

Investment Market Trend Forecasting and Strategy Optimization based on Big Data Analytics

Yuyao Zhang

Jinan Foreign Language School, Jinan Shandong, 250000, China

2102243618@qq.com

Abstract. The financial market is an important part of the global economy, and there exists a certain special internal regularity in its changes. With the rapid development of big data technology, the huge amount of data in the financial market provides new opportunities for the prediction of investment trends and strategy optimization. Based on big data analytics technology, combined with machine learning models, the study explores how to construct investment market trend prediction models through effective data preprocessing and feature engineering, and evaluates and compares the performance of different prediction models. How to accurately predict the direction of the financial market will be an important reference value for relevant practitioners as well as researchers. Traditional financial time series research mainly focuses on linear trend fitting based on statistical knowledge, however, the trend of the financial market is not limited to pure mathematical and rational function curves, the financial market is affected by a combination of various factors from all walks of life. Big data analytics has significant application potential in the investment market, providing investors with more scientific and accurate decision support.

Keywords: Financial Time Series Forecasting; Investment Markets; Big Data Analytics; Trend Forecasting; Strategy Optimization; Machine Learning.

1. Introduction

In today's context of globalization and information technology, the volume of data in the investment market has exploded[1]. Market participants are looking to analyze large amounts of historical and real-time data to predict future market trends in order to achieve better investment returns[2]. Traditional investment analytics methods are often inadequate in the face of complex market environments, and the rise of big data analytics provides a new way to solve this problem. In the information society, big data analysis technology has become an important basis for enterprise decision-making[3]. Through efficient data processing process, it extracts valuable information for enterprises from massive, diversified and fast-changing data, and then guides practice and optimizes decision-making.

Big data technology allows investors to make decisions in a broader context by collecting and analyzing market data from multiple sources (e.g. stock exchanges, news, social media, etc.). Compared with traditional data analysis methods, Big Data is able to handle massive amounts of multi-dimensional, unstructured data, which provides a richer basis for predicting market trends[4]. The acceleration of hardware upgrading and iteration makes the upgrade of computer software technology with hardware addition as well as technical guarantee, which improves the performance of computer operation[5]. People are not satisfied with manual search for regular features among data, which promotes the development of deep learning technology, people gradually began to try to use computers instead of human resources to extract the features that exist among the huge amount of data. Deep learning technology can mimic the human audiovisual and thinking mode, learn the characteristics of the sample data law, has a certain learning and analyzing ability, and based on the current big data era of explosive growth of source data, so that deep learning has a large number of data sets can be trained and verified.

Research on investment market forecasting based on big data has achieved some results[6]. Many studies have utilized machine learning models (e.g., Support Vector Machines, Random Forests, Neural Networks, etc.) to predict market trends with good results[7]. These models have some

limitations when facing data noise, heterogeneity and complex market dynamics[8]. How to apply the prediction results to the optimization of the actual investment strategy, especially to maximize the return under the premise of controlling the risk, is still an urgent problem to be solved[9].

In this paper, we first introduce the application of big data in the investment market and its impact on trend forecasting, followed by constructing a variety of market trend forecasting models and providing a detailed comparative analysis of the performance of these models[10]. Next, it investigates how to optimize investment strategies based on big data, especially dynamically adjusting portfolios in response to market volatility. The use of deep learning can correlate finance, psychology, and statistics in financial market forecasting. Therefore, it is of great theoretical significance and practical application value to carry out theoretical explorations as well as technical research on the prediction of financial market trends.

2. Big Data Analytics in the Investment Market

With the rapid development of the global financial market and the advancement of digitalization, the volume of data in the investment market has increased dramatically. The field of "financial time series forecasting" has many branches, including stock forecasting, index forecasting, exchange rate forecasting, fund forecasting and so on. However, with the rapid development of Internet technology in modern society, the speed of news dissemination has begun to accelerate, and the economic model of "Internet +" has led to an explosive growth of economic data, which makes it difficult for traditional statistical methods to meet the demand for in-depth characterization of the exploding volume of data under the Internet + model. Big data analytics can help investors mine valuable information from massive data, improve the science and accuracy of decision-making, and thus take the lead in market competition.

Big data technology enables investors to gain a more comprehensive understanding of market dynamics through the collection and analysis of structured and unstructured data. Compared to traditional data analytics, where investors rely on historical financial market data to accumulate "experience" to make investment decisions, with the popularization of statistics, investors have begun to use mathematical tools to judge and verify the accuracy of their strategies. The ability to analyze real-time data allows investors to make dynamic adjustments to their strategies in response to rapid changes in the market, avoiding lagging decision-making errors and thus improving investment yields and risk control.

In recent years, investment market research based on big data has made great progress, and many scholars have utilized machine learning models, such as support vector machines, random forests, and neural networks, to make market trend predictions and improve the accuracy of the predictions to a certain extent. However, current research still faces problems such as data noise, feature selection, and insufficient model generalization ability. The stochastic differential equation evaluation experiments conducted on 7 stock price index data and 10 simulated stock prices from NASDAQ and London Stock Exchange by feature extraction of stock data through wavelet variation show that the model results are better than the selected benchmark model and can effectively detect price manipulation patterns.

3. Investment Market Trend Forecasting Model

Data preprocessing, feature engineering, and model selection and construction are critical in investment market trend forecasting. In order to achieve accurate market forecasts, raw data need to be preprocessed to remove noise and extract useful features. Based on different data characteristics, suitable prediction models are constructed with machine learning methods such as time series models, support vector machines, random forests, and neural networks. By evaluating the performance of different models, the optimal prediction method is selected to ensure the generalization ability and application effect of the model.

3.1 Data Preprocessing and Feature Engineering

In the investment market, the raw data collected often contains a lot of noise, such as missing values, outliers, or repetitive data, etc. In the process of stock price prediction based on RNN networks, the phenomenon of gradient explosion or gradient disappearance occurs gradually as the amount of stock historical data increases, resulting in its inability to feature mine and find long-term contextual relationships under larger samples, which inspires the gated storage unit use, which requires the integration of data from different sources in order to construct a unified dataset for subsequent model analysis.

As the magnitude and units of data may vary significantly from market to market, the direct use of unprocessed data can lead to a decrease in the predictive performance of the model. To solve this problem, data need to be standardized or normalized. Feature selection techniques have gradually been widely used in stock forecasting, but existing forecasting methods tend to use a single feature selection technique, which may ignore some important assumptions, such as the regression function implied between the input and output variables. With these methods, the sensitivity of the model to the data can be improved and the prediction accuracy can be enhanced. Mean Squared Error (MSE) for model evaluation:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

When dealing with large-scale investment data, not all features are useful for market trend prediction. Indicator-oriented stock selection and investment operations are a challenging decision for investors because different sections of stocks have different specific attributes and different affinities to technical indicators. Feature extraction further improves the expressive power of the data and enhances the learning effect of the model by combining multiple related features into new variables. Sharpe Ratio for strategy optimization:

$$S = \frac{E[R] - R_f}{\sigma_R} \quad (2)$$

Due to the time-dependent nature of data in investment markets, processing time series data requires consideration of its dynamic characteristics. For time series data, commonly used processing methods include sliding window, difference and periodicity analysis. Build an intelligent financial sharing platform. Using cutting-edge technologies such as big data and artificial intelligence, an intelligent financial sharing platform is built. The platform can automatically collect, process and analyze financial data, thereby significantly improving work efficiency and data accuracy. Through the processing of these time characteristics, the model can better grasp the laws of market changes and improve the accuracy of forecasts.

3.2 Time Series-based Market Forecasting Model

Time series data are widely available in the investment market, such as stock prices and trading volume, and their main characteristics include trend, seasonality and cyclicity. Traditional event research in finance focuses on news headlines in finance and its related fields to assess the impact of events on listed companies, with slow dissemination and a restricted audience. However, with the rapid development of the Internet, breaking

The trend has led to faster access to information and the financial sector has begun to attract more and more attention.

Autoregressive Integral Sliding Average (ARIMA) model is a classical approach in time series forecasting. ARIMA model characterizes the dynamics of time series data by combining the autoregressive (AR), difference (I) and sliding average (MA) components. The autoregressive and sliding average components of the data are fitted by choosing appropriate orders for model parameter estimation and testing. The ARIMA model is suitable for short-term forecasting, and is able to effectively capture the linear relationship in the time series.

For time series data with significant seasonal fluctuations, traditional ARIMA models may not adequately capture these seasonal patterns. The Seasonal ARIMA (SARIMA) model deals with this problem by introducing seasonal components. The SARIMA model adds seasonal autoregressive,

difference and sliding average components to ARIMA, allowing the model to better fit data with seasonal fluctuations. By accurately modeling seasonal cycles, the accuracy of forecasts can be improved, showed in Figure 1 :



Figure 1. Risk vs Return by Strategy

Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) capable of dealing with long time dependencies and is particularly suitable for time series prediction tasks. LSTM solves the gradient vanishing and exploding problems of traditional RNNs in long time series learning by introducing a gating mechanism. LSTM is capable of memorizing long time dependency information and capturing complex time series patterns. By training the LSTM network, the model learns from historical data and predicts future market trends, which is suitable for dealing with time series data with complex nonlinear relationships.

3.3 Model Evaluation and Performance Comparison

In model evaluation, the selection of appropriate evaluation indexes is crucial. Commonly used regression model evaluation metrics include mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). MSE and RMSE reflect the magnitude of the error between predicted and actual values, with RMSE being often used due to its sensitivity to large errors. The MAE, on the other hand, provides an average of the prediction errors, and the R^2 metric is used to measure how well the model explains the variance of the data.

To ensure the reliability of model evaluation, cross validation techniques are commonly used to evaluate the performance of models. k-fold cross validation divides the dataset into k subsets, where k-1 subsets are used to train the model at a time and the remaining subsets are used for testing. The overall performance of the model can be obtained by calculating the average evaluation metrics of different folds. Cross-validation can effectively avoid the overfitting problem of the model and provide more stable and reliable performance estimates. Trend prediction techniques for financial markets were mainly based on statistical models in the early days, looking for the characteristics of trend changes in the historical data of financial markets. However, with the increase of financial market historical data, financial market trend prediction based on statistical modeling requires more and more labor and time costs, which becomes a major drawback, showed in Figure 2:

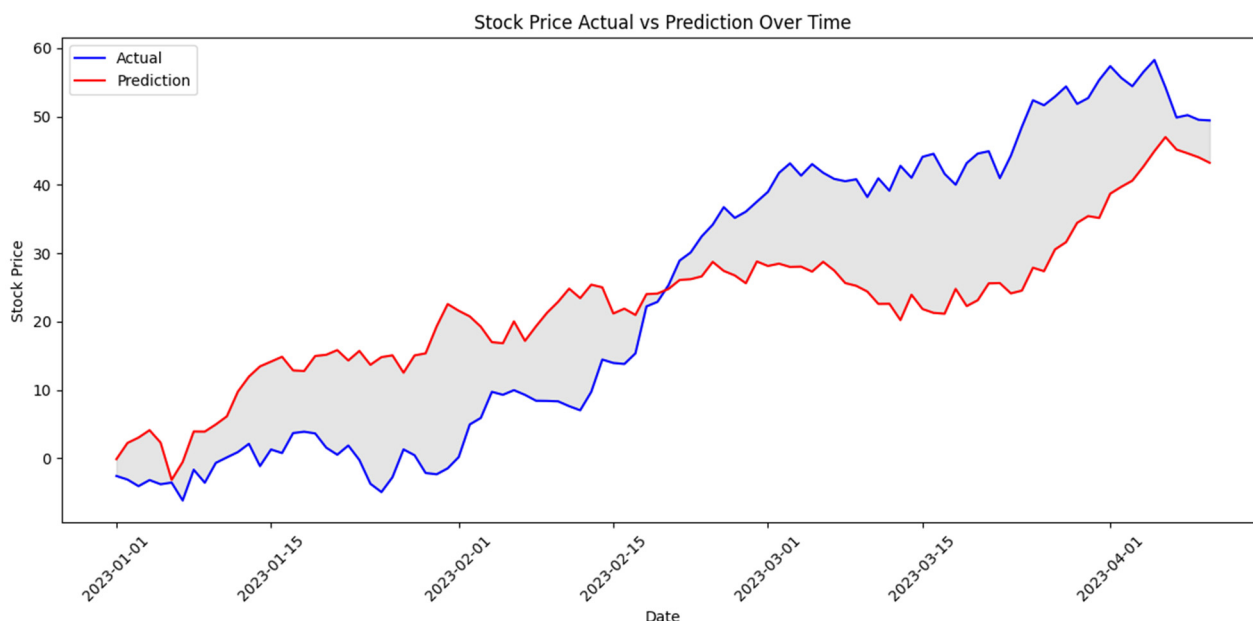


Figure 2. Stock Price Actual vs Prediction Over Time

When forecasting market trends, multiple models are usually applied for comparison to select the best forecasting method. For each model, a performance evaluation is conducted through the aforementioned evaluation metrics to compare the performance of each model in terms of prediction accuracy, computational complexity, and stability, etc. The ARIMA model is suitable for short-term linear trend prediction, while the LSTM model excels at capturing complex nonlinear patterns. By comparing the actual performance of these models, the model that best suits the current market data characteristics and forecasting needs can be selected.

After completing the model evaluation and performance comparison, the results are analyzed and discussed in depth. The strengths and weaknesses of each model are analyzed, the performance under different market conditions is explored, and factors that may affect forecast accuracy are identified. Discuss the applicability of the models in real-world applications, including data requirements, computational resource consumption, and the interpretability of the prediction results. In order to give better practical significance to the stock price prediction technique, the model combining Markowitz portfolio theory and the DDPG model in reinforcement learning is used to simulate stock market trading operations under the premise that the simulation of trading operations will generate commission, and in order to secondly reduce the investment risk of the simulation of stock market trading, customized integrated empirical boxes are utilized to open the second confirmation of trading operations and trading shares. and buy and sell shares.

4. Strategy Optimization and its Application to Investment

The core of investment strategy optimization lies in the scientific approach to adjust the portfolio to achieve the best risk-return ratio. A reinforcement learning-based financial market trend analysis method is used to find suitable time points for buying and selling in a simulated stock market based on daily closing prices. Markowitz portfolio theory is used as the basis for determining portfolio ratios, and the closing price, total number of shares, and market capitalization of a stock are used as feature inputs to determine the portfolio weight allocation; subsequently, a model combining DDPG and custom integrated experience boxes is introduced. Modern strategy optimization also incorporates big data analytics and machine learning techniques to dynamically adjust strategies through a data-driven approach in response to complex market changes.

Utilizing big data technology, it is possible to obtain the latest information about the market in real time and make dynamic adjustments to investment strategies. By analyzing market data and trends in real time, strategies can be optimized in different market environments. Based on market sentiment analysis and event-driven modeling, investment portfolios can be quickly adjusted to capture short-term opportunities. Machine learning algorithms can optimize strategies based on historical and real-time data, automate the adjustment of investment weights to adapt to market changes, and improve the flexibility and responsiveness of strategies.

In strategy optimization, risk control is key to ensure portfolio stability and long-term returns. A small-sample financial market trend prediction method based on sentiment analysis. In order to enrich the dataset for financial market trend prediction by mining sentiment features from stockholders' comments, sentiment features are extracted using a sentiment analysis technique based on a combination of word2vec and LSTM. In terms of return optimization, big data analysis can be used to evaluate the historical performance and risk-return ratio of different strategies in order to select the optimal investment plan.

Validating the effectiveness of strategy optimization through real-world use cases is key. In this paper, we will show several examples of investment strategy applications based on big data analytics and analyze the performance of these strategies in the actual market. The effectiveness and stability of the strategies are evaluated by comparing the actual returns of the strategies with the expected returns. Combine the reinforcement learning DDPG model with integrated customized experience boxes to judge the buying and selling operations under the simulated stock market, and finally conduct the simulated stock market buying and selling under the condition of setting commission, and make a comparative analysis with other traditional models and baseline returns.

5. Conclusion

This paper explores the application of big data analytics in investment market trend forecasting and strategy optimization, proposing a series of solutions for data preprocessing, model construction and strategy optimization. By comparing different time series forecasting models, ARIMA, SARIMA and LSTM, the advantages and disadvantages of each model in capturing market trends under different market conditions are demonstrated. Dynamic strategy adjustment based on big data can help investors optimize their investment decisions in the fast-changing market environment, effectively improving returns and controlling risks.

The introduction of big data technology has opened up new paths for analyzing and forecasting the investment market. Its powerful data processing and real-time analysis capabilities have led to more accurate market trend forecasts and more flexible strategy adjustments. With the ever-changing market environment, future research still needs to further improve the model's generalization ability, cope with data noise and heterogeneity, and deeply combine artificial intelligence technology with big data analysis to achieve more intelligent investment strategy optimization. Big data analysis has revolutionized the investment market, providing investors with more scientific and effective decision support. With the continuous progress of technology and deeper application, big data-driven investment strategy optimization will play an important role in the future.

References

- [1] Sardellitti, S, Barbarossa, et al. Joint optimization of collaborative sensing and radio resource allocation in small-cell networks[J]. IEEE Transactions on Signal Processing, 2013, 61(18):4506-4520. DOI: 10.1109/TSP.2013.2267737.
- [2] Chen S M, Chen S W. Fuzzy Forecasting Based on Two-Factors Second-Order Fuzzy-Trend Logical Relationship Groups and the Probabilities of Trends of Fuzzy Logical Relationships[J]. IEEE Transactions on Cybernetics, 2017, 45(3):391-403.Doi:10.1109/TCYB.2014.2326888.
- [3] Ku J, Zhen H, Gong W. Offline data-driven optimization based on dual-scale surrogate ensemble [J]. Memetic computing, 2023, 15(2):139-154.

- [4] Jiang Y, Wang T, Zhao H, et al. Big Data Analysis Applied in Agricultural Planting Layout Optimization[J]. Applied Engineering in Agriculture, 2019, 35(2):147-162.DOI:10.13031/aea.12790.
- [5] Li M, Du W, Qian F, et al. Total plant performance evaluation based on big data: Visualization analysis of TE process[J]. Chinese Journal of Chemical Engineering, 2018, 26(08):1736-1749. DOI:10. 1016/ j. cjche. 2018.06.009.
- [6] Liang M. Optimization of Quantitative Financial Data Analysis System Based on Deep Learning [J]. Complexity, 2021, 2021(1):1-11.DOI:10.1155/2021/5527615.
- [7] Wu Z , Zhang Y , Singh V ,et al. Automating Cloud Network Optimization and Evolution[J].IEEE Journal on Selected Areas in Communications, 2013, 31(12):2620-2631.DOI:10.1109/JSAC.2013.131204.
- [8] Nametala C A L, Souza J V D, Pimenta A, et al. Use of Econometric Predictors and Artificial Neural Networks for the Construction of Stock Market Investment Bots[J]. Computational Economics, 2022, 61(2):743-773.DOI:10.1007/s10614-021-10228-0.
- [9] Gu Z Y, Zhu Y Y, Xiang J L, et al. A prediction method of operation trend for large axial-flow fan based on vibration-electric information fusion[J].Journal of Central South University, 2021, 28(6):1786-1796.DOI:10.1007/s11771-021-4629-6.
- [10] Qin L, Li W. A combination approach based on seasonal adjustment method and echo state network for energy consumption forecasting in USA[J].Energy Efficiency, 2020, 13(7):1-20.DOI:10.1007/s12053-020-09897-x.