Unsupervised Discovery of Significant Candlestick Patterns for Forecasting Security Price Movements

Karsten Martiny
Institute for Software Systems (STS), Hamburg University of Technology, Hamburg, Germany
karsten.martiny@tu-harburg.de

Keywords: Candlestick Patterns, Time Series Analysis, Market Forecasting, Pattern Discovery, Information Extraction, Unsupervised Learning, Hierarchical Clustering

Abstract: Candlestick charts are a visually appealing method of presenting price movements of securities. It has been developed in Japan centuries ago. The depiction of movements as candlesticks tends to exhibit recognizable patterns that allow for predicting future price movements. Common approaches of employing candlestick analysis in automatic systems rely on a manual a-priori specification of well-known patterns and infer prognoses upon detection of such a pattern in the input data. A major drawback of this approach is that the performance of such a system is limited by the quality and quantity of the predefined patterns. This paper describes a novel method of automatically discovering significant candlestick patterns from a time series of price data and thereby allows for an unsupervised machine-learning task of predicting future price movements.

1 INTRODUCTION

Candlestick charts are a visually appealing method of presenting price movements. The method of analyzing candlestick charts in order to predict price movements of a security has been developed in Japan centuries ago. In fact, the use of candlesticks amongst ancient Japanese rice traders dates back to the early 18th century and thereby, this method is established significantly longer than modern financial markets themself exist. However, this technique remained unknown to the western world until the early 1990s, when it was first made accessible in (Nison, 1991). It has gained popularity amongst western analysts ever since.

Due to the increasing importance of machine-driven trading systems some approaches (as explained below) have been developed to exploit candlestick patterns in algorithmic trading systems. However, previously proposed methods depend on an initial manual definition of significant candlestick patterns. This paper describes an unsupervised learning method that is able to identify significant patterns without any a-priori knowledge and can therefore be used to develop an adaptive trading system.

This paper is structured as follows. Section 2 provides the preliminaries for analyzing candlestick charts. In Section 3, a method for an unsupervised learning process of inferring significant candlestick patterns is described and it is shown how these results can be used to infer price movement prognoses. Section 4 presents an evaluation of the proposed techniques' prediction performance. The paper finishes with a summary in Section 5.

2 CANDLESTICK ANALYSIS

To create a candlestick chart, the key points of every day are used to construct a candle. A day's key points consist of its opening, highest, lowest, and closing price. These data sets are sometimes also referred to as *OHLC-data* and will form the actual machine input subsequently.

As depicted in Figure 1, the construction of a candle works as follows: a rectangle is drawn between opening and closing price (called the candle's body), a line is drawn from the body's upper edge to the highest value (called upper shadow or upper wick) and, accordingly, another line is drawn from the body's lower edge to the lowest value (called lower shadow or lower wick). Additionally, the candle's body is colored according to the situation: "rising days" (i.e., the closing price is above the opening price) are depicted with hollow white bodies, while "falling days" are marked through solid black bodies.

Note that a candle does not necessarily have to exhibit all of these features: since the opening or closing

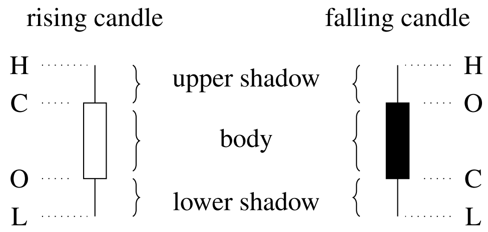

Figure 1: Depiction of time intervals as candles: O = opening price, H = highest price, L = lowest price, C = closing price; if the closing price is above the opening price, the candle's body is white, otherwise it is colored black.

values may coincide with a day's high or low values, there can be candles with only one or even no shadow at all. The candlestick's main advantage is that it allows for an instant perception of the market participants' attitude. For instance, as explained in (Nison, 2003), large shadows indicate significant movements in both directions and are therefore usually associated with uncertainty. This could give hints that no direction is currently in favor or that a turning point is imminent.

While candlesticks are able to provide valuable insight into a market's situation, single candlesticks are usually regarded as too fragile to allow for a prognosis of required reliability. Hence, instead of relying only on a single candle's shape, forecasts are based upon constellations of successive candlesticks. These so-called *Candlestick Patterns* usually consist of a series of three candlesticks with certain properties. In addition to the candles' shapes, their positions relative to each other, as well as the prevailing direction of movement, are taken into consideration. For a thorough guide to candlestick patterns see (Nison, 1991) or (Morris, 2006).

In general, candlestick patterns can be classified into two categories: reversal patterns indicate an imminent turning of the current movement's direction while continuation patterns confirm the current movement. Opposed to other established analysis methods, candlestick patterns only require a succinct time period to form a characteristic pattern and thereby emit a signal. Consequently, it is only natural to utilize these patterns for short-term forecasts. In fact, distinguished patterns have the means to forecast the following day's direction with a rather high certainty, but by extending the forecast further into the future, its reliability will quickly diminish. In (Morris, 2006), statistical correlations between a forecast's length and its quality have been analyzed and it was shown that a feasible hit ratio can be achieved for a maximum of three days, while after a maximum of seven days a pattern's prognosis is hardly able to outperform a random guess.

Due to their short-term nature, candlestick patterns may be employed to facilitate strategies with a rather high trading frequency. Also, since other analyzing methods tend to forecast movements on a larger scale, they can be augmented with a candlestick pattern analysis to pinpoint the exact positions of turning points and thereby result in an increased quality of the inferred predictions.

3 UNSUPERVISED DETECTION OF CANDLESTICK PATTERNS

Previous research in the field of candlestick pattern analysis aimed at an automatic detection of predefined significant patterns. This task can be seen as a form of stream-based query answering. Several approaches of inferring future price prognoses, based on querying streams for candlestick patterns and incorporating a variety of machine learning techniques, have been proposed: In (Chou et al., 1997) fuzzy membership functions are combined with induction trees to infer predictions, (Lee and Jo, 1999) employs a rule-based expert-system to produce trading recommendations. (Ng et al., 2011) proposes to train a radial basis function neural network in order to derive investment decisions. In (Lee et al., 2011) descriptions of candlestick patterns through fuzzy linguistic variables are combined with genetic algorithms in order to obtain investment decisions. In (Lin et al., 2011), an autonomous trading system is described, which learns a trading strategy based on candlestick patterns and other technical indicators through an echo state network.

While these works differ in the proposed data mining techniques, all of them share the same basic approach to the task: before any of these systems are able to infer predictions, they rely on an initial definition of significant candlestick patterns provided by a domain expert. These manual definitions are then used to obtain training data sets. Though all works were able to show that their respective approach leads to the obtainment of valuable analysis results, their dependencies on a domain expert bear a major drawback. The potential performance of all of these systems is strictly limited by the quality and quantity of the predefined set of significant patterns. It is especially not possible that these approaches are able to find any supplementary significant patterns that may be contained in training data sets but are unknown to the domain expert. This work describes an alternative approach of classifying candlestick patterns that facilitates an unsupervised learning task. Hence this approach has the means of discovering all significant latent patterns contained in the training data set without having to rely on the (possibly limited) knowledge of a domain expert and thus enables the development of truly adaptive trading systems.

3.1 Domain Model

3.1.1 Evidence

As explained before, a particular candlestick is described through its OHLC-data and thus these values are used to form the observation space. In order to prevent the system from inferring dependencies of certain price levels, it is necessary to avoid absolute values. Instead of directly using the OHLC-data, the values are scaled such that they denote a relative change with respect to the opening value. In addition to resulting in universally applicable descriptions of candlesticks' shapes, this normalization has the convenient side effect that the evidence space is reduced because the opening value can be omitted. Consequently, the observation of a particular candlestick CS_t for a time interval at t is described by:

$$CS_t = \left(\frac{H}{O}, \frac{L}{O}, \frac{C}{O}\right) = (H'_t, L'_t, C'_t) \tag{1}$$

Next to the included candlestick shapes within each pattern, it is also necessary to provide information about the candlestick's positions relative to each other. This information is based on the midpoint M_t of a candlestick:

$$M_t = \frac{H_t - L_t}{2} + L_t$$

Using these midpoints, the relative position of a candle j with respect to a preceding candle i is defined as the relative change Δ_{ij} of their respective midpoints:

$$\Delta_{ij} = \frac{M_j - M_i}{M_i} \tag{2}$$

A candlestick pattern comprises three consecutive candlesticks and therefore, by using Equations 1 and 2, a particular instance of the evidence e_t^T formed by a complete candlestick pattern in a certain trend state T (as explained below) is described by

$$e_t^T = \left(CS_{t-2},\ CS_{t-1},\ CS_t,\ \Delta_{t-2,t-1},\ \Delta_{t-1,t}\right) \tag{3}$$

To complete the description of a particular pattern it is also required to provide information about the pattern's context (i.e., the current trend state). In order to identify this trend state, a very short weighted moving average (denoted by MA) is used. This indicator calculates the average of the last three candles' closing values. An upward movement is assumed if this moving average has been strictly monotonically increasing for at least two days before a pattern was formed. Accordingly, a downward movement is assumed for a strictly monotonically decreasing average. If neither condition is fulfilled the context does not exhibit a clear trend. Thus the trend state T_t for a certain pattern CP_t is defined as

$$T_t = \begin{cases} 1 & \text{if} \quad MA_{t-4} < MA_{t-3} < MA_{t-2} \\ -1 & \text{if} \quad MA_{t-4} > MA_{t-3} > MA_{t-2} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

3.1.2 Prognosis

The prognosis model employed in this work predicts the direction of price movements for a short time period. For a prediction period of p days into the future, the prediction (also called hypothesis) h_t^p is defined as[1]

$$h_t^p = \operatorname{sgn}\left(\Delta_{t,\ t+p}\right). \tag{5}$$

3.2 Pattern Detection

The central goal of this work is the automatic detection of significant patterns (i.e., patterns that tend to indicate similar subsequent price movements) solely from training data without any manually defined a-priori information. While the task of inferring predictions based on unknown pattern memberships resembles the process of learning memberships of unknown Gaussian mixture components, this domain bears the additional challenge that the number of mixture components (i.e., the number of significant patterns) is unknown a-priori, too. Thus, conventional clustering procedures such as EM-clustering, SVM-clustering, or k-means clustering are not applicable for this task as all of these methods require an a-priori definition of the number of clusters. Since the feature values of observed candle instances are distributed roughly evenly, density-based clustering methods are neither applicable because although they are able to cope with an unknown number of clusters, the feature distribution would result in a single large cluster (with the

[1] Strictly speaking, this definition would yield three different forecasting states because next to rising and falling days it may also happen that midpoint positions are exactly equal. However, since a pair of candles virtually never exhibits exactly the same midpoint values in practice, the hypothesis space is modeled in a binary way. In order to achieve an exhaustive definition, the forecast is implemented with the assumption $\operatorname{sgn}(0) = -1$. Due to the practical irrelevance one could also keep the third unneeded state or use other assumptions as well without impairing the results.

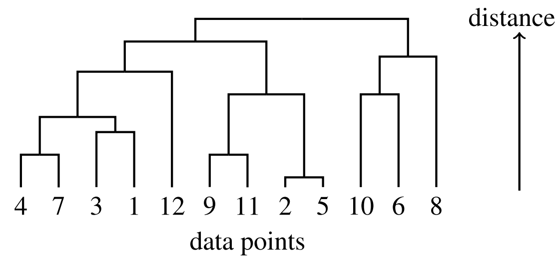

Figure 2: Schematic depiction of a hierarchical cluster structure (this type of graph is called *dendogram*)

exception of occasional outliers), provided that the training set is sufficiently large. In order to cope with both the unknown shapes and numbers of significant patterns, a hierarchical clustering structure is used in this work that allows to query for significant patterns based on characteristic properties.

3.2.1 Hierarchical Clustering

Hierarchical clustering methods construct an ordered structure of clusters such that the distance according to some distance measure between respective cluster members increases with the hierarchical order of a cluster. To illustrate this, Figure 2 depicts an example of a hierarchical cluster structure: in a complete structure, the topmost cluster contains all elements of the data set and its children are split such that they form two clusters with a minimal distance between all respective cluster elements. This work uses *agglomerative hierarchical clustering* ("bottom-up clustering"), as explained in (Ma; and Wu, 2007) to construct the required cluster hierarchy.

A useful distance measure to define similarities between different observations in this context is the euclidean distance. Hence, the distance between two patterns x and y (of the form described in Equation 3, containing 11 properties) is defined as:

$$d(x,y) = \|x-y\|_2 = \sqrt{\sum_{i=1}^{11}(x_i-y_i)^2}$$

After merging two observations into a cluster, the feature values of the resulting cluster are set to the respective average values of both merging candidates. Applying agglomerative hierarchical clustering with this distance measure to the training set of input patterns results in a structure that provides information about the similarity of different observations and therefore can be used to identify recurring patterns.

Once this structure has been constructed, it can easily be used to query for significant patterns. To identify significant patterns, two properties for each cluster (i.e., each possible pattern) are defined: First, it is necessary to ensure that a cluster indeed represents a recurring pattern and not only some random effect. In order to ensure recurrence, a cluster must contain at least a certain minimum number of members (representing instances of a particular pattern). This is denoted by the instance count I_P of a pattern P and is defined as

$$\begin{aligned} I_P &= \#\text{ instances for a pattern } P \\ &= \#\text{ elements in the corresponding cluster.} \end{aligned}$$

Also, a pattern should exhibit a certain reliability R_P, i.e., a certain ratio of all cluster members need to be succeeded by the same directional development. The reliability of a pattern P is defined as

$$R_P = \frac{\max \#\text{ instances with the same outcome}}{I_P}.$$

This stresses the importance of considering the instance count: each of the initial clusters only contains a single instance and thus exhibits a reliability of 100%. Without requiring a certain number of occurrences, all of these clusters would be considered significant. By adjusting this parameter, the generalization capabilities of the system can be tuned: a low value will lead to a high amount of discovered distinct patterns with only few instances each, while increasing this value will result in fewer, but more general, descriptions of significant patterns.

Based upon these properties a pattern is considered significant if both values exceed a certain threshold (i.e., $I_P \geq I_{\min}$ and $R_P \geq R_{\min}$). Search for these pattern starts at the bottom of the hierarchical structure. As soon as a cluster is found that meets the specified criteria, it is added to the set of significant patterns and its parents are omitted from the search. The rationale of this can be understood as follows. If a pattern is already considered as significant, an addition of further instances bears the danger of diluting the pattern's properties by considering auxiliary insignificant instances.

As a result of this procedure, one obtains an emergent set of significant pattern descriptions. This knowledge is elicited solely from latent information of the training data and therefore transcends the limitations imposed by relying on the knowledge of a domain expert. Once the descriptions of significant patterns have been obtained, they can be incorporated in various analysis approaches in order to infer predictions of future price movements. To evaluate the quality of detected patterns, a naive Bayesian classifier is used in the following.

3.3 Inferring Prognoses from Candlestick Patterns

Once the descriptions of significant candlestick patterns have been obtained through agglomerative hierarchical clustering, this information can be used with a naive Bayesian classifier in order to infer predictions regarding future price movements and thereby evaluate the quality of the previously discovered patterns. The Bayesian network used for this purpose is depicted in Figure 3. The evidence node in this network represents a random variable for a particular observation as described in Equation 3. This evidence is first used to determine whether the observation resembles an instance of some known significant pattern. Thus, node P is a discrete node that indicates the cluster membership of the observation. Note that obviously not every observation can be assigned to a known pattern. Hence, in addition to the set of known patterns, an additional "absorbing pattern" is introduced, which indicates that an observation is not an instance of a known pattern. Once the pattern membership is determined, this information is used to predict the direction of future price movements.

The conditional probabilities $P(e|p)$, denoting that a particular observation e is due to a certain pattern p, are modeled as Gaussians. Hence, the probability of this observation is a mixture of Gaussians:

$$P(e) = \sum_{p=1}^{n+1} P(p)P(e|p) = \sum_{p=1}^{n+1} w_p \cdot \mathcal{N}(\mu_p, \sigma_p^2)$$

w_p denotes the mixing weight for each component and can be determined by the prior $P(p)$ which in turn is determined by the ratio of p's number of occurrences and the size of the training set. Since both the pattern membership and the associated forecast are discrete variables, the corresponding conditional probability $p(h|p)$ is defined through a conditional distribution table. The values of these probabilities are acquired automatically as described in the following.

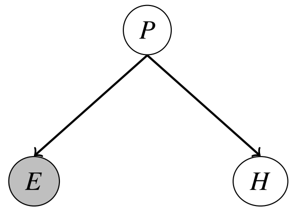

Figure 3: Bayesian classifier for candlestick pattern matching: The evidence node E (shaded) denotes the actually observed pattern, P denotes the unknown pattern assignment and H the inferred hypothesis (i.e., the predicted movement according to Equation 5).

method	reliability ($R_{\min}$)	# patterns	# matches	matching ratio	success rate
unsupervised	60%	162	6387	0.43	0.66
	70%	103	3961	0.26	0.70
	80%	70	2690	0.18	0.72
	90%	26	1044	0.07	0.81
	100%	3	105	0.01	0.90
supervised	N/A	27	989	0.07	0.67
static	N/A	27	202	0.01	0.53

Table 1: Results for a one-day forecast with $I_{min} = 30$ and varying $R_{\min}$ parameters. For each $R_{\min}$ the number of identified patterns, the number of observations that could be matched to one of the patterns, the corresponding matching ratio and the resulting success ratio of the forecasts are listed.
To compare these results to conventional approaches, the table also lists the forecasting results for applying the static rule descriptions and using supervised learning, respectively.

4 EVALUATION

In order to to evaluate the performance of the proposed techniques, a training scenario has been set up with historical data from the *Dow Jones Industrial Average* comprising OHLC-data from February 26, 1932 until October 12, 2011, thus containing 20000 pattern instances. This large data set hast been selected to ensure that the results are not influenced by an overfitting to short-term effects. To cope with this large amount of data, the system has been evaluated with a rolling forecasting procedure (as described for example in (Tsay, 2010)): initially, the first 5000 data sets including the subsequent directional development have been used to detect latent patterns and to estimate the unknown parameters of the Bayesian classifier afterwards. After this initialization, the system is used for inferring forecasts for the following 1000 data sets (the "rolling forecasting window", this time without subsequent developments). The inferred forecasts are then compared to the actual developments so that a hit ratio of the prognoses can be determined. Upon completing this test procedure, the tested data is added to the training set, the system is retrained and the rolling forecasting window is advanced to the next 1000 datasets. This procedure allows for both a long-term performance evaluation with the available data and an adherence to possible additional latent patterns in the test sets.

Evaluations with varying minimum instance count values have shown that requiring a minimum of 30 observed instances of any pattern in order to consider it as significant provides a feasible compromise between an accurate fitting and appropriate generalization capabilities. The corresponding results for a one-day directional forecast with varying $R_{\min}$ parameters are listed in Table 1. A comparison of the used reliability thresholds and the according success rates of the forecasts shows that the resulting success rate reflects a rather close tracking of the respective reliability parameter. Thus, by deciding upon a certain $R_{\min}$, one can tune the success rate to the extreme. However, there is a clear trade-off between success rate and matching ratio: e.g., configuring the system such that it yields a 90% success rate significantly decreases the matching ratio to 0.01. Consequently, a signal is generated only approximately twice per year and is concerned with only forecasting movements of the following day. Obviously such a sparse existence of very short-term signals renders this result virtually useless in practice, although its quality is outstanding. If instead one would for example use a reliability of 70% (i.e., intentionally accepting occasional false signals), the resulting success rate would still be highly satisfactory but, on average, signals are generated more than once per week. Thus, assuming that a correct prediction yields profits, the overall gain of any trading strategy employing this parameter setting would result in significantly higher gains.

To compare these results to conventional approaches, the table also lists the results of applying the pattern descriptions from (Morris, 2006) to the test data set (i.e., static pattern matching without any adaptive component) as well as the results of using these manual descriptions for training the network (i.e., a supervised learning task). A comparison of these results clearly shows that the proposed unsupervised method is able to outperform conventional approaches with respect to both the number of detected patterns as well as the resulting success rate.

Tests with forecasts of various durations were able to confirm the statements from (Morris, 2006) regarding the forecasting ability: a forecast three days into the future with $R_{\min} = 0.7$ resulted in a success rate of roughly 60% while the success rate of a four-day forecast was not able to exceed 50% significantly and therefore can hardly be of any use in practice.

5 SUMMARY

For candlestick-based prognoses, well-defined knowledge regarding significant patterns is required, as discussed in Section 2. In accordance with previous work, a probabilistic learning approach is used to infer prognoses based on a set of known significant patterns. Rather than merely using the inference procedures for computing prognoses based on predefined patterns, the learning process is extended to an earlier stage of the data processing. By pursuing the approach of unsupervised pattern detection, a self-contained knowledge base can be created which allows for truly adaptive forecasting systems. Through the ability of directly identifying latent patterns from training data, the resulting system is able to cope with securities that may exhibit varying behavior. Most importantly, the system's performance is independent of any manually specified knowledge and is therefore insusceptible to any knowledge deficiencies in both quantity and precision, which are virtually inevitable if a human domain expert is tasked with specifying his expertise in a machine-readable form. By replacing the manually specified pattern sets of the approaches discussed in Section 3, the presented procedure of identifying significant candlestick patterns can be combined with various methods of inferring the actual forecasts and thus can be used as a an extension for existing systems. Additionally, it is interesting to note that the feature selection used by human analysts is obviously a suitable choice for a machine-learning approach as well.

REFERENCES

Chou, S., Hsu, H., Yang, C., and Lai, F. (1997). A stock selection DSS combining AI and Technical Analysis. *Annals of Operations Research*, 75:335–353.

Lee, C.-H. L., Liaw, Y.-C., and Hsu, L. (2011). Investment decision making by using fuzzy candlestick pattern and genetic algorithm. In *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, pages 2696 –2701.

Lee, K. and Jo, G. (1999). Expert system for predicting stock market timing using a candlestick chart. *Expert Systems with Applications*, 16(4):357 – 364.

Lin, X., Yang, Z., and Song, Y. (2011). Intelligent stock trading system based on improved technical analysis and Echo State Network. *Expert Systems with Applications*, 38(9):11347 – 11354.

Ma;, G. G. C. and Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. SIAM, Society for Industrial and Applied Mathematics.

Morris, G. L. (2006). *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures*. McGraw-Hill Professional, 3rd edition.

Ng, W. W. Y., Liang, X.-L., Chan, P. P. K., and Yeung, D. S. (2011). Stock investment decision support for Hong Kong market using RBFNN based candlestick models. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, volume 2, pages 538 –543.

Nison, S. (1991). *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East*. New York Institute of Finance.

Nison, S. (2003). *The Candlestick Course*. Wiley & Sons.

Tsay, R. S. (2010). *Analysis of Financial Time Series (Wiley Series in Probability and Statistics)*. Wiley, 3rd edition.