International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 2, Mar – Apr 2017

RESEARCH ARTICLE OPEN ACCESS

Stock Selection using Principal Component Analysis with Differential Evolution

Dr. Balamurugan .A [1], Arul Selvi. S [2], Syedhussian .A [3], Nithin .A [4]

Professor [1], Assistant Professor [2], UG Scholar [3] & [4]

Department of Computer Science and Engineering

Sri Krishna College of Technology, Coimbatore

Tamil Nadu - India

ABSTRACT

A stock selection model with both discrete and continuous resolution variables is proposed, in which a novel sigmoid-based mixed discrete-continuous differential evolution algorithm is used 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 contrive 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 analogous weights (continuous variables), the original differential evolution algorithm focusing only on continuous problems is enlarged 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 synthetic dataset of share market as the study sample, the results show that the novel stock selection model can make a profitable portfolio and significantly outperform in terms of investment return.

Keywords:- Artificial intelligence, constrained optimization, evolutionary computing, portfolio analysis

I. INTRODUCTION

Quantitative asset management involves a set of processes, i.e., ideas proposal, returns forecast, portfolios construction and performance evaluation. Amongst them, stock selection for further portfolio formulation may be one of the most crucial but demanding issues, due to the complexity of financial markets. Traditional statistical regression models are relatively easy to implement and understand due to their simple forms, nonetheless they often appear relatively poor performance 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 organizational risk minimization and were employed to model stock markets. For model optimization, the DE algorithm, a typical evolutionary algorithm (EA), has widely been applied to financial market analysis. These above studies all demonstrated that the CI techniques significantly outperformed the traditional statistical regression approaches in modeling financial markets. 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 assess 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 suitable features (discrete decision variables) and optimizing the corresponding weights (continuous decision variables), the powerful CI resource 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 suggest a stock selection model with a novel sigmoid-based DE algorithm for the mixed discrete-continuous optimization, and to verify its supremacy over benchmark models with other model designs (in terms of different decision variables and fitness functions) and other popular optimization techniques.

The literature review of the project about existing, proposed techniques are discussed in Chapter 2. Chapter 3 is being discuss about the design methodology and modules present in the proposed system. The implementation of each module can be referred in Chapter 4. The classification results are discussed in Chapter 5. Chapter 6 provides the overall conclusion about the project and also discusses about the future scope of the project.

II. LITERATURE REVIEW

Neural networks are used to forecast the future stock prices and develop a suitable trading system. Wavelet analysis is used to de-noise the time series and the results are compared with the raw time series prediction without wavelet de-noising. Quality and Poor 500 (S&P 500) is used in experiments. In this paper use a gradual data sub-sampling technique, i.e., training the network mostly with recent data, but without abandon past data. In addition, effects of NASDAQ 100 are studied on prediction of S&P 500. A

daily trading strategy is employed to buy/sell according to the predicted prices and to calculate the directional effectiveness and the rate of returns for different periods. The purpose of this paper is to examine rigorously the arbitrage model of capital asset pricing developed in Ross. The arbitrage model was proposed as an alternative to the mean variance capital asset pricing model, introduced by Sharpe, Lintner, and Treynor, that has become the major analytic tool for explaining phenomena notice in capital markets for risky assets.

2.1 Novel Stock Selection Model

The quantitative asset management involves a set of processes, i.e., ideas proposal, returns forecast, portfolios construction and performance evaluation. Amongst them, stock selection for further portfolio formulation may be one of the most crucial but demanding issues, due to the complexity of financial markets. Traditional statistical regression models are relatively easy to instrument and understand due to their simple forms, nevertheless they often appear relatively poor performance despite of the wide application, ANNs often suffer from over-fitting and local best problems. To avoid such problems to some degree, SVMs were proposed based on the principle of constitutional risk minimization and were employed to model stock markets. For model optimization, the DE algorithm, a typical evolutionary algorithm (EA), has far apart been applied to financial market analysis. In particular, two main steps are involved in this novel stock selection model: stock scoring and stock ranking. First, a stock scoring apparatus is designed, in which stocks are evaluated based on various fundamental and technical features. Second, the top-ranked stocks are selected to compose an equal-weighted portfolio as the model output. For choosing proper features (discrete decision variables) [2] and optimizing the corresponding weights (continuous decision variables), the powerful CI resource technique of DE is especially introduced and improved to a novel mixed discrete-continuous variant with sigmoid-based transmutation for the discrete part, i.e., the novel sigmoid-based DE algorithm. The main aim of this study is to propose a stock most suitable model with a novel sigmoid-based DE algorithm for the mixed separate-continuous optimization, and to verify its advantage over benchmark models with other model designs (in terms of different decision variables and fitness functions) and other popular optimization techniques.

2.2 Predicting Stock Price Using Neural Networks Optimized by Differential Evolution with Degeneration

Structural learning, in which the structures of estimation systems are optimized, has been actively studied in researches on supervised learning of neural networks and fuzzy rules. GAd(Genetic Algorithm with degeneration)[6] is the structural learning methods, which are modeled on genetic harm and degeneration. In the algorithms, a gene is defined by a pair of a normal value and a damaged charge that shows how much the gene is damaged. Simple one-point crossover and Gaussian mutation are adopted to deal with the pair. However, it was very difficult to incorporate more efficient crossover working than one-point crossover, because the pair of the value and the rate must be treated. Recently, a new evolutionary algorithm, Differential Evolution (DE), has been proposed and successfully applied to the optimization problems including non-linear, non-differentiable, non-convex and multi-modal functions. The next chapter is discussing about the design methodology and modules present in the proposed system.

III. IMPLEMENTATION

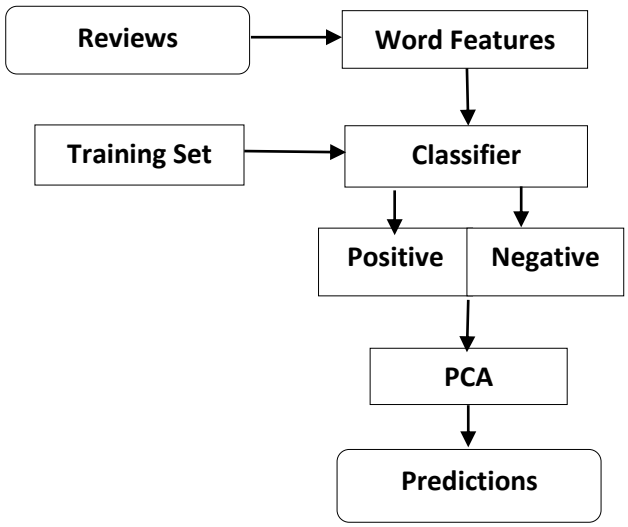

Fig. 2.1: Flow diagram

3.1 Dataset Processing

In this module a synthesis stock market dataset for performing for the processes mentioned in the following modules are built. This module contains high dimensional data as a synthesis dataset as its contains additional information with several attributes along huge records in difference time factors to analyze for providing accurate predictions in future cases.

3.2 Designing of Stock Selection Model

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. However, standard measures such as Euclidean distance is the most common use of distance, inspect the root of square differences between coordinates of a pair of objects.

3.3 Designing of Stock Ranking Objects

Here the homogeneous tensor with the utilities of the objects to calculate the probability of each value of the data to be clustered with the Centroid is used. After calculating the probabilities of the values, we binarize the values that have high probabilities. Let denote the score of stock is assigned by feature j at time t, i.e., the Z-score normalization. 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.

3.4 Establishing The Forecasting Technique to Optimal Centroids

The need of analyzing and grouping of data is required for better understanding and inspection. This can be solved by using the clustering technique which groups the similar kind data into a particular cluster. One of the most commonly and widely used clustering is K-Means congregate because of its simplicity and performance.

3.5 Selection of Dimensions Through Actionable Weight Using Principle Component Analysis

If the dataset used is large, then the performance will be reduced and also the time complexity is increased. To overcome this problem, this method focuses on altering the initial cluster Centroid explicitly, for this purpose; Principal Component Analysis (PCA) is nearly new here. Principal component analysis (PCA) is a widely used statistical technique for unsupervised dimension reduction. Clustering is a commonly used data clustering for unsupervised learning tasks. Thus the principal components are the continuous solutions to the discrete cluster membership indicators. The classification result of this project will be discussed in next chapter.

IV. CLASSIFICATION RESULT

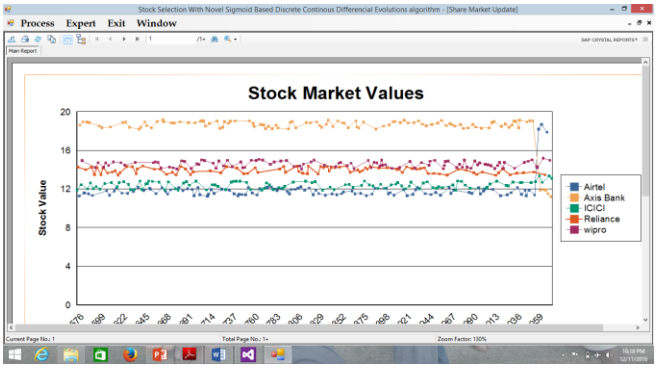

Fig. 4.1: Stock market comparison

Fig. 4.1 shows the stock market comparison of various different stocks available which automatically synchronizes based on the time period. The positive and negative words are extracted from the dataset as represented in fig. 4.2.

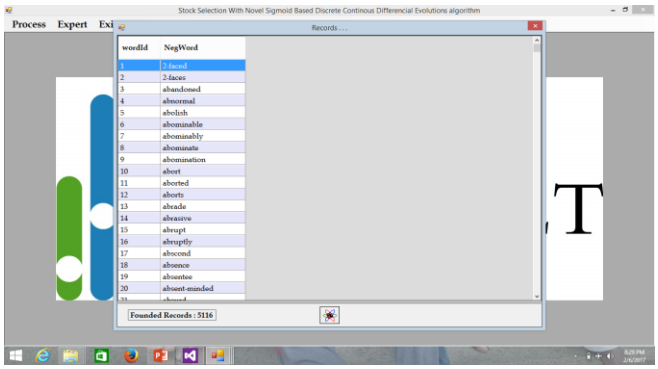

Fig. 4.2: Dataset processing

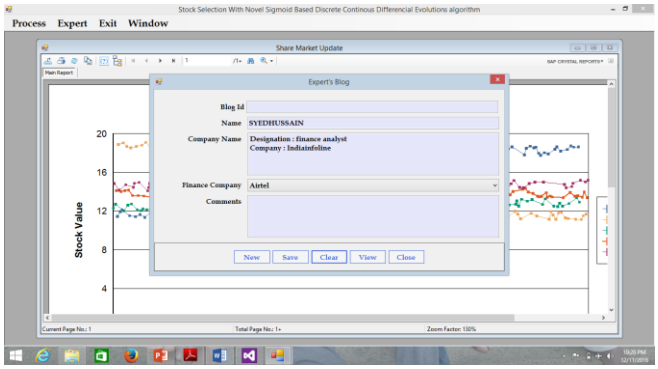

Fig. 4.3: Stock market registration

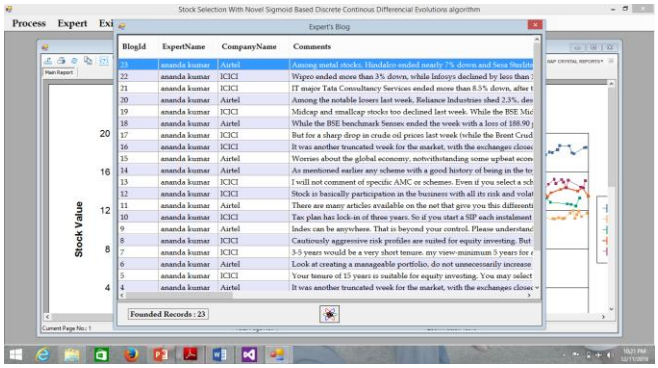

Fig. 4.4: Review extraction

A new stock product can be added to the system as shown in fig. 4.3. The fig. 4.4. shows the review given by the stock market analyst about the different stocks on various time

periods is extracted. The conclusion and future scope of this paper will be discussed in upcoming chapters.

V. CONCLUSION

By proposing a novel stock selection model with discrete and continuous variables by using principal component analysis, 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 main perspectives. First, the proposed stock selection model can obtain much higher returns than the market average performance, for both the whole market and different industries. 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.

VI. FUTURE SCOPE

Designed the novel stock selection model and the sigmoid based mixed discrete-continuous DE algorithm with PCA. In terms of sample data, benchmark models and evaluation criteria. The empirical results and verifies the effectiveness of the proposed stock selection model and the novel sigmoid based DE algorithm are the main directions for future research.

REFERENCES

[1] L. Yu, S. Wang, and K. K. Lai, "**Mining stock market tendency using ga-based support vector machines**" in *Internet and Network Economics*,. Lecture Notes in Computer Science, X. Deng and Y. Ye, Eds. Springer Berlin Heidelberg, 2007, vol. 3828, ch. 33, pp. 336–345.

[2] Y. L. Becker, P. Fei, and A. Lester, "**Stock selection: An innovative application of genetic programming methodology**" in *Genetic Programming Theory and Practice IV*, ser. Genetic and Evolutionary Computation, R. Riolo, T. Soule, and B. Worzel, Eds. Springer US, 2008, ch. 19, pp. 315–334.

[3] Y. L. Becker, H. Fox, and P. Fei, "**An empirical study of multiobjective algorithms for stock ranking**" in *Genetic Programming Theory and Practice V*, ser. Genetic and Evolutionary Computation Series, R. Riolo, T. Soule, and B. Worzel, Eds. Springer, 2008, pp. 239–259.

[4] T. Takahama, S. Sakai, A. Hara, and N. Iwane, "**Predicting stock price using neural networks optimized by differential evolution with degeneration**" *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 12, pp. 5021–5031, 2010.

[5] C. F. Huang, T. N. Hsieh, B. R. Chang, and C. H. Chang, "**A comparative study of stock scoring using regression and geneticbased linear models**" in *Proceeding of 2011 IEEE International Conference on Granular Computing (GrC 2011)*. IEEE, Nov. 2014, pp. 268–273.

[6] L. Wang and S. Gupta, "**Neural networks and wavelet de-noising for stock trading and prediction**" in *Time Series Analysis, Modeling and Applications*, ser. Intelligent Systems Reference Library, W. Pedrycz and S. M. Chen, Eds. Springer Berlin Heidelberg,2015, vol. 47, ch. 11, pp. 229–247.

AUTHOR PROFILE

Dr. A.Balamurugan is currently working as Professor in Department of Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. He has 25 years of teaching experience and published more than 40 papers in national and international journals and conferences. His area of interest includes Wireless Sensor Networks.

Mrs. S.Arul Selvi is currently working as Assistant Professor in Department of Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. She has 2 years of teaching experience and her area of interest includes Networks and Data Mining.

Mr. A.Syedhussain is currently pursuing Bachelor's degree in Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. His area of interest includes Data Structure and Data Mining.

Mr. A.Nithin is currently pursuing Bachelor's degree in Computer Science and Engineering at Sri Krishna College of Technology, Coimbatore. His area of interest includes Data Mining and Artificial Intelligence.