MATH5470: Paper Reproduction: (Re-)Imag(in)ing Price Trends

HUANG Zhanmiao, SHEN Xuanyu {zhuangdj,xshenar}@connect.ust.hk

Department of Mathematics, HKUST

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Introduction

We reconsider the idea of trend-based predictability using methods that flexibly learn price patterns that are most predictive of future returns, rather than testing hypothesized or pre-specified patterns (e.g., momentum and reversal). Recent work [1] explores machine learning techniques to predict future returns of stocks using past price data, leveraging convolutional neural networks (CNNs) to analyze two-dimensional images encoding past price dynamics. This project replicates and extends experimental results from the paper [1]. Extensions include providing Grad-CAM visualizations [3] for model interpretation.

Dataset

The sample runs from 1993-2019 based on the fact that daily opening, high, low prices. In the original paper, authors construct datasets consisting three scale of horizons (5-day, 20-day, 60-day). Here we just collect the 20-day version. The total size of data is 8.6G in a zipped file (802.9MB). The download link of data is: https://dachxiu.chicagobooth.edu/download/img_data.zip.

Data Preprocessing

The dataset has transfer 1-d time series to 2-d images of shape 64×60 (Fig. 1), which contains the information of price (daily open, close, high, low and a moving average price) and volume bar (daily trading volume), and a table with information about stock ID, date and its future return in 5 days, 20 days and 60 days (Fig. 2). We choose the return in 20 days as our label, when return is positive, then it's labeled as 1, otherwise 0. To reduce the time cost, we chose data from year 1993, 1998, 2003, 2008, 2013, 2018 as training data, data in year 1999, 2010 as validation and data in year 1994, 2000, 2004, 2009, 2014, 2019 as testing data. Note that such data separation pattern is different from the paper [1] after we verified that this would trigger higher classification ability on test set.

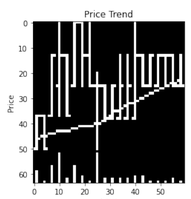


Fig. 1: Example input image of Price Trend

	Date	StockID	EWMA_vol	Retx	Retx_5d	Retx_20d	Retx_60d
0	2017-01-31	10001	0.000450	0.000000	4.370390e-07	-0.000002	-0.011860
1	2017-02-28	10001	0.000180	-0.003937	3.951997e-03	-0.003162	0.003953
2	2017-03-31	10001	0.000064	0.007936	-7.874612e-03	-0.015749	0.015748
3	2017-04-28	10001	0.000030	0.000000	9.999881e-03	0.016001	0.032002
4	2017-05-31	10001	0.000015	0.000000	4.370390e-07	0.015748	NaN

Fig. 2: Table of future return

Methodology

A Deep Convolution Neural Network (CNN) is trained to do the binary classification of future 20-day return prediction. The workflow follows from basic procedure of training, model tuning, and finally prediction.

- Input: 20-day OHLC in image of 64×60 grey scale.
- Output: binary label in {0, 1} with 1 represents positive return in future 20 days otherwise 0 for negative return as prediction.

Training Details

Network Structure

We use a Convolution Neural Network to make such binary classification, since the input is an image. The details of the network is summarized in the following figure. For each orange block, the number in the left $a \times b$ is the size of kernel, number in the right is the number of output channels. For Max-Pool layer, the number in the left is the kernel size. Then the output of the last convolution block is flattened to a vector, then passed a fully-connected layer and Softmax activation to become the probability of the binary labels.

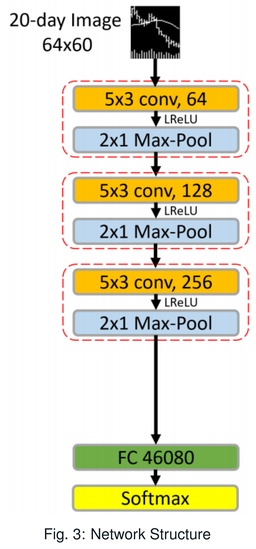
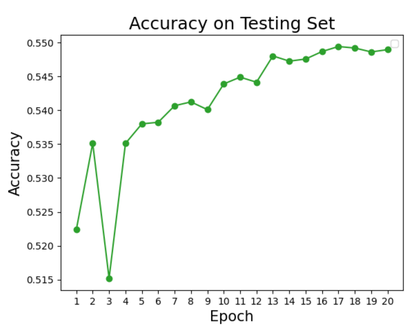


Fig. 3: Network Structure

Prediction Performance

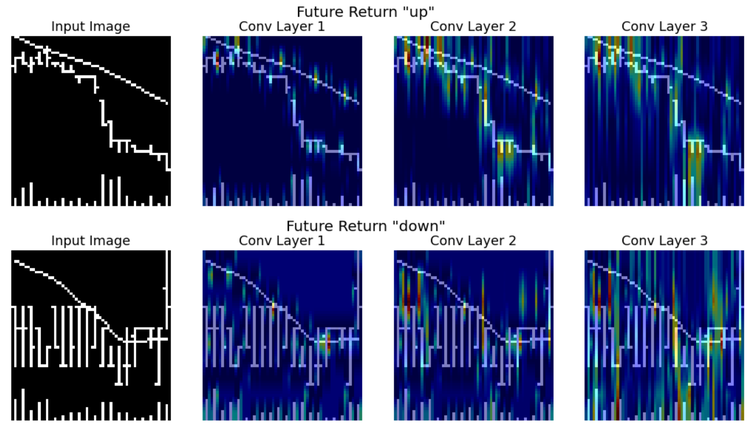
The loss on the training data steadily decreases over 20 epochs, while the loss on the validation data remains relatively constant. We then achieve prominently higher accuracies on the total testing set, around 52% before training and 55% after training, an increase of 3%. In particular, the accuracy can be as high as 57% on some randomly selected subsets of the testing data.

Our prediction accuracy (55%) is observed to be higher than other open source results (on Github etc. [2]) which are only 53% or so after 20–30 epochs training. It is reasonable to owe to be our data separation method that crosses the years.



Visualization and Interpretation of CNN

To visualize and interprete the CNN model, we employ Grad-CAM (Gradient-weighted Class Activation Map, [3]), which can produce images processed by each layer of the CNN and illustrate the most important features in triggering up or down return predictions. In the figures below, the images in each row are represented in the order of CNN layers. Input image is shown as greyscale as processed in CNN, while the images in convolutional layers are are shown in RBG format. The brighter regions correspond to higher activation. By observation, the special transition in price trend and high volume bar are commonly striking in colors, which may imply noticeable state transition or policy change in the market and influence the next movement.



Conclusion

Our network is predicting the sign of future return in 20 days, which may help investor in decision-making on investment. To make general predictions during a long range of time, we select the training data that is evenly distributed in time, one can find our model can achieve nearly 55% accuracy on testing data which is also evenly distributed in time, this result can be compared with previous result [2] which achieve 53% on their testing data. To make the result more solid, some additional simulation like cross-validation should be done, these additional simulations are canceled due to limitation of time.

References

[1] Jingwen Jiang, Bryan Kelly, and Dacheng Xiu. "(Re-) Imag (in) ing Price Trends". In: *The Journal of Finance* 78.6 (2023), pp. 3193–3249.

[2] Aoran Li et al. "Paper Reproduction: (Re)Imag(in)ing price trend". In: *MATH5470* (2023).

[3] Ramprasaath R Selvaraju et al. "Grad-CAM: Why did you say that?" In: *arXiv preprint arXiv:1611.07450* (2016).

Contribution: Neural Network Training: Xuanyu Shen, Grad-CAM: Zhanmiao Huang, Poster: All members