Information Sciences 411 (2017) 151–175

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A formal approach to chart patterns classification in financial time series

Yuqing Wan, Yain-Whar Si*

Department of Computer and Information Science, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 13 November 2016
Revised 8 March 2017
Accepted 18 May 2017
Available online 19 May 2017

Keywords:
Chart patterns
Formal specification
Pattern matching
Financial time series
Technical analysis

ABSTRACT

Classifying chart patterns from input subsequences is a crucial pre-processing step in technical analysis. In this paper, we compile comprehensive formal specifications of 53 chart patterns reported in the literature. A first-order logic representation is chosen to describe the shape and corresponding constraints of each pattern. These formal specifications are formulated in such a way that data mining algorithms can use them for classification without significant modification. These formal specifications are also intended to serve as a reference model for future research in the chart patterns classification area. Using these formal specifications, we perform extensive experiments using real datasets from NYSE Composite (NYSE), Hang Seng Index (HSI), and Amazon (AMZN). The performance of the proposed method is compared against Template Based (TB), Euclidean Distance (ED), and Dynamic Time Warping (DTW) approaches. The experimental results show that the rules translated from the specifications can be effectively used to identify chart patterns from real datasets.

1. Introduction

In technical analysis, the appearance of chart patterns in financial time series is often considered to be a sign of possible change in the price trend. Various pattern matching approaches can be used to measure the similarities between chart pattern templates and the subsequences extracted from financial time series. For instance, the template-based (TB) [11] approach measures the temporal and amplitude distance between a chart pattern template and the query subsequence. The rule-based (RB) [11] approach relies on a set of predefined rules for classifying the query subsequence. Euclidean distance (ED) and dynamic time warping (DTW) [4] measure the distance between a re-scaled chart pattern template and the original subsequence. Regardless of the choice of method used for pattern matching, there is currently no industry-wide standard for how these patterns should be defined in an unambiguous way. For instance, some of the well-known patterns, such as “Head-and-Shoulders”, “Triple Tops” and “Cup with Handle” are frequently used to evaluate the effectiveness of various pattern matching methods. However, there is no documented work on the precise definitions of these chart patterns. Although the shapes of 53 known chart patterns and their application in predicting future price trends have been documented in [5,9], these reports are written in natural language. Therefore, it is certainly possible that the shape of one pattern may differ slightly from one definition to another. As there is no agreed-upon definitions, researchers and financial analysts may

* Corresponding author. Present address: Faculty of Science and Technology, University of Macau, E11-4022, Avenida da Universidade, Taipa, Macau, China.
E-mail addresses: qinggor@qq.com (Y. Wan), fstasp@umac.mo (Y.-W. Si).

http://dx.doi.org/10.1016/j.ins.2017.05.028
0020-0255/

specify these patterns in whatever way they think is suitable for their applications. Such discrepancy may lead to inconsistent and ambiguous shapes.

Although new approaches were proposed to solve the chart pattern matching problem in recent years, one of the key questions remains unsolved is how these chart patterns should be defined in an unambiguous way. To alleviate this problem, we compile formal specifications of 53 known chart patterns reported in [5]. Formal specifications have several advantages over other representations. First, the patterns defined in the formal specifications are mathematical representations. Therefore, there is no ambiguity in defining the shapes of the patterns. Second, formal specifications can be easily translated into rules that can be used to identify patterns from a given time series. Third, automated tools can be designed to verify the correctness of the formal specifications. Due to such advantages, a first-order logic representation is chosen to describe the shape and corresponding constraints of each pattern.

In financial trading, chart patterns are normally represented by either line charts or candlesticks. Therefore, formal specifications for each pattern can be viewed as the constraints defined on the relative positions of the data points (or candlesticks) within a given pattern. In addition, some of the specifications may be considered to be constraints on the temporal aspects of the patterns. Therefore, these specifications can be easily formulated into rules. Rules translated from formal specifications are not only useful for the RB method [11], but they can also be used to create standard templates in other pattern matching approaches. For example, the TB, ED or DTW methods can identify chart patterns using input templates. Normalized templates can also be used to generate synthetic data [11] [27] for experiments. To further illustrate the usefulness of formal specifications, we use rules translated from these specifications to classify patterns from the datasets of NYSE Composite (NYSE), Hang Seng Index (HSI), and Amazon.com (AMZN). To apply the RB approach, the original time series is first segmented to reduce the number of data points until the length is the same as the number of points defined in the rules. We use the perceptual important point (PIP) [7] method as a pre-processing step in the RB approach. The number of chart patterns found in the real datasets and several interesting observations are discussed in the experiment section.

The contributions of this paper can be summarized as follows:

- To the best of authors' knowledge, no one has ever defined the formal specifications of all chart patterns frequently used in the financial trading community. In this work, we endeavour to compile comprehensive formal specifications of 53 chart patterns reported in the literature. We believe that the specifications put forward in this paper can be used together as a reference model for future research in this area.
- In this paper, we show that the formal specifications can be directly translated into rules to classify chart patterns in real datasets. The experimental results show that the rules translated from the specifications can be effectively used to identify chart patterns from real datasets.

In Section 2, we review recent work on chart pattern matching. The criteria for categorizing the 53 chart patterns are detailed in Section 3. The formal specifications of these patterns are given in Section 4. In Section 5, we present our experimental results. Finally, in Section 6, we conclude our paper with directions for future work.

2. Related work

Technical analysis is the study of historical stock prices to predict future price trends. In [1], Achelis introduced the basic concepts, terminology and tools of technical analysis. One of the essential tools widely used in technical analysis is stock pattern charting. In *Encyclopedia of chart patterns* [5], Bulkowski described the shape of 53 chart patterns. In this paper, we define formal specifications of the 53 chart patterns based on the descriptions given in [5]. Edwards et al. [9] also introduced technical theories, various chart patterns and other indicators in financial time series. The patterns introduced in [9] are also included in [5].

In general, pattern matching approaches require the calculation of the similarity between a chart pattern template and a query subsequence. In the TB [11], RB [11] and hybrid (HY) [27] approaches, the query subsequence and the template pattern should contain the same number of data points. In the TB approach, the similarity between the query pattern template and the segmented time series can be measured by calculating their point-to-point amplitude distances and temporal distances. In the RB approach, a set of predefined rules is used to identify the segmented time series. In the HY approach, the Spearman's correlation coefficient of the sequence is first used to rank the segmented sequence. A set of predefined rules is then used to identify the segmented sequence.

When using these methods, segmentation is a necessary pre-processing step for decreasing the number of points in a time series. The PIP method [7] is one of the most commonly used segmentation methods. In the PIP method, a set of critical points, called "PIPs", is extracted from the original subsequence. In [11], it is reported that the vertical distance PIP (PIP-VD) performs better than other variants of the same method. In [24], the effects of different segmentation methods on financial time series are evaluated. It has also been reported that the RB approach recognizes the "Head-and-Shoulders Tops" pattern with high accuracy when the PIP approach is used for segmentation. There are other segmentation methods which extract important points from the original time series, such as, the piecewise linear approximation (PLA) method [14] and the turning points (TP) method [22]. Some other segmentation methods represent the original time series in various ways. For example, the piecewise aggregate approximation (PAA) method [15] represents a time series by the mean values of consecutive fixed-length segments. Symbolic Aggregate approXimation (SAX) [18] transforms a time series into a sequence of symbols. With SAX, a time series is segmented using PAA into several parts and then each part is represented by a