An Integrated Machine Learning Framework for Stock Price Prediction

Quanzhi Bi[1][0000-0002-4085-9704], Hongfei Yan[1,3][0000-0001-5914-8585], Chong Chen[2][0000-0002-9704-1575], and Qi Su[1][0000-0002-4769-2812]

[1] Peking University, Beijing, P.R. China
{biquanzhi, yanhf, sukia}@pku.edu.cn
[2] Beijing Normal University, Beijing, P.R.China chenchong@pku.edu.cn
[3] National Engineering Laboratory for Big Data Analysis and Application Technology, Center for Big Data Research, Peking University, Beijing, P.R.China

论文摘要

Abstract. Predicting the future price of financial assets has always been an important research topic in the field of quantitative finance. This paper attempts to use the latest artificial intelligence technologies to design and implement a framework for financial asset price prediction. The framework we use is divided into three modules: Feature Engineering, Regressor, and Hyper Optimizer. The Feature Engineering module extract multiple features using technical indicators, FinBERT, FFT, ARIMA, stacked auto-encoder, PCA and XGBoost. The Regressor module consists of a generative adversarial network, where the generator is Seq2Seq and the discriminator is GRU. The HyperOptimizer module will tune the parameter in GAN using the Bayesian optimization algorithm. Finally, we conducted numerical experiments on our framework, which shows that the framework implemented in this paper performs better than the benchmark method.

Keywords: Stock Prediction · Stacked Autoencoder · NLP · GAN · Bayesian Optimization.

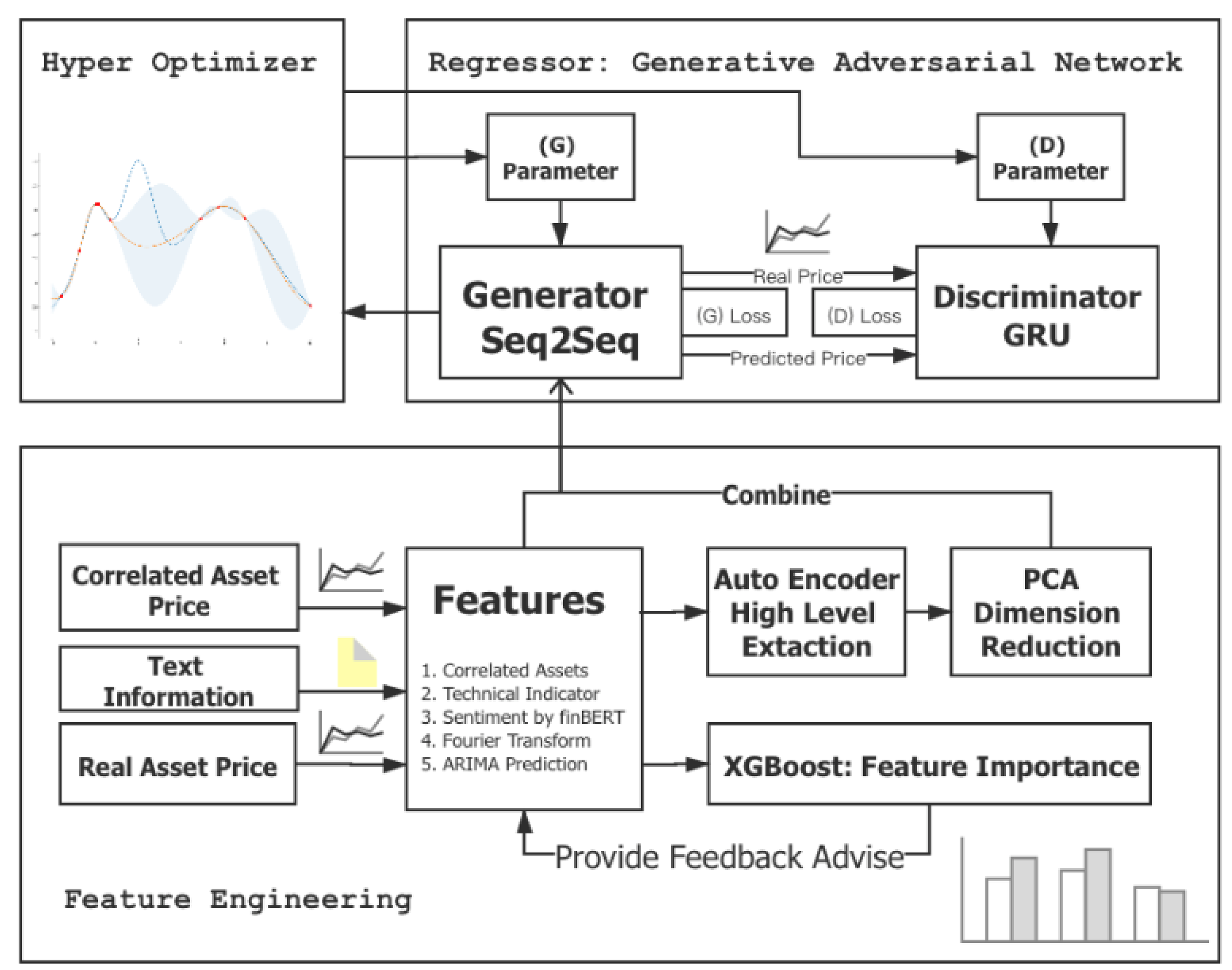
系统模型

1. Implemented and improved the framework for financial asset price forecasting, the raw idea was originally proposed by [3].
2. Integrates multiple traditional technical indicators
3. Combines a characteristic mix of underlying assets, including commodities, currencies, indices, VIX, and more
4. Uses the latest NLP model FinBERT[2] to conduct sentiment analysis of financial-related news in the market.
5. Uses Fourier transform techniques to extract the overall trend of price changes
6. Extract of asset price similarity using traditional ARIMA model
7. Uses Stacked Autoencoders to identify other advanced features
8. Uses principal component analysis PCA to create feature portfolios to reduce the number of dimensions of features created by the self-coder
9. Uses the currently most popular statistical learning model, XGBoost[4] to analyze feature importance
10. Forecasts asset prices using the generative adversarial network GAN, where the generator is the latest Seq2Seq and the discriminator is GRU
11. Uses the Bayesian optimization algorithm BOA for GAN tuning, resulting in a significant reduction in tuning time
12. tests 147 select US stocks

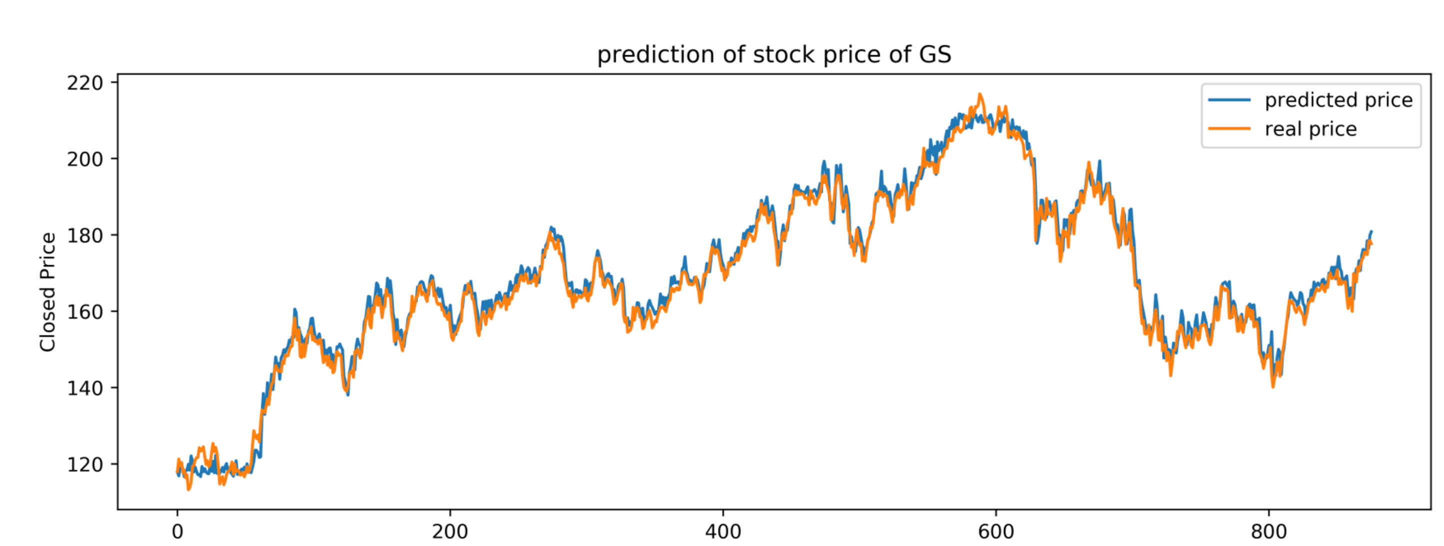
论文简介

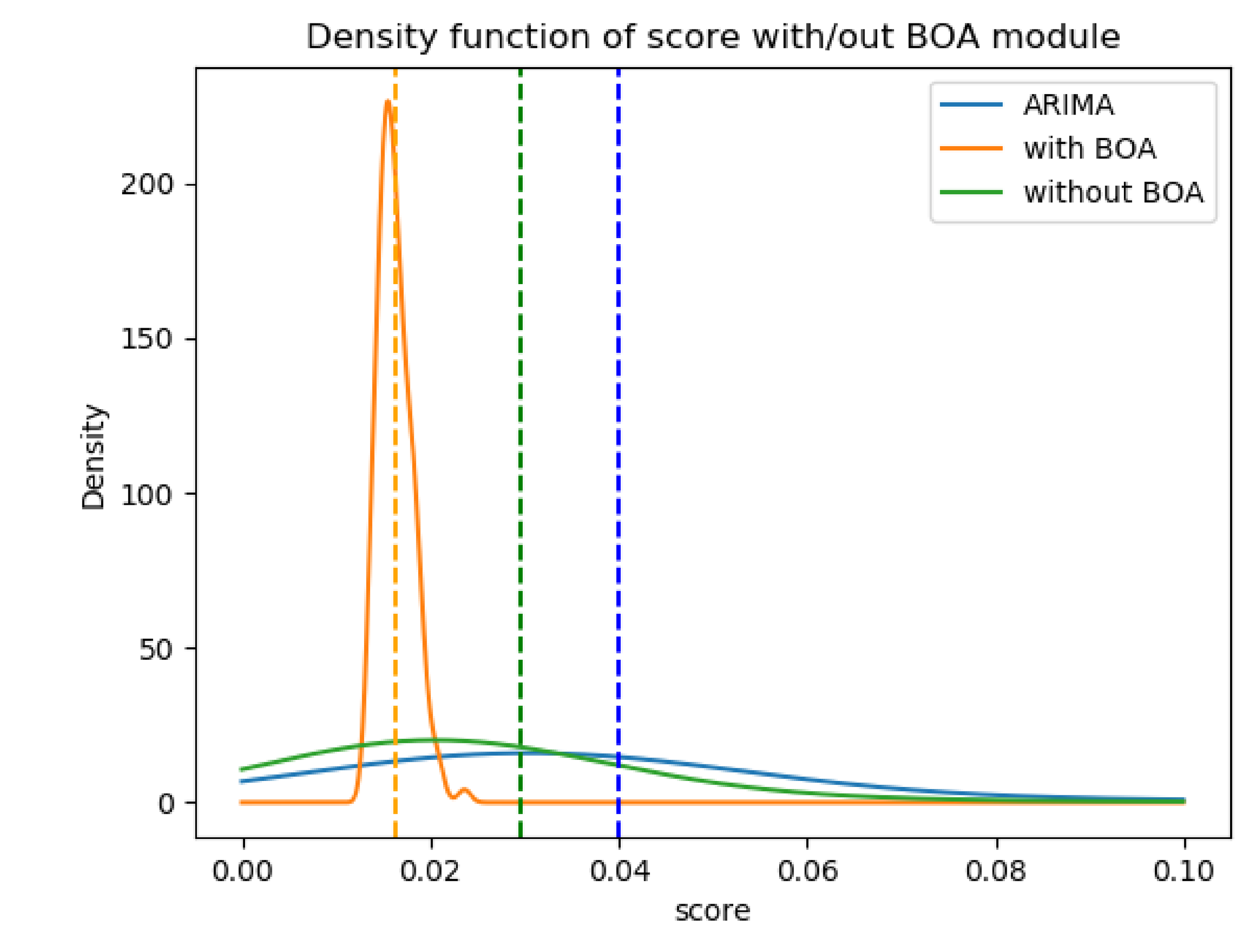
* Supported by: NSFC Grant 61772044; MSTC Grant 2019YFC1521203: research, development and demonstration of key technologies for knowledge organization and services for Antiques based on Knowledge Graph; Peking University Grant 2020: "New Ideas for Teaching 2.0" key project.

算法原理



实验仿真





论文结论

The main work and conclusions of this paper are as follows.

1. Implements and improves upon the framework proposed by [3], where the improvements are: using FinBERT, using SeqSeq, running multiple underlying assets, etc.
2. Uses data from Yahoo Finance and SeekingAlpha, combined with the latest AI technology and traditional methods, extracted a number of artificial and deep-seated characteristics, including related assets, technical indicators, sentiment values, DFT, ARIMA, and used SdA and principal component analysis for dimension reduction.
3. Uses GAN to learn and predict the generated feature sequences, incorporating sequence processing techniques such as SeqSeq and GRU, which have become more popular in recent years.
4. Uses a Bayesian optimization technique to perform global black-box tuning of the GAN parameters, achieving a rapid implementation of hyperparameter tuning.
5. Numerical experiments on 147 U.S. stocks were conducted to verify the validity of our proposed feature engineering module, regressor module, and hyper optimizer module. Our module has been proved to out-perform the benchmark.

西安電子科技大學
XIDIAN UNIVERSITY