A New Methodology to Exploit Predictive Power in (Open, High, Low, Close) Data

Andrew D. Mann and Denise Gorse

Department of Computer Science, University College London,
London WC1E 6BT, UK
{A.Mann,D.Gorse}@cs.ucl.ac.uk

Abstract. Prediction of financial markets using neural networks and other techniques has predominately focused on the close price. Here, in contrast, the concept of a mid-price based on an Open, High, Low, Close (OHLC) data structure is proposed as a prediction target and shown to be a significantly easier target to forecast, suggesting previous works have attempted to extract predictive power from OHLC data in the wrong context. A prediction framework incorporating a factor discovery and mining process is developed using Randomised Decision Trees, with Long Short Term Memory Recurrent Neural Networks subsequently demonstrating remarkable predictive capabilities of up to 50.73% better than random (75.42% accuracy) on hourly data based on the FGBL German Bund futures contract, and 42.5% better than random (72.04% accuracy) on a comparison Bitcoin dataset.

Keywords: Machine Learning, LSTMs, Decision Trees, Factor Mining, OHLC Data, Financial Forecasting, Mid-Price

1 Introduction

The accurate prediction of an asset's direction has long been the goal of many academics and industry practitioners, with predictive methodologies ranging from the use of traditional technical analysis (TA) to more recent machine learning (ML) techniques. This paper utilises ML technology in the form of Randomised Decision Trees (RDTs) [1] and Long Short Term Memory Recurrent Neural Networks (LSTM RNNs) [2] as a key component in a process for trend detection which takes advantage of the relative ease of prediction of the *mid-price* (defined in terms of OHLC candlestick levels in Section 2.2) when compared to the traditional close price prediction target. RDTs are used to identify the most important factors from a rich factor universe generated from all possible combinations of OHLC lagged levels given L lags, using differences, ratios, and pairwise operations. Within this context it is demonstrated that OHLC levels have a remarkably high predictive potential, in contrast to the negative view espoused by a majority of academics and some practitioners [3][4][5].

2 Background

2.1 Literature Review

There exists no prior literature relating to a mid-price based on a candlestick structure, as proposed here. The common definition of a mid-price is the price halfway between the bid and ask; this has no relevance to the current work. There have, however, been many studies focusing on the predictive power of candlestick patterns. These studies have reported varying results, with most evidencing little or no value in these patterns as predictors of close price movements.

On the negative side, Marshall, Young, and Rose (2005) [3] find that the relationships between OHLC levels have no useful information when applied to stocks in the Dow Jones Industrial Average. Horton (2009) [4] confirms there is little to no value in candlestick charting. Interestingly, Fock, Klein, and Zwergel (2005) [5] present negative results for both the DAX stock index and the FGBL German Bund futures contract, which latter the current work conflicts with (though it should be noted both that our target is different—mid-price rather than close price—and that our OHLC-derived patterns are not traditional candlesticks but data-mined constructions).

On the positive side, Xie et al. (2012) [6] find that candlestick patterns have significant predictive power to forecast US equity returns. Lu et al. (2015) [7] find predictive power in several patterns, but these are rare and the research in addition did not sufficiently address the distinction between candlestick patterns being able to yield profit and their being able to predict trends. One study, that of Lu (2014) [8], finds that traditional patterns have little value but that novel ones may do so; this finding is in line with observations made in the current work, though it should again be emphasised that our use of the mid-price as target creates a very different context.

Overall the evidence in the literature favours the dominant academic belief that candlestick patterns have little value. The results presented below, albeit in the context of mid-price prediction and utilising novel OHLC patterns as input factors, may thus be somewhat of a surprise.

2.2 A Mid-Price Definition and Motivation

Two definitions of mid-price are used in the current work. Mid-price-1 is defined as the price mid-way between a time interval's high and low,

$$\textit{mid-price}\text{-}1 = \frac{\textit{high} + \textit{low}}{2}, \tag{1}$$

while mid-price-2 focuses on the real body of a candlestick (area of the candlestick between open and close) and is defined as

$$\textit{mid-price}\text{-}2 = \frac{\textit{open} + \textit{close}}{2}. \tag{2}$$

The predominant reason for investigating the use of a mid-price as a prediction target was the observation that mid-price time series display far less noise than close price series. As an example, the time series of close price, mid-price-1, and mid-price-2 were examined for 27,927 samples of the German Bund futures data set used here. The standard deviation of price movements in this example data set shows the close price has a standard deviation of 13.71 ticks[1] compared to 11.64 and 10.52 ticks for mid-price-1 and mid-price-2 respectively. Similar results were obtained for many other examples of financial time series data, confirming the mid-price (in particular mid-price-2) as a less noisy target.

2.3 Machine Learning Models Used

Factor Importance Mining. Randomised Decision Trees [9] are used to rank the importance of a factor to its target using the Gini impurity metric which measures the frequency of an incorrect classification of an element in a feature set if it was randomly allocated a classification; a higher value is thus a measure of a more significant level of correlation between factor and target.

Mid-price Prediction. LSTM RNNs are selected as the prediction model due to their ability to detect persistent statistical patterns in sequences while avoiding issues with vanishing gradients[2]; the addition of LSTM units to a RNN allows the network to selectively remember and forget information while retaining long and short term dependencies. The LSTM RNN is here trained to minimise a mean square error loss function using residual back-propagation (RPROP) [10]. RPROP is a first-order optimisation algorithm acting independently over each weight and accounting only for the sign of the partial derivative (ignoring magnitude); this results in a computationally cheap locally adaptive scheme allowing fast convergence in binary classification (here, to predict whether a price movement is up or down).

2.4 Performance Metrics

Normalised Percentage Better than Random (NPBR) and a simple accuracy were used as evaluation metrics. The latter measures the proportion of correctly predicted directional movements; it has the advantage of simplicity but the weakness of being an unreliable indicator of performance in a strongly trending market, where there may be a tendency to overpredict the majority class. NPBR (also known as the Kappa Statistic [11]) is a more robust performance metric for imbalanced data sets, with a range of -100% to 100%, a score of 0% being equivalent to chance. The metric is formalised as

$$t = n_{00} + n_{01} + n_{10} + n_{11}, \tag{3}$$

[1] A tick is the minimum movement in a price series, which for the FGBL futures contract is equivalent to 10 EUR.

[2] Gradient calculations in layers further from the output accumulate progressively more fractional derivative factors, which results in weight changes tending to zero in lower layers and thus vanishing.

$$R_{total} = \frac{(n_{11}+n_{01})(n_{11}+n_{10})+(n_{00}+n_{01})(n_{00}+n_{10}))}{t}, \quad (4)$$

$$NPBR = \frac{(n_{11}+n_{00})-R_{total}}{t-R_{total}}. \quad (5)$$

In this n_{00} represents true negatives, n_{01} false positives, n_{10} false negatives, and n_{11} true positives, these four quantities summing to the total number of predictions, t. This measure allows a comparison against random, which is a valuable metric to state.

3 Methodology

3.1 OHLC Factor Mining

All possible combinations are generated of one hour OHLC bars using differences and ratios given L lags. This rich factor universe is then ranked for importance in relation to a target (mid-price direction at t+1) using Randomised Decision Trees deriving their importance values from the Gini metric. The top N factors are then selected. In this instance N=100 as beyond the top 100 factors the Gini metric curve flattens, as can be observed in Figure 1.

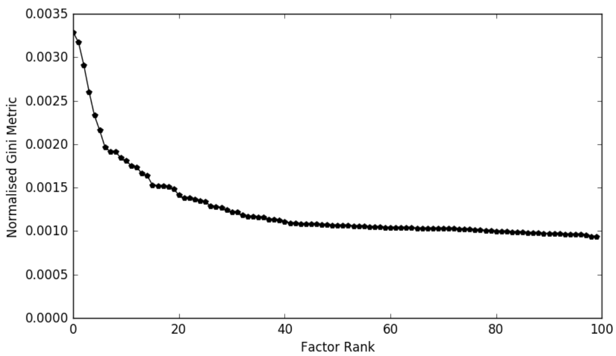

Fig. 1. Ranked Factor Importance Curve

It is notable that the top ranked factors using this machine learning methodology do not include simple lags of the kind considered in the baseline experiments of Section 4.1; in fact conventional lagged inputs do not appear anywhere in the top 100 factors. This supports the later observation that factor mining in itself, without further filtering as described below, gives rise to a large improvement in prediction performance over that seen in the baseline experiments.

The top N factors are then filtered based on correlation to target and factor-to-factor correlation, selecting factors which pass the tests $|corr_{\mathrm{ft}}| \leq c_1$ and $|corr_{\mathrm{ff}}| \geq c_2$ respectively, with c_1 and c_2 optimised on the training set.

3.2 Mid-price Directional Prediction

Once the optimal factors have been selected they were standardised and used to train the LSTM RNN with outputs in the range $[-1, +1]$ and targets of -1 (down) and +1 (up). The net used to produce the results of the next section had eight hidden units and a 2% weight decay; experiments were carried out using other numbers of hidden units and differing amounts of weight decay, but results were found to be robust to reasonable variations of these parameters. It was decided to avoid the risk of overfitting by not optimising these network parameters; the results below, for an out-of-sample dataset, may thus be regarded as generally indicative of the level of predictive power that can be achieved.

4 Results

4.1 Baseline Performance: Use of Close and OHLC Lags as Inputs

A baseline performance was established by investigating the prediction of both close and mid (see Table 1) from close price lags and OHLC lags (defined as a full set of OHLC lags, for two preceding time steps, a total of eight factors in all). Lagged inputs are defined by the equation below,

$$\delta_i = \frac{(p_i - p_{i-1})}{p_{i-1}}, \tag{6}$$

where p_i is the current price and $p_{i\text{-}1}$ is the previous price.

Table 1. Baseline Performance Results

I/O Configuration	Accuracy	NPBR
Close from Close Lags	51.74%	1.89%
Mid-1 from Close Lags	66.15%	32.27%
Mid-2 from Close Lags	69.64%	39.25%
Close from OHLC Lags	51.44%	0.60%
Mid-2 from OHLC Lags	71.34%	42.69%

The first line of Table 1 corresponds to traditional directional prediction; as can be seen from the table results are poor, with only a 51.74% accuracy. However it should be noted that the poor performance derives primarily from the use of close price as a target rather than as a single lagged input. Replacing the target at t+1 by either of the mid-prices, but retaining the simple close lag as input, results in an immediate and large improvement in directional accuracy, with an accuracy of 66.15% and 69.64% for mid-price-1 and mid-price-2 respectively. It is thus possible to predict a mid-price to a high accuracy while continuing to use traditional baseline close price lags as factors.

It can also be seen from the table that using additional open, high, and low (OHL) lagged inputs has only a very small effect on the network's ability to predict close direction; this may well explain why many traditional candlestick patterns appear not to be predictive [3][4][5]. There is however a somewhat more noticeable improvement in mid-price-2 prediction when additional OHL lagged inputs are used; this suggests that mid-price-2 predictions might be improved by a more intelligent selection of OHLC based factors.

At this point only two lags have been considered. The number of lags of OHLC data could have an impact on predictive power and certainly has an impact on the complexity of the model (fewer parameters being preferred).

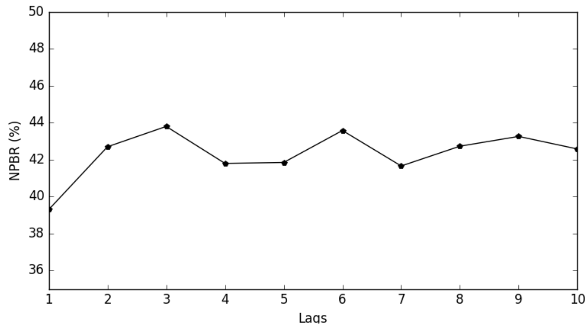

Fig. 2. Factor Lag Experimental Results

Figure 2 shows training data NPBR peaks at three, six and nine OHLC lags. However the maximum is reached at three, implying three lags of OHLC data is sufficient in this context. (Interestingly, many candlestick patterns are created from three lags of OHLC data, such as the Three Line Strike.)

4.2 Use of Mined OHLC Factors as Inputs

Table 1 was suggestive of the possibility that suitably configured OHLC data might enhance mid-price-2 prediction. In the experiments below mined data as described in Section 3.1 were used. The term *Importance Mining* in Table 2 refers to test results using the top 100 importance-ranked factors, and *Correlation Subset* to a reduced input set with those same factors now filtered.

Table 2. Factor Mining Performance Results

I/O Configuration	Accuracy	NPBR
Mid-2: Importance Mining	74.48%	48.75%
Mid-2: Correlation Subset	75.42%	50.73%

It can be seen from Table 2 that factor importance mining does substantially improve the LSTM RNN net's performance, resulting in an increase in NPBR from 42.69% (see OHLC input result in Table 1) to 48.75%. However from Table 2 correlation based filtering adds only a further 1.98% to the NPBR. In addition the optimal values of the correlation thresholds c_1 and c_2 (see Section 3.1) were found to be 0.2 and 1.0, respectively. These observations indicate both that it is the use of the mined factors per se that is predominantly leading to the improvement in performance, and that the LSTM RNN is able to operate effectively without correlation based input screening.

At this point it might appear that the mid-price predictive power could be an artifact of the FGBL futures contract. To allay this concern we apply the same methodology to predict mid-price-2 (with no additional parameter optimisation). Bitcoin was chosen due to its having very different dynamics, being an emerging market highly sensitive to news, exhibiting high volatility, showing the effects of price manipulation, and with low liquidity constraints.

Table 3. German Bund vs. Bitcoin Performance

	Close from Close Lags		**Mid-2 from Correlation Subset**	
I/O Configuration	**Accuracy**	**NPBR**	**Accuracy**	**NPBR**
FGBL Futures	51.74%	1.89%	75.42%	50.73%
Bitcoin	51.47%	0.64%	72.04%	42.5%

As can be seen in Table 3 the performance of Close from Close Lags is similarly poor for Bitcoin as for FGBL futures. However the factor mining methodology (incorporating correlation based filtering with the same thresholds c_1 and c_2 as for FGBL futures) produces a remarkable 42.5% NPBR on Bitcoin, even though it was threshold-optimised on FGBL futures. Thus the predictive value of the mid-price appears to be consistent across vastly different markets.

5 Discussion

It has been shown that use of the proposed mid-price (Equation 2) as target can result in up to a 75.42% prediction accuracy (50.73% NPBR) using appropriate machine learning techniques. OHLC data was used to generate candlestick factors via Randomised Decision Trees which increased the predictive power of an LSTM RNN from an initial 39.25% (Mid-2 from Close Lags) to this maximum of 50.73% NPBR, showing OHLC data does have a high predictive value in relation to the mid-price. However it was demonstrated also that OHLC data did not increase predictive power when forecasting the traditional close price target, which is in line with [3][4][5]. Hence the results here, while they may be surprising, are not at odds with the conclusions drawn in other work. The usefulness of OHLC data is not in predicting the close price, but predicting the mid-price, which has been neglected in past research.

The discovery of the high predictive power of the mid-price is in itself a significant result given the prevailing sentiment that no aspect of an asset's price behaviour can be predicted substantially above random. It is not immediately obvious how to harness this high predictive power within a trading strategy, as a mid-price prediction is not located at a specific moment in time but only within an interval. However a trading strategy built around the mid-price is by no means impossible, though it would necessarily require more for its execution than the simple prediction of this value.

References

1. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning, vol. 63, pp. 3-42 (2006)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation (1997)
3. Marshall, B., Young, M., Rose, L.: Candlestick technical trading strategies: Can they create value for investors. Journal of Banking & Finance, vol. 30, pp. 2303-2323 (2005)
4. Horton, M.: Stars, crows, and doji: The use of candlesticks in stock selection. The Quarterly Review of Economics and Finance, vol. 49, pp. 283-294 (2009)
5. Fock, J., Klein, C., Zwergel, B.: Performance of Candlestick Analysis on Intraday Futures Data. The Journal of Derivatives, col. 13, no. 1, pp. 28-40 (2005)
6. Xie, H., Zhao, X., Wang, S.: A comprehensive look at the predictive information in Japanese candlesticks. International Conference on Computational Science (2012)
7. Lu, T., Chen, Y., Hsu, Y.: Trend definition or holding strategy: What determines the profitability of candlestick charting. Journal of Banking & Finance, vol. 61, pp. 172-183 (2015)
8. Lu, T.: The profitability of candlestick charting in the Taiwan stock market. Pacific-Basin Financial Journal, vol. 26, pp. 65-78 (2014)
9. Breiman, L., Friedman, R.A., Olshen R.A., Stone, C.G.: Classification and Regression Trees. Pacific Grove, CA: Wadsworth. (1984)
10. Riedmiller, M., Braun, H.: A direct adaptive method for faster backpropagation learning: The rprop algorithm. Neural Networks, IEEE International Conference, pp. 586-591 (1993)
11. Smeeton, N.C.: Early History of the Kappa Statistic. Biometrics. 41: 795. JSTOR 2531300 (1985)