MAFS 6010Z Project2: Paper Replication—Reimaging of Price Trend

Aoran Li, Yijia Ma, Tianying Zhou, Langting Weng

October 29, 2023

Abstract

In this experiment, We have re-implemented the main idea of Jiang, Jingwen and Kelly, Bryan T. and Xiu, Dacheng in their paper (Re-)Imag(in)ing Price Trends. For the data-set, We only use part of the original paper's data, which include the 20 days OHLC data image and the corresponding labels (ret20). The images are two dimensions array, which represent the candle-list figures of previous 20 days when displayed in grey-figures. For the models, we rigorously follow the architecture and hyper-parameters of the original paper. We have implied the 20-day CNN model proposed by the authors. Apart from the baseline 2 classification label, we implement the same model on further 2 different classification labels, which are ret20 up/down/na and ret5 up/down. So there are total 3 kinds of models in our whole project, i.e. 20D3C, 20D2C, 5D2C. For the training and inference workflow, we mainly follow the setting of original paper. However, due to limitation of computation resource, we only tried 3-fold model combination. In the end, we apply our models on the test data-set and discussed the further improvement of our study. The project code is available on My Github Repo. Please feel free to give a star if you find it useful.

1 Data Set

In the original paper, authors construct datasets consisting 3 scale of horizons(5-day, 20-day, 60-day). In this project, We mainly focused on the 20-day image data(data resource). Followed by the original paper's design, data from 1993 to 2001 were utilized as the training set for model training and validation purposes. Subsequently, the data from 2002 to 2019 were employed as the test set in order to predict and evaluate the efficacy of the model. The training set encompasses a total of 3.16G with 885004 items, while the test set comprises a total of 4.69G with 1311990 items.

The input image shape is (64, 60) with only one channel. Some example image pictures is shown in Figure 1.And the corresponding image label is binary classification number that represents the positive or negative return in future periods. We have trained 3 models on different labels.For a detailed label design, please refer to the model construction part.

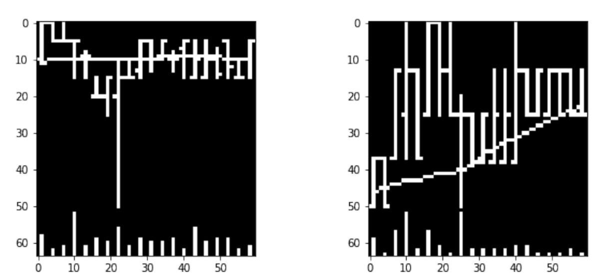


Figure 1: Examples of 20-day Image with volume bar and moving average line

2 Model Construction

The model architectures are exactly the same as in the original paper, which is shown in Figure 2.

Apart from the basic layer architecture, we also customize some parameters in CNN layer in order to have same input shape in Linear layer, such as *padding parameter* in *nn.Conv2d*. The detailed construction is shown in Figure 3. The total parameters of our model is 1,417,732.

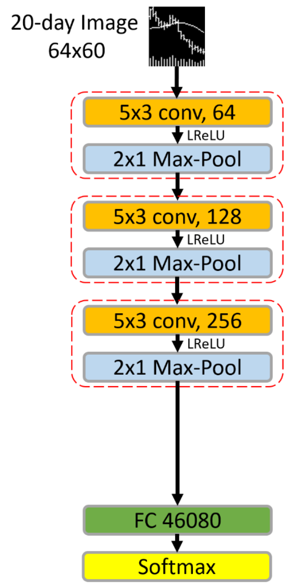


Figure 2: 20-day CNN model design

```
CNN20DModel(
  (block1): Sequential(
    (0): Conv2d(1, 64, kernel_size=(5, 3), stride=(3, 1), padding=(3, 1), dilation=(2, 1))
    (1): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): LeakyReLU(negative_slope=0.01, inplace=True)
    (3): MaxPool2d(kernel_size=(2, 1), stride=(2, 1), padding=0, dilation=1, ceil_mode=False)
  )
  (block2): Sequential(
    (0): Conv2d(64, 128, kernel_size=(5, 3), stride=(1, 1), padding=(2, 1))
    (1): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): LeakyReLU(negative_slope=0.01, inplace=True)
    (3): MaxPool2d(kernel_size=(2, 1), stride=(2, 1), padding=0, dilation=1, ceil_mode=False)
  )
  (block3): Sequential(
    (0): Conv2d(128, 256, kernel_size=(5, 3), stride=(1, 1), padding=(3, 1))
    (1): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): LeakyReLU(negative_slope=0.01, inplace=True)
    (3): MaxPool2d(kernel_size=(2, 1), stride=(2, 1), padding=0, dilation=1, ceil_mode=False)
    (4): Flatten(start_dim=1, end_dim=-1)
    (5): Dropout(p=0.5, inplace=False)
  )
  (fc): Linear(in_features=46080, out_features=2, bias=True)
```

Figure 3: Detailed design

3 Model Training

3.1 Training Workflow

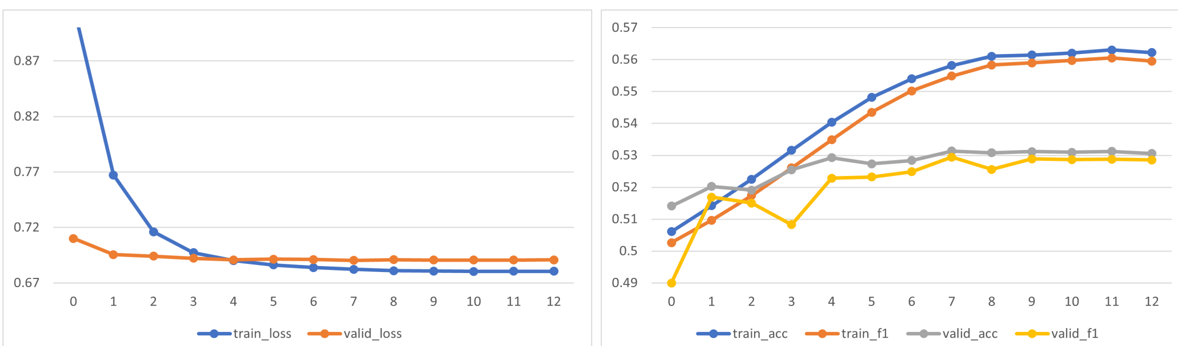
In the model training phase, we adhere to the procedure outlined in the original paper, such as the **xavier initiation**, **batch-normal&drop out design** and **early-stopping mechanism**. However, due to constraints in computing resources and the limited GPU usage time on Kaggle, we have simplified some of the initial training design. Furthermore, in order to enhance training stability and bolster model robustness, we have implemented certain process improvements. The disparities between our approach and that described in the original article are summarized in Table 1.

	Our Design	**Original Paper**
Learning rate	1e-4, with Cosine decay function	1e-5
Optimizer	AdamW, with 0.001 weight decay	Adam
Train-Valid Split	stratified 3-fold splitting, each fold take 33.3% data as validation set	randomly select 30% of training data as validation, rest 70% to train model
Model Stacking	each model do the 3-fold training and average the 3 sub-models results to predict	Each CNN model retrained 5 times and do average forecast
Label	- binary classification of future 20-day return; - binary classification of future 5-day return; - three-classification of 20-day return(2 for NA return)	binary classification of future 20-day return –1 positive, 2 negative

Table 1: Difference of Training Design

3.2 Training Results

Based on the model architecture mentioned above, our final model boasts a total of 1,417,732 parameters. It requires 16 minutes to train a single epoch on the Kaggle P100 GPU and moreover 12 hours for the entire 3-fold modeling workflow. Here we use the 20D2C(2 classification label of future 20-days return) model as an example. Figure 4 demonstrates the changing of the training&validation loss over the epochs, as well as its accuracy and macro f1 score on the valid set. It is clear that the model is close to convergence after 8, 9 epochs.



(a) loss (b) acc & f1 score

Figure 4: Training Results of Model 20D2C

4 Model Performance

4.1 Testing Results

To assess the model's performance on the test set, we predicted the three models on the test data and evaluated the results using ACC, F1_SCORE, and other metrics. The metrics results are shown in Table 2. Our findings indicate that the addition of NA values to the labels does not significantly impact the accuracy of the model. However, the model utilizing future 5-day earnings as labels appears to exhibit higher accuracy.

Furthermore, we examined the model's performance across different years and observed a notable decline in accuracy for the time period far removed from the training set (2011-2019). This may suggests that the model's predictive power diminishes over time, indicating limited long-term forecasting capabilities.

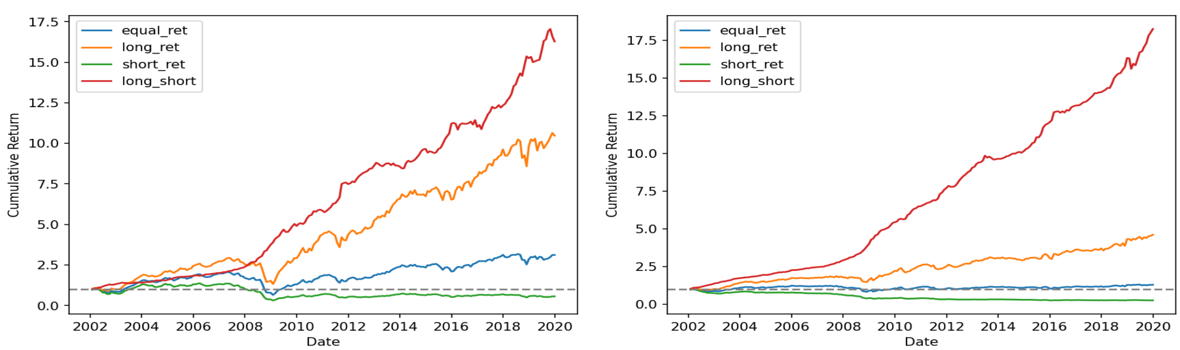
	20D2C	20D3C	5D2C
Loss	0.693	0.728	0.693
Accuracy	0.5274	0.5215	0.5339
Macro f1 score	0.527	0.3569	0.5232
True Positive Rate(1/1)	0.5305 (362083/682552)	0.5 (341275/682552)	0.673 (447003/664242)
True Negative Rate(0/0)	0.524 (325009/620300)	0.5526 (342784/620300)	0.3905 (251472/644052)
Acc: 2002-2005	0.5237	0.5159	0.5438
Acc: 2006-2011	0.5435	0.5383	0.5346
Acc: 2011-2019	0.5163	0.511	0.5273

Table 2: Model Testing Results

4.2 Backtesting Performance

In addition to model evaluation metrics such as the acc and f1 score, we build the stock portfolio based on the model prediction signal. For each timestamp, we select the top 10% stocks with the strongest uptick signal as the long portfolio(orange line), while the short portfolio comprises the top 10% stocks with the strongest downward signal(green line). Finally, the returns of the model long-short portfolio(red line) and the market average portfolio(blue line) were compared. The long-short return plots of the groups are shown in Figure 5.

As evident from the figure, the model-based portfolio exhibits significant return ability in the long run. However, it appears that the majority of the returns come from the long end, indicating that the model's short-selling prediction ability is relatively weak. This suggests that there may be room for improvement in the model's short-side predictions, which could potentially enhance the overall performance of the portfolio.



(a) CNN20D2C

(b) CNN5D2C

Figure 5: Back-testing Results of Model 20D2C & 5D2C

5 Conclusion and Future work

It's clearly that CNN network does have some ability to learn from the stock's K-line plot and give some plausible predictions on future return. Although there is obviously ample room for further improvement of the model's accuracy, this paper has provided us with valuable inspiration on how to apply image models to time-series data learning. Due to the time limitation and the restrained computing resources, we can not fully exploit our research on this direction. However, we have outlined some potential challenges and avenues for future exploration in this paper.

- **Model Design**

- This paper's architecture is basically derived from the VGG model. This design makes it harder for the network to go deeper. Therefore, we suggest try some ResNet type architecture(like RepVGG) and multi-parallel channel structure to build the model, which may enhance the model ability.

  - Our replication model's total parameters is about 140k, and the training data-set's length is only 885k. The low parameter-to-sample ratio can probably cause the model over-fitting, which may partially explain the poor performance on the test data-set. Combining a vision model with a sequential model such as LSTM or Transformers through a seq2seq structure may help mitigate this issue.

- **Label Design**

  - Only using future return as an indicator may not be a good choice for model's learning. Because the stock's future price may be go up intensively and then falls rapidly. In this scenario, although the future return may ends up positive, it's hard to say is good to holding this stock right now. Therefore, simple binary classification may not be a good choice for model's learning and it may even confuse the model. Thus, we may take the risk and max draw down into consideration in future label design.

- **Loss Design**

  - Simple Cross Entropy loss function may be too simple for model's learning. Maybe we can take the boosting loss design into consideration, which means give more weight to the wrong-predicted sample after each training.

- **Training Design**

  - Periodical Recursive training may be helpful to the future prediction since the model can learn some recent information.

6 Contribution

- Coding & Modeling: Aoran Li

- Report Writting: Aoran Li, Yijia Ma, Tianying Zhou, Langting Weng

7 References

- (Re-)Imag(in)ing Price Trends. Jiang, Jingwen and Kelly, Bryan T. and Xiu, Dacheng