MARKETPSYCH RESEARCH

Literature Review on An Ensemble of LSTM Neural Networks for High-Frequency Stock Market Classification

Author of paper: Svetlana Borovkova and Ioannis Tsiamas, Vrije Universiteit Amsterdam

Date of paper: 16 Jul 2018

link: https://poseidon01.ssrn.com/delivery.php?ID=988100070082068118001097087027116111029016091027091055028016021026125123091103106126123010021012119029003107087009092

1 Proposed Methodology

1.1 Main Idea

An ensemble of LSTM models was proposed to predict intra-day price directions of 22 large-cap US stocks. The weighting for each LSTM model in the ensemble was based on its performance on the validate set.

Data was aggregated at 5 minutes interval and the goal was to correctly predict the price direction at each 5 minutes interval. As such, one time step will refer to a 5 minutes interval. Due to further details not provided, my estimate is 84 time steps/datapoints per day.

For each 1 of the 22 stocks to predict, a primary competitor stock was identified and the compeitor stock's data was also used for predictling the stock's price direction.

1.2 Pre-Processing and Feature-Engineering (x and y variables)

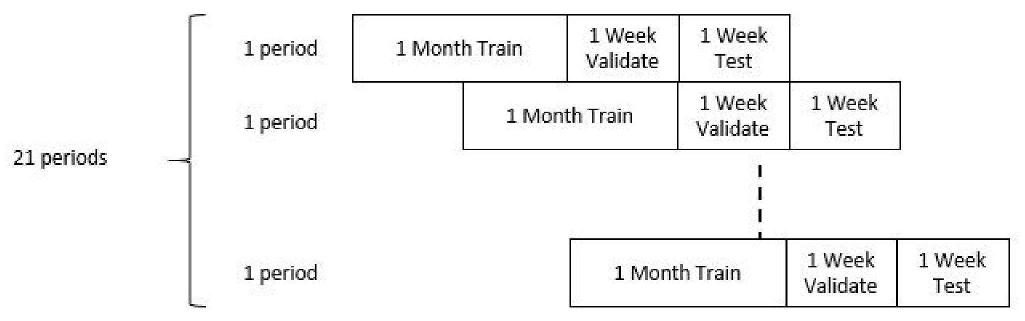
At each time step, the y variables are manually labelled Buy or Sell based on whether the price goes up or down respectively at the next time step.

Many technical indicators were engineered. The stock's technical indicators and the competitor stock's technical indicators were used as x variables. Some of the x variables are as follows:

1. Open
2. Close
3. High
4. Low
5. Volume Weighted Average Price
6. Total number of Trades
7. Percentage change in Price
8. Price Difference (on raw level)
9. Double and Triple Exponentially Smoothed Returns
10. Moving Average Convergence/Divergence

1.3 Train-Validate-Test Set Up

The paper used a rolling window of Train-Validate-Test periods to evaluate performance as follows:



A total of 1 year of data (2014) were used for the 21 periods and 21 weeks (approximately 6 months) were tested.

1.4 Emsemble Modelling and Weighting Method

For every period, 12 LSTM models were trained on the Train set and each of the 12 LSTM model were evaluated by the Area Under the Curve (AUC) score of the Receiver Operating Characteristic (ROC) on the Validate set. The trained 12 LSTM models were then applied on the Test set, and the 12 results were averaged with weights proportional to each AUC score on the validate sets to give rise to 1 result for the Test set.

Finally, over 21 periods, 21 weeks (last 6 months of 2014) were tested, and the average AUC score for the 21 test weeks was used as the final result for each stock.

2. Results and Evaluation

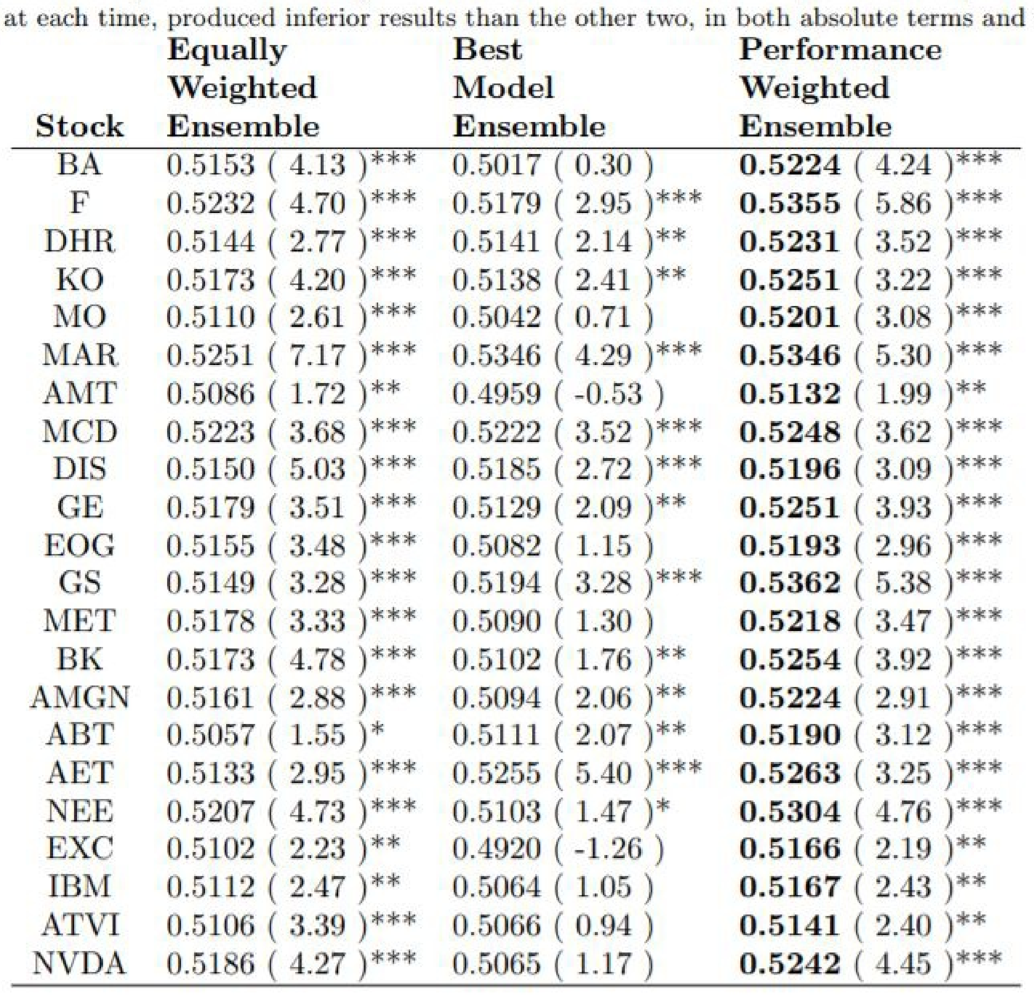
2.1 Proposed Results

The tabulated results (Table 2 below) show the following three comparisons for each stock:

1. use equal weights to weigh the 12 models
2. use only the best out of the 12 models
3. use the proposed performance weights to weigh the 12 models

The performance weighted ensemble consistently produced the best result across all 22 stocks, although only with a fair average AUC score of around 0.52.

Table 2. AUC scores for the LSTM ensembles. Average AUC scores for the 22 stocks. The t-statistic is used to measure whether a AUC score is significantly better than random (0.50). In bold is the best model for the particular stock. Nearly all AUC scores for the Equally Weighted and Performance Weighted ensembles are significantly better than random at $\alpha = 0.01$. One asterisk indicates significance only at $\alpha = 0.1$, double asterisk indicates significance at $\alpha = 0.05$, while the triple one indicates significance at $\alpha = 0.01$. The ensemble, which is comprised from only the best model at each time, produced inferior results than the other two, in both absolute terms and significance.

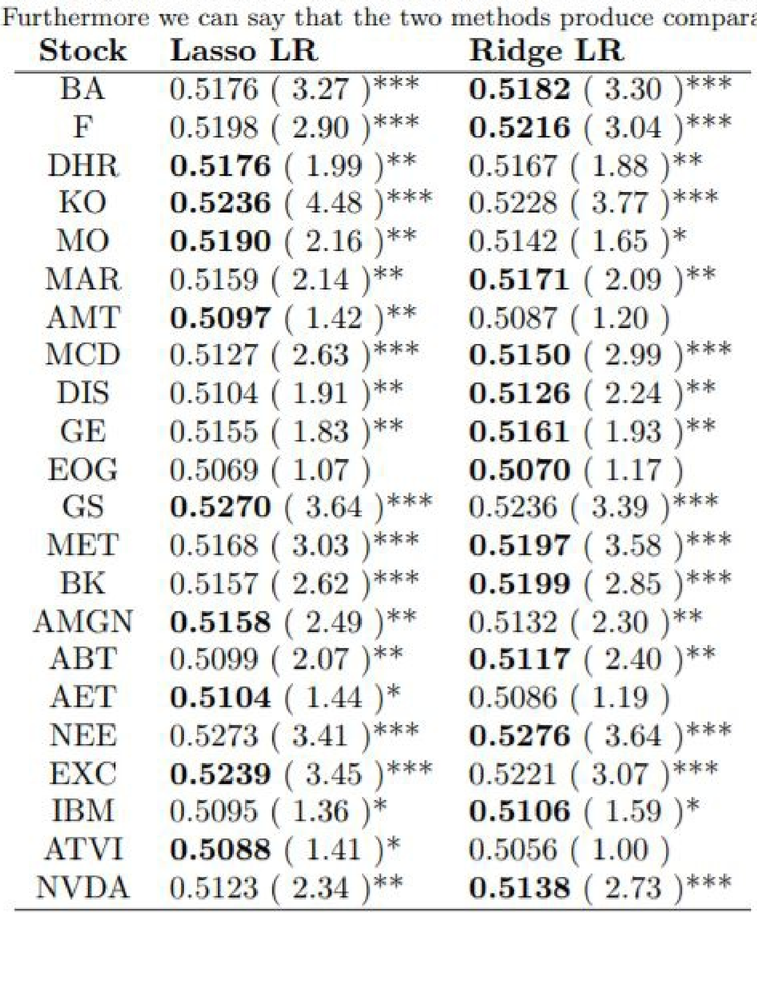


Stock	Equally Weighted Ensemble	Best Model Ensemble	Performance Weighted Ensemble
BA	0.5153 (4.13)***	0.5017 (0.30)	**0.5224** (4.24)***
F	0.5232 (4.70)***	0.5179 (2.95)***	**0.5355** (5.86)***
DHR	0.5144 (2.77)***	0.5141 (2.14)**	**0.5231** (3.52)***
KO	0.5173 (4.20)***	0.5138 (2.41)**	**0.5251** (3.22)***
MO	0.5110 (2.61)***	0.5042 (0.71)	**0.5201** (3.08)***
MAR	0.5251 (7.17)***	0.5346 (4.29)***	**0.5346** (5.30)***
AMT	0.5086 (1.72)**	0.4959 (-0.53)	**0.5132** (1.99)**
MCD	0.5223 (3.68)***	0.5222 (3.52)***	**0.5248** (3.62)***
DIS	0.5150 (5.03)***	0.5185 (2.72)***	**0.5196** (3.09)***
GE	0.5179 (3.51)***	0.5129 (2.09)**	**0.5251** (3.93)***
EOG	0.5155 (3.48)***	0.5082 (1.15)	**0.5193** (2.96)***
GS	0.5149 (3.28)***	0.5194 (3.28)***	**0.5362** (5.38)***
MET	0.5178 (3.33)***	0.5090 (1.30)	**0.5218** (3.47)***
BK	0.5173 (4.78)***	0.5102 (1.76)**	**0.5254** (3.92)***
AMGN	0.5161 (2.88)***	0.5094 (2.06)**	**0.5224** (2.91)***
ABT	0.5057 (1.55)*	0.5111 (2.07)**	**0.5190** (3.12)***
AET	0.5133 (2.95)***	0.5255 (5.40)***	**0.5263** (3.25)***
NEE	0.5207 (4.73)***	0.5103 (1.47)*	**0.5304** (4.76)***
EXC	0.5102 (2.23)**	0.4920 (-1.26)	**0.5166** (2.19)**
IBM	0.5112 (2.47)**	0.5064 (1.05)	**0.5167** (2.43)**
ATVI	0.5106 (3.39)***	0.5066 (0.94)	**0.5141** (2.40)**
NVDA	0.5186 (4.27)***	0.5065 (1.17)	**0.5242** (4.45)***

2.1 Proposed Benchmarks

Lasso and Ridge Logistic Regression (Table 3 below) with the same feature selections for each stock were used as the proposed benchmarks to the proposed LSTM ensemble.

Table 3. AUC scores for Lasso and Ridge Logistic Regressions. Average AUC performance for the 22 stocks by employing Lasso and Ridge logistic regressions with a weight decay of $\lambda = 0.1$. In bold is the best model for the particular stock. The t-statistic in the parenthesis indicates whether the result is significantly better than random (AUC = 0.5). One asterisk indicates significance only at $\alpha = 0.1$, double asterisk indicates significance at $\alpha = 0.05$, while the triple one indicates significance at $\alpha = 0.01$. The majority of the scores are significant at $\alpha = 0.05$ or $\alpha = 0.01$. Furthermore we can say that the two methods produce comparable results.



Stock	Lasso LR	Ridge LR
BA	0.5176 (3.27)***	**0.5182** (3.30)***
F	0.5198 (2.90)***	**0.5216** (3.04)***
DHR	**0.5176** (1.99)**	0.5167 (1.88)**
KO	**0.5236** (4.48)***	0.5228 (3.77)***
MO	**0.5190** (2.16)**	0.5142 (1.65)*
MAR	0.5159 (2.14)**	**0.5171** (2.09)**
AMT	**0.5097** (1.42)**	0.5087 (1.20)
MCD	0.5127 (2.63)***	**0.5150** (2.99)***
DIS	0.5104 (1.91)**	**0.5126** (2.24)**
GE	0.5155 (1.83)**	**0.5161** (1.93)**
EOG	0.5069 (1.07)	**0.5070** (1.17)
GS	**0.5270** (3.64)***	0.5236 (3.39)***
MET	0.5168 (3.03)***	**0.5197** (3.58)***
BK	0.5157 (2.62)***	**0.5199** (2.85)***
AMGN	**0.5158** (2.49)**	0.5132 (2.30)**
ABT	0.5099 (2.07)**	**0.5117** (2.40)**
AET	**0.5104** (1.44)*	0.5086 (1.19)
NEE	0.5273 (3.41)***	**0.5276** (3.64)***
EXC	**0.5239** (3.45)***	0.5221 (3.07)***
IBM	0.5095 (1.36)*	**0.5106** (1.59)*
ATVI	**0.5088** (1.41)*	0.5056 (1.00)
NVDA	0.5123 (2.34)**	**0.5138** (2.73)***

3 Personal Takeaway

3.1 Short Train period

I guess the argument for shorter train time frame is because LSTM model should only remember the sequence of events that most related to predicting the future. If LSTM were to train from too far a period behind, its memory may become noise. Aggregating data at shorter time interval like 5 minutes can help to create more datapoints for training.

3.2 More experiments before ensembling

Ensembling LSTMs seems to be a way to consistently improve the overall results but only by 1 %. More experiments into feature engineering and other set ups could be investigated to produce a substantial improvement in results before using ensembling to gain that extra 1% improvement

Generated on: 2018-12-18 05:49:51.466959
Code revision: 56a7a39361b685f2df4513b989f4dcea22c2c0e7

DISCLAIMER Your use of this the content contained in this publication is at your own risk. Research reported by MarketPsych Data, LLC ("MarketPsych") are based on sources believed to be reliable and are written in good faith, but no representation or warranty, expressed or implied, is made as to their accuracy or completeness. Any opinions expressed or results published in these documents are subject to change without notice. MarketPsych is not responsible for errors or omissions, or responsible to keep any information up to date or to correct any past information. Any recommendations contained in this material are for educational and information purposes only and should not be construed as a solicitation to enter into a commercial relationship or to purchase any security or other financial instrument. MarketPsych does not provide customized or personally tailored advice or recommendations through this publication. Research results and trading signals contained in this material are of a generic nature and are not tailored to the specific circumstances of any visitor or reviewer. You should use any information gathered from this publication only as a starting point for your own independent research and decision making process. Any specific investment recommendations or advice contained or referred to in this material may not be suitable for all investors. You should carefully consider whether any specific recommendation is suitable for you in light of your financial circumstances. You are advised to conduct your own due diligence and seek advice of a qualified professional when it comes to making investment decisions for your own account. The risk of loss in trading securities and other financial instruments can be substantial. No trading program can offer the potential for profit without a corresponding risk of loss. The ability to withstand losses and to adhere to a particular trading program in spite of trading losses are material points which can adversely a ffect investor performance. You may sustain a total loss of you investment. Past performance is not necessarily indicative of future results. MarketPsych is not responsible for the success or failure of your investment decisions relating to information or services presented herein. MarketPsych reserves the right to refuse its services to anyone, either current subscriber or potential subscriber, with or without any reason or for no reason. MarketPsych reserves the right at any time to modify this disclaimer and risk disclosure without notice.