



Webinar “Making Deep Neural Nets Work for Forecasting Stock Prices” Q&A

Q: AAPL has had an "upward bias" for quite a while. Have you tested this system on other stocks with more "choppy" performance?

A: Yes we have, we are actively pursuing strategies predicated on deep learning and CNN in Futures, S&P stocks and other more volatile assets.

Q: How do you use this for index forecasting where ranking is not possible?

A: Good question:For Index Forecasting were ranked feature may not always be possible, we apply other types of market-regime-agnostic features, by which we are not capturing the absolute value but the rate of change overtime (and you are correct, these are not ranked features). We look at change % over time, we look at velocity and speed of change over time for index specific features.

Q: How do you use this for index forecasting where relative ranking is not possible or does not make sense as it does for relative ranking for stocks?

A: See answer above -- in addition, ETFs can be ranked relative to the SP500

Q: Does the 5 to 1 rule mentioned for the out of sample data imply that when putting this strategy to work, the model should be retrained every "n" periods? Have you combined this approach into a Deep Reinforcement Learning Model?

A: Yes the notion of retraining every "n" period works in most cases. However, there are actually more sophisticated ways to determine when a retraining is necessary -- for example we can rate the recent performance of the network and when the quality of the prediction falls below a certain threshold we can force a retraining.

Q: What's the output being predicted? Is that a buy/sell recommendation or the real return?

A: It is a classification network. So a 1 label indicates a buy and zero indicates don't do anything or sell. The network is trained to buy and hold for 21 days or until a target gain or a stop loss levels are met.

Q: What's the majority voting accuracy/precision? If we predict everything to be "buy", what would the accuracy be, or if we predict everything to be "sell", what would the accuracy be



A: Accuracy refers to how correctly each example is classified both 1 (buy) and zero (either sell or do nothing). Precision refers to of the examples we classified as a "1" how much of the time are we right? We considered that as well. We look for our system to have higher precision than all "buy."

Q: For the back-testing, how is a trading strategy implemented based on the output?

A: If the classification is "buy" enter the position. If it is "do nothing" exit or if you're not in a position then do nothing.

Q: In the historical data, are you including news released about the company prior to the price rising or falling?

A: We do have news sentiment as a factor. It is not used in this particular system but can easily considered along with other factors.

Q: Is OHLC data used in transformation or only close data?

A: Only close data for backtesting the strategy but the technical indicators are mainly predicated on OHLC since we don't capture intraday data historically.

Q: Given the new research in evolutionary optimization performing at par with NNet reinforcement learning from 2017 is there any testing on your side testing CNN vs simple evolutionary optimization?

A: We compare a wide array of machine learning disciplines. The concept behind Deep Learning is most exciting because it makes so much sense and it works very well on stationary problems (mainly around image classification). We have uncovered exciting results but since this is a relatively new concept compared to other methods of machine learning technology used, we need more time to assess the efficacy of our finding in the real world.

Q: With just one stock (Apple), it could be a coincidence. Did you run similar studies on other stocks? All stocks in the SP500, for instance?

A: We used only APPL to simplify the presentation -- we actually apply this concept to other stocks including the S&P, Futures and soon against FX.

Q: Do you use a "global model" (one neural network applied to all companies) or is there a separate network for each company? Or a separate network for different subsets (sector, etc)?



A: This is one of the questions that was presented in the Q/A, we prefer a model per stock but in reality we need more data to train on and thus we identify similarly behaving stocks to gather more training data.

Q: How many features did you use in the backtest?

A: In this particular example we used 32 features (mainly technical and fundamental). In general, the # of features depends on the application, but anywhere from 10-100s

Q: Did you use Bayesian optimisation for the Hyperparameter search?

A: We are actually using Bayesian Optimization in our LSTM Hyperparameter search but not in this particular example. The reason being is that the sample size is relatively small to employ Bayesian Optimization which samples parts of the training data vs all of it.

Q: Do you collect your own data or use 3rd party vendors for Fundamental data?

A: We don't source any of the data but rather partner with other specialized alternative data providers.

Q: Do you find feature importance to be similar among different tickers? Or is there a separate model for AAPL, a separate model for INTC, and separate model for NFLX, and so forth..

A: Separate models for each asset

Q: Do you have an actual strategy running live using this ideas?

A: We are in the process of deploying a Futures strategy based on this but we are also expecting more to come soon. How important (or how contributive) a feature is in a strategy is very hard to determine in a Deep Neural Network framework, which is one of the most difficult aspect to sell to the skeptics since its not easily explainable. The only close match to scaling which features are appropriate is through a process called Feature Selection (which is somewhat similar to the Hyperparameter search described in the presentation)

Q: With the backtest, you've shown an approximately 52% return for a period of 10 years; why would an investor opt for Lucena instead of buying S&P500 ETF (for instance), which in the same period had a greater return?

A: The backtest was not meant to represent a strategy but rather to underscore the accuracy of the model out of sample. With a 69% accuracy as was shown you make fewer bets but with higher degree of success expectations. I am sure that smart investors like yourself can exploit these odds better than we can.



Just a few examples: Leverage? Derivatives? Extending the frequency of trades with more stocks (not just AAPL) etc...

Q: What is the optimal number of training examples, how this number depends from the price forecasting forecasting horizon? What's the best network structure?

A: There is no hard rule unfortunately, the market is so dynamic and configurations are so situation specific that it takes quite a bit of practice to get an intuitive feel. More importantly, the best way to answer any of these questions is to conduct a lot of tests (i.e. the Hyperparameter search I mentioned)

Q: Do you use the same features for all the stocks or each stock has its specific set of features?

A: We use the same type of features across all the applicable stock however the feature values would be widely different from one stock to another. In addition, we have special features for Forex and for Futures.

Q: The input is the previous 63 days data and the output is the max/min return over the next 21 days?

A: The output is a classification: 1 for "buy" 0 for "do nothing." In addition, the backtest is conducted based on how we classify a 1 which is that the stock will move higher and will reach the upper bound of the ATR sometime in the next 21 days

Q: Will Transfer Learning work for this and if so can I get pre-trained nets from Lucena?

A: Transfer learning may work but I wouldn't even attempt to provide pre-trained nets just because some of our research is proprietary and it's not our business model. On the other hand we are happy to share the concepts for others to take upon themselves a Transfer Learning model.

Q: Must factor time series be in visual form? Or can they be in tabular (quantitative) form?

A: The form considered is a matrix of values (a table). But we can choose to view it as an image.

Q: With the completed model are you able to get a table of relative factor importance?

A: One of the hardest things for us as technologists is to explain which factors were more important than others in a decision. Obviously a feature selection process is done mainly over trial and error (with no real smarts to it). But in many cases with such



complex models its very hard to wrap one's head around and thus not all Neural Network decisions are fully explainable. I know it's scary but that's the reality!

Q: As we are dealing with time series information, how often must the model be retrained to ensure the model and hyperparameters properly apply?

A: In our back tests we usually retrain monthly. But retraining daily in practice would be ideal. In addition, dynamic retraining based on some recent accuracy rating may also be a good approach.

Q: Are you typically starting with the same convolutional layers on different stocks and then allow the machine learning to change over the training period so the end results and signals are different on each stock, or does this stay more uniform regardless of which stock is being run through the system. Basically how different do the models become as you continue to train on individual stock data.

A: We prefer a separate model per stock but in reality one stock doesn't hold enough training data and so we extend the model with additional highly correlated constituents.

Q: Was a dataset generated for supervised learning and classification?

A: Classification vs. regression you mean? This is a typical classification problem 1 for a buy or 0 for sell or do nothing. It's still a supervised learning however.

Q: Will the slides cover dropout and layer by layer description of a conv net ?

A: In this particular example we use L2 regularization in order to reduce overfitting. A dropout method could have been used just as well but our Hyperparameter search uncovered L2 regularization to be more effective.

Q: Was there a research paper/papers that you based your work on, that we can have a look at?

A: Yes <https://arxiv.org/pdf/1710.00886.pdf>

Q: To generate a Recurrence Plot you need a time delay and dimension embedding parameter. How did you calculate these ? Or did you use 1 for each parameter (i.e., no embedding at all)

A: There is no intentional time delay apply to using a signal. Time delay could be a function of the data provider and we are at their mercy as to how fast new data is made available to us. In this example, we use 63 days of daily data. 1 for each parameter, no embedding.



Q: How much data is required to successfully build a model? How long does the model function before requiring retraining?

A: It depends on how many training examples per day we are able to collect. For an S&P 500 model we look for at least 3 years of daily data.

Q: I understand that Dr Balch has use Q Learning methods for trading, is this method applied to production models?

A: We are not yet using Q learning for strategies at Lucena. We tested Q Learning on intraday signals with some success but weren't able to overcome slippage and transaction cost. We will be revisiting Q Learning in the future I am sure.

Q: How does this approach to modeling the financial data compare with competition models? Is this method unique?

A: It's unique in the context of us testing it as a new concept. We are developing this approach from the bottom up relying mainly on white paper and academia research and not based on a proven commercial success or winners in quant competitions.

Q: Are you manually creating features, or do you have an automatic method to generate features? Something like an Attentional NN?

A: There is a manual configuration by a data scientist but the features are created and calculated algorithmically

Q: What is intuitive of using recurrence plot?

A: RP is just simple to implement and sort of makes sense by which we are not only considering the basic formation of a trend but also the Euclidean distance between every two points.

