Frontiers in Economics and Management Volume 5 Issue 12, 2024
ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

Portfolio Optimization of Intelligent Investment Advisors based on Deep Learning

Yucheng Zhuang[1, a], Mo Zhang[2, b, *], Tongtong Li[3, c], and Tao Wang[4,5, d]

[1] Department of Financial Technology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong 999077, China

[2] School of Accoutancy, Shandong University of Finance and Economics, Jinan 250014, China

[3] School of Marxism, Shandong Jianzhu University, Jinan, 250014, China

[4] School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, 250014, China

[5] Chinese Studies of Hanyang University, Graduate School of International Studies, Hanyang University, Seoul 04763, Republic of Korea

[a]chengyuzh3@163.com, [b, *]beethmo@163.com, [c]litongtong19@sdjzu.edu.cn, [d]oucwangtao@163.com

These authors contributed equally and are considered co-first authors

Abstract

This research proposal aims to develop an optimal fund portfolio allocation plan by integrating deep learning ensemble models into intelligent investment advisors. Through layer-by-layer data processing, we seek to ensure the optimal output of fund portfolio allocation plans, while maintaining a database system to manage diverse fund data generated during model training. Furthermore, economic cycle analysis will be incorporated to enhance the stability and accuracy of the artificial intelligence solution applied to fund portfolio allocation.

Keywords

Intelligent Investment Advisor; Fund Portfolio; Deep Learning.

1. Introduction

Smart advisors, an emerging financial service in the fintech domain, have garnered significant attention in recent years. These services utilize advanced data analysis and technological advancements to provide investors with portfolio allocation recommendations. Numerous traditional financial institutions, such as securities firms and investment banks, have integrated smart advisory services into their offerings. Concurrently, domestic Internet companies like Ant Financial and Baidu Finance have ventured into this field. Researchers globally are dedicated to enhancing the accuracy and efficiency of intelligent advisors using technologies such as machine learning algorithms, big data analysis, blockchain, and natural language processing. With ongoing technological advancements, these updates will further propel the evolution and innovation of robo-advisers.

There are notable differences between domestic and foreign markets in the realm of smart advisors. One primary distinction lies in investment choices. Smart advisors often offer clients a portfolio comprising multiple asset classes, including stocks, bonds, currencies, and other options. However, due to market variations, investment preferences differ across regions. For instance, in the United States, passively managed funds have consistently outperformed actively managed funds, leading mainstream advisory platforms to prefer passive index ETFs

ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

as the underlying investment. Another significant difference concerns regulatory requirements, which vary across countries, impacting the functionality of robo-advisors.

Overall, the robo-adviser market is poised for growth, fueled by technological advancements and increasing investor participation. Furthermore, the scope of smart advisors' functionalities is anticipated to expand gradually.

2. Scheme Design

We plan to utilize deep learning models to evaluate the historical performance of all public funds, aiming to derive an optimal fund portfolio allocation scheme. Clustering models will initially categorize the acquired fund data, and predictive models will be trained under each classification. The introduction of economic cycle analysis will provide a comprehensive market understanding, assisting intelligent advisors in managing risks more effectively and optimizing asset allocation.

Our research is structured into the following five steps:

2.1. Data Acquisition

2.1.1. Data Sources

(1)Fund Data: Primarily sourced from the Choice database, focusing on non-money funds classified by the U.S. Securities and Exchange Commission (SEC), namely equity funds, hybrid funds, and bond funds. Money market funds were excluded due to their distinct portfolio structure and evaluation criteria. The dataset spans a three-year period from September 30, 2020, to September 30, 2023, encompassing 5,468 funds that have been operational for at least five years as of the screening date.

(2)Economic Data: Monthly data from China's National Bureau of Statistics for the period from January 2000 to October 2023, including the year-on-year growth rate of industrial added value and the CPI (Consumer Price Index) growth rate, to measure economic growth and inflation levels.

2.1.2. Data Storage

(1)Data Processing: The fund data, contained in an Excel file, records the daily adjusted net asset value (Adjusted NAV) information for 5,468 funds. A Python script extracts the fund code, fund name, date, and daily percentage change in Adjusted NAV. The data is then split into CSV files, each containing 800,000 records, to facilitate efficient database storage and subsequent data handling.

(2)Database: Python scripts are used to import the CSV files into a MySQL database. The script creates a local MySQL database, establishes a database connection, and defines a table to store the extracted fund data.

Table 1. Table Definitions for Extracted Fund Data

Column Names	Data Type	Nullable
Code	varchar(20)	FALSE
NameFund	varchar(255)	TRUE
Date	datetime	FALSE
Daily Return	double	FALSE

2.2. Fund Classification

2.2.1. Classification Model

(1) Introduction to the Model

ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

Clustering, an unsupervised learning method, involves grouping data points into clusters based on their similarity. The K-Means clustering algorithm, proposed by J.B. Macqueen in 1967, classifies a sample set into multiple clusters. The algorithm uses the Euclidean distance to measure similarity and iteratively refines cluster centers until convergence.

(2) Model Construction

To address data dispersion issues, the K-means model uses a Standard Scaler function. Missing values are handled using the last non-null value before the missing entry. The funds are classified into six categories using the K-means algorithm.

(3) Training and Evaluation of the Model

The K-means model is trained using 5,468 funds and classified into six categories. Cross-validation ensures model accuracy, yielding an average silhouette coefficient of 0.806, indicating good cluster matching.

2.2.2. Classification Results

Through the K-means model, 5,468 funds were successfully classified into six categories. The average silhouette coefficient of 0.806 confirms the model's robustness.

3. Fund Forecast

3.1. Model Construction

A neural network time series prediction model is used to forecast fund performance. The model incorporates data preprocessing, splitting, standardization, neural network construction, and training/validation. The model structure comprises a multi-layer perceptron with two hidden layers, each containing 64 and 32 neurons, respectively, using the ReLU activation function.

3.2. Forecast Results

The model predicts the top 10 best-performing funds in each category for the next month, demonstrating high predictive accuracy on the validation set.

4. Business Cycle Analysis

4.1. Methods and Indicators for Dividing Business Cycles

The Bry-Boschan algorithm is employed to identify business cycle turning points. The Merrill Lynch Investment Clock divides the economic cycle into four phases: trough, expansion, contraction, and peak, based on economic growth and inflation levels measured by the year-on-year growth rate of industrial added value and CPI.

4.2. Construction Steps of Economic Cycle Segmentation

The process involves seasonal adjustment using the X12 method, noise reduction through HP filtering, determination of turning points using the Bry-Boschan model, and phase assignment based on duration thresholds.

Table 2. The Results of the Different Stages of China's Economic Cycle from 2000 to 2023

Start Time	End Time	EconomicGrowth	Inflation Levels	Stages	Number of Months
20-Jan	21-Jan	down	up	Contraction	13
21-Feb	21-Apr	down	down	Rock Bottom	3
21-May	22-Feb	down	up	Contraction	10
22-Mar	22-Oct	up	up	Peak	8
22-Nov	23-Apr	up	down	Expansion	6
23-May	23-Oct	down	up	Contraction	6

ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

4.3. Forecast of the Economic Cycle

A Convolutional Neural Network (CNN) model predicts CPI and industrial value added, facilitating business cycle forecasting. The model structure includes convolutional layers, activation functions, and linear layers. Predictions from March 2023 to October 2023 demonstrate high accuracy compared to actual results.

4.4. Fund Analysis based on Economic Cycle

Table 3. Comparison between Actual and Predicted Economic Cycle

Date	The Actual Economic Cycle	The Predicted Business Cycle
23-Oct	Contraction	Shrinkage
23-Sep	Shrinkage	Shrinkage
23-Aug	Shrinkage	Shrinkage
23-Jul	Shrinkage	Shrinkage
23-Jun	Shrinkage	Shrinkage
23-May	Shrinkage	Shrinkage
23-Apr	Dilation	Expansion
23-Mar	Expansion	Expansion

Historical data analysis identifies the best-performing fund categories across different economic cycle phases.

Table 4. Best Performing Fund Category within Each Economic Cycle

Business Cycle	Top Performing Fund Category
Expansion	0
Peak	5
Contraction	2
Valley Bottom	1

5. Constructing the Optimal Portfolio

5.1. Timeline Assumptions

Based on the forecasting models, we construct an optimal recommended portfolio for October 2023.

5.2. Fund Category Selection

Given the business cycle forecast, October 2023 is anticipated to be in the contraction phase, with Category 2 funds identified as the top performers.

5.3. Fund Selection within the Selected Category

The top 10 best-performing funds in Category 2 are selected based on expected returns.

ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

Table 5. Top 10 Funds in Category 2

Code
161032.SZ
002910.OF
003321.OF
003323.OF
004685.OF
378546.OF
003322.OF
161129.SZ
003717.OF
002834.OF

5.4. Actual Performance of the Fund Portfolio

The constructed portfolio returned -0.33% in October 2023, outperforming the average return of -3.36% for 18 CSI 300 index funds, despite the overall market's negative performance.

6. Summary

This document elaborates on a research plan for optimizing investment portfolio allocation by a smart investment advisory firm leveraging deep learning techniques. The study first procured extensive historical data on non-money market funds from the Choice database and processed the data using Python scripts for storage. By employing the K-Means clustering algorithm, the research successfully categorized these funds into six distinct groups, with a robust model performance indicated by an average silhouette coefficient of 0.806. Subsequently, a neural network time series prediction model was utilized to accurately forecast the future performance of funds within each category, identifying the top-performing funds. Additionally, the study integrated economic cycle analysis, utilizing the Merrill Lynch Investment Clock methodology and a Convolutional Neural Network (CNN) model to predict future economic cycles, providing a macroeconomic context for portfolio optimization. Ultimately, based on economic cycle predictions and fund performance analysis, the research constructed an optimal fund investment portfolio for a specific time period, which outperformed the average return rate of the CSI 300 Index funds in the actual market, demonstrating the effectiveness and practicality of the research methodology. The entire research plan showcases the potential application of artificial intelligence in the field of smart investment advising, offering investors a more scientific and systematic approach to investment portfolio optimization.

References

[1] Pattarin, F., Paterlini, S., & Minerva, T. (2004). Clustering financial time series: An application to mutual funds style analysis. Computational Statistics & Data Analysis, 47(2), 353-372.

[2] Li, L., Wang, J., & Li, X. (2020). Efficiency analysis of machine learning intelligent investment based on K-Means algorithm. IEEE Access, 8, 20203011366.

[3] Marathe, A., & Shawky, H. (1999). Categorizing mutual funds using clusters. Advances in Quantitative Analysis of Finance and Accounting, 7.

[4] Chavan, S., Kumar, P., & Gianelle, T. (2021). Intelligent investment portfolio management using time-series analytics and deep reinforcement learning. SMU Data Science Review, 5(2), Article 7.

[5] Lam, J. W. (2016). Robo-advisors: A portfolio management perspective (Senior Thesis). Yale College.

ISSN: 2692-7608 DOI: 10.6981/FEM.202412_5(12).0026

[6] Sun, Q. Y., & Zhao, Y. Q. (2017). Exploration of the development of intelligent investment advisory business in traditional financial institutions in China – Taking Capricorn intelligent investment as an example. International Finance, (09), 34-39.

[7] Wang, Y., & Pan, W. (2022). A dynamic allocation study of global stock market sectors based on investment clock and Black-Litterman model. The Journal of Quantitative Economics, 13(3), 37-53.