1.51	0.58	−7.95	−5.79
10	1.89	0.63	−0.58	−5.40	25	−15.51	0.80	−16.54	−15.78
11	−25.61	0.77	−22.31	−41.58	26	−14.25	0.94	−17.70	−7.00
12	−28.67	1.09	−35.68	−23.86	27	9.03	0.51	3.92	2.82
13	−23.44	1.15	−23.62	−17.51	28	5.18	0.65	0.07	−1.68
14	−13.24	1.88	−10.35	−23.11	29	−2.38	0.83	−7.57	−6.46
15	44.67	0.85	41.41	26.45	30	3.96	0.77	1.91	8.40

Noticeably, all algorithms will terminate when (1) Iterations reach the given maximum $G = 100$ or (2) The difference between the best objective value amongst all generations and the average objective value of the latest 15 generations can be controlled under the tolerance $tol = 10^{-5}$ [60].

The novel stock selection model and its benchmarks are performed to evaluate all sample stocks in terms of scores, and then to select m top-ranked stocks to formulate an equal-weighted portfolio. Here, m is set to 20 percent of the number of all candidate stocks [40]. Moreover, due to the randomicity stemming from initial solutions and some random parameters, all models run thirty times for each case, and the average values are calculated as the final results. The empirical study is conducted via the software $Matlab$ R2010b on a computer with CPU 2.50 GHz.

5.2 Model Effectiveness

The proposed model is employed to select stocks in the Shanghai A share market of China for formulating equal-weighted portfolios. Table 5 and Fig. 3 give the corresponding results of the portfolio returns and the accumulative returns in percentage, where R_t^p denotes the portfolio return of the proposed model in period t, and $Std.$ is the standard deviation of the proposed model (i.e., model deviation) across thirty independent runs. To check the effectiveness of the proposed model, the average return of all candidate

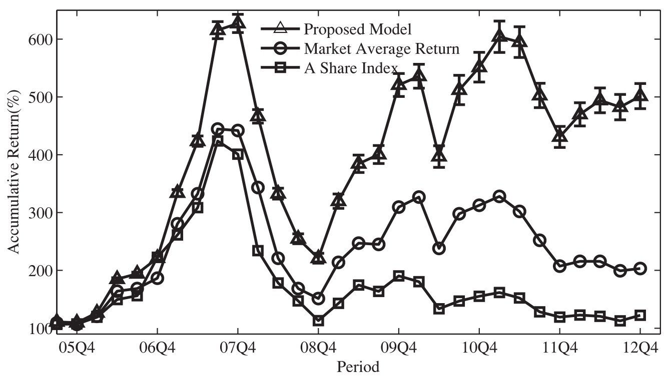


Fig. 3. Comparison results between the proposed model and the market performance in terms of accumulative returns.

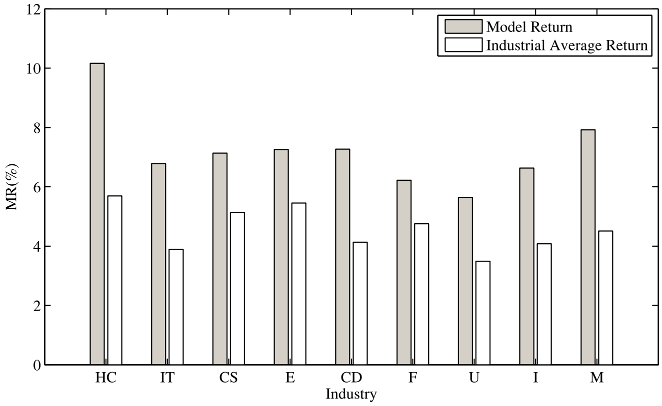


Fig. 4. Performance of the proposed model in different industries in terms of MRs.

stocks (marked as $R1$) and the return of Shanghai A share index $(R2)$ are also calculated as the market performance for comparison.

Table 5 and Fig. 3 show that the portfolios formulated by the novel method can obtain a very satisfactory result, in terms of significantly higher returns and accumulative returns compared with the market average returns $R1$ (without stock selection) and the returns of Shanghai A share index $R2$. In particular, the proposed model defeats the market average performance $R1$ in 26 out of 30 cases (accounting for approximately 86.44 percent) and the Shanghai A shares index $R2$ in 22 out of 30 cases (approximately 72.78 percent). Moreover, focusing on standard deviations of portfolio returns, the maximum dropdown of the proposed model is approximately 28.67 percent, much less than the figures for $R1$ (35.68 percent) and $R2$ (41.58 percent), implying the robustness of the proposed method. From Fig. 3, it is obvious that the proposed model significantly outperforms both the market average return and Shanghai A share index, in terms of much larger accumulative returns across all testing periods.

To verify the effectiveness of the proposed model for different industries, all sample stocks in the Shanghai A share market are divided into different sectors according to Wind Database, including health care (HC), information technology (IT), consumer staples (CS), energy (E), consumer discretionary (CD), financials (F), utilities (U), industrials (I) and materials (M). Particularly, the stocks in financial industry are also considered here. Fig. 4 displays the returns of the proposed model, together with sectoral average returns for comparison. From the results, it can be seen that the returns of the proposed model are much larger than the corresponding sectorial average returns in all cases, which repeatedly indicates the effectiveness and the robustness of the proposed model.

To explore the importance of features, Fig. 5a displays the selecting frequencies of all candidate features. It can be found that Feature 3 (price-to-sales ratio, P/S), Feature 5 (return on asset, ROA), Feature 23 (net profit growth, NPG) and Feature 25 (raw stochastic value, RSV) are the most important features for stock evaluation in the case of China's stock market, with the highest frequencies to be chosen in the proposed model. These four features analyse stocks from different perspectives. In particular, P/S reflects whether a stock is undervalued, ROA shows how profitable

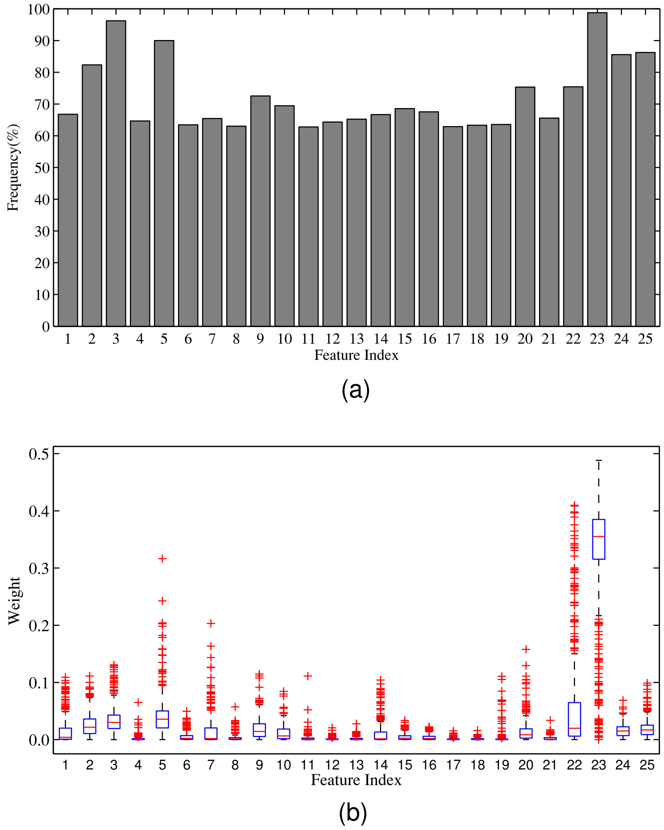


Fig. 5. Feature importance exploration in terms of selecting frequencies (a) and weights (b).

the occupied assets of a company are, NPG gives a good picture of stock profits, and RSV compares the current stock price to its short-term range. Moreover, Fig. 5b provides the box plots of feature weights. Similarly, NPG is tested as the most important feature, with a relatively larger average weight (approximately 0.674). The results further imply that the movement of a stock price is closely dependent on its profit growth, and the stock with a higher NPG value is prone to higher potential of price increase.

5.3 Superiority over Benchmark Models

To statistically prove the superiority of the proposed model, a set of benchmark models are performed for comparison, with other decision variable designs, fitness functions, and model optimization algorithms.

5.3.1 Comparisons with Benchmark Decision Variable Designs

To test the effectiveness of the model design for decision variables, a series of benchmarks with other decision variable designs have been formulated (see Table 3). It is worth noticing that except decision variable design, other parts of the models (including fitness function and optimization tool) follow a similar path to the proposed model. Table 6 reports the corresponding comparison results, in which model performance is evaluated via MR.

From Table 6, the proposed model (A0) can be shown to be the best model with the highest MR and the least model deviations (marked as *Std.*) for both training and testing periods. The effectiveness of the proposed model implies

TABLE 6
Comparison Results Amongst Stock Selection Models with Different Decision Variable Designs

Model	Training Period		Testing Period				Statistical Tests (p-value)	
	$MR(\%)$	$Std.(\%)$	$MR(\%)$	$Std.(\%)$	$SharpeRatio(\%)$	$Std.(\%)$	Normality test	One-tailed t test
A0	11.48	0.06	7.70	0.14	31.66	0.68	0.424	N.A.
A1	N.A.	N.A.	6.51	0.00	19.82	0.00	N.A.	<0.001
A2	9.99	0.06	6.59	0.24	20.33	0.20	0.500	<0.001
A3	10.02	0.07	6.55	0.25	20.42	0.28	0.305	<0.001
A4	11.00	0.08	7.10	0.30	21.38	0.42	0.500	<0.001
$R1$	N.A.	N.A.	4.39	0.00	17.70	0.00	N.A.	<0.001
$R2$	N.A.	N.A.	2.46	0.00	9.14	0.00	N.A.	<0.001

that both feature selections and the corresponding impacts (in terms of weights) are the most important decision variables in stock evaluation, which should not be neglected. Amongst benchmark models, Model A4 considering all decision variables of feature selections, directional indicators and weights ranks the best, whereas Model A1 with no decision variable is the poorest model, which highlights the importance of optimization process in stock selection. When comparing Benchmarks A1 and A2, the results of MR and $SharpeRatio$ indicate that adding the variables of feature selections can largely improve the performance of stock selection models. However, the comparison result of MR between Models A2 and A4 (or Models A0 and A4) implies that the variables of directional indicators might be otherwise unnecessary. The main reason lies in that the directional indicator for a given feature is relatively stable and invariable in the short term, and there is no need to tune it too often. Finally, the results also show that the variables of feature weights are indispensable, since the MRs of Model A4 are much higher than those of Model A3. These above results further arrive at the conclusion that both feature selections and the corresponding weights are significantly crucial factors in stock selection models. Unsurprisingly, the proposed model using both of them performs the best of all.

Furthermore, a statistical test is conducted on the superiority of the proposed model (A0) over the benchmark models, via a one-tailed t test with the null hypothesis that the MRs of the proposed model are no larger than those of other benchmarks. First, a normality test is conducted for all MRs of different models with the null hypothesis that the MRs follow a normal distribution, as the results listed in Table 6. And the results show that all MR data can pass the test at the confidence level of $99\,\text{percent}$, which meets the basic assumption of t test [61]. The results of t test are listed in the last column of Table 6, and the superiority of the proposed model over all benchmarks with other decision variable designs can be statistically confirmed under the confidence level of $99\,\text{percent}$, with all p-values far below $1\,\text{percent}$.

5.3.2 Comparisons with Benchmark Fitness Functions

First, the fitness functions with different penalties on the number of selected features (see (11)) are investigated. Fig. 6 illustrates the impacts of the penalty parameter α on the number of selected features and portfolio return. From Fig. 6a, it can be obviously seen that the total number of selected features monotonically decreases as the penalty α increases. However, the penalty parameter α has a negative impact on the MR value, as shown in Fig. 6b. A linear regression analysis is conducted to capture the relationship between penalties and average returns, and the result presents an obvious negative influence of α on MR, in terms of slope (-0.026), R^2 (0.880) and F-value (65.814). The results also imply that the proposed method with $\alpha = 0$ can not only produce an informative feature subset but also obtain the most satisfactory results in terms of investment returns, defeating all other benchmark models with positive penalties.

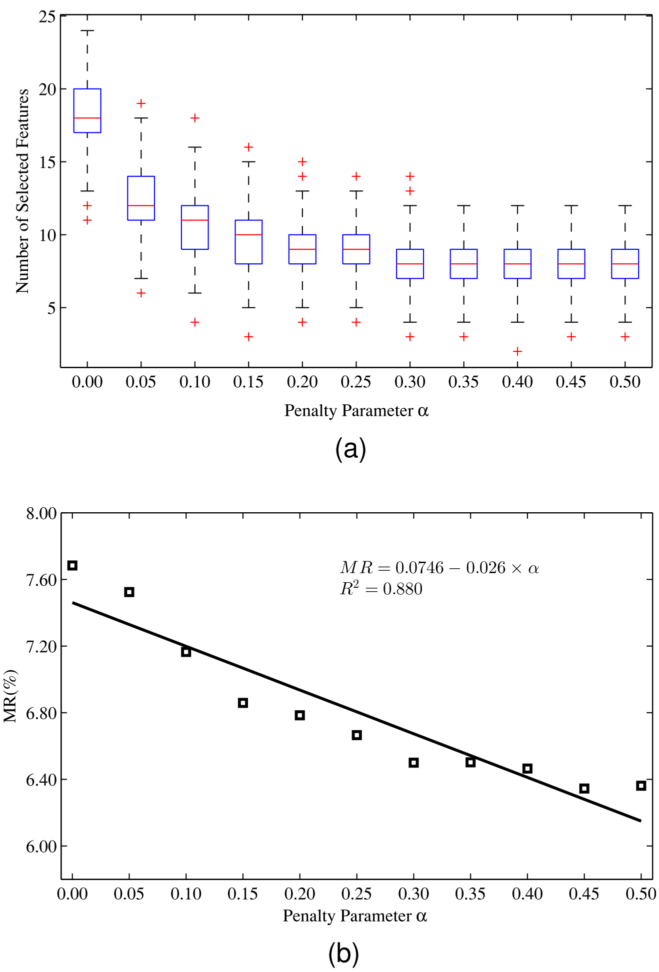


Fig. 6. Influence of penalty on the total number of selected features (a) and model performance in terms of MR (b).

TABLE 7
Comparison Results Amongst Stock Selection Models with Different Fitness Functions

Panel A: Model Performance					
Model	*IC*	*SPREAD*	*IFHR*	*CR*	*MV*
$MR(\%)$	7.70	7.47	7.22	7.21	7.15
	(0.14)	(0.29)	(0.21)	(0.22)	(0.27)
$SharpeRatio(\%)$	31.66	31.83	30.96	30.41	21.76
	(0.68)	(1.11)	(0.93)	(0.92)	(0.39)
Panel B: Statistical Tests					
Model	*IC*	*SPREAD*	*IFHR*	*CR*	*MV*
Normality (p-value)	0.42	0.07	0.50	0.38	0.50
One-tailed t test	t-value	10.24	3.83	10.53	9.81
H0:$MR_{IC} \leq MR_B$	p-value	< 0.001	< 0.001	< 0.001	< 0.001

Note: MR_{IC} indicates the MR of the proposed model with IC as the fitness function, and MR_B is the MR of the corresponding benchmark model with another fitness function.

Second, other popular fitness functions are also introduced, i.e., *IFHR*, *SPREAD*, *CR* and *MV* (see (12)-(15), respectively). Panel A of Table 7 displays the performance of different models with different fitness functions in terms of MR, $SharpeRatio$ and the corresponding standard deviations (in bracket). The results show that the proposed model with the *IC* fitness function performs the best, in terms of the highest MR and $SharpeRatio$ and the least corresponding deviations. Moreover, all stock selection models with different fitness functions can obtain higher returns than the market average return without stock selection $(R1)$ and Shanghai A share index $(R2)$.

To statistically prove the superiority of the *IC* fitness function, one-tailed t tests are also conducted, with the null hypothesis of the inferiority of the proposed model to other benchmarks. Panel B of Table 7 reports the results of statistical tests. First, the normality tests show that all MR data follow a normal distribution at the significance level of $1\,\mathrm{percent}$. Furthermore, according to the results of one-tailed t tests, the superiority of the proposed model can be statistically proved under the confidence level of $99\,\mathrm{percent}$. In particular, the proposed model with the *IC* fitness function can be proved statistically better than all benchmarks with other popular fitness functions, as its p-values of one-tailed t tests against the corresponding counterparts are all far less than $1\,\mathrm{percent}$.

5.3.3 Comparison with Benchmark Optimization Algorithms

In this section, the proposed sigmoid-DE method is compared with other optimization algorithms, including typical feature selection algorithms and CI-based optimization algorithms. As for typical feature selection methods, the Pearson correlation method (a baseline) and LASSO (a state-of-the-art algorithm) are performed. The CI-based benchmark models are diverse DE-variants (e.g., Round-DE, AMDE, Res-DE and Any-DE) and other popular optimization algorithms (e.g., FA, PSO and GA).

Table 8 reports the comparison results between the proposed sigmoid-based DE model (Sig-DE) and the two typical feature selection models of Pearson correlation method

TABLE 8
Comparison Results between the Proposed Model and Linear Feature Selection Algorithms

Model	Sig-DE	Pearson	LASSO	$R1$	$R2$
$MR(\%)$	7.70	6.52	6.64	4.39	2.46
$SharpRatio(\%)$	31.66	28.96	27.99	17.70	9.14

(Pearson) and LASSO, in terms of MR and $SharpRatio$. It can be obviously found from the table that the MR and $SharpRatio$ values of the novel sigmoid-based DE are much higher than those of the Pearson correlation method and the LASSO method, which demonstrates that the proposed sigmoid-based DE performs much better than these two typical feature selection models in stock selection.

As for various CI-based benchmarks, Table 9 reports the comparison results, in which $Max.$ and $Min.$ refer to the maximum and the minimum returns of the portfolios in all cases, respectively, $Prob.(R1)$ and $Prob.(R2)$ are the ratios of cases with higher returns respectively than $R1$ and $R2$ in total cases, and $HitRate$ is the ratio of the cases with positive returns [3]. Both the corresponding average values and standard deviations (in bracket) are listed, and the best results are highlighted. As shown in Table 9, the proposed model with the sigmoid-based DE algorithm appears its distinct effectiveness and robustness, in terms of the highest MR and $SharpeRatio$ and the least deviations.

Compared with DE variants, the proposed sigmoid-based DE algorithm performs the best in terms of the highest MR, $SharpeRatio$, $Prob.(R1)$ and $HitRatio$. Furthermore, the results of one-tailed t tests with the null hypothesis that the MR of the proposed model is not higher than those of benchmarks further confirm the superiority of the proposed sigmoid-based DE algorithm, since that all p-values are far less than the significance level of $1\,\mathrm{percent}$ except Round-DE (see Panel B in Table 9). Actually, though the proposed sigmoid-based DE algorithm and Round-DE can fall into one category using the posterior conversion operators both following statistical probability distributions, the former does outperform the latter in terms of MR and $SharpRatio$, although such superiority is not so significant as those to other DE variants.

The main reasons for the superiority of the novel sigmoid-based DE algorithm can be generally summarized into the three following aspects. First, different from Res-DE and Any-DE, the novel sigmoid-based DE algorithm remains the mutation operator of the original DE form, which effectively guarantees the diversity of population. Second, different from AMDE using a searching process with an unnecessary great computational burden, the sigmoid-based DE algorithm resorts to a simple but effective way directly manipulating the bit-strings. Third, when comparing the two similar DE variants using the posterior conversion operators both based on statistical probability distributions, the novel algorithm with the most popular approach in binary classification (i.e., the sigmoid model following a logistic probability distribution) defeats Round-DE with a relatively simple conversion process (rounding the continuous value $x_{p,d,g} \in [0,1]$ to the nearest integer 0 or 1 based on a Bernoulli distribution). Therefore, the proposed sigmoid-based DE algorithm with the simple but

TABLE 9
Comparison Results Amongst Stock Selection Models with Different Model Optimization Algorithms

Panel A: Model Performance								
Model	Sig-DE	Round-DE	AMDE	Res-DE	Any-DE	FA	GA	PSO
$MR(\%)$	**7.70**	7.67	6.90	6.97	7.00	7.23	7.17	7.62
	(0.14)	(0.20)	(0.32)	(0.26)	(0.22)	(0.41)	(0.18)	(0.22)
$SharpeRatio(\%)$	**31.66**	31.26	29.54	29.60	29.67	30.45	30.49	31.50
	(0.68)	(0.79)	(1.32)	(1.13)	(0.81)	(1.73)	(0.78)	(0.88)
$Max.(\%)$	50.89	**51.00**	46.20	46.05	46.69	49.95	46.29	50.08
	(1.34)	(1.58)	(3.30)	(2.04)	(2.16)	(3.21)	**(0.99)**	(1.64)
$Min.(\%)$	−28.72	−31.84	−27.74	−27.42	−27.67	−28.56	**−27.21**	−28.59
	(1.02)	(0.77)	(1.23)	(0.96)	(0.85)	(1.72)	**(0.59)**	(1.09)
$Prob.(R1, \%)$	**86.44**	86.11	84.22	84.78	84.22	86.13	84.33	86.11
	(3.27)	(2.78)	(3.60)	(2.99)	(2.89)	(4.17)	**(2.50)**	(3.04)
$Prob.(R2, \%)$	72.78	72.78	73.22	72.78	72.00	**74.33**	72.78	73.44
	(2.33)	(1.97)	(2.55)	(2.49)	**(1.88)**	(2.63)	(1.54)	(2.70)
$HitRatio(\%)$	**66.67**	66.56	66.67	66.67	66.67	66.40	66.67	66.67
	(0.00)	(0.61)	(0.00)	(0.00)	(0.00)	(1.13)	(0.00)	(0.00)
Panel B: Statistical Test								
Model	Sig-DE	Round-DE	AMDE	Res-DE	Any-DE	FA	GA	PSO
Normality test (p-value)	0.135	0.424	0.184	0.500	0.500	0.298	0.224	0.012
One-tailed t tests	t-value	0.873	12.32	12.83	16.05	6.41	13.55	1.54
H0:$MR_S \leq MR_B$	p-value	0.195	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.067

Note: MR_S indicates the MR of the proposed model with the sigmoid-based DE algorithm as the optimization algorithm, and MR_B is the MR of the corresponding benchmark model with another optimization algorithm.

efficient binary conversion method of sigmoid model outperforms other existing DE variants for discrete or mixed discrete-continuous problems.

Comparing with other three popular CI tools of FA, GA and PSO, the sigmoid-based DE algorithm also performs the best in terms of the highest MR, $SharpeRatio$, $Max.$, $Prob.(R1)$ and $HitRatio$. GA ranks the best in terms of Min and FA performs the best in terms of $Prob.(R2)$. A one-tailed t test is also performed, as the results listed in Panel B of Table 9. The corresponding results statistically confirm that the novel method with the sigmoid-based DE algorithm can be statistically proved to be better than most benchmarks (except the model with PSO) in terms of MR, under the confidence level of 99 percent. Das and Suganthan [18] similarly observed the effectiveness of the DE algorithm compared with PSO and GA.

Generally, all above mentioned results demonstrate that the proposed model with the novel sigmoid-based DE algorithm for model optimization can achieve a significantly better performance than the considered benchmarks with both the existing modified DE variants for discrete or mixed discrete-continuous problems and popular CI algorithms.

5.4 Summarizations

According to above analyses, it can be concluded that the proposed stock selection model with the sigmoid-based DE algorithm can be statistically proved to be significantly more powerful and efficient than other designed benchmarks, in terms of investment returns and model robustness.

On the one hand, the portfolio formulated by the proposed stock selection model can obtain much higher returns than the market average performance (i.e., the portfolio based on all candidate stocks without selection) and Shanghai A share index in the Chinese stock market. Moreover, such effectiveness can be also confirmed for different sectorial subsets.

On the other hand, by comparing with series of benchmark models with different model designs and model optimization techniques, statistical tests further prove the superiority of the novel model under the confidence level of 99 percent. First, for decision variable design, the proposed model considering feature selection and weight optimization greatly outperforms all benchmark models with other decision variable designs. Second, for fitness functions, the proposed model with *IC* can be statistically proved to be much better than all benchmarks with penalties on number of selected features or other popular fitness functions. Finally, for model optimization, the proposed method with the sigmoid-based DE algorithm defeats the benchmarks with linear feature selection models, other existing DE variants (modified for discrete or mixed discrete-continuous problems) and other popular CI optimization algorithms (FA, PSO and GA).

6 Conclusions

This paper proposes a novel stock selection model with discrete and continuous variables, i.e., feature selection and weight optimization, in which the traditional DE algorithm is introduced and extended to a sigmoid-based DE algorithm for this mixed discrete-continuous problem. Compared with the existing DE variants for discrete or mixed discrete-continuous optimization, the novel sigmoid-based DE algorithm makes contributions from two main perspectives. First, it might be the first try to introduce the simple but efficient sigmoid-based conversion into the traditional DE algorithm for mixed discrete-continuous optimization. Second, this novel sigmoid-based DE algorithm is then incorporated into the proposed stock selection model for feature selection and weight optimization, which finely verifies the effectiveness of the novel mixed discrete-continuous DE variant in stock selection, different from the previous studies based on simple numerical examples.

Using the Shanghai A share market of China as the study sample, the empirical results indicate that the proposed stock selection model can be used as a very powerful and efficient tool for stock selection. First, the proposed stock selection model can obtain much higher returns than the market average performance, for both the whole market and different industries. Furthermore, statistical tests prove the superiority of the novel model to benchmark models with different model designs and model optimization techniques, under the confidence level of 99 percent. Particularly, the proposed method with the sigmoid-based DE algorithm defeats the benchmarks with typical linear feature selection model (e.g., Pearson and LASSO), other existing DE variants (modified for discrete or mixed discrete-continuous problems) and other popular CI optimization algorithms (e.g., FA, PSO and GA).

However, the proposed stock selection model can be further improved from the following four perspectives. First, by introducing some other important objectives, the proposed model can be extended into multiple objective models to provide different satisfactory portfolios according to different goals. For instance, investment risk is another essential issue in stock selection, which can be also considered in the proposed model. Second, stock market timing is also a crucial task in stock investment, and the model can be improved not only to select promising stocks but also to give helpful advices for the buying and selling points. Third, besides the data splitting strategy used in this study, other popular strategies (e.g., the regular sliding window procedure with a fixed training sample size) can be also employed to test the robustness of the proposed model. Finally but the most importantly, the proposed model should be extended to other capital markets for further testing its generalizations. We will look into these interesting issues in the near future.

Acknowledgments

This work is supported by grants from the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC No. 71301006 and 71433001), the National Program for Support of Top-Notch Young Professionals, and the Fundamental Research Funds for the Central Universities in BUCT. L. Tang is the corresponding author.

References

[1] Y. L. Becker, H. Fox, and P. Fei, "An empirical study of multi-objective algorithms for stock ranking," in *Genetic Programming Theory and Practice V*, R. Riolo, T. Soule, and B. Worzel, Eds. New York, NY, USA: Springer, 2008, pp. 239–259.
[2] C. F. Huang, T. N. Hsieh, B. R. Chang, and C. H. Chang, "A comparative study of stock scoring using regression and genetic-based linear models," in *Proc. IEEE Int. Conf. Granular Comput.*, Nov. 2011, pp. 268–273.
[3] Y. L. Becker, P. Fei, and A. Lester, "Stock selection: An innovative application of genetic programming methodology," in *Genetic Programming Theory and Practice IV*, R. Riolo, T. Soule, and B. Worzel, Eds. New York, NY, USA: Springer, 2007, pp. 315–334.
[4] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.
[5] S. A. Ross, "The arbitrage theory of capital asset pricing," *J. Econ. Theory*, vol. 13, no. 3, pp. 341–360, Dec. 1976.
[6] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *J. Financial Econ.*, vol. 33, no. 1, pp. 3–56, Feb. 1993.
[7] L. Yu, S. Wang, and K. K. Lai, "Neural network-based mean–variance–skewness model for portfolio selection," *Comput. Oper. Res.*, vol. 35, no. 1, pp. 34–46, 2008.
[8] L. Wang and S. Gupta, "Neural networks and wavelet de-noising for stock trading and prediction," in *Time Series Analysis, Modeling and Applications*, W. Pedrycz and S. M. Chen, Eds. Berlin Germany: Springer, 2013, pp. 229–247.
[9] L. Yu, S. Wang, and K. K. Lai, " Mining stock market tendency using ga-based support vector machines," in *Proc. 1st Int. Workshop Internet Netw. Econ.*, 2005, pp. 336–345.
[10] T. Takahama, S. Sakai, A. Hara, and N. Iwane, "Predicting stock price using neural networks optimized by differential evolution with degeneration," *Int. J. Innovative Comput. Inform. Control*, vol. 5, no. 12, pp. 5021–5031, 2009.
[11] N. Hachicha, B. Jarboui, and P. Siarry, "A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics," *Inform. Sci.*, vol. 181, no. 1, pp. 79–91, 2011.
[12] J. Wang, M. Wang, P. Li, L. Liu, Z. Zhao, X. Hu, and X. Wu, "Online feature selection with group structure analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3029–3041, Nov. 2015.
[13] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
[14] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learning Res.*, vol. 3, pp. 1157–1182, 2003.
[15] K. Javed, H. Babri, M. Saeed, et al., "Feature selection based on class-dependent densities for high-dimensional binary data," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 465–477, Mar. 2012.
[16] W. Sheng, X. Liu, and M. Fairhurst, "A niching memetic algorithm for simultaneous clustering and feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 7, pp. 868–879, Jul. 2008.
[17] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Berlin, Germany: Springer, 2006.
[18] S. Das and P. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Trans. Evol. Comput.*, vol. 15, no. 1, pp. 4–31, Feb. 2011.
[19] R. M. Alguliev, R. M. Aliguliyev, and C. A. Mehdiyev, "Sentence selection for generic document summarization using an adaptive differential evolution algorithm," *Swarm Evol. Comput.*, vol. 1, no. 4, pp. 213–222, 2011.
[20] J. Tvrdík and I. Křivỳ, "Differential evolution with competing strategies applied to partitional clustering," in *Swarm and Evolutionary Computation*. New York, NY, USA: Springer, 2012, pp. 136–144.
[21] S. Das, A. Abraham, and A. Konar, "Automatic clustering using an improved differential evolution algorithm," *IEEE Trans. Syst. Man Cybern. Part A, Syst. Humans*, vol. 38, no. 1, pp. 218–237, Jan. 2008.
[22] D. Datta and J. R. Figueira, "A real–integer–discrete-coded differential evolution," *Appl. Soft Comput.*, vol. 13, no. 9, pp. 3884–3893, 2013.
[23] J. Lampinen and I. Zelinka, "Mixed integer-discrete-continuous optimization by differential evolution," in *Proc. 5th Int. Conf. Soft Comput.*, 1999, pp. 71–76.
[24] R. Angira and B. Babu, "Optimization of process synthesis and design problems: A modified differential evolution approach," *Chem. Eng. Sci.*, vol. 61, no. 14, pp. 4707–4721, 2006.
[25] G. Pampara, A. P. Engelbrecht, and N. Franken, "Binary differential evolution," in *Proc. IEEE Congr. Evol. Comput.*, 2006, pp. 1873–1879.
[26] T. Gong and A. Tuson, "Differential evolution for binary encoding," in *Soft Computing in Industrial Applications*, A. Saad, K. Dahal, M. Sarfraz, and R. Roy, Eds. New York, NY, USA: Springer, 2007, pp. 251–262.
[27] C. Deng, B. Zhao, Y. Yang, and A. Deng, "Novel binary differential evolution without scale factor F," in *Proc. 3rd IEEE Int. Workshop Adv. Comput. Intell.*, 2010, pp. 250–253.
[28] L. Wang, X. Fu, M. I. Menhas, and M. Fei, "A modified binary differential evolution algorithm," in *Proc. Int. Conf. Life Syst. Modeling Intell. Comput.*, 2010, pp. 49–57.

[29] S. Patra, S. Goswami, and B. Goswami, "A binary differential evolution algorithm for transmission and voltage constrained unit commitment," in *Proc. Joint Int. Conf. Power Syst. Technol. IEEE Power India Conf.*, 2008, pp. 1–8.

[30] R. Balamurugan and S. Subramanian, "Hybrid integer coded differential evolution-dynamic programming approach for economic load dispatch with multiple fuel options," *Energy Convers. Manage.*, vol. 49, no. 4, pp. 608–614, 2008.

[31] C. S. Deng, B. Y. Zhao, A. Y. Deng, and C. Y. Liang, "Hybrid-coding binary differential evolution algorithm with application to 0-1 knapsack problems," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Dec. 2008, pp. 317–320.

[32] Y. C. Lin, K. S. Hwang, and F. S. Wang, "A mixed-coding scheme of evolutionary algorithms to solve mixed-integer nonlinear programming problems," *Comput. Math. Appl.*, vol. 47, no. 8, pp. 1295–1307, 2004.

[33] H. K. Kim, J. K. Chong, K. Y. Park, D. Lowther, et al., "Differential evolution strategy for constrained global optimization and application to practical engineering problems," *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1565–1568, Apr. 2007.

[34] J. Krause, J. Cordeiro, R. S. Parpinelli, and H. S. Lopes, "A survey of swarm algorithms applied to discrete optimization problems," *Swarm Intelligence and Bio-inspired Computation: Theory and Applications*, Philadelphia, PA, USA: Elsevier, 2013, pp. 169–191.

[35] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst. Man Cybern. Comput. Cybern. Simul.*, 1997, pp. 4104–4108.

[36] S. Palit, S. Sinha, M. Molla, A. Khanra, and M. Kule, "A cryptanalytic attack on the knapsack cryptosystem using binary firefly algorithm," in *Proc. 2nd Int. Conf. Comput. Commun. Technol.*, Sep. 2011, pp. 428–432.

[37] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using z score transformation," *J. Molecular Diagnostics*, vol. 5, no. 2, pp. 73–81, May. 2003.

[38] S. H. Penman and S. H. Penman, *Financial Statement Analysis and Security Valuation*. New York, NY, USA: McGraw-Hill, 2007.

[39] C. F. Huang, B. R. Chang, D. W. Cheng, and C. H. Chang, "Feature selection and parameter optimization of a fuzzy-based stock selection model using genetic algorithms," *Int. J. Fuzzy Syst.*, vol. 14, no. 1, pp. 65–75, 2012.

[40] R. Tortoriello, *Quantitative Strategies for Achieving Alpha*. New York, NY, USA: McGraw-Hill, 2009.

[41] X. Panos, M. George, K. Theodore, P. John, and C. Zopounidis, *Multicriteria Portfolio Management*. New York, NY, USA: Springer, 2012.

[42] J. N. Carpenter, F. Lu, and R. F. Whitelaw, "The real value of China's stock market," Center for Global Economy and Business Coordinator, Stern China Initiative, Working paper, Apr. 2014.

[43] B. Tian, Q. Gong, Y. Yang, Z. Shang, and Y. Feng, " Stock selection using support vector machine in Chinese securities exchange," *J. Harbin Inst. Technol.*, vol. 14, no. 3, pp. 378–384, 2007.

[44] L. Tan, T. C. Chiang, J. R. Mason, and E. Nelling, "Herding behavior in Chinese stock markets: An examination of A and B shares," *Pacific-Basin Finance J.*, vol. 16, no. 1-2, pp. 61–77, 2008.

[45] X.-W. Ai, T. Hu, X. Li, and H. Xiong, "Clustering high-frequency stock data for trading volatility analysis," in *Proc. IEEE 9th Int. Conf. Mach. Learning Appl.*, 2010, pp. 333–338.

[46] C. A. Vitt and H. Xiong, "The impact of patent activities on stock dynamics in the high-tech sector," in *Proc. IEEE Int. Conf. Data Mining*, 2015, pp. 399–408.

[47] C. F. Tsai and Y. C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decision Support Syst.*, vol. 50, no. 1, pp. 258–269, 2010.

[48] W. Fan, E. A. Fox, P. Pathak, and H. Wu, "The effects of fitness functions on genetic programming-based ranking discovery for web search," *J. Amer. Soc. Inform. Sci. Technol.*, vol. 55, no. 7, pp. 628–636, May 2004.

[49] M. Kuhn and K. Johnson, "An introduction to feature selection," in *Applied Predictive Modeling*, New York, NY, USA: Springer, 2013, pp. 487–519.

[50] H. Markowitz, "Portfolio selection*," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[51] I. Guyon, "Practical feature selection from correlation to causality," in *Mining Massive Data Sets for Security*, Amsterdam, The Netherlands: IOS Press, 2008, pp. 27–44.

[52] N. Meinshausen and P. Bühlmann, "High-dimensional graphs and variable selection with the lasso," *Ann. Statist.*, vol. 34, pp. 1436–1462, 2006.

[53] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4, 1995, pp. 1942–1948.

[54] X. S. Yang, *Nature-Inspired Metaheuristic Algorithms*. Frome, BA11 6TT, U. K., Luniver Press, 2010.

[55] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor, MI, USA: Univ. Michigan Press, 1975.

[56] A. W. Lo, "The statistics of sharpe ratios," *Financial Analysts J.*, vol. 58, no. 4, pp. 36–52, 2002.

[57] A. Chipperfield, P. Fleming, H. Pohlheim, and C. Fonseca, *Genetic Algorithm Toolbox for Use with Matlab*, 1st ed., Sheffield, South Yorkshire, U. K.: Department of Control and Systems Engineering, University of Sheffield, 1994.

[58] A. H. Gandomi, X. S. Yang, and A. H. Alavi, "Mixed variable structural optimization using firefly algorithm," *Comput. Struct.*, vol. 89, no. 23, pp. 2325–2336, 2011.

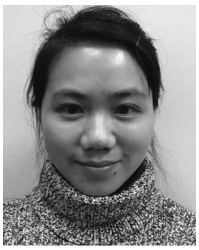
[59] C. X. Guo, J. S. Hu, B. Ye, and Y. J. Cao, "Swarm intelligence for mixed-variable design optimization," *J. Zhejiang University Sci.*, vol. 5, no. 7, pp. 851–860, 2004.

[60] B. J. Jain, H. Pohlheim, and J. Wegener, "On termination criteria of evolutionary algorithms," in *The Genetic and Evolutionary Computation Conference*, San Mateo, CA, USA: Morgan Kaufmann, 2001, pp. 768–776.

[61] J. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA, USA: Duxbury Press, 1995.

Lean Yu received the PhD degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a professor in the School of Economics and Management, Beijing University of Chemical Technology, Beijing, China. His research interests include big data mining, business intelligence, safety and emergency management, decision support systems, and knowledge science and engineering. He is a member of the IEEE.

Lunchao Hu received the BS degree from Beijing Normal University, Beijing, China, in 2012. He received the MS degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2015. His research interests include evolutionary computation and financial market modeling.



Ling Tang received the PhD degree from the Institute of Policy and Management, Chinese Academy of Sciences, Beijing, China, in 2012. She is currently an associate professor in the School of Economics and Management, Beijing University of Chemical Technology, Beijing, China. Her research interests include business intelligence, data mining, computer simulation, and energy economics.