Quantitative Trading Strategy Based on LSTM Model

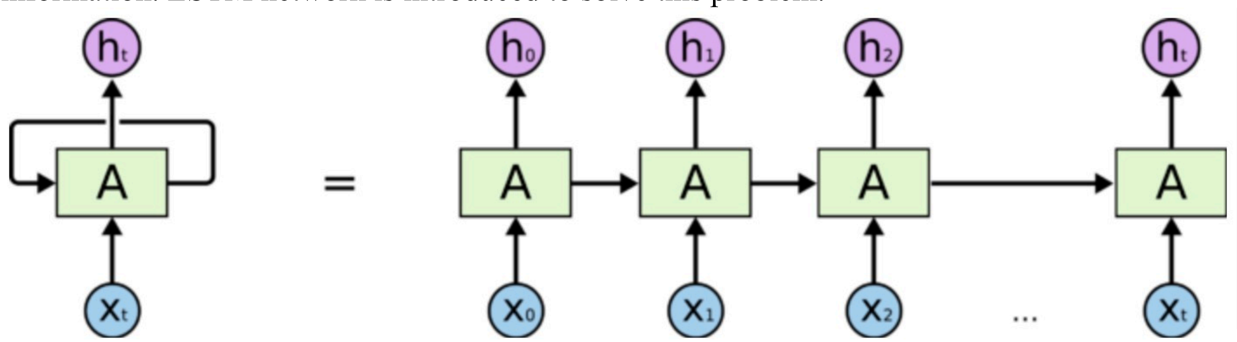
Tiancheng Hou Mengrui Zhang Xutao Chen

Abstract: LSTM is well known for its application of prediction on time series data, especially in classification. We use LSTM model to classify stocks in our asset pool into winning/losing group by their future performance. Based on rolling prediction results, we construct portfolios and rebalance them each month. Compared with benchmark, they achieve much higher annualized return under acceptable maximum drawdown,

Keywords: LSTM Model, Quantitative Strategy

Introduction

While reading a paper, we will use previous word and understanding, instead of a blank mind, to speculate the meaning of current word. Traditional neural network does not consider the sequence, while recurrent neuron network (RNN) can solve this problem. Through cyclic connection, RNN allows persistence of information. To better understand the cyclic structure, we could expand it. As figure 1 shows, current state is affected by previous states. However, as lookback window becomes large, RNN encounters vanishing problem—unable to link to distant information. LSTM network is introduced to solve this problem.

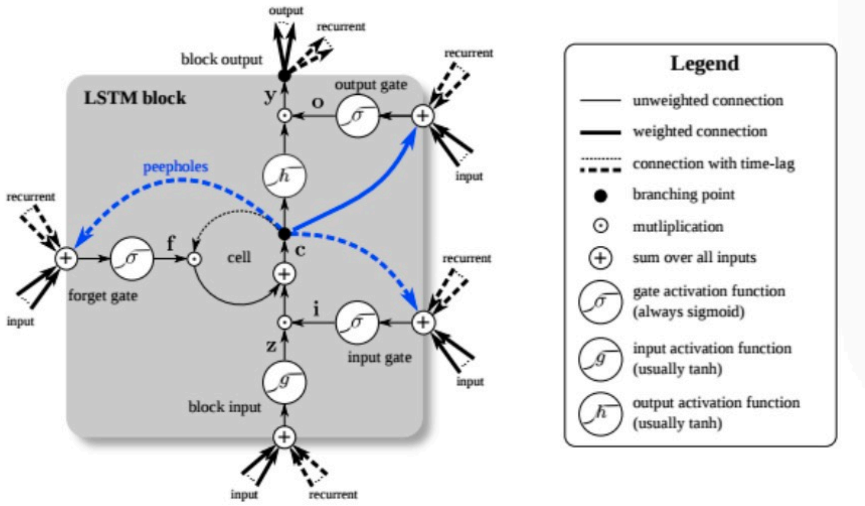


[Figure 1: Unfold LSTM Network]

Proposed in 1997 by Sepp Hochreiter and Jürgen Schmidhuber, LSTM network is a kind of RNN composed of LSTM units (or blocks). It is a model capable of having short-term memory and long-term memory working together. An LSTM is well-suited to classify, process and predict time series given time lags of unknown size and duration between important events.

A common architecture of LSTM block is composed of a memory cell, an input gate, an output gate and a forget gate, as figure 2 shows. An LSTM (memory) cell stores a value (or state), for either long or short time periods. This is achieved by using an identity activation function and plus instead of multiply for the memory connection. In this way, when an LSTM network (that is an RNN composed of LSTM units) is trained with backpropagation through time, the gradient does not tend to vanish.

The LSTM gates compute an activation, often using the logistic function. Intuitively, the input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit.



[Figure 2: Structure of LSTM unit]

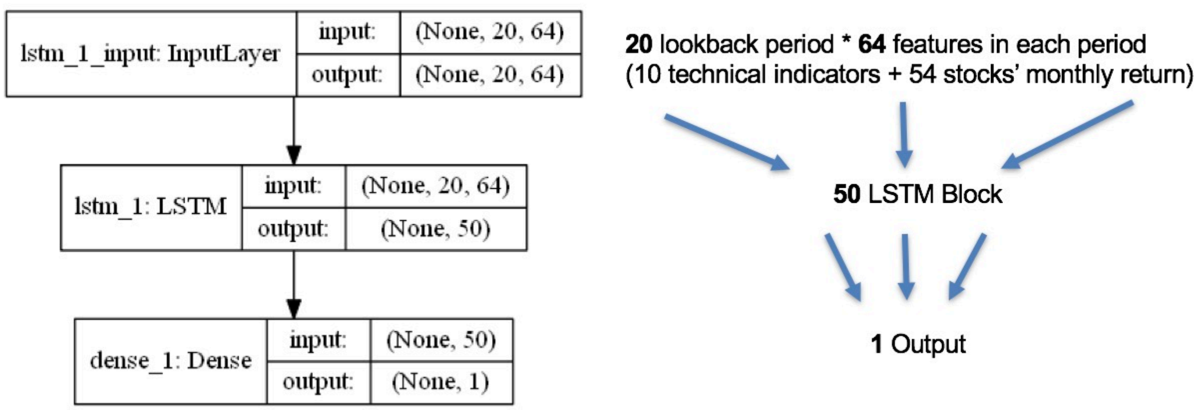
Therefore, we use LSTM network to process stocks' time series data and predict their future performance. In the next section, we will briefly explain how to construct our LSTM network, including data prepossessing, feature selection, parameter setup and program workflow. Based on prediction results, we will construct different portfolios and compare their performance with benchmark. In the last section, we conclude that LSTM is well-suited for finance data sine its prediction performs well. We also demonstrate possible direction for further studies.

LSTM Implementation

While it's very hard to directly predict values (i.e. stock price), LSTM performs well in classification problems, so we set stock's relative future performance as LSTM target output. In addition, we try to decrease prediction errors by predicting on an asset pool, instead of individual stock, and only making use extreme values. Specifically, a big asset pool is constructed and if stocks' monthly return is above/below the asset pool's median return, it's classified into winning/losing group.

Due to LSTM's high requirement on computation ability and our limited resource, only 54 stocks are included in the asset pool for backtest. There is selection bias, but we try to decrease it by selecting 6 stocks from each main industry with market cap ranging from big to small.

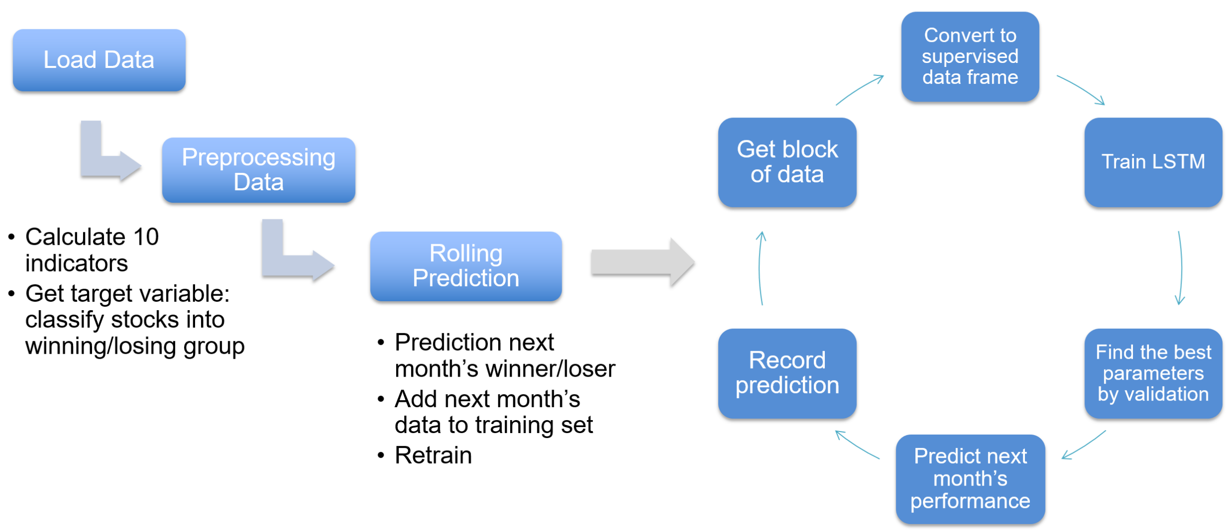
For input features, we choose 10 technical indicators and all stocks' daily return. Technical indicators include Median return, 20 days Moving Standard Deviation, Stochastic Oscillator D, MACD: Macd, sign and diff, RSI, CCI, ADX, and TRIX. We set 20 timesteps as lookback window and use 50 LSTM units to finish the prediction. In other words, there are 1280 (20*64) features in each sample and 50 LSTM output are used for every prediction.



[Figure 3: LSTM network structure]

For each prediction, we slice 200-days data block, separating it to 80% for training and 20% for validation. Through validation, we find the best epoch and corresponding model weights, and set them as weights for prediction of the last day in the data block.
Following is the program workflow. One thing worth mentioning is that all LSTM network is inherited from previous one except the first one which needs initializing.

Program Workflow



Prediction Results

After rolling prediction for each month, a probability matrix is generated. The higher/ lower the probability, the more likely stocks will be classified into winning/losing group. The predicted classification ranges from 2010/02/23 to 2017/09/08, total 96 months. In the next section, our back testing will base on this result.

	AAPL	ABB	ADBE	AEP	AMGN	AMZN	BA	BBD	BP	BTI	...	RIO	
Date													
2010-02-23	0.538847	0.353620	0.058308	0.941635	0.400793	0.130008	0.997141	0.322500	0.775085	0.975656	...	0.073927	0.48
2010-03-23	0.427872	0.996303	0.757463	0.717296	0.314615	0.988614	0.981497	0.331736	0.056342	0.869988	...	0.997664	0.11
2010-04-21	0.821159	0.138314	0.086946	0.034132	0.695820	0.496179	0.035443	0.929267	0.362070	0.100414	...	0.620284	0.91
2010-05-19	0.781184	0.029718	0.833219	0.942879	0.033495	0.012112	0.718003	0.870113	0.592619	0.960161	...	0.995544	0.45
2010-06-17	0.998168	0.985907	0.332878	0.012854	0.092629	0.886154	0.482254	0.841069	0.868410	0.995379	...	0.978568	0.76

Back testing

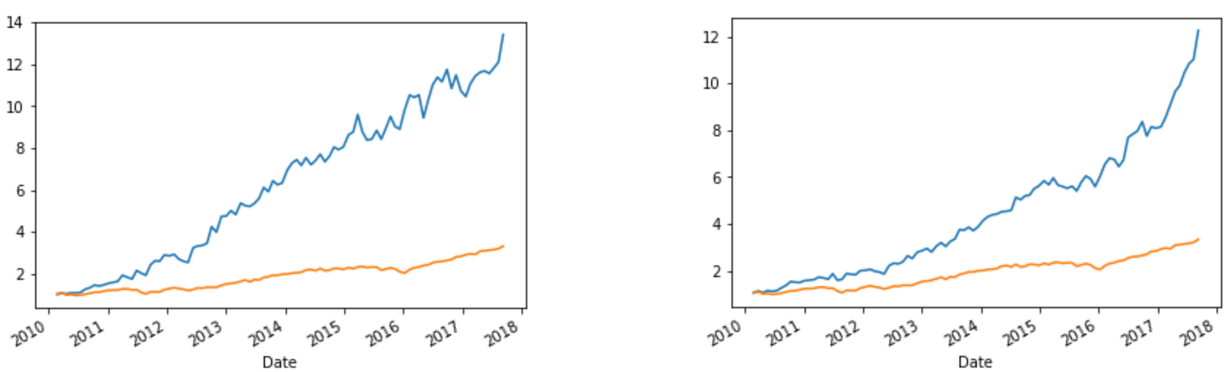
A natural strategy based on the prediction result is to long/short stocks in the winning/losing group. Due to limited size of current asset pool and constraints of prediction error, we tend to ignore stocks located in the middle range, and only select stocks with highest/lowest probabilities. To measure strategies' performance, we construct a benchmark portfolio that equally-weighted long all stocks in the asset pool.

Strategy 1

Long top N stocks in the winning group. The Sharpe Ratio and maximum drawdown are better than those of benchmark. In terms of return, strategy 1 performs much better, achieving a 38.33% annualized return.

Table 1: Performance of strategy 1

	LSTM (Long 1)	LSTM (Long 2)	Benchmark
Sharpe	1.3518	1.6136	1.10
Maximum Drawdown	-13.21%	-14.82%	-18.67%
Annualized Return	38.33%	36.79%	16.23%
Cumulative Return	1241%	1126%	233%

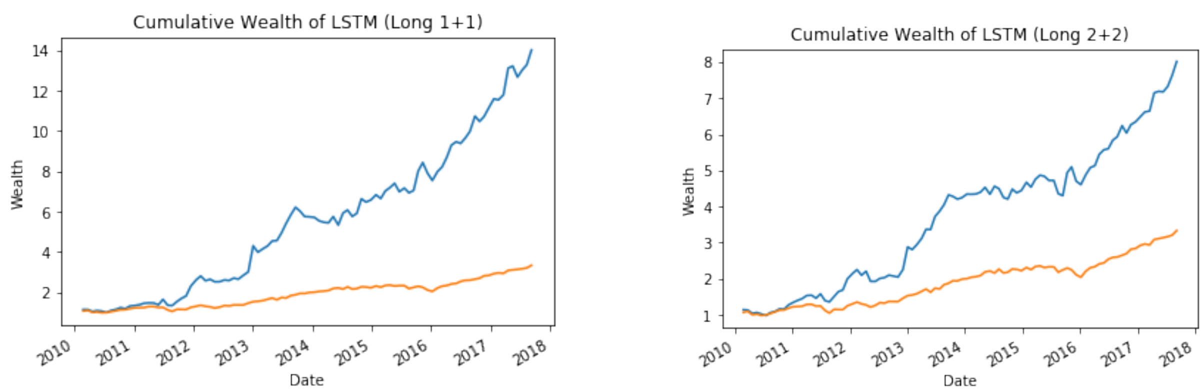


[Figure 4 & 5: Cumulative wealth of strategy 1]

Portfolio 2
An unexpected interesting result is that longing, instead of shorting, bottom N stocks performs well. The most possible reason is mean-reverting effect. Since the lookback window is 20 days, which may be a mean-reverting period, the stock that performs worst in the last month is more like to recover in the next month.
Therefore, we long top N stock both in wining/losing groups. As the Table 2 shows, all performance metrics are better than those of benchmark. The best portfolio achieves almost 1000% 8-year cumulative return and 34% annualized return.

Table 2: Strategy 2

	LSTM (Long 1+1)	LSTM (Long 2+2)	Benchmark
Sharpe	1.3010	1.2891	1.10
Maximum Drawdown	-18.51%	-14.38%	-18.67%
Annualized Return	39.11%	29.70%	16.23%
Cumulative Return	1302%	701%	233%



[Figure 6 & 7: Cumulative wealth of strategy 2]

Conclusion

1. LSTM performs well in classification on time series data.
2. Portfolios constructed based on LSTM prediction outperform benchmark portfolio.

Extension

Because of limited computation resource, we only construct a small asset pool, which causes problems including selection bias and constraints on number of stocks for each portfolio. Future studies could expand the asset pool to include all S&P 500 stocks. As stocks pool become large, more stocks could be included while constructing portfolio, leading to diversification advantage – higher Sharpe Ratio and lower risk.