Convolutional Neural Network for Predicting Company Performance Based on Historical Financial Statement Information

Justin Amezquita (jaa35@stanford.edu)

MOTIVATION

Investors in public markets generally fall into two camps: fundamental investing and quantitative investing. Fundamental investors read through historical financial statements, such as a firm's 10-K and 10-Q reports, to come up with a holistic view of the firm's long-term future performance.

In contrast, quantitative investors use advanced mathematical techniques and large amounts of data to predict stock movement, usually in the near-term.

This project attempts to combine the long-term investment horizon and insight of fundamental investors with the data analysis techniques used by quantitative investors. The goal is to introduce a deep learning framework that can aid fundamental investors in their research.

APPROACH

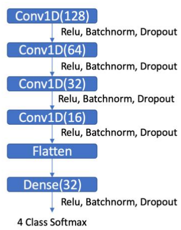
The approach uses key financial ratios and metrics, such as return on assets, interest coverage, and working capital turnover ratios, to predict which quartile a particular company's returns will be one year forward. A convolutional neural network with four convolutional layers was used for prediction. This model is compared to two baseline models, a linear regression and a random selector. The random selector picks returns randomly from a normal distribution that was fit to the actual sample set returns for a particular time period.

The CNN is also built to take in a timeseries of input data. Returns are calculated using a look-back period of 1Q, 1yr, 2yr, and 8yr.

DATASET AND INPUT FEATURES

The dataset consists of historical financial information for a selection of 523 companies that have been listed on the S&P1500 Super Composite Index January 2010 and January 2018. Each training example consists of 103 key ratios and metrics calculated from information contained in the firm's 10-K and 10-Q reports. The label for each example is the one-year forward stock returns for the firm. Lookahead bias is avoided by introducing a three-month lag on the date of the input features.

CNN MODEL

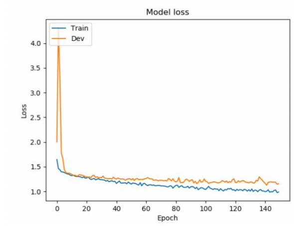


A CNN architecture was used because of its flexibility in folding in more historical time periods for training. The default CNN uses input data from only one fiscal quarter to predict 1 year forward returns. However, by using the depth channel in the CNN as the time dimension, we can instead choose to use any number of prior quarters as input data to predict 1 year forward returns.

The CNN has a 4-class softmax output layer. The categorical cross-entropy loss function was used, shown below.

$$L(y,\hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{C}(y_{ij}\cdot\log(\hat{y}_{ij}))$$

RESULTS



Accuracy

Model	Train	Test
Random Selector	N/A	0.23
Linear Regression	0.26	0.26
CNN	0.34	0.32
CNN 1-yr priors	0.52	0.37
CNN 2-yr priors	0.75	0.48
CNN 8-yr priors	1.00	0.54

10,041 samples were used for training, 3,347 samples were used for development, and 3,348 samples were used for testing.

DISCUSSION

The CNN outperformed the random selector baseline model by 9% and the linear regression baseline model by 6%. However, the CNN model accuracy of 32% does not provide confidence in using the model as an actual trading strategy. Interestingly, as more historical data is used to make predictions, the accuracy goes up. When 1, 2, and 8 years worth of prior financial statement information is used, the model accuracy increases to 37%, 48%, and 54%, respectively. It makes sense that using a time series as input data would improve accuracy, since the model is able to identify trends over time via the depth channel in the CNN input.

FUTURE WORK

Given improvement in results from using a time series of historical data, other recurrent NN architectures such as LSTM can be studied. Additionally, the input sample set can be expanded to include more companies over a longer period of time.

REFERENCES

[1] Fama, Eugene F. and French, Kenneth R., Long-Horizon Returns (November 20, 2017). Chicago Booth Research Paper No. 17-17; Fama-Miller Working Paper.

[2] Chen, K., Zhou, Y., and Dai, F. (2015). A lstm-based method for stock returns prediction: A case study of china stock market. In *Big Data (Big Data0, 2015 IEEE International Conference on*, pages 2823-2824. IEEE.