Deep learning in stock market based on long short term memory technique

Mahdis Hosseini

Deep Learning Course(2021)

Abstract

Stock market prediction has always been crucial for stakeholders, traders and investors. Investors make decisions based on various factors, including consumer price index, price earning ratio,….here we developed a Long Short Term Memory (LSTM) model in order to predict next day opening price (one step ahead).

Introduction

To assist investors' decisions, machine learning approaches have long been studied to automatically analyze vast amounts of financial information, such as past stock prices. Here we use a Long Short Term Memory method with two input sequences, a sequence of daily features and a second sequence of annual features, in order to predict the next day closing price and make a better decision in trading as the article did this. But to compare we also fit another lstm model to 5 columns of prices, without using other finance methods .

Data

The database choose for this research contains New York Stock Exchange S&P500 data, publicly available at Kaggle[1]

For this Project I used "prices-split-adjusted.csv", rows with symbol "YHOO" which belongs to yahoo company. The data frame consists of prices has shown below:

date	symbol	open	close	low	high	volume
2010-01-04	YHOO	16.940001	17.100000	16.879999	17.200001	16587400.0
2010-01-05	YHOO	17.219999	17.230000	17.000000	17.230000	11718100.0

Fig.1.Dataset Features

[1] https://www.kaggle.com/dgawlik/nyse

To compute Indicators, package “stockstats” is used. Indicators which add to input are: (ema[2]12,ema26 and macd[3]). The formulas can be seen below:

$$\begin{cases} MACD = EMA_{12} - EMA_{26} \\ EMA = (Value_{today} * \frac{smoothing}{(1+days)}) + (Value_{yesterday} * (1 - \frac{smoothing}{(1+days)})) \\ smoothing = \frac{2}{(selectedtimeperiod+1)} \end{cases} \quad (2)$$

Model

The model lstm which is used here is an improvement of RNNs , is suitable for sequences[4] and can store and retrieve information using its gates[5]

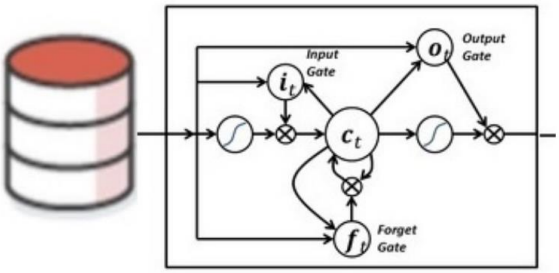

Fig.2.LSTM block

PreProcessing

Data preprocessing is an important step before fitting any model. One crucial step in every Project is to check if there is any missing values and replaced them by proper methods. Here we did not face any missing data situation. Another important step is to split data set to train and valid (and sometimes and test). Because we have sequential data here, we must do this splitting in order. The training set consists of consecutive observations from 2010 to 2015 while the test set consists of observations from 2015 to 2016 which represents sequences of 1395 time series for the training set and 339 for the test set. For this specific model, we also need to prepare a set of windows of

[2] Exponential moving average
[3] Moving average convergence Divergence
[4] Sustsker et al(2014)
[5] Bengio et al (2009)

any desirable length (here,14) of sequences. Last important step is to scale sequences to help the learning procedure of model.

Methodology

The topology chosen is 2 hidden nodes for the first LSTM and an initial learning rate of 0.001 with total 8 features. We also use a dropout layer and a fuly connectet layer at the end because of the many to one structure of the model. the optimizing method which is used here is adam optimizer which is a very good optimizer to handle noisy datasets according to other related works. Parameters chosen after testing some values which also had been suggested from other related works.

Results

Fitting the Lstm model to size8 multivariate time series of yahoo stock data, we achieve below results of MSE [6]loss, as we can see, the MSE is equal to 0.0004 which is a better result than MSE of Fitting the Lstm model to size5 multivariate time series of yahoo stock data, which is equal to 0.0005.

```
epoch :  10 loss :  0.0004646712332032621
epoch :  20 loss :  0.000409039668738842
epoch :  30 loss :  0.00040265743155032396
epoch :  40 loss :  0.000405897502787411 2
epoch :  50 loss :  0.00041151134064421 06
epoch :  60 loss :  0.00041817521560005844
epoch :  70 loss :  0.00042477011447772384
epoch :  80 loss :  0.00042932573705911636
epoch :  90 loss :  0.00042919270345009863
```

Fig.3.Output

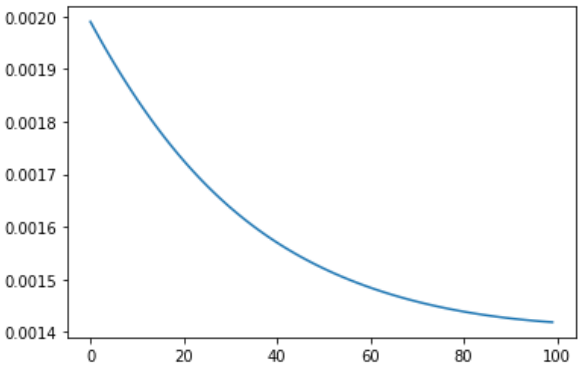

Fig.4. loss per step

[6] Mean squred error

Conclusion

In this study, we established a method based on LSTM algorithm to predict the next day Opening price. This model takes into account two frequencies with two types of variables, daily and annual variables. It confirms then our hypothesis that the combination of these two databases gives a better performance. This e method can be a useful tool for traders and stakeholders to determine their trading strategy.

Refrences

1- Lanbouri.Z, Achchab.S “A new approach for Trading based on Long-Short Term memory technique”
2- Gonçalves.A , Ribeiroc.V.M , Pereiraa.F.L , Paula Rocha.A “Deep learning in exchange markets”
3- Yoshihara.A, Fujikawa.K, Seki.K, Uehara.K “Predicting Stock Market Trends by Recurrent Deep Neural Networks