Deep Candlestick Mining

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. A data mining process we name *Deep Candlestick Mining* (DCM) is developed using Randomised Decision Trees, Long Short Term Memory Recurrent Neural Networks and k-means++, and is shown to discover candlestick patterns significantly outperforming traditional ones. A test for the predictive ability of novel versus traditional candlestick patterns is devised using all significant candlestick patterns within the traditional or deep mined categories. The deep mined candlestick system demonstrates a remarkable ability to outperform the traditional system by 75.2% and 92.6% on the German Bund 10-year futures contract and EURUSD hourly data.

Keywords: Machine Learning, LSTMs, RNNs, Decision Trees, Clustering, Factor Mining, OHLC Data, Candlestick Patterns

1 Introduction

The ability to predict the movement of financial markets has been a longstanding aim of academics and industry practitioners, using a variety of techniques from technical analysis (TA) to machine learning (ML) and pattern recognition methodologies.

Japanese candlesticks are one of the oldest forms of pattern recognition techniques used to attempt to predict markets. They were first proposed by Munehisa Homma around 1750 for charting the price behaviour of rice markets. Candlestick charts visualise an asset's price by aggregating period specific bars (e.g. 1 hour bars) consisting of open, high, low and close (OHLC) price levels, and frequently-sequential patterns are used as a tool to predict future market direction. Many industry practitioners believe candlestick patterns are an effective predictive tool, though there is much debate in the academic world as to their effectiveness [1][2][3].

In this paper a new process referred to as Deep Candlestick Mining (DCM) is proposed as a means to discover asset-specific predictive candlestick patterns using ML techniques such as Randomised Decision Trees (RDT) [4], Long Short Term Memory Recurrent Neural Networks (LSTM RNNs) [5] and k-means++ [6]. DCM-based prediction is shown to substantially outperform the use of traditional candlestick patterns on hourly data for the German 10-year futures contract (FGBL) and the EURUSD markets.

2 Background

2.1 Literature Review

There have been many academic studies focusing on the power of candlestick patterns, reporting varying results. Most studies conclude there is little or no value in using candlestick patterns to predict future directional price movements.

Marshall, Young and Rose (2005) [1] find that candlestick Open, High, Low, Close (OHLC) levels contain no useful information in the case of the Dow Jones Industrial Average. Further negative findings are reported by Horton (2009) [2] and by Fock, Klein, and Zwergel (2005) [3]. The latter applied candlestick charting techniques to both the DAX and the FGBL futures contract—interestingly this study presents positive findings on FGBL, but only by using the proposed deep mining process.

On the positive side significant directional prediction power is found in candlestick charting by Xie et al. (2012) [7] on US equity returns. Notably Lu (2014) [8] finds evidence of statistically significant candlestick patterns, three of which are novel, found using a simple four-price-level approach (although the rules were defined by Lu and not data mined as here). The results presented here show further evidence, through an exhaustive mining process, that novel candlestick patterns can be an effective tool to predict future directional price movement.

2.2 Machine Learning Models Used

Factor Importance Mining. The importance of a factor to its target is analysed using Randomised Decision Trees (RDT) [9]. To produce a ranked dictionary of factors (with most important at the top) the RDT uses the Gini impurity metric to measure the frequency of incorrect classification if a classification were to be randomly allocated; higher values indicate a greater correlation between the factor and its target.

Directional Prediction. A Long Short Term Memory Recurrent Neural Network (LSTM RNN) is used as the directional prediction model taking factors influenced from the factor importance mining step as input. The LSTM RNN is trained using RPROP [10], a first-order optimisation algorithm that uses only the sign of the partial derivative, ignoring magnitude, and acts independently on each weight. RPROP is beneficial in data-intensive applications as it provides a computationally cheap and fast-converging locally adaptive method for binary classification (here, into price movements predicted to be up or down).

Candlestick Mining. K-means++ is used to cluster the LSTM RNN test set factors. K-means++ is a data mining clustering algorithm which improves on k-means by providing an approximate solution to the NP-hard problem of selecting initial cluster centroids. We will later analyse these clusters to find out what directionally predicting OHLC patterns they represent.

2.3 Dataset Usage

Eleven years of hourly data are used, as shown in Figure 1. The LSTM RNN is trained using five years of data. A dataset of two years is then used to assess the LSTM RNN's performance. This performance is then analysed and the factors used clustered to extract meaning—this is where the deep candlestick mining occurs. A further dataset of two years is used to select those candlestick patterns most effective in prediction. A final two years of data is used as an out of sample test set to assess the effectiveness of the developed candlestick prediction system.

[5 YEARS]	[2 YEARS]	[2 YEARS]	[2 YEARS]
LSTM RNN TRAIN SET	LSTM RNN TESTING & K-MEANS++ (CANDLESTICK MINING)	FILTER CANDLESTICKS FOR SIGNIFICANCE	TEST CANDLESTICK PREDICTION SYSTEM
[DATASET 1]	**[DATASET 2]**	**[DATASET 3]**	**[DATASET 4]**

Fig. 1. Dataset Usage

2.4 Performance Metrics

Directional accuracy and Normalised Percentage Better than Random (NPBR) are used as evaluation metrics in this study. The former is simply the proportion of correct predictions whereas NPBR (also known as the Kappa Statistic [11]), as used by us previously [13], is more appropriate in trending markets, where there would be a tendency to overpredict the majority class. NPBR, which ranges from -100% to 100%, heavily penalises such overprediction and would assign a value of 0%—equivalent to random chance—to the case in which all instances were assigned to the majority class.

3 Methodology

3.1 OHLC Factor Mining

All possible combinations of ratios and differences of one hour OHLC data are calculated given L lags. Randomised Decision Trees are then used to rank the importance of each factor to a target (in this case the future close price directional change), deriving the importance value from the Gini metric. The top N factors are selected by inspection of the Gini metric curve. As can be seen in Figure 2 the Gini metric curve noticeably flattens for FGBL beyond N=100, though this is not the case for EURUSD. N should ideally be optimised for each asset when selecting a factor universe. However to keep a consistent approach in demonstrating the Deep Candlestick Mining process we chose to use a constant N=100 here with no further optimization; the results presented below are therefore a general indication of the process's utility.

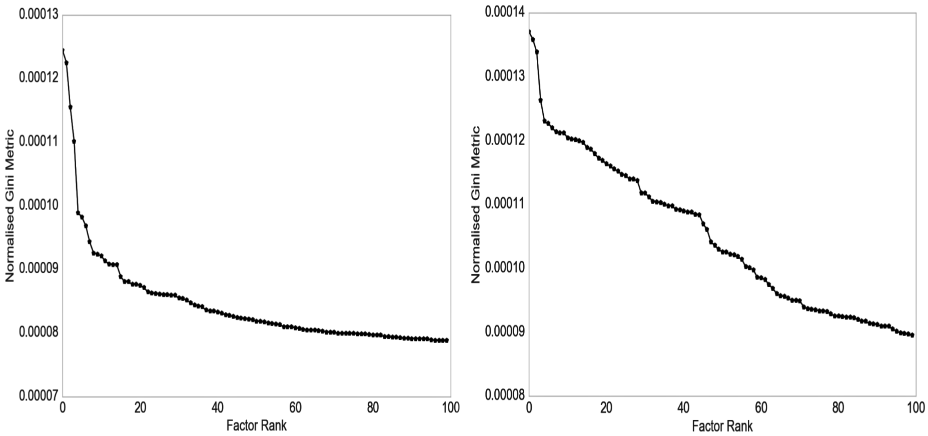

Fig. 2. Three-Lag Importance Mining Curve: FGBL (left) and EURUSD (right)

Using the top N factor universe a further filtering is then applied focusing on correlation of factor-to-target (*ft*) and factor-to-factor (*ff*). Factors that pass the tests $|corr_{\mathrm{ft}}| \leq c_1$ and $|corr_{\mathrm{ff}}| \geq c_2$ make the optimal factor universe, with c_1 and c_2 being optimised on Dataset 1 (see Figure 1).

3.2 Close-price Directional Prediction

The optimal factors are then standardised and used as inputs to the LSTM RNN with targets of -1 (down) and +1 (up). The network architecture used 8 hidden LSTM units with a weight decay factor of 2%. Other architecture configurations were tested but results were found to be robust to reasonable variations of these quantities. It was decided not to optimise the network parameters to avoid the risk of overfitting. As with the decision in Section 3.1 to use a constant N=100, results can therefore be viewed as a performance indicator where there is scope to improve the process.

3.3 Clustering

The LSTM RNN factors which were used to perform the directional prediction (Dataset 2) are now clustered using k-means++, where k is selected by maximising the Silhouette Coefficient [14]. An initial (parent) clustering revealed an interesting split in the data structure at k=2, which was verified as real by plotting the magnitude of each factor dimension and verifying the clusters had very different structure. This clustering was then re-clustered into child clusters with the aim of revealing more interesting candlestick patterns. The optimal parent and child clustering configuration was found at k=2,6 and k=2,9 for FGBL and EURUSD respectfully. Other clustering techniques and k-selection criteria could have been used; however the optimal selection of a clustering algorithm and associated selection criteria is outside the scope of this paper.

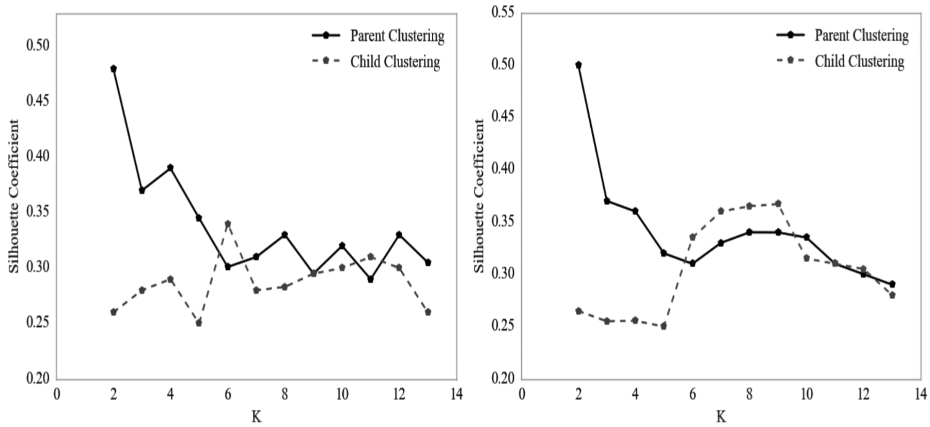

Fig. 3. Three-Lag Silhouette Coefficients: FGBL (left) and EURUSD (right)

3.4 Candlestick Mining

For each cluster we look at: (1) the LSTM RNN's NPBR; (2) the direction the cluster represents. The latter is done by indexing each candidate in a cluster and computing an up-movement ratio (defined as proportion of up movements at $t+1$). It is important to confirm the LSTM RNN predicts the same direction the cluster is representing. If the LSTM RNN's NPBR is greater than 0%, the percentage of up movements deviates from 50% (indicating a directional bias) and the LSTM RNN's majority prediction direction agrees with the direction the cluster represents, then the cluster is valid. Clusters are then further validated by for each member identifying the OHLC patterns it corresponds to, in order to ensure the clustering did indeed group together patterns of similar shape; in all instances this was found to be the case. The mined candlestick patterns will be the centroids of the clusters. These patterns are essentially what the LSTM RNN would have seen if it had been looking at OHLC data as a human might look at a candlestick chart when a prediction was made.

4 Results

4.1 Traditional Candlestick Patterns

To assess the power of the deep mined candlestick patterns against an appropriate baseline an assessment of 100 bull (predicting up) and bear (predicting down) traditional candlestick patterns (50 candlestick types)[1] were tested on

[1] 2 Crows; 3 Black Crows; 3 Inside; 3 Line Strike; 3 Outside; 3 Stars in South; 3 White Soldiers; Abandoned Baby; Advance Block; Belt Hold; Break Away; Closing Marubozu; Conceal Baby Swell; Counter Attack; Dark Cloud Cover; Down Side

FGBL and EURUSD hourly data (Dataset 3). Significance levels were calculated using a binomial distribution (as in [12]), where the null hypothesis was candlesticks are no better than guessing, which translates to 50% directional accuracy. *Significant candlesticks* (see Table 1) are patterns with a directional predictive power significant at 10% or better.

Table 1. Significant Traditional Candlestick Patterns

Candlestick Pattern	**Asset**	**Number of Candlesticks**	**N**	**Accuracy**	**Type**	**Significance Level**
Advanced Block	FGBL	3	83	54.21%	Bear	$*$
3 Outside	FGBL	3	74	54.05%	Bull	$*$
3 Inside	FGBL	3	20	55.00%	Bear	$**$
Harami	FGBL	2	104	52.88%	Bear	$*$
Harami	EURUSD	2	238	57.14%	Bull	$*$
Inverted Hammer	EURUSD	1	76	55.26%	Bull	$*$
Matching Low	EURUSD	2	221	55.20%	Bull	$**$
Advanced Block	EURUSD	3	130	53.80%	Bear	$*$

($*$: significant at 10%; $**$: significant at 5%; $***$: significant at 1%)

It should be noted from the above that only four patterns were significant at the 5% and 10% levels and no pattern was significant at the 1% level. Hence while there is some predictive ability in traditional candlestick patterns it appears not to be widespread, in line with the negative results of the majority of academic studies into candlestick charting.

4.2 Deep Mined Candlestick Patterns

Deep mined candlestick patterns are dataset-specific, being mined from the dataset the LSTM RNN predicted on. For FGBL eight candlesticks were found to be significant at 10% or better; for EURUSD this number was five. The significant patterns for both datasets are listed in Table 2.

Interestingly there were two significant candlestick patterns on EURUSD at the 1% level, while in contrast no patterns were found to be significant at this level for traditional patterns. Moreover the significant deep mined candlesticks have an average accuracy of 57.04% and 57.7% on FGBL and EURUSD respectively, while the average accuracy for the significant traditional patterns was in comparison 54.04% on FGBL and 55.35% on EURUSD, showing the deep mined patterns outperformed the traditional patterns by 3% and 2.35% respectively.

Gap 3 Methods; Downside Gap 2 Crows; Engulfing; Evening Star; Gap Side White; Hammer; Hanging Man; Harami; High Wave; Hikkake; Hikkake Mod; Homing Pigeon; Identical 3 Crows; In Neck; Inverted Hammer; Ladder Bottom; Long Line; Marubozu; Mat Hold; Matching Low; Morning Star; Piercing; Rise Fall 3 Methods; Separating Lines; Shooting Star; Short Line; Spinning Top; Stalled Pattern; Stick Sandwich; Takuri; Tasuki Gap; Thrusting; Tri Star; Unique 3 River

Table 2. Significant Deep Mined Patterns

Candlestick Pattern	Asset	Number of Candlesticks	N	Accuracy	Type	Significance Level
Pattern 0,2	FGBL	2	565	53.09%	Bull	$*$
Pattern 3,1	FGBL	2	83	61.44%	Bear	$**$
Pattern 1,0	FGBL	3	30	60.00%	Bull	$*$
Pattern 0,6	FGBL	3	92	59.78%	Bear	$**$
Pattern 5,6	FGBL	2	178	58.43%	Bear	$**$
Pattern 2,1	FGBL	2	563	55.06%	Bear	$**$
Pattern 5,2	FGBL	2	150	55.33%	Bear	$*$
Pattern 4,1	FGBL	2	408	53.19%	Bear	$*$
Pattern 1,1	EURUSD	4	312	58.01%	Bull	$***$
Pattern 1,3	EURUSD	4	47	57.44%	Bull	$*$
Pattern 1,7	EURUSD	4	156	58.33%	Bull	$**$
Pattern 1,0	EURUSD	3	73	62.64%	Bull	$*$
Pattern 1,6	EURUSD	2	470	57.02%	Bear	$***$

($*$: significant at 10%; $**$: significant at 5%; $***$: significant at 1%)

Figures 4 and 5 show examples of novel patterns discovered by the deep candlestick mining (DCM) process. It is notable that clusters 0,2 and 2,1 in Figure 4 (FGBL) look very similar to the bullish and bearish Engulfing candlestick pattern in reverse. Deep mined candlestick pattern 0,2 (leftmost) is in fact a traditional candlestick pattern called Bearish Harami which was identified as being significant when the traditional candlestick patterns were analysed. This is an important point as it shows the deep mining process can both find new candlestick patterns and identify significant traditional ones.

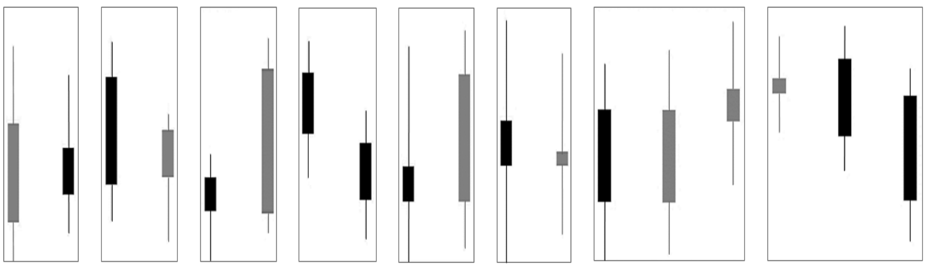

Fig. 4. FGBL Candlestick Patterns: 0,2; 2,1; 3,1; 4,1; 5,2; 5,6; 0,6; 1,0

Candlestick patterns for EURUSD (examples in Figure 5) appear on average to require more lags to be significant, implying a greater level of information content is required to make correct predictions. For EURUSD there were no discovered correspondences between mined and significant traditional patterns.

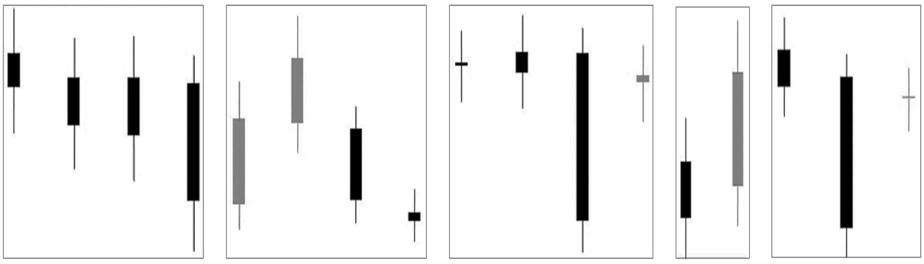

Fig. 5. EURUSD Candlestick Patterns: 1,1; 1,3; 1,7; 1,6; 1,0

4.3 Traditional Vs. Deep Mined Candlestick System

Often a practitioner will use multiple candlestick patterns for making decisions. A comparison in this spirit between traditional and deep mined candlesticks was carried out by using all the patterns available in either category. Dataset 4 was used to assess the predictive power of both systems, in terms of NPBR. As can be seen in Table 3 the DCM system outperformed the traditional system by 75.2% and 92.6% on FGBL and EURUSD respectively.

Table 3. Traditional Prediction System vs. Deep Mined Prediction System

Asset	Number of Traditional Patterns	Traditional NPBR	Number of Deep Mined Patterns	Deep Mined NPBR
FGBL	4	3.48%	8	6.10%
EURUSD	4	6.52%	5	12.56%

5 Discussion

The deep candlestick mining (DCM) process introduced here has been shown to be remarkably effective at discovering statistically significant OHLC patterns. This is not in conflict with the many academic studies which claim candlestick patterns have no, or limited, predictive power [1][2][3] because the patterns the DCM process discovers are largely novel (though for FGBL some interesting correspondences with traditional candlestick patterns were discovered). DCM-derived patterns outperformed the best-discovered traditional patterns by 75.2% and 92.6% on FGBL and EURUSD respectively in relation to their ability to forecast directional movement better than random. The DCM process has many parts that could be further optimised to produce potentially better results. It would be expected these optimisations would be both asset and time period granularity (daily, hourly, minute, etc.) dependent. The results here are therefore only an early indication of the promise of deep candlestick mining.

References

1. 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)
2. 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)
3. Fock, J., Klein, C., Zwergel, B.: Performance of candlestick analysis on intraday futures data. The Journal of Derivatives, vol. 13, no. 1, pp. 28-40 (2005)
4. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning, vol. 63, pp. 3-42 (2006)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation vol. 9, no. 8, pp. 1735-1780 (1997)
6. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: 18th annual ACM-SIAM symposium on Discrete algorithms, pp. 1027-1035. Society for Industrial and Applied Mathematics Philadelphia, PA, USA (2007)
7. Xie, H., Zhao, X., Wang, S.: A comprehensive look at the predictive information in Japanese candlesticks. International Conference on Computational Science (2012)
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)
12. Prado, A.H., Ferneda, E., Morais R.C.L., Luiz, B.J.A., Matsura, E.: On the effectiveness of candlestick chart analysis for the Brazilian stock market. In: 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, pp. 1136-1145, Procedia Computer Science vol. 22 (2013)
13. Mann, A.D., Gorse, D.: A new methodology to exploit predictive power in (open, high, low, close) data. In: 26th International Conference on Artificial Neural Networks, LNCS, Springer Heidelberg (in press)
14. Rousseeuw, J.P.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Computational and Applied Mathematics, vol. 20, pp. 53-65