Entropy

CAO KY HIEU

August 2019

Contents

1 DATA EXPLORATION **1**
- 1.1 Price path of ABT ticker and its Simple Moving Average(SMA) . . . 1
- 1.2 VNINDEX SMA path . . . 1
- 1.3 Relative Strength Index(RSI) . . . 2
- 1.4 Price path and 60-day Sharpe Ratio of ABT . . . 2
- 1.5 Seasonal Effect . . . 3
- 1.6 Test the daily return of ABT and VNINDEX . . . 4

2 FEATURE ENGINEERING **5**
- 2.1 Feature extraction . . . 5
- 2.2 Feature Selection . . . 7

3 MODEL SELECTION **7**
- 3.1 Define the problem . . . 7
- 3.2 Loss Function . . . 8
- 3.3 Architecture of Model . . . 8

4 Tuning Parameters **10**

5 Discussion **10**

6 SECTION 2 - UPDATE MODEL **10**
- 6.1 Innovation . . . 10
- 6.2 Improvement . . . 11
- 6.3 Weakness of model . . . 11
- 6.4 Conclusion . . . 14

1 DATA EXPLORATION

1.1 Price path of ABT ticker and its Simple Moving Average(SMA)

Price path of ABT ticker and its Simple Moving Average(SMA) in Figure 1

Using SMA makes the path smoother. Apperently, the SMA does not fluctuate like the price path. SMA just simply take average precious days (5 days) values for the next day.

1.2 VNINDEX SMA path

Similarly the SMA and price path of VNINDEX is illustrated in figure 2.

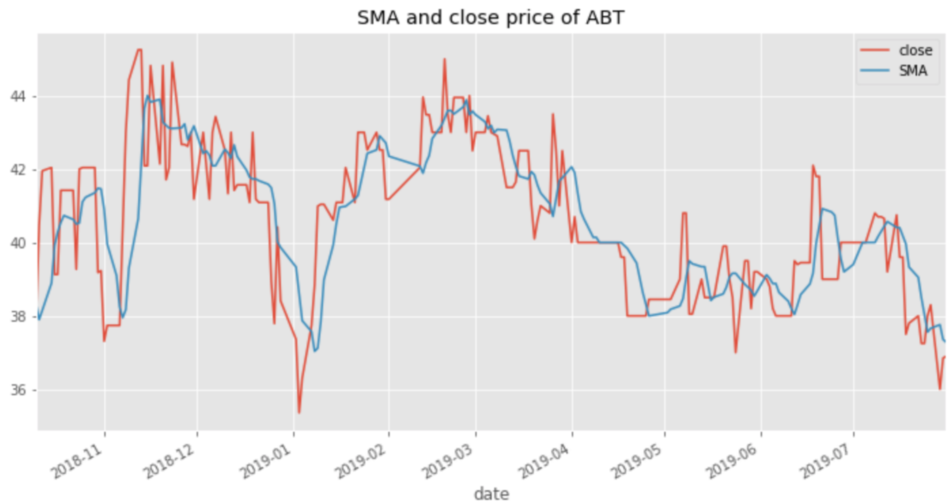

Figure 1: window = 5 for SMA and data in 200 nearest days

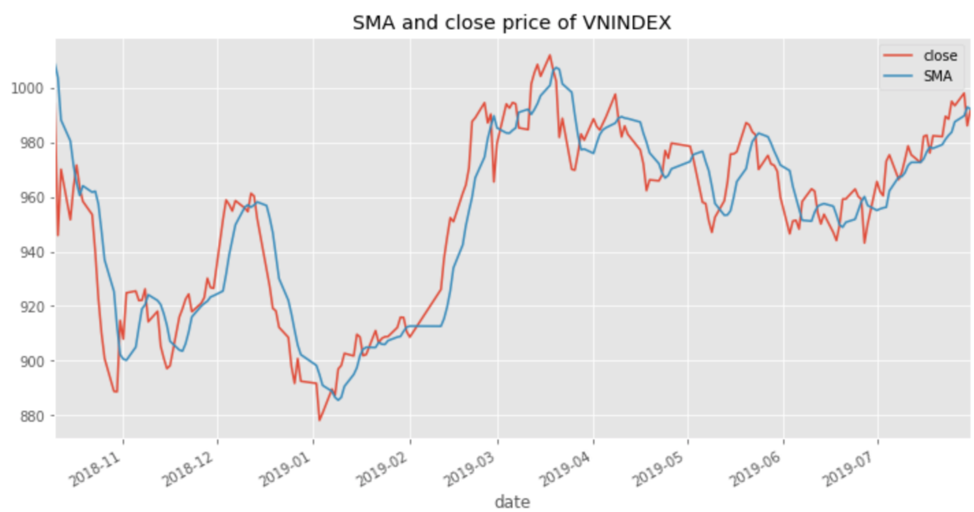

Figure 2: window = 5 for SMA and data in 200 nearest days

1.3 Relative Strength Index(RSI)

stocks which have had more or stronger positive changes have a higher RSI than stocks which have had more or stronger negative changes .This index was developed by J. Welles Wilder, who believed that when RSI above 70 or below 30, it is the strong indication of market reversals. We can see that in figure 3. In this picture, I use exponentially weighted moving average(EWMA) for smoothing.

1.4 Price path and 60-day Sharpe Ratio of ABT

Price path and 60-day Sharpe Ratio of ABT in figure 4. The maximum value of Sharpe ratio in this figure is 2.22. Sharpe ratio is the measure ratio the **Expected Return** and **volatility**. It helps to balance between Risk-return trade off to chose the best portfolios.

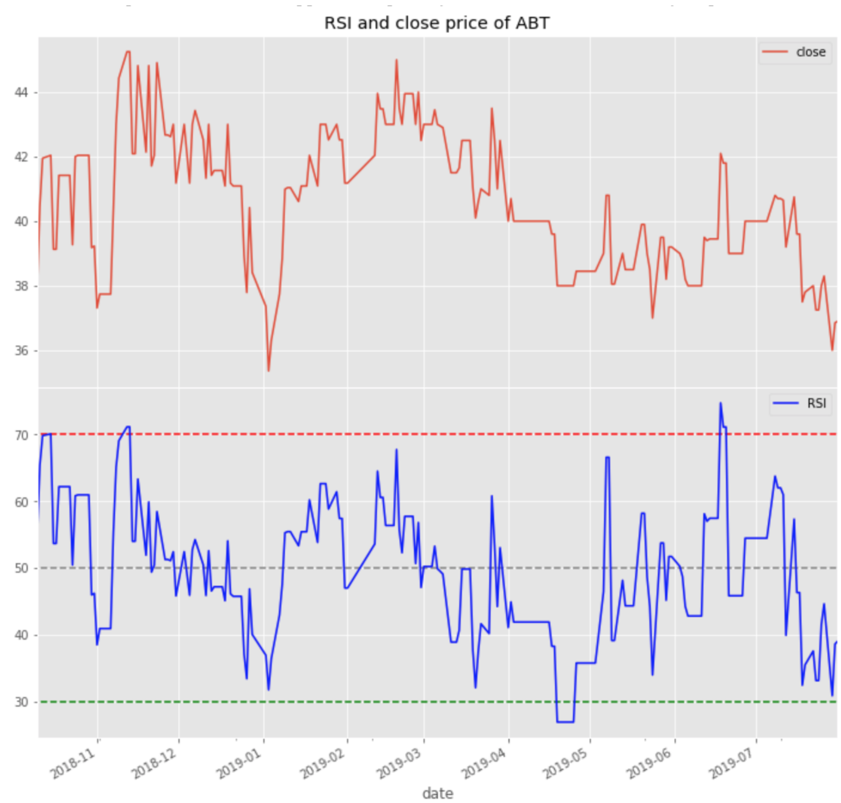

Figure 3: 22-day SRI of ABT ticker

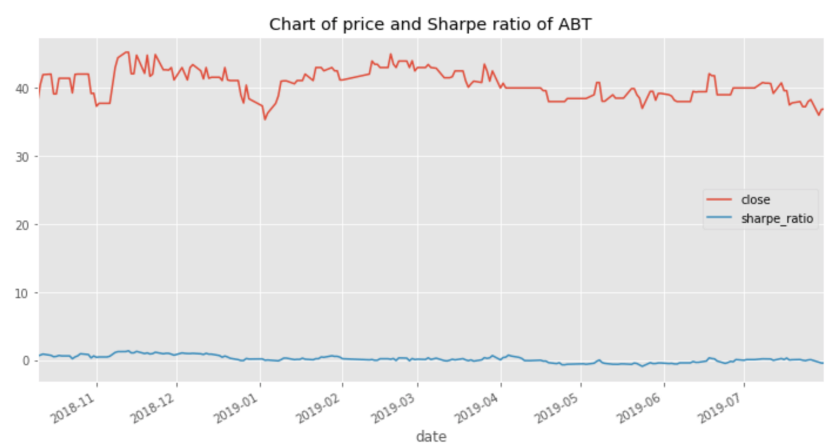

Figure 4: 60-day Sharpe ratio of ABT ticker

1.5 Seasonal Effect

In order to evaluate the seasonal effect in price path of ABT ticker, I use Seasonal and Trend decomposition using Loess(STL) to decompose $y_t = T_t + S_t + R_t$ like figure 5.

The magnitude of seasonal part is just in range from -0.5 to 0.5, smaller than residual (-2 to 2), so the effect of seasonal is not much to price.

Similarly, to evaluate the seasonal effect to vietnam market, I evaluate VNINDEX and VN30. The results in figure 6 and figure 7

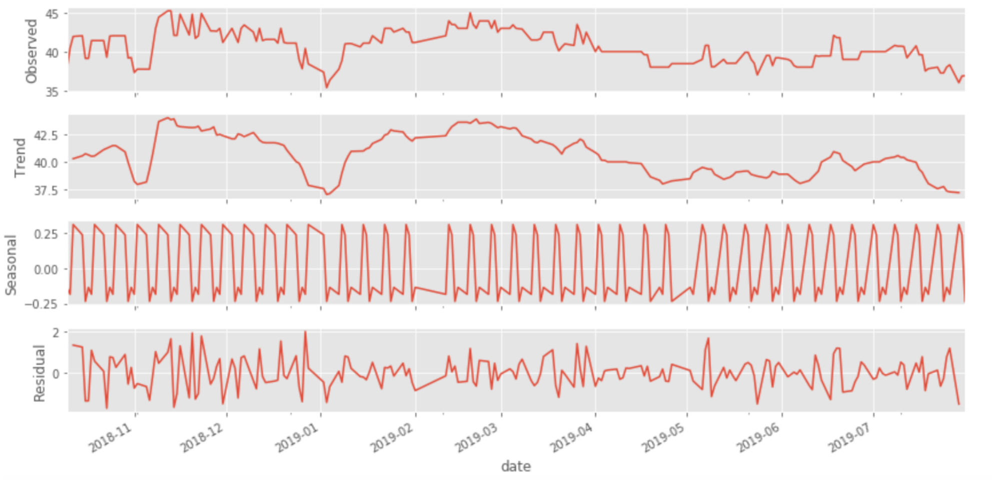

Figure 5: STL ABT ticker price with frequency = 5

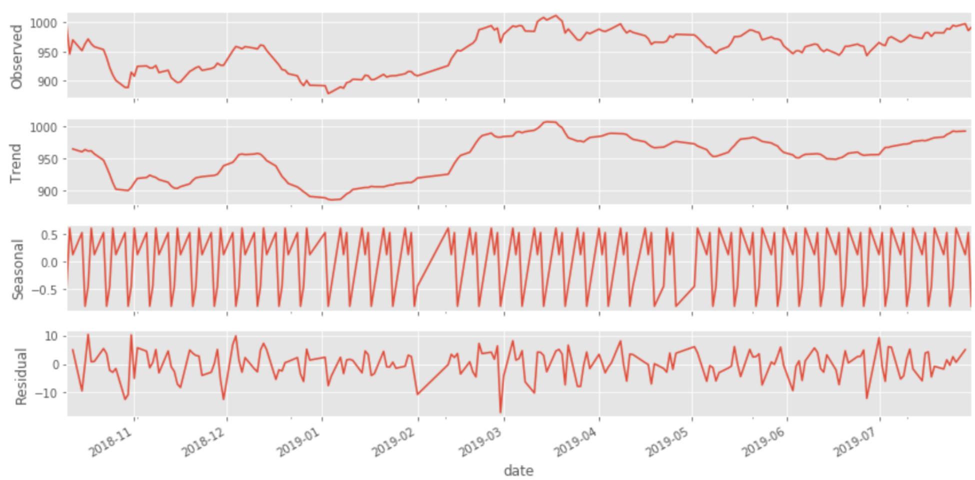

Figure 6: STL VNINDEX with frequency = 5

1.6 Test the daily return of ABT and VNINDEX

To compare the ABT daily return with VNINDEX daily return, I do one-side t-test for two mean which unknown the means and unknown the variance.

$$H_0 : \mu_{ABT} = \mu_{VNINDEX}$$

$$H_1 ; \mu_{ABT} > \mu_{VNINDEX}$$

The p-value of this test is 0.55. Because p-value is high so we can not reject H_0. It means there is no significant difference between the two mean daily returns. We can check the distribution of return in this two tickers in the figure 8

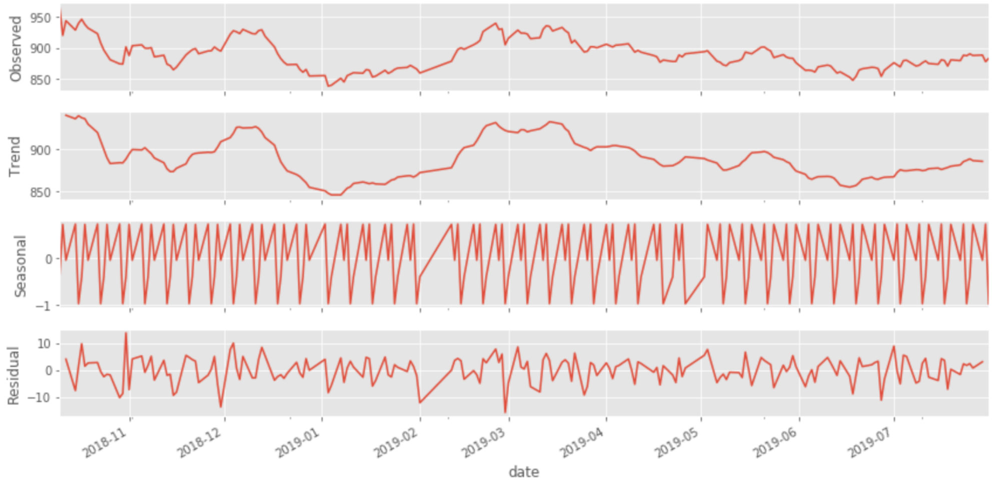

Figure 7: STL VN30 with frequency = 5

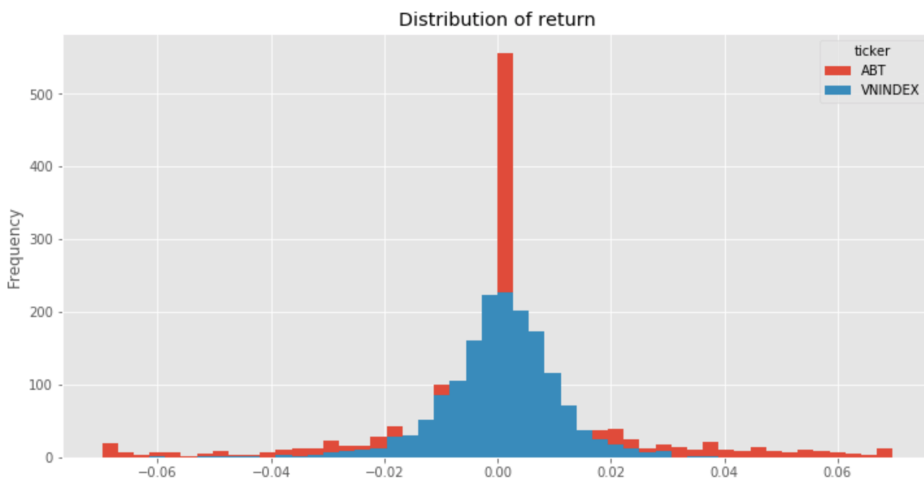

Figure 8: Distribution of daily return of ABT and VNINDEX

2 FEATURE ENGINEERING

To evaluate the effect of features, I use VN30 ticker to calculate the correlation between features and Sharpe Ratio. Note that the Sharpe Ratio is calculated 17-working day rolling(because the test set from 1/8/2019 to 23/8/2019 there is 17 working day in this period) and I shift the result 1 day. Because I want to check the correlation between present features with "future" Sharpe Ratio.

2.1 Feature extraction

- Continuous Features
  With "adjusted close price" and "volume" features, I calculate the daily return for VN30. The correlation map in figure 9

- Nominal features

  Beside that, I also consider the nominal features: "Day of week"("dow") and "Month". I change these features to one hot features to measure the correlation with Sharpe ratio. The correlation is illustrated

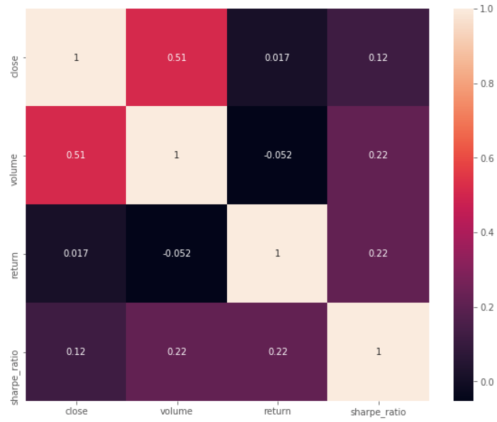

Figure 9: Features Correlation map

in figure 10.

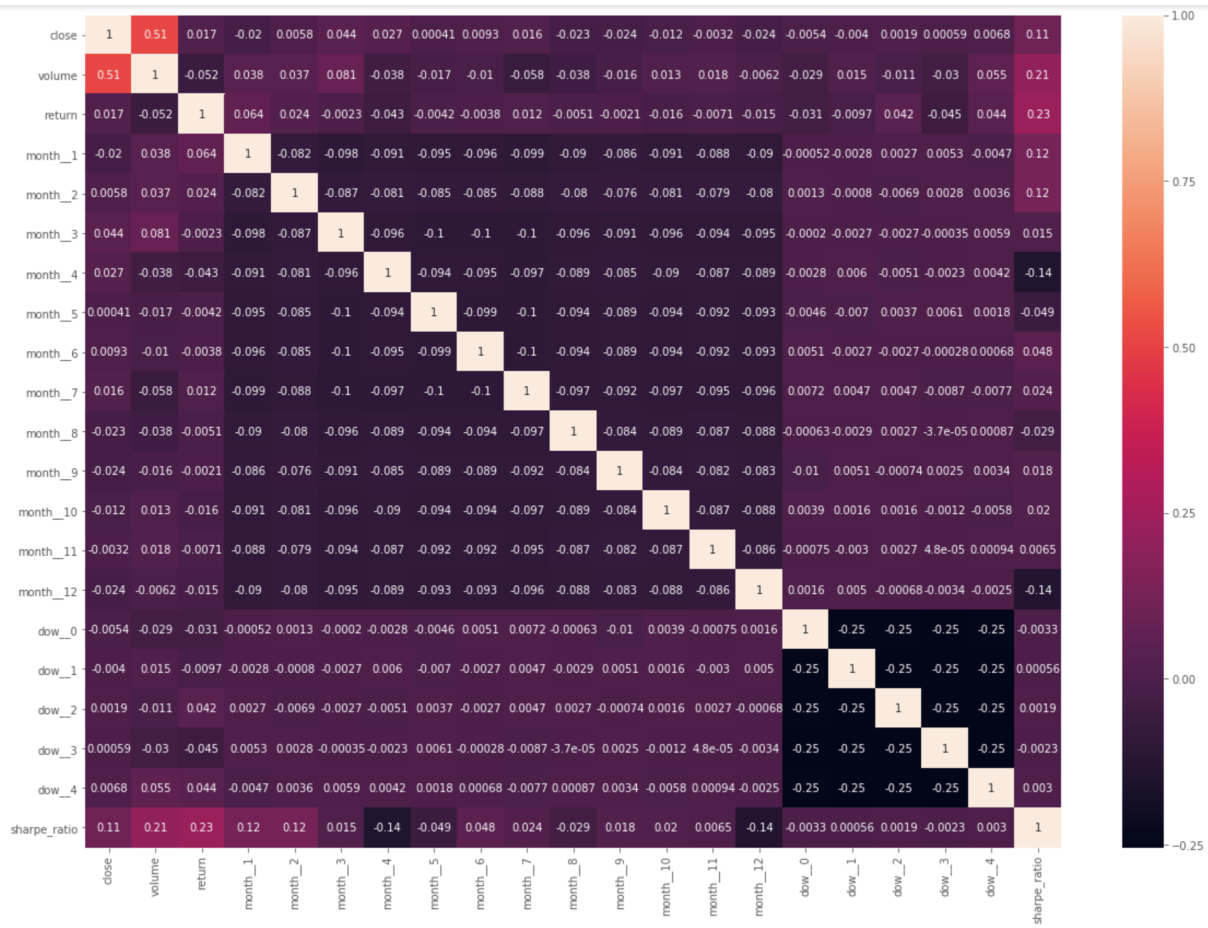

Figure 10: Correlation map with more features

- Past event
  I try to evaluate the correlation between the Sharpe Ratio with the past features by shift these value

one day a head. The new features added "s" suffix in the end. The new correlation matrix in figure 11

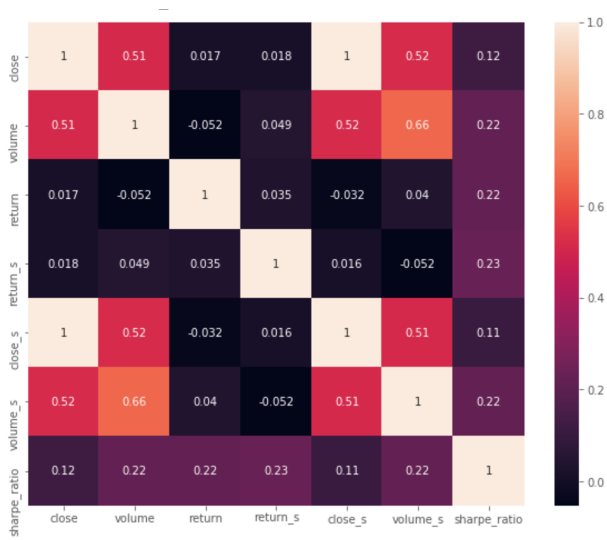

Figure 11: Correlation map with more features

2.2 Feature Selection

The correlation between Sharpe ratio with Day of Week is small, we can eliminate this features.
We can see that the Sharpe ratio even have higher correlation with the past features compared with original features.
So, to capture the past effect of features to the Sharpe Ratio, I choose the Convolution Neural Network to solve this problem. It can capture the past effect and generate the new features through each layers of networks. So the features I use is "volume" and "close price". Although "return" have high correlation with Sharpe ratio like figure 9, but it also have high correlation with "volume" and indeed when train CNN model it does not improve because in the formula we can easily calculate the return from price so the deep network can extract this feature.
In figure 10, just Month 4 and 12 have high correlation with Sharpe ratio, however my model is overfiting the training set because have small training data. So I also not use Month feature.

3 MODEL SELECTION

3.1 Define the problem

In this part, I want to redefine the problem. My target is to find the list of tickers will have highest Sharpe Ratio in next 17 working days (1/8/2019 to 23/8/2019).

SO my model will take input the available data(the data before 1/8/2019) here I choose 65 days (13 weeks) for the input. The output of my model is a binary vector with 1 for chosen tickers and 0 for not chosen.
So I define a loss function to maximize Sharpe ratio and also push the variable to 0 or 1 because the portfolios have same weight for each tickers.

3.2 Loss Function

The formula :

$$L = -\frac{E(DailyReturns)}{\sqrt(var(DailyReturns))} + M * (1-w)^T.w$$

- Here w is the binary vector returned by model
- $Dailyreturns$ calculated by $\frac{w^T}{|w|} * R$ with R is the vector of daily return of each ticker.
- M is a hyper parameter for tuning.

,

The first term in right hand side of equation is the Sharpe ratio, and the second part is the constraint that every ticker in portfolios have same weight(uniform). So we will minimize this Loss function in training process.

3.3 Architecture of Model

: **Pipeline for training**

- Data: Here I group data according date, 65 consecutive days for one sample (X), and the next 19 days is the data for maximize the Sharpe Ratio in it (y). The data have shape X(438,65,2), y(438,19). I just slide this pair to the next day to get new pair.
  I devide dataset to training and validate set.
  Training set compose Validate set is the newest 45 pair data.
- Input layer: have shape like in figure 12
- First Conv2d Layer: I use kernel size (1,15): It means I choose 15 days in the past will be effect to the current day. I use 4 filters at this layer.
- BatchNormalization: I use 5 batch normalization layer to reduce overfitting of model.
- Average Pooling: Reduce the parameters for model.
- ACtivation function: Almost Relu function except the last layer
- Last Conv2d layer: I use kernel size (438,2) to calculate the correlation between the 438 tickers.
- Dense layer: shape of output is 438, equivalent to the number of tickers. The previous Flatten layer have output shape is 6, so the number of parameters in Dense layer will not too many.
- Final Activation: Sigmoid function: I want the output will be binary vector so I use sigmoid and threshold 0.5 for decisde output in this layer

  Finally, I use Adam Optimizer to optimize loss function had define above with w is the output of this network. I also use L2 regularizer to regularize this network.
- When predict the tickers, I use one more filter to catch the available tickers at that time by matching which one are collected in last day because some tickers in time have been eliminated from market.
- Finally, I tuning top k by tuning in the validate set and the result is k =10. The Sharpe ratio in validate set is 9.2

  The final result is ['HVH', 'TCH', 'HVN', 'TCB', 'VND', 'TGG', 'EVG', 'SJF', 'KPF', 'STK']

  Sharpe Ratio = 8.736833779957259

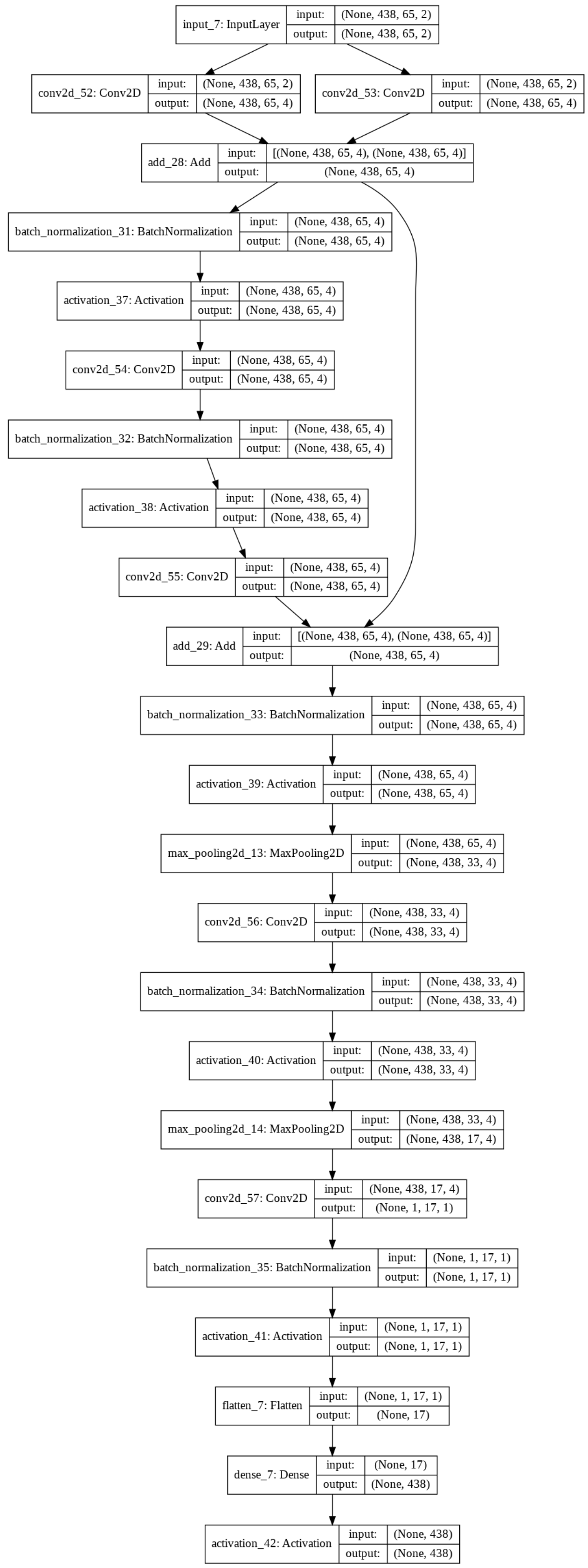

Figure 12: Detailed network

4 Tuning Parameters

I use hyperopt library to tuning hyper parameters for this network.

The hyper-parameter for tuning

- Learning rate
- Activate function
- Weight decay coefficient
- Number of filters of each Conv2d

This library helps to tuning parameters using Bayesian optimization. Each hyper parameters we will define the space for them. Almost the parameters I set uniform distribution, like Learning rate Uiform(0.001,0.1) . It will search in the region where the lower bound (95 percent confident interval) is highest after each iteration.

Because the model is overfitting, I also use small batch size is 16.

5 Discussion

First, the number of data point of dataset is quite small. With the way I divide for X and y, I just have 1636 data point to train the network. We can get larger dataset by collecting the price after each hour. With these way we can have more data by slide the pair each hour. We can get approximately 24 times bigger dataset.

Because using neural network so more data the performance will be better. Beside of that, we can use another features like "Month' in the Feature extraction have mention. I don't use this feature because my model have approximately 12 thousand parameters, if I use more features , the number of parameters will raise, and the variance of model will be high.

6 SECTION 2 - UPDATE MODEL

6.1 Innovation

- **Processing data**

  Because I use CNN model to solve this problem, the input and output had to have fixed size. It means X have the shape (438,64,2), is number of tickers , the number of day and number of features respectively. However, some tickers in the time have disappeared (the company is corrupted..etc), and in this time, also have some new tickers appear in the market.

  To deal with this problem, I fill the data X where np.nan by 0.0, It means with the missing data, our model will not learn anything from these data. Meanwhile, with Y is the daily return in 19 days, I fill nan by -1e4 with meaning that the missing data is the tickers have been eliminated or not disappeared yet, so we set its daily return low so when the chosen portfolios have content these tickers value will have negative mean(Sharpe ratio will negative). So in optimization process, the optimizer will learn to eliminate these tickers from portfolios automatically.

  In the formula to calculate Sharpe ratio, I also use clip function for the variance, because some missing value tickers filled with -10^2 so the variance in this time will be 0, it can make Sharpe ratio return negative infinity, the gradient also infinity so model can not learn. I clip 0 by small positive value (10^{-6}), so the Sharpe ratio for each ticker with missing data in this time large negative number.

```
sharpe_ratio: 0.36%
0.34% (+/- 0.03%)
```

Figure 13: Result to do 10-Fold cross validation.(Have not multiply for sqrt(n))

This process is natural with the mind that we will not invest in the tickers will be eliminated in near features(19 days) and also will not invest in the tickers have not appeared yet.

- **Model**

  I design model based on residual block of Resnet model. So the size of input data have change a little compared with Session 1. I choose size of input is (438,64,2). In many tries, I realize when increasing the depth of model, the prediction will be well discriminated. It means all the values in 438 dimension vector output of model will close to 0 and 1. It is the indicator for well classify ability of model.

6.2 Improvement

- **Loss Function**

  Because the portfolios with the numbers of tickers below 10 is too little and larger than 50 is too many so management price is high, so I modify the loss function like below:

$$L = -\frac{E(DailyReturns)}{\sqrt(var(DailyReturns))} + M * (C - w)^T.w$$

  - M after tuning manually, I set fix by 3e-6
  - The constant C is larger than 1, after tuning manually, I set C = 1.6 .
  - I also clip the value of Daily return for eliminate the portfolios with all missing data.

  By setting C bigger than 1, the constraint helps the outcome of model will be more value equal 0. It just like we plus the constraint on the 1-norm of W. It means we don't want to have too much value equal 1, so it can helps the optimizer just try to optimize the portfolios with small number of tickers.

- **New architect**

  Inherit from the Resnet model, I design the model like 14. When I increase the depth of network, the ability to push output value to 0 and 1 of model is higher, but because the number of parameters is also increase so I stop at 7 blocks. I choose the size of X is 64 because when go through a block, the output size will reduce a half. I tune the architecture by using hyperas library.

  My best model perform well in train and validate data. I do 10 Fold cross validation with 100 epoch for each fold, we can see in figure 13 the variance of model is small. The metric I use to validate is Sharpe ratio. It means the model can capture data well and don't overfit. The Sharpe ratio here I haven't multiply for $\sqrt(n)$ (19 days).

- **Constraint in model**

  With the constraint part in loss function, the output of model become discriminated between the value like figure 15, and with tuning C constant in loss function helps the set numbers larger than 0.5 of output vector is small . It satisfies with the constraint that should not have many tickers(larger than 50) in the portfolios.

6.3 Weakness of model

- The model is too large. The number of trainable parameters approximate 23.000 parameters meanwhile the the sample from data set just have 1555 samples.

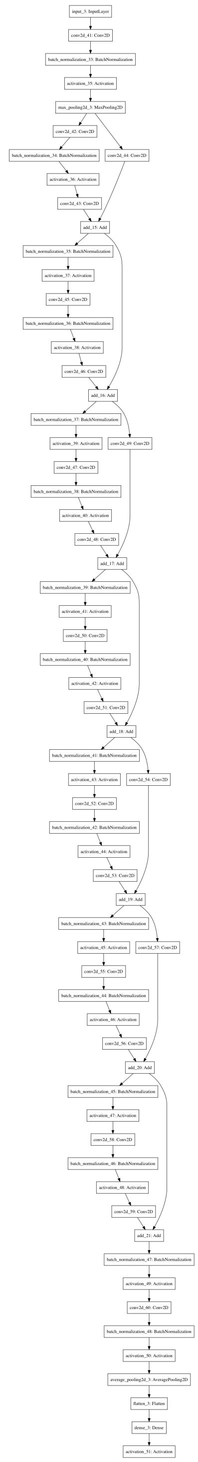

Figure 14: Model with residual block.

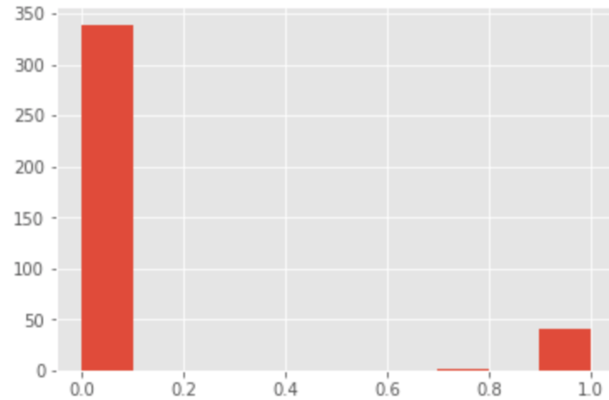

Figure 15: Distribution of the output in latest input data(until 31/07/2019)

Figure 16: The eliminated tickers but also chosen by model

- I try to regularize model by batch normalization layer after each convolution layer, use L2 regularizer for each layer, and train model with small batch size, 16. I also use Early Stopping callbacks functions to avoid overfiting.
- Although I has set the nan value to -10^2, but the model still have missing when choose the tickers have been eliminated from market like figure 16. To not choose these tickers I have to filter before submit the result to API. But in this time just have 381 tickers in total 438.

  To overcome this problem, I think I can choose larger value to fill nan in daily return.

The Sharpe ratio I get from API just approximate 0.64 for this model. meanwhile, when do cross validation, the model get the mean 0.34 $*\sqrt{19}$, approximately 1.48, and the standard deviation just 0.038$\sqrt{\sqrt{19}}$,approximately 0.06 . So ,we can see that the test data in August is different compared to what model perform in the data set. From that, we can see the model may be the market are effected by outside factor, not from inside of market. Because my model just use information of market(volume and close price), so model can not capture the outside effect like news, policy,.. . And usually, these information is the most important indicators for the predict the market.

From that, to get more strong indicators for this problem, I think we can use some information outside market, like statistics from social networks. How many tweet about this ticker in that day,...

6.4 Conclusion

Because the features and samples are small so we need more data, more indicator for training model. Beside of that, my model is almost automatically chosen the tickers with the well-defined constraint in the loss function.

It helps we choose suitable number of tickers and also solve the optimize the portfolios in training time. So we don't have to separate this problem to two sub problems, predict return and optimize portfolios.