Application of Machine Learning in Financial Asset Price

Zixuan Pan[1](✉), Zhiqiu Wang[2], and Junming Chu[3]

[1] Department of Economics, University of California, 530 Evans Hall, Berkeley, CA, USA
Pzx0317@berkeley.edu

[2] Metropolitan College, Boston University, 1010 Commonwealth Ave, Brookline, MA 02215, USA
zhiqiu00@bu.edu

[3] University of California, Berkeley, USA
junming0119@berkeley.edu

Abstract. Asset pricing is an important part of modern financial theory. How to reveal its pricing law is an important research topic in the current financial field. Starting from asset pricing, expounds the application and development of machine learning technology in asset pricing, and classify it as a machine learning method based on feature processing, thus derive the machine learning from manual extraction features to modeling and solving application process based on hypothesis model, improve the diversity of data sources, and gradually use agents to interact with the environment, realize the reasoning path and reasoning logic, promote the application of machine learning in financial asset pricing.

Keywords: machine learning · financial investment · asset pricing · application

1 Introduction

Financial market is a complex and changeable field, and its operation law has always been valued by people [2]. Researchers have always used many methods to explore the factors affecting the stock market, resulting in the concept of "factor zoo". (Xu, 2022) There are mainly linear regression, nonlinear fitting, traditional methods, and machine learning, etc. Because there are a lot of noise and uncertainty factors in the financial market, in the analysis of complex, high dimensional and noise complex financial market data, due to the influence of nonlinear factors, the complexity of the prediction function greatly improved, so the traditional measurement and linear analysis method are not suitable for the complex, high dimensional, the noise of data analysis [1]. Machine learning, as a new technology, has been widely used in the computer, biology, medical care, media, finance and other industries. In these fields, the research of asset pricing methods based on machine learning provides new ideas for enterprise decision-making due to its advantages of high algorithm efficiency, strong applicability and easy to handle big data.

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 389–394, 2023.
https://doi.org/10.2991/978-94-6463-198-2_40

2 Overview of Asset Pricing

Asset pricing is a method of revaluation of an asset under uncertainty. The asset discussed in this paper is a financial instrument or a kind of negotiable securities, and the price is a price that reflects the comprehensive action of many factors, such as fundamentals, risks, emotions and other factors. (Jiang, 2022) Many scholars have studied such assets from different perspectives, including random walk theory, effective market assumptions, and behavioral finance. Random walk theory means that the market response to random events is the randomness of Brownian movement, namely the stock price is unpredictable, and it is regarded as a "fool game"; the effective market assumption divides the market into weak effective market, semi-strong effective market, strong effective market, and the stock price can fully reflect all effective information about an asset, but with the market reversal, momentum effect and market value effect, the effectiveness of the effective market assumption is lower and lower. The behavioral finance theory holds that stock prices are influenced not only influenced by the company's own value, but also by individual and group behaviors of investors [3].

3 Machine Learning Asset Pricing Methods

The asset pricing model described in this paper is a broad sense of asset pricing, which includes a risk-return ratio model and a direct forecast asset pricing model. Asset pricing data is divided into two types: structural and non-structural, whose quality is the maximum value affecting the model performance. Structural data includes securities trading data, futures trading data, foreign exchange trading data, macro data, market factors, scale, value, momentum, profit, turnover rate, etc. Non-structured data includes data from social media, including Facebook, Twitter, and YouTube, as well as quarterly reports, annual reports, financial intermediary analysis, financial pictures, and corporate knowledge maps of listed companies [4]. It should be noted that structural and non-structural data cannot be completely opposite and there will be crossover between them. Based on this model, machine learning models are mainly divided into machine learning methods based on feature processing and machine learning methods of end-to-end processing depth. The infrastructure of the machine learning model is shown in Fig. 1. Based on this, the author mainly discusses the application of machine learning method based on feature processing in financial investment pricing.

4 Application of Machine Learning Method in Financial Asset Pricing

Feature engineering is to use domain knowledge to extract more input features from the original data, reduce feature loss, and provide a basis for future modeling. Data features are the limit of machine learning ability. Their characteristics include data cleaning, noise processing, abnormal sample processing, data imbalance processing, data standardization processing, data discretization, and missing completion. Using machine learning technology for investment pricing can be divided into the following aspects:

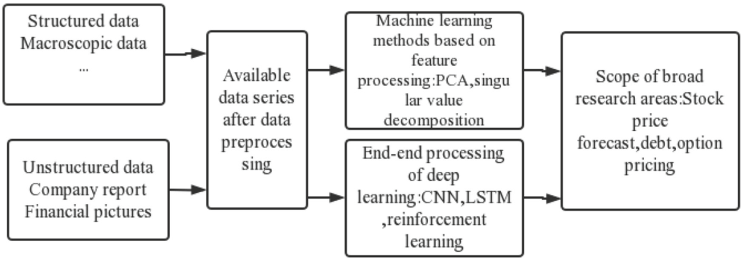

Fig. 1. Process of machine learning asset pricing algorithm

4.1 High-Dimensional Data Dimension Reduction Algorithm

The dimensional reduction algorithm for high-dimensional data mainly includes component analysis, singular value decomposition and independent component analysis. In the investment pricing, On the one hand, it combines the main component analysis with arbitrage pricing, The non-arbitrage factor in the data is illustrated. A model of predicted excess reward for the high-dimensional financial panel is also presented, Explain the causes of predicted yield and covariance structure, and introduces the arbitrage penalty term on this basis, Solving the problem of the low signal-to-noise ratio of financial data, Information on core prices was obtained (see Fig. 2, The X-axis represents the number of enhanced iterations, The Y-axis represents the error rate; Deep lines indicate that within the sample. Light color lines indicate the performance outside the sample); on the other hand, Singular value decomposition is similar to the PCA, It is also an effective way to handle the dimensionality reduction, It is widely used in machine learning, Such as feature decomposition, compression and noise reduction, recommendation system, natural language processing, etc.; besides, Singularity decomposition is not meant to decompose the matrix into a square matrix, Instead, obtained by using the elbow sub * transformation, U is a matrix of m m order, The y is the diagonal matrix of the m n order, 3 Is a singular value matrix, At a low rank, The left singularity vector and the right singularity vector array can be obtained. Singular value decomposition is often used in the dimension reduction and compression of financial data. The application of this algorithm analyzes the securities and financial markets, and finds the largest independent signal, which can minimize the risk, and get the best combination. Similarly, basic processing such as nonoisy financial data is necessary before performing ICA (PSARADAKISZ, 2018).

4.2 Mathematical and Statistical Algorithm

The proposed algorithm has a finite sample inference power on parameters and underlying conditions in financial investment pricing, and monitors the market dynamics in real time, with low sensitivity, training speed, and high interpretation. But Markov networks, as derived networks of directed graphs, are simply structured, robust, and well interpretive, largely applicable to the difficulty of labeling stock price sequences. In the literature (Li, 2019). The long-term equilibrium deviation Markov error correction model due to

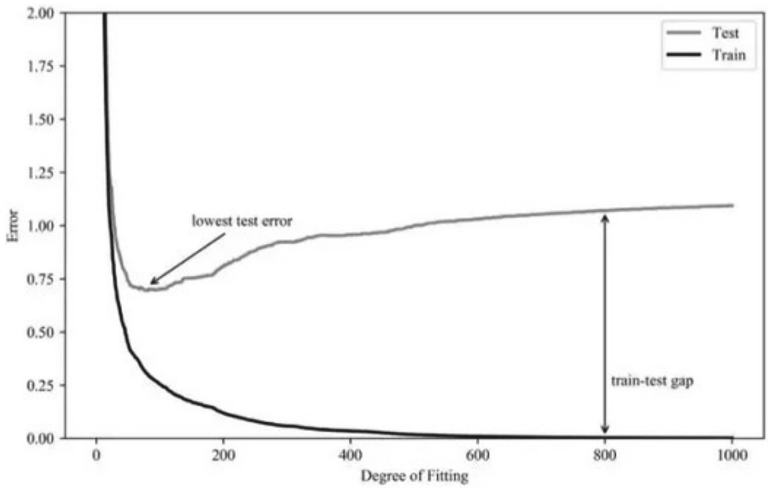

Fig. 2. Gradient-enhanced within and out of sample error rates in the regression tree classifier

different interest rate regulation is constructed. In addition, the excess return of an asset can be formulated as an additive prediction error model:\

$$\begin{aligned} & r_{i,t} + 1 = E_t\left(r_{i,t+1}\right) + \in_{i,t+1}, \\ & \text{among } Et\left(r_{i,t+1}\right) = g\star\left(z_{i,t}\right) \end{aligned} \tag{1}$$

In other words, our purpose is to separate an E t (ri, t + 1) The representation of, which is a function of the predictor, and for ri, t + 1There is the largest out-of-sample explanatory power available. Where the predictor is the high-dimensional zi, t, The functional form can be a flexible conditional expected yield function g ⋆, which here assumes that the function form is independent of neither i nor t [5].

To capture the properties of these data, a two-stage error correction model based on Markov correction, which combines high price odds ratios and long-term equilibrium properties, provides the accuracy of market data for asset pricing.

4.3 High and Low-Dimensional Conversion Algorithm

Data in the financial market has a large dimension, and it is usually the practice to convert the high-dimensional data into low-dimensional data, classify and predict the low-dimensional data, and then restore it to high-dimensional data, thus solving the problem of dimensional disaster and over-learning. Autoencoders and S VM are one of them. Unlike the P C A, the autoencoders and the SV M employ a nonlinear approach to reduce the dimensionality. The method includes the model encoding and the decoder in two parts, using the neural network for feature extraction, and to input it into other classification models. Automated encoder is suitable for processing high-dimensional data, it represents high-dimensional features as low-dimensional features, is an unsupervised algorithm for compressed data, by a small number of neurons in the hidden layer to form input variables, and by the nonlinear activation function g ⋆ Implement the transformation of variables, similar to the P C A algorithm, but it can use the neural network to achieve a nonlinear mapping, to make it more flexible and widely used. Traditional genetic coding algorithm cannot be both features and risk-benefits, so adopted the automatic encoder, revenue into low-dimensional factor set, and introduced conditional automatic encoder model, allowing the nonlinear effect on factor exposure, and using covariant term neural network to nonlinear feature information into beta format, to establish an automatic encoder, with economic guidance for prediction, and can "prepare" other algorithms, such as the processed data into other algorithms. (Wang, 2020).

4.4 Traditional Machine Learning Classification Theory Algorithm

Decision trees and random forests are widely used analysis in financial markets. A decision tree is a set of systems consisting of multiple decision rules that divide multiple attribute spaces into several non-cross subregions, while the leaf nodes represent different classification effects. Random forest is a decision tree synthesis algorithm suitable for UHF transactions in large-scale data. In the case of a large number of candidate variables, the random forest method is used to optimize the threshold selection of each decision tree, thus obtaining a tree-based state portfolio ranking, thus smoothing the boundary parameters of the multidecision tree, and thus improving the effectiveness of the prediction. The empirical results show that random forest is better than gradient enhancement trees and deep neural networks. Capital market abnormalities are mainly based on monthly data, based on recent income and daily data, based on historical closing price, dividends and earnings, and the decision tree is used to predict the sensitivity index of buying stocks.

4.5 Heuristic Algorithm

Particle swarm algorithm is a group intelligence algorithm, which imitates the foraging behavior of birds. Within a certain range, it uses the sharing of information to find the largest food source (optimization), and then it is iteratively solved by the random solution. Traditional financial algorithms do not fully take into account the realistic constraints such as short selling mechanism, transaction cost, investment ratio and investment friction. If linear or nonlinear equations are used to solve the constraints, the efficiency will be reduced, and heuristic microgroups and genetic algorithms can be used to solve these problems. Usually, both methods are combined with other models, such as deep learning. The main function of this method is to find the best parameters to match the model, to reduce the adjustment of the parameters, and to get a good application in the financial investment pricing.

5 Conclusion

By comparing the above machine learning algorithms, we can conclude that:

The ① financial time series is unbalanced and has strong noise. Due to reduce data noise and uncertainty, manual extraction of financial features are generally used, which requires a deeper understanding of the financial market. Data features extracted from the technical, fundamental and information aspects lead to different results, which restrict the effectiveness of the model. In addition, the actual external constraints, such as transaction costs, liquidity, market sentiment, and investor psychology, cannot be described by the model.

② uses the method of multivariate analysis to achieve better results by using preprocessing (PCA and ICA) for the multidimensional collinearity problems caused by matrix decomposition and dimension reduction by concentrating macroeconomic variables such as exchange rate, interest rate and money supply in one model.

③ price of multifactorial assets; excessive duplication of factors can lead to factor failure. The factors selected by different financial data will make the different factors selected by different portfolios, thus affecting the effect of calculation.

④ In the traditional machine learning prediction methods, if the market occurs in a bull-bear conversion, the algorithm using a fixed mode for data fitting will fail.

References

1. Xu J , Zhu Y K, Xing C X. A Review of the Application of Machine Learning in Financial Asset Pricing [J]. Computer Science, 2022,49 (6): 276-286.
2. Jiang F W, Xue H, Zhou M. Does big data improve the pricing power of multi-factor models?—— research on China's A-share market based on machine learning method [J]. System Engineering Theory and Practice, 2022, 42 (8): 2037-2048.
3. PSARADAKISZ, S0LA M, SPAGNOLOF. On Markov error correction models, with an application to stock prices and divi—dends [J]. Journal of Applied Econometrics, 2018, 19(1): 69–88.
4. Li Y L, Wu C, Wang K S, Yuan W J. A Unified Perspective of Financial Asset Pricing and Its Application in Risk Management [J]. Practice and understanding of mathematics, 2019,40 (24): 60-67.
5. Wang X F. Comparative study of financial asset pricing and ordinary commodity pricing [J]. Journal of the Central University of Finance and Economics, 2020 (03): 28-32.