Portfolio Construction Based on Combination of Individual Stocks and LSTM

Mingxuan Chen[1, †], Zhanming Shen[2, *, †] and Wanting Xu[3, †]

[1]College of Art & Science, Boston University, Boston MA, USA

[2]School of Software, Shandong University, Jinan, China

[3]School of Foreign Language, Chengdu University of Technology, Chengdu, China

*Corresponding author: 202100161057@mail.sdu.edu.cn

†These authors contributed equally.

Abstract. With the development of the idea of portfolio optimization, it has become one of the important topics in the field of modern finance for achieving the balance between maximizing the return of assets and minimizing the risk. This study selected ten stocks from WIND and used LSTM Neural Network as the fundamental forecasting model. Based on the method of cyclic prediction, we predicted the asset's return forecast value of the next day by rolling multi-dimensional historical features. Five stocks with the best outcomes were selected for optimization, using both the return component and the volatility, covariance matrix of the stocks' returns as the covariance matrix to build the efficient frontier and to optimize the weighting. The Monte Carlo Simulation was used to build the maximum Sharpe ratio portfolio and minimum volatility portfolio. The maximum Sharpe ratio portfolio showed that the Midea and Mindray had the potential to profit dramatically at the risk of high volatility. However, relatively high volatility means the possibility of profiting is questionable, so a Minimum Volatility Portfolio was also conducted, which indicated conservative investors usually reduce the risk of their investment as much as possible. Whereas the max Sharpe ratio portfolio showing that risk-seeking investors are more willing to take risks if it can generate more profit. Overall, a hybrid intelligent algorithm based on cyclic prediction, Monte Carlo Simulation and artificial neural network was proposed, and the new applications of machine learning and mathematical methods in asset allocation were promoted.

Keywords: Machine Learning; Portfolio Construction; LSTM Neural Network.

1. Introduction

The key of asset allocation is that the performance of individual stocks that one holds is less important than the overall behavior of a broader combination of these stocks. Asset allocation and the idea behind portfolio optimization has grown dramatically over the past century, and various methods as well as models have been widely used by managers from all sorts of funds and trusts [1, 2]. One of the revolutionary models that is used in the field of modern finance, the Capital Asset Pricing Model, was developed in the early 1960s by William Sharpe, Jack Treynor, John Lintner and Jan Mossin. the fundamental idea of this model provided the first coherent framework for relating the required return on an investment to the risk of that investment [3]. However, the idea of portfolio has a further history, American economist Harry Markowitz pioneered Modern Portfolio Theory in his paper "Portfolio Selection," which was published in the Journal of Finance in 1952 [4]. He was also awarded a Nobel Prize for his unprecedented understanding in this field [5].

Contemporarily, portfolio management is an extremely complicated process that every single portfolio manager has one's own ways of thinking, but the fundamental basics originated from Modern Portfolio Theory and Capital Asset Pricing Model. Modern portfolio management has been widely thought to consist of 3 phases [6]. Among all of the portfolio managing strategies, the all weather strategy is one of the most popular as well as the most flexible strategies that are known today. Bridgewater is a hedge fund that is known for its famous all weather investing strategy, its co-chief investment officer, Raymond Thomas Dalio, developed Bridgewater's All Weather Strategy in the 1970s after observing market changes and potential return scenarios surrounding the political turmoil from Richard Nixon's presidency [7]. The essential benefit of an all weather fund is that it tends to behave reasonably well under both bearish and bullish economic environments, and one of the commonly used strategies is the long/short strategy. This long/short strategy involves taking both long and short positions when creating a portfolio. After selecting stocks, securities, and other investing opportunities from the universe, a portfolio of baskets is calculated based on each investment candidate's return, and the portfolio manager typically rank all candidates in his/her universe from the lowest factor exposure to the highest, and long on the top percentile because those candidates are most likely to profit in the future, while short on the bottom percentile because those candidates are more likely to be losing. This all weather investing strategy is generally used by portfolio managers today along with their own alpha factor, because holding both long and short position while the portfolio's components are diversified, the loss of individual investment candidate's impact on the portfolio as a whole can be negligible, so that the portfolio can perform relatively well under any economic environments [8].

The purpose of this research paper was using machine learning models to forecast the returns of selected stocks, and using the forecasted returns as the return component to simulate possible combinations of the portfolio, which then resulted in specified weightings corresponding to each prerequisite (Max Sharpe Ratio and Min Volatility).

2. Data & Method

For this study, all the data used in this research paper came from WIND and Yahoo Finance. The original dataset contained open price, close price, adjusted closed price, and volume. For the purpose of forecasting as well as portfolio optimization, data cleaning did not seem to be necessary due to the fact that there were no missing values. The very first step of this research included randomly selecting ten stocks to train the price prediction model respectively, and the price of the next day was predicted by rolling multi-dimensional historical features. The only standard used when selecting stocks is the market value ranking, thus the industries of candidates were relatively diversified. According to the market value ranking of stocks in 2022, 10 stocks belonging to different industries were selected as shown in Table. 1. The Return of the stocks can be calculated as

$$Return_t = (Price_t - Price_{t-1})/Price_{t-1} \tag{1}$$

Four fundamental indicators and six technical indicators were selected as shown in Table. 2.

Table 1. List of selected stocks.

600519.SH	Kweichow Moutai
601398.SH	ICBC
300750.SZ	CATL(Contemporary Amperex Technology Co., Limited)
601318.SH	PING AN INSURANCE
601857.SH	CNPC(China National Petroleum Corporation)
002594.SZ	BYD
600900.SH	CHINA YANGTZE POWER
000333.SZ	Midea
002415.SZ	HIKVISION
300760.SZ	Mindray

Secifically, the four fundamental indicators helped reflect the company's financial standing, to a certain extent, could reflect the future return situation; as for the six technical indicators, they were based on the performance of the past return performance, reflecting the stock performance at technical level. Using these ten indicators could help one better understand the situation of stocks and better predict future trends. In order to speed up the gradient descent to find the optimal solution, as well as to prevent the problem that the loss function is difficult to converge, normalization was needed. The preprocessed data was limited to a certain range, thereby eliminating the adverse effects caused by outlier sample data, so as to improve the accuracy of the model. In addition, stability of training could be ensured. Through min-max standardization, the price data was converted into a decimal number in the (0, 1) interval.

Table 2. Description of selected indicators.

Type	Features	Definition
Fundamental Indicators	PE	Ratio of a stock 's market price to earnings per share
	PB	Ratio of stock price divided by net assets per share
	PS	Ratio of stock price divided by sales per share
	PCF	Ratio of stock price to cash flow per share
Technical Indicators	MACD	Technical Indicators for Judging the Timing of Buying and Selling
	RSI	A popular momentum indicator, which determines whether the stock is high or low.
	WR	It is used to measure whether the stock is overbought or oversold
	BOLL	An indicator designed according to the principle of standard deviation which plays an important reference role in predicting the trend of future market.
	ADTM	Using the difference between the upward fluctuation range and the downward fluctuation range of the opening price to describe the popularity.
	MI	Based on the principle of constant speed of stock, this indicator examines the speed of rise and fall of stock price and analyzes the trend of stock price with its change.

This research used LSTM Neural Network as the fundamental forecasting model [9-11]. For a given sequence, after applying LSTM, the hidden layer sequence h and output y could be calculated by the following iterative equations:

$$h_t = f_a(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{2}$$

$$y_t = W_{hy}h_t + b_y \tag{3}$$

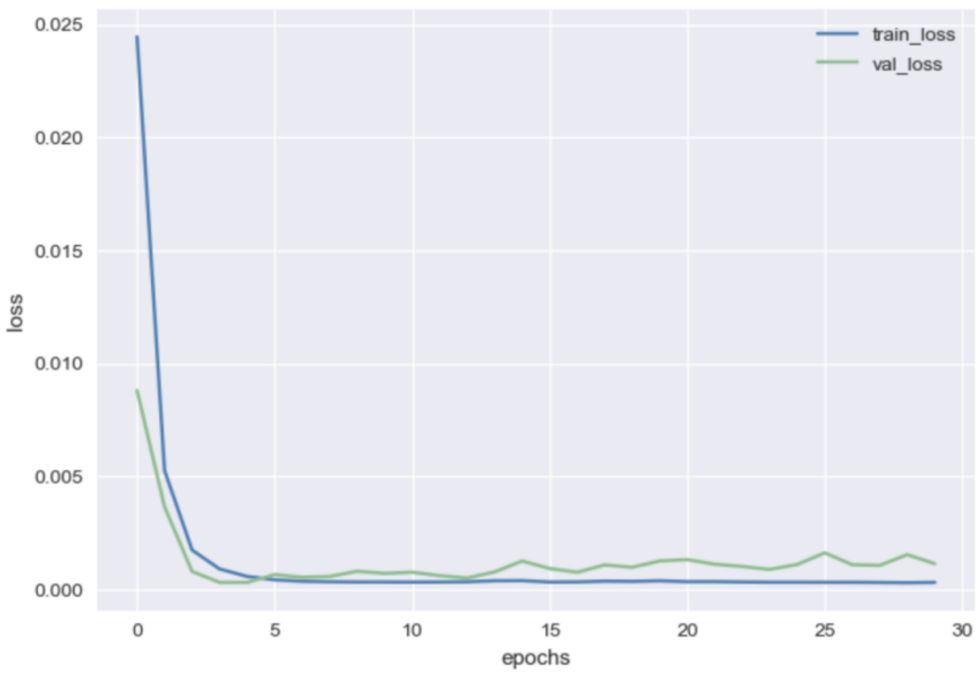

Figure 1. Training set.

Using the method of cyclic prediction, 30 days were taken as the time slice of each round of prediction, and a neural network containing two LSTM layers and a fully connected layer was finally constructed. Here, input X is daily pushback 30-day feature sequence and output Y is the asset's return forecast value of the next day. The loss function selected in the training process was the 'mean square error' (MSE), and the optimization method was the 'Adam' method. As the traditional stochastic gradient descent method (SGD) only updates the current parameter through the current gradient, a more comprehensive Adam method was selected, which comprehensively considered the gradient of the loss function with respect to the current parameter, as well as the first-order momentum and second-order momentum calculated based on the historical gradient, and performed deviation correction. In terms of model training, the ratio of 4 : 1 (alpha = 0.8) was chosen to divide the data set, the first 80 % of the data was used as training data, and the last 20 % of the data as test data, so as to construct the model. The batch size (batch _ size) was set to 16, and the model was trained for 30 epochs (seen from Fig. 1).

3. Results & Discussion

The results on the training and test set are given in Fig. 2. It can be seen that the trained LSTM network could basically judge the general trend of return, but it could not predict the fluctuation of daily frequency. The results on the test set are similar, with the model being able to predict certain trends, but not daily frequency fluctuations. In order to start portfolio optimization process, five of the ten stocks were selected to put into the portfolio. The criteria of selection were based on the forecasted results. The original goal was to forecast the returns of our ten stocks for the next two weeks from the last date of our dataset, and to compare the sums of each stock's return. Thus, the five stocks with the best outcomes could be used to start portfolio simulations. However, the forecasting of the returns for the future two weeks failed. Hence , the sums of the last fifteen forecasted results were used to select the stocks with the best potential to profit. The Table. 3 lists five stocks had the best potential to profit in the future.

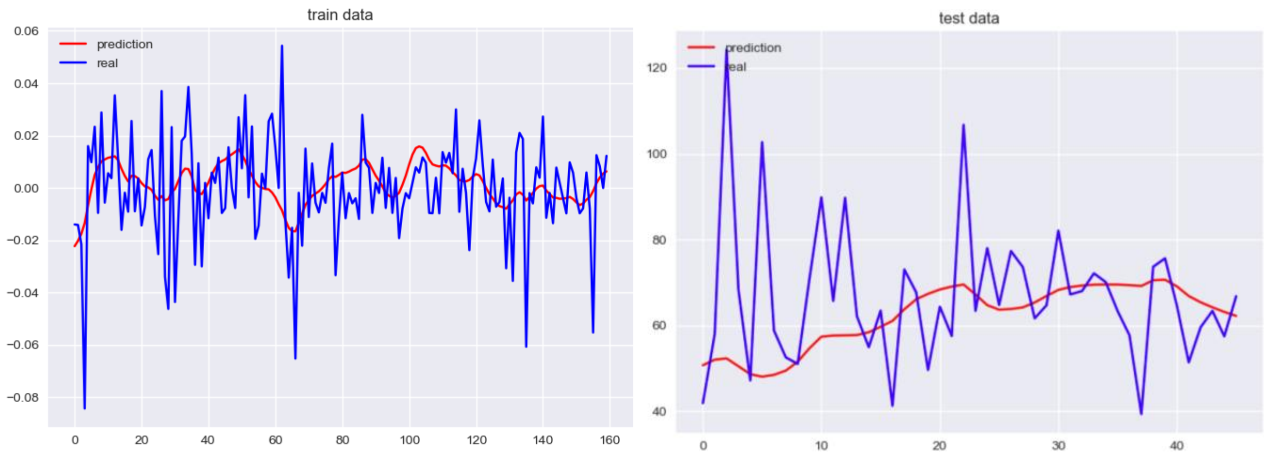

Figure 2. Results of training and test set.

Table 3. List of potential stocks.

601398.SS	ICBC
300760.SZ	Mindray
000333.SZ	Midea
601857.SS	CNPC(China National Petroleum Corporation)
600900.SS	CHINA YANGTZE POWER

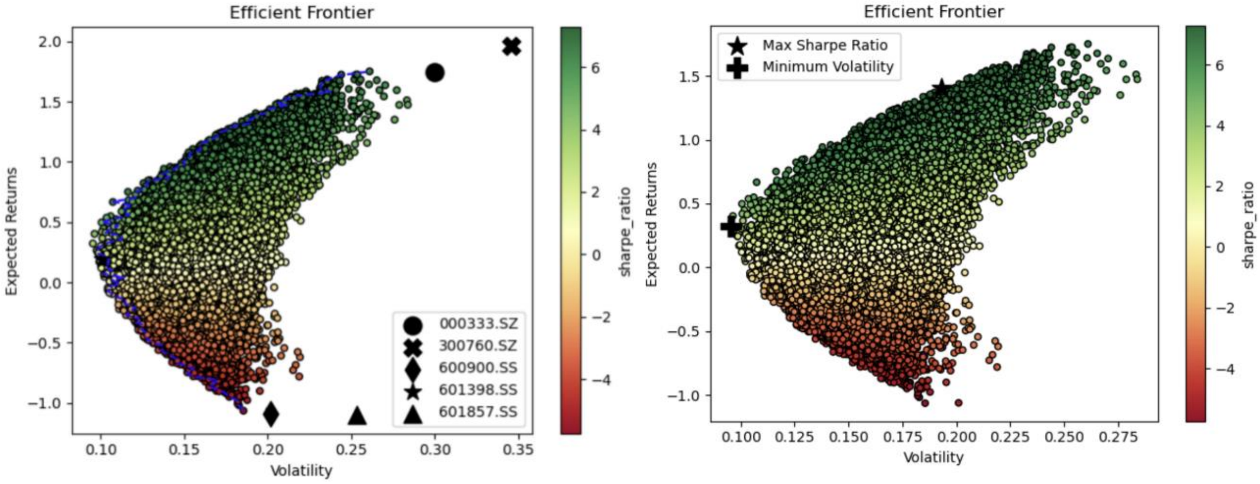

Figure 3. Simulated efficient frontier of portfolio.

The ideal process of portfolio simulation is using both the return component and the risk metric as the covariance matrix to build the efficient frontier and to optimize the weighting as the way which is selected. However, due to the current inaccessibility of Barra Beta, the volatility, covariance matrix of the stocks' returns, were used to represent the risk metric in simulations. Undoubtedly the results would be somewhat problematic because the covariance of stock's returns could not explain the risks of stocks perfectly, so this project still has a lot of space for improvement. The return component used for portfolio simulations was the previously forecasted returns. The purpose of this procedure was to predict the future returns of selected stocks as best as possible so that the weighting of each stock in the portfolio would be able to be set according to the expected returns. However, as the forecasting of the future returns of the selected stocks failed, the forecasted returns that were able to attain were used, from 03/10/2022 to 12/31/2022, for the purpose of weighting determination, which is another part that can be improved.

As shown in Fig. 3, 10,000 possible combinations of the five stocks were simulated. According to the efficient frontier graph, CHINA YANGTZE POWER (600900.SS) and CNPC (China National Petroleum Corporation) (601857.SS) appeared to be losing despite their relatively high volatility. Industrial and Commercial Bank of China (601398.SS), as one of the largest in the world as well as the first place in The Banker's Top 1,000 World Banks Ranking, would most likely continue to profit at a very low risk. Midea (000333.SZ) and Mindray (300760.SZ) had the potential to profit dramatically at the risk of high volatility. Sharpe ratio measures a portfolio's risk-adjusted returns. The higher the sharpe ratio is, the more returns one will get for extra risks that have been taken. A portfolio with a high sharpe ratio means one will be compensated very well for the risk one is taking.

Table 4. Optimal portfolio performances and weight parameters.

Maximum Sharpe Ratio portfolio		
Performance	Name	Weights
Returns 140.31% Volatility 19.29% Sharpe ratio 727.38%	000333. SZ	33.37%
	300760. SZ	39.43%
	600900.SS	0.02%
	601398.SS	26.88%
	601857.SS	0.30%
Minimum Volatility portfolio		
Performance	Name	Weights
Returns 32.25% Volatility 9.54% Sharpe ratio 338.13%	000333. SZ	3.67%
	300760. SZ	6.80%
	600900.SS	3.43%
	601398.SS	86.08%
	601857.SS	0.02%

In this case, as given in Table. 4, the weightings of the five stocks were 33.37% for Midea, 39.43% for Mindray, 0.02% for CHINA YANGTZE POWER, 26.88% for ICBC, and 0.30% for China National Petroleum Corporation. Finally, the outcome ended up being a Sharpe ratio of 727.38%, ideally, this means one would be compensated greatly for the 19.29% volatility that is taken, but relatively high volatility means the possibility of profiting is questionable, so a Minimum Volatility Portfolio was conducted in the following session. As shown in the right panel of Fig. 3, two portfolios on the efficient frontier graph were plotted, unsurprisingly the minimum volatility portfolio lied on the left side of the frontier, because conservative investors usually reduce the risk of their investment as much as possible, while max Sharpe ratio portfolio fell relatively to the right side of the frontier, showing that risk-seeking investors are more willing to take risks if it can generate more profit.

4. Conclusion

In summary, this study builds a portfolio optimizing process using the method of machine learning and forecasting. Ten stocks were selected for forecasting, where five of the ten were picked up for optimization. The maximum sharpe ratio portfolio and minimum volatility portfolio were built. The overall process was fluent, but the result could not be told as successful due to insufficient knowledge of financial engineering. There are many flaws of this project that have to be improved if it is aimed to be used in practical financial market. The first improvement is the forecasting process. Using Monte Carlo Simulation requires to use the best forecasting returns that could possibly obtain as the return component, and there are still many aspects need to be refined (e.g., selecting indicators that are more accordant to selected stocks), because each company of each industry in most cases is totally different from the others. The process of selecting stocks should be dealt with more caution. Moreover, optimization process should have more information for each stocks, e.g., tear sheet should be created in pyfolio to retrieve fundamental information of each stocks that could have been missed, and conclusions could be drawn from these details that whether this specific stock is appropriate for the portfolio, or whether this stock will be able to profit. Last but not least, Fama French models should be adopted to evaluate the results. By taking out all of the common factors, the problems of whether the alpha of certain portfolio is successfully found, and if the portfolio has the unique strategy or methodology to capture value could be more efficiently solved. Overall, these results offer a guideline for portfolio construction.

References

[1] Breuel T M. Benchmarking of LSTM networks. arXiv preprint arXiv: 1508.02774, 2015.

[2] Bonate P L. a Brief Introduction to Monte Carlo Simulation. Clin Pharmacokinet, 2011, 40: 15 – 22.

[3] Perold A F. The Capital Asset Pricing Model. Journal of Economic Perspectives, 2004, 18 (3): 3 - 24.

[4] Harry M. Portfolio Selection. The Journal of Finance, 1952, 7 (1): 77 - 91.

[5] The Nobel Prize. This Year's Laureates Are Pioneers in The Theory of Financial Economics and Corporate Finance, Retrieve from: https://www.nobelprize.org/prizes/economic-sciences/1990/press-release/.” Accessed Jan. 28, 2023.

[6] Kim D, Jack C F. Modern portfolio theory: Foundations, analysis, and new developments. John Wiley & Sons, 2013.

[7] Bridgewater. The All-Weather Story, Retrieve from: https://www.bridgewater.com/research-and-insights/the-all-weather-story. Accessed Jan. 28, 2023.

[8] Choueifaty Y, Froidure T, Reynier J. Properties of the most diversified portfolio. Journal of investment strategies, 2013, 2 (2): 49 - 70.

[9] Troiano L, Mejuto V E, Loia V. Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications. IEEE Transactions on Industrial Informatics, 2018, 1: 3226 - 3234.

[10] Tino P, Schittenkopf C, Dorffner G. Financial volatility trading using recurrent neural networks. IEEE Transactions on Neural Networks, 2001, 12 (4): 865 - 874.

[11] Budiharto W. Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). Journal of Big Data, 2021, 8 (1).