

**FULL-LENGTH REPRINT – differs from the original in layout but not in contents**

Chmielewski L., Janowicz M., Orłowski A.,  
Prediction of trend reversals in stock market by classification of Japanese candlesticks.  
In R. Burduk et al., editors, Proc. 9th Int. Conf. on Computer Recognition Systems CORES 2015,  
Wrocław, Poland, 25-27 May 2015  
vol. 403 of *Advances in Intelligent Systems and Computing*, pages 641-647.  
The copyright owner of the paper is Springer. The final publication is available at SpringerLink  
via [doi:10.1007/978-3-319-26227-7\\_60](https://doi.org/10.1007/978-3-319-26227-7_60).

## Prediction of trend reversals in stock market by classification of Japanese candlesticks

Leszek J Chmielewski, Maciej Janowicz, and Arkadiusz Orłowski

Faculty of Applied Informatics and Mathematics (WZIM),  
Warsaw University of Life Sciences (SGGW), Poland  
ul. Nowoursynowska 159, 02-775 Warsaw, Poland  
{leszek\_chmielewski,maciej\_janowicz,arkadiusz\_orlowski}@sggw.pl  
<http://www.wzim.sggw.pl>

**Abstract.** K-means clustering algorithm has been used to classify patterns of Japanese candlesticks which accompany the prices of several assets registered in the Warsaw stock exchange (GPW). It has been found that the trend reversals seem to be preceded by specific combinations of candlesticks with notable frequency. Surprisingly, the same patterns appear in both *bullish* and *bearish* trend reversals. The above findings should stimulate further studies on the problem of applicability of the so-called technical analysis in the stock markets.

**Keywords:** clustering, K-means, trend reversals, Japanese candlesticks

### 1 Introduction

The technical analysis of the stock market assets [1,2] belongs to the most controversial approaches to investigations of data series, especially having economical and financial meaning. This is because the aim of technical analysis is, basically, no less than the approximate prediction of specific values of the data which appear as realizations of a random process. On one hand, it has been declared a kind of pseudoscience, which, because of the incorrectness of its most important principles, cannot lead to any sustainable increase of returns above the market level [3,4]. On the other hand, it has also been applied without any serious knowledge about the market dynamics. Considerable revision of the ultra-critical stand of the many experts regarding technical analysis has been influenced by more recent publications, e.g. [5,6,7,8]. As a part of a common knowledge about the stock market dynamics let us notice that time series generated by the prices of stocks are *not* random walks, and at least the short-time correlations *are* present. Whether they can indeed be exploited with the purpose of maximization of returns is an open question. What we investigate here is meant to be a very small contribution to answer it. It is well known that the very important part of the technical analysis is the localization of possible ends of a given trend – upward or downward. Many technical-analysis indicators (like MACD – Moving Average Convergence-Divergence or RSI – Relative Strength

Index) are used by technical analysts for this purpose. Another tool used to this end are the graphical patterns made by the sequences of Japanese candlesticks. A Japanese candlestick is a four-element sequence containing the opening, maximal, minimal, and closing prices of an asset during a given trade session. We do not analyze here the patterns themselves but we attempt to obtain an answer to the following rather humble question: are the trend reversals accompanied more often by some types of candlesticks than by others? For that purpose we perform first the classification of candlesticks to have the above *types* well defined. For the purpose of classification we employ a well-established clustering algorithm called K-means [9,10,11,12,13]. Let us notice that clusterization of candlesticks for a given asset allows one to ascribe labels to them. This, in turn, makes it possible to investigate how their sequences with given labels have performed in the past and what is the predictive power (if any) of sequences with particular labels.

The main body of this work is organized as follows. In Section 2 we recall and modify the definition of the Japanese candlesticks which form our working example. In Section 3 we provide our results about the relation between candlestick types and trend reversals in prices of several stocks. Finally, Section 4 comprises some concluding remarks.

## 2 Japanese candlesticks as a representation of value of assets in stock market

The Japanese candlestick is a sequence of four numbers  $(O(a, t), X(a, t), N(a, t), C(a, t))$ , where  $O$  denotes the opening value of the asset  $a$  at the trading day  $t$ ,  $X$  is the maximum value (*high*) reached during the trading session,  $N$  is the minimum (*low*), and  $C$  is the closing value. There exists a well-known graphical representation of the candlestick [14] often considered important in what is called the technical analysis of stock markets.

In what follows below we employ a sequence of five elements  $(O, X, N, C, V)$  which we call an *augmented Japanese candlestick* where  $V$  represents the transaction volume associated with the asset and the trading day. An augmented candlestick of the asset  $a$  on the day  $t$  can be denoted as a 5-tuple

$$\mathbf{y}(a; t) = (O(a, t), X(a, t), N(a, t), C(a, t), V(a, t)) . \quad (1)$$

In the following we shall call it simply a candlestick. The time series of  $n + 1$  candlesticks, called otherwise a sequence, can be written down as

$$S_n(a; t) = (\mathbf{y}(a; t), \mathbf{y}(a; t + 1), \dots, \mathbf{y}(a; t + n)) . \quad (2)$$

Each sequence has its own starting time  $t$  and ending time  $t + n$ .

We define the (Euclidean) distance between two candlesticks (differently than in our previous work [15]) as

$$d(\mathbf{y}_1, \mathbf{y}_2) = \sum_i (y_{1,i} - y_{2,i})^2 , \quad (3)$$

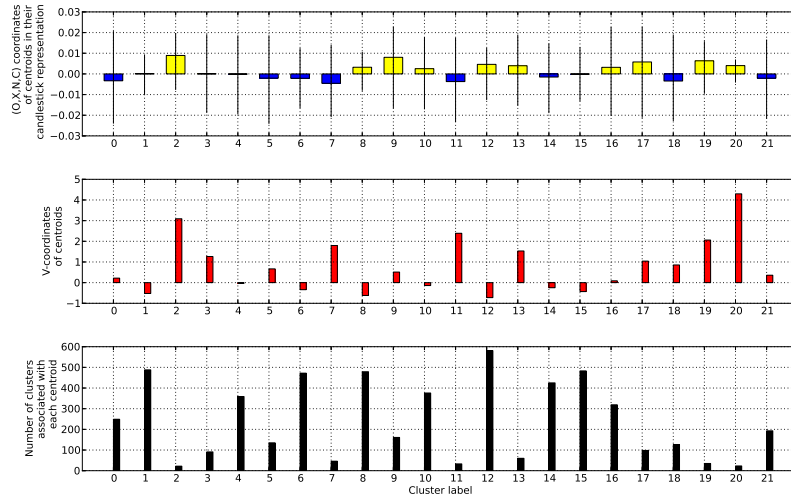
where  $y_{1,i}$  and  $y_{2,i}$  are corresponding components of  $\mathbf{y}_1$  and  $\mathbf{y}_2$ , respectively, i.e., they run through the elements of appropriate sets  $\{O, X, N, C, V\}$ . In order to consider this formula meaningful, the values of the asset and the transaction volume must be comparable. To achieve this, we normalize our time series by subtracting the average of closing values  $C$  from  $O$ ,  $X$ ,  $N$ , and  $C$ , and then dividing these differences by the standard deviation of  $C$ . Similarly, the volume  $V$  is also normalized by subtracting its average value and dividing this difference by the standard deviation of original  $V$ . This way, the standard deviations of new renormalized  $C$  and  $V$  are exactly 1. All candlesticks analyzed further are normalized in the above sense.

Using the K-means algorithm we have classified the Japanese candlesticks as they appear in the group of twenty biggest and most powerful stocks (WIG20) in the Warsaw stock exchange. This has been done for each of the twenty stocks separately. We have assumed that there are 32 clusters. This should have corresponded to the 32 semi-quantitative features of the candlesticks: a candlestick can be *black* (close value lower than open value) or *white*, its body length (i.e. absolute value of the difference between the close and open value) can be larger or smaller than average, its upper *shadow* (i.e. the difference between the maximum and the larger of open and close value) can be longer or shorter than average, the lower *shadow* can also be long or short with analogous meaning; finally, the corresponding volume can be larger or smaller than average. All this gives  $2^5$  features, hence 32 clusters. An example of the coordinates of the centroids associated with each cluster for BZWBK stocks is shown in Fig. 1. We have used the implementation of the K-means algorithm as given in the module `Scikit-learn` [16,17]. To improve presentation, for every centroid with coordinates  $(O, X, N, C, V)$  we subtracted the first coordinate from the first four to obtain  $(0, X - O, N - O, C - O, V)$  and displayed its candlestick representation. This is the reason why the candles in the upper part of Fig. 1 have the same level of opening values.

### 3 Trend reversals and candlesticks

It is by no means self-evident what the trend in the data obtained from the random process really means qualitatively even though the intuitive meaning is understandable. In particular, it is not clear when the trend actually starts and when it ends. To quantify these concepts, we have employed the following simple procedure. To every closing value of a given asset we have ascribed the slope of the straight line obtained from the preceding 10 (ten) closing values. As an indicator of the start and end of the trend we have chosen the change of sign of the above slope. A justification for using the number 10 is that the period of two trading weeks is usually considered important by the technical analysts.

To every change of sign of the slope as defined above which appears between  $n$ th and  $(n + 1)$ th trading sessions we have ascribed one of the 32 cluster labels (from 0 to 31) of the candlestick which appeared in the session  $n$ . If, however, a cluster with a given label contained less than  $N/240$  candlesticks, where  $N$  is



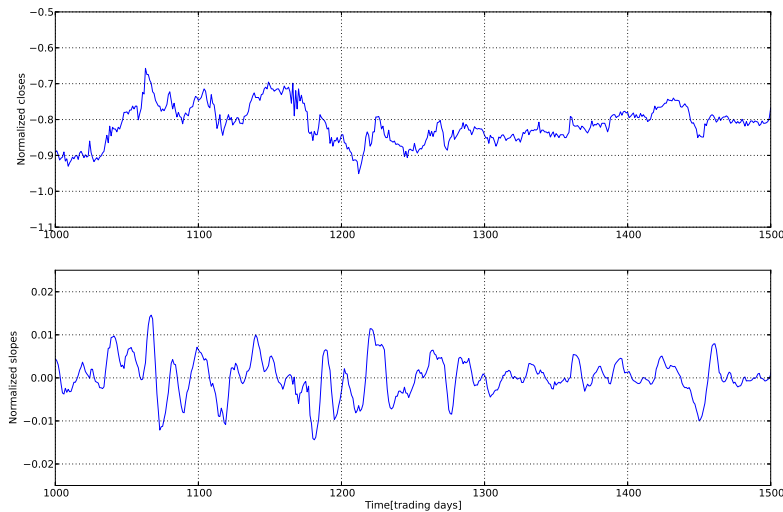
**Fig. 1.** Coordinates of centroids corresponding to clusters of Japanese candlesticks obtained using the K-means algorithm for the BZWBK stocks (5327 trading days with equal number of candlesticks). First four coordinates of centroids shown in upper part of the figure are displayed as candlesticks, the fifth coordinates (i.e. volume) are given as red horizontal bars in the middle part of the figure while the black horizontal bars in the figures illustrate the numerical amount of elements in the clusters corresponding to the centroids.

the number of trading days considered, it has been discarded. For instance, in the case of BZWBK stocks ( $N = 5327$ ) we have kept only 22 clusters (and 22 labels). We have applied such a filter to exclude candlesticks which appear too rarely, less than once per year.

In Fig. 2 we have plotted an example of the time series generated by a stock in GPW (BZWBK stock, trading session No. 1000-1500) together with the associated time-dependence of the slopes generated by ten preceding closing values.

In Fig. 3 we have plotted the relative frequencies of appearances of the cluster labels for the change from downward to upward trend and from the upward to downward for BZWBK stocks. It has been calculated as the ratios  $M_{j,du}/L_j$  and  $M_{j,ud}/L_j$ , where  $M_{j,du}$  is the number of appearances of a candle belonging to the  $i$ th cluster near the down-to-up trend reversal,  $M_{j,ud}$  is the number of appearances of a candle belonging to the  $j$ th cluster near the up-to-down trend reversal, and  $L_j$  is the total amount of appearances of any candle belonging to  $j$ th cluster.

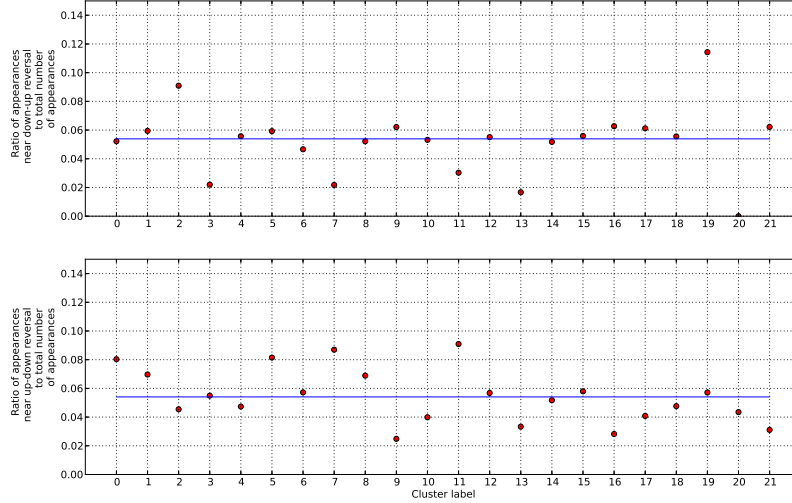
This and similar figures which we have obtained for other stocks traded at GPW have been somewhat surprising to us. Indeed, it seems that there are such types of candlesticks which appear frequently close to the trend reversals (as



**Fig. 2.** Example of the time series generated by a stock in GPW (BZWBK stock, trading session no. 1000-1500) – upper part of the figure, together with the associated trading-session dependence of the slopes generated by ten preceding closing values – lower part of the figure.

defined above) and relatively rarely outside the regions close to the zeros of the slope series. Since the total number of those zeros has been of the order of  $N/10$  (575 for BZWBK stocks) where  $N$  is the number of trading sessions, one may be tempted to say that, in fact, candlesticks of some types are concentrated in the trend-reversal regions. We have observed that this behavior is more pronounced in the stocks which are traded longer in GPW and much less visible for stocks relatively new in the market. It is quite evident that the most significant from the point of view of the trend reversals are those types of candlesticks which appear quite infrequently out of the *reversal moments*. This is rather intuitive from the point of view of technical analysis.

What is more, as for the down-to-up reversal, the most significant candlestick is the one with relatively long and light body (i.e. closing price is larger than the opening price), relatively short *shadows* and quite large volume. This means that during the trading day there is almost a steady grow and the interest of investors to buy an asset does not diminish. The fact that, with such investors' mood, the new *bullish* trend is likely, seems to be intuitive. Similarly, the sticks which appear relatively likely near the up-to-down reversals are those with closing prices lower than the opening ones but with relatively long *shadows* and large volumes. It seems that the trade days on which the prices are characterized with such candlesticks are very turbulent but such that the pessimistic moods finally settle down and the *bearish* reversal becomes likely.



**Fig. 3.** Relative frequencies of appearances of the cluster labels for the change from downward to upward trend (upper part of the figure) and from the upward to downward (lower part of the figure) for BZWBK stocks. The horizontal lines represent the ratio of the total number of trend reversals to the number of trading days.

#### 4 Concluding remarks

In this work we have performed classification of the Japanese candlesticks which appear in the stocks traded in the Warsaw stock exchange using the standard K-means clustering algorithm. With each cluster we have associated a label which specifies a *type* of candlesticks. We have paid particular attention to those types (labels) of candlesticks which are close to the trend reversals. We have found that there exist types of candlesticks which frequently tend to appear close to the trend-reversal regions and others which cannot be found in such regions.

Needless to say, the above results are very preliminary and require careful reexamination. What is especially important is to find a much more convincing definition of the trend and the trend reversal than that given in this work in terms of linear regression slopes. What is more, as often stressed by technical analysts, the candlesticks are to be considered within the specific market context. This can be done using other technical-analysis indicators. We hope to report results of such improved analysis in a forthcoming publication.

## References

1. Murphy, J.: *Technical Analysis of Financial Markets*. New York Institute of Finance, New York (1999)
2. Kaufman, P.: *Trading Systems and Methods*. John Wiley and Sons, New York (2013)
3. Malkiel, B.: *A Random Walk Down the Wall Street*. Norton, New York (1981)
4. Fama, E., Blume, M.: Filter rules and stock-market trading. *Journal of Business* **39** (1966) 226–241
5. Brock, W., Lakonishok, J., LeBaron, B.: Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* **47**(5) (1992) 1731–1764
6. Lo, A., MacKinley, A.: Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* **1** (1988) 41–66
7. Lo, A., MacKinley, A.: *A Non-Random Walk down Wall Street*. Princeton University Press, Princeton (1999)
8. Lo, A., Mamaysky, H., Wang, J.: Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance* **55**(4) (2000) 1705–1765
9. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In Cam, M.L., Neyman, J., eds.: *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*. Volume 1., Berkeley, University of California Press (1967) 281–297
10. Steinhaus, H.: Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci.* **4**(12) (1957) 801–804
11. Lloyd, S.: Least square quantization in PCM (1957) Bell Telephone Laboratories Paper.
12. Forgy, E.: Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* **21**(3) (1965) 768–769
13. Hartigan, J.: *Clustering Algorithms*. John Wiley and Sons, New York (1975)
14. Wikipedia: Candlestick chart — Wikipedia, The Free Encyclopedia (2014) <http://en.wikipedia.org/w/index.php?title=Candlestick.chart>. [Online; accessed 19-December-2014].
15. Chmielewski, L., Janowicz, M., Orłowski, A.: Clustering algorithm based on molecular dynamics with Nose-Hoover thermostat. Application to Japanese candlesticks. In Rutkowski, L., et al., eds.: *Artificial Intelligence and Soft Computing: Proc. Int. Conf. ICAISC 2015*. Lecture Notes in Artificial Intelligence series Vol. 9120, Springer (2015) 330–340 doi:10.1007/978-3-319-19369-4\_30.
16. Scikit-learn Community: Scikit-learn – machine learning in Python (2015) <http://scikit-learn.org>. [Online; accessed 10-February-2015].
17. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12** (2011) 2825–2830