Stock price prediction with LSTM network
Project Progress Report for COMP 562

Siyang Jing SIYANGJ@LIVE.UNC.EDU
Jiyu Xu XUJIYU@LIVE.UNC.EDU
Jiacheng Tian JCTIAN@LIVE.UNC.EDU
Yuhui Huang YUHUI97@LIVE.UNC.EDU

Abstract

The time series of stock prices are non-stationary and nonlinear, making the prediction of future price trends much challenging. Inspired by Convolutional Neural Network (CNN), we make convolution on the time dimension to capture the long-term fluctuation features of stock series. To learn long-term dependencies of stock prices, we combine the time convolution with Long Short-Term Memory (LSTM), and propose a novel deep learning model named Time Convolution Long ShortTerm Memory (TC-LSTM) networks. TC-LSTM can obtain the stock longer data dependence and overall change pattern. The experiments on two real market datasets demonstrate that the proposed model outperforms other three baseline models in the mean square error.

1. Introduction

In the financial industry, stock price prediction has constantly been a popular field of research, because stock price predictability is one of the most important concerns for investors. According to many widely accepted studies, the financial markets, and particularly stock markets, have been proved to be predictable in some scenarios. While different features are available for prediction, such as general economic climate and social media effects, most studies focused on analysis of past trading patterns.

1.1. Comparison with Previous Researches

Our research assumes minute-level price fluctuation pattern is independent of corporate fundamentals and macro economy. Thus, unlike the studies of (Chiang et al., 2016), (Chourmouziadis & Chatzoglou, 2016), and (Zhong & Enke, 2017) in which daily price data are used as input, we seek to develop a predictive model based on minute-level input price data. The prediction of future stock price had also been understood as both classification and regression problems in previous studies. (liang Chen & yu Chen, 2016) and (Zhong & Enke, 2017) provided prediction of market direction as either up or down. In more complicated cases, (Chourmouziadis & Chatzoglou, 2016) specified cash and stock within the optimal portfolio composition. Our study intends to give a prediction of the stock return in the next minute compared to the current time point. Due to certain limitation and just for preliminary testing of our strategy, we are currently using 10 days of minute-level price data of 50 stocks, but our aim is to obtain data of 10 years for model training. The input will be 500 dimensional lagged stock returns, which are the returns of 50 stocks in the previous 10 minutes. We may adjust the number of lagged periods for better performance later.

There have been linear and nonlinear models to predict stock price movement with varying degrees of success. (Chong et al., 2017) noted a multilayer artificial neural network might be particularly suitable with such time-series data, due to its higher computational power and sophistication of algorithm. Such model selects features based on raw input price data automatically and does not require understanding or providing data from the side of fundamentals or macro economy, which fits our assumption about minute-level price fluctuation pattern. Our model will be composed of two parts, with the first part being unsupervised learning with traditional ML techniques like RBM, PCA, etc. The second part takes advantage of recurrent neural network (RNN) model, especially its variant LSTM. For performance measurement, previous studies have used trade simulation or various MSE methods (Chiang et al., 2016); (Chourmouziadis & Chatzoglou, 2016); (Zhong & Enke, 2017); (Chong et al., 2017). We plan to use MSE in the assessment.

University of North Carolina at Chapel Hill, COMP 562 Course Project.

Algorithm 1 Bubble Sort

```
Input: data x_i, size m
repeat
  Initialize noChange = true.
  for i = 1 to m − 1 do
    if x_i > x_{i+1} then
      Swap x_i and x_{i+1}
      noChange = false
    end if
  end for
until noChange is true
```

1.2. Final Poster

In addition to the final report, you are also required to prepare a poster overview of your project. Details on the poster are available in the project description.

2. Optional Suggestions for Your Paper and Formatting Guidance

2.1. Figures

You may want to include figures in the paper to help readers visualize your approach and your results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, be sure to include a caption describing your figure.

You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment `figure*` in LaTeX), but always place two-column figures at the top or bottom of the page.

2.2. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 1 shows an example.

2.3. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title *above* the table, as in Table 1.

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	$\surd$
CLEVELAND	83.3± 0.6	80.0± 0.6	$\times$
GLASS2	61.9± 1.4	83.8± 0.7	$\surd$
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	$\times$
META	67.1± 0.6	76.5± 0.5	$\surd$
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	$\surd$

Tables contain textual material that can be typeset, as contrasted with figures, which contain graphical material that must be drawn. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns, but place two-column tables at the top or bottom of the page.

2.4. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2014.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (1959). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (Samuel, 1959). List multiple references separated by semicolons (Kearns, 1989; Samuel, 1959; Mitchell, 1980). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (Michalski et al., 1983).

The references at the end of this document give examples for journal articles (Samuel, 1959), conference publications (Langley, 2000), book chapters (Newell & Rosenbloom, 1981), books (Duda et al., 2000), edited volumes (Michalski et al., 1983), technical reports (Mitchell, 1980), and dissertations (Kearns, 1989).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Acknowledgments

If you did this work in collaboration with someone else, or if someone else (such as another professor) had advised you on this work, your report must fully acknowledge their contributions. If you received external help or assistance on this project, you must cite these sources here in the acknowledgements section. If you do not have anything to list in this section, write simply "None."

References

Chiang, Wen-Chyuan, Enke, David, Wu, Tong, and Wang, Renzhong. An adaptive stock index trading decision support system. *Expert Systems with Applications*, 59:195 – 207, 2016. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2016.04.025. URL http://www.sciencedirect.com/science/article/pii/S0957417416301919.

Chong, Eunsuk, Han, Chulwoo, and Park, Frank C. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187 – 205, 2017. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2017.04.030. URL http://www.sciencedirect.com/science/article/pii/S0957417417302750.

Chourmouziadis, Konstandinos and Chatzoglou, Prodromos D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Systems with Applications*, 43:298 – 311, 2016. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2015.07.063. URL http://www.sciencedirect.com/science/article/pii/S0957417415005230.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, Pat (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

liang Chen, Tai and yu Chen, Feng. An intelligent pattern recognition model for supporting investment decisions in stock market. *Information Sciences*, 346-347:261 – 274, 2016. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2016.01.079. URL http://www.sciencedirect.com/science/article/pii/S0020025516300159.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.

Zhong, Xiao and Enke, David. Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, 67:126 – 139, 2017. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2016.09.027. URL http://www.sciencedirect.com/science/article/pii/S0957417416305115.