A hidden semi-Markov model for chart pattern matching in financial time series

Article *in* Soft Computing · July 2017

DOI: 10.1007/s00500-017-2703-7

2 authors:

Yuqing Wan
University of Macau
6 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Yain Whar Si
University of Macau
88 PUBLICATIONS **344** CITATIONS
SEE PROFILE

Noname manuscript No.
(will be inserted by the editor)

A hidden semi-Markov model for chart pattern matching in financial time series

Yuqing Wan · Yain-Whar Si

Received: date / Accepted: date

Abstract Many pattern matching approaches have been applied in financial time series to detect chart patterns and predict price trends. In this paper, we propose an extended hidden semi-Markov model for chart pattern matching (HSMM-CP). In our approach, a hidden semi-Markov model is trained and a Viterbi algorithm is used to detect chart patterns. The proposed approach not only simplifies the traditional way of training an HSMM but also reduces potential biases in parameter initialization. We compare the proposed model with current approaches on a set of templates selected from 53 chart patterns. Experiments on a synthetic dataset show that the proposed approach has the highest average accuracy and recall among other pattern matching approaches. Specifically, the HSMM-CP approach achieves highest accuracy for "Triangles, Ascending", "Head-and-Shoulders Tops", "Triple Tops" and "Cup with Handle" patterns. Moreover, experiments results show that the HSMM-CP performs significantly better than other approaches in distinguishing patterns with similar shapes such as "Head-and-Shoulders Tops" and "Triple Tops". Experiments are also conducted on a real dataset comprising the historical prices of several stocks.

Keywords pattern matching · hidden semi-Markov model · chart patterns · financial time series

Yuqing Wan
Department of Computer and Information Science, University of Macau
E-mail: qinggor@qq.com

Yain-Whar Si
Department of Computer and Information Science, University of Macau
Tel:+853-88-224454
Fax: +853-88-222426
E-mail: fstasp@umac.mo

1 Introduction

Technical analysis is commonly conducted to extract useful information from historical stock prices. When forecasting price trends, traders usually analyse past market data to identify interesting patterns. Detecting certain patterns that are considered helpful in predicting future price trends in time series of historical stock prices is a common practice in technical analysis. The characteristics of these patterns, which are also known as chart patterns, have been extensively studied by stock market experts and are recognised as signals that forecast the price movement. Bulkowski [2] compiles a detailed description of 53 chart patterns and examines their unique characteristics and relationships with price movement in detail. In this paper, these 53 chart patterns are grouped into 5 categories based on their shape.

Many pattern matching approaches have been used to locate patterns in financial time series. As the data size of a financial time series is large, some pattern matching approaches including the template-based (TB) and rule-based (RB) approaches [8] pre-process a time series with segmentation methods to decrease the number of points in the series. Therefore, the pattern matching results of the TB and RB approaches are influenced by the segmentation results.

In contrast, other pattern matching methods such as the dynamic time warping (DTW), Euclidean distance (ED) and support vector machine (SVM) methods detect patterns in time series without segmentation. Although these approaches are not affected by the segmentation methods, they require more processing time in similarity computations.

Although TB and RB approaches are efficient in calculating the similarity measure of the time series, segmentation during pre-processing step could result in

lost of information from the original time series. TB calculates the temporal and amplitude distance between the templates and the segmented time series. RB defines rules to identify a chart pattern. Since templates and rules are fixed in advance, these two approaches often lack the flexibility in identifying slight variations from the chart patterns. On the other hand, ED and DTW measure similarity without segmentation. DTW can measure the similarity of two time series which differ in length. SVM can be trained to find a maximum-margin hyperplane to separate two kinds of patterns. Therefore, SVM can recognize more variations in chart patterns. For two similar patterns, the positions of these two patterns in a space can be so close to prevent SVM from separate them accurately. Likewise, ED and DTW cannot accurately separate two similar patterns. The hidden semi-Markov model(HSMM) can detect more variations of a chart patterns. It is also capable of separating two similar patterns because HSMM does not calculate similarity measure based on the distance.

The HSMM has been applied in many different areas and is commonly used in speech recognition, speech synthesis, human activity recognition and handwriting recognition. The HSMM is also used to detect specific waveforms in time series in [9] and [16]. An expectation-maximisation (EM) [7] algorithm is used to estimate the parameters in an HSMM. In EM, the forward-backward procedure [17] is used to calculate the re-estimation formulas and sum probability. In this paper, we propose a Hidden semi-Markov model for chart patterns (HSMM-CP) in financial time series. In the HSMM-CP, an HSMM is trained to identify a chart pattern in a different way than EM, using the template pattern information. This paper makes two contributions to the literature.

- We classify the 53 chart patterns in [2] into 5 categories according to their shape features. Fourteen patterns from these categories can be used as benchmarks for future pattern matching research.
- We propose a novel way to train an HSMM for chart pattern matching. The proposed model, which is known as the HSMM-CP, has a higher accuracy than other pattern matching approaches (TB, RB, ED, DTW and SVM).

Section 2 reviews the related work. Classification of chart patterns is given in Section 3. We briefly discuss the current pattern matching approaches in Section 4. A detailed explanation of the HSMM-CP is given in Section 5. The results of experiments involving synthetic and real data are discussed in Section 6. Section 7 concludes the paper.

2 Related work

The TB and RB pattern matching approaches [8] require the number of data points in an input time series to be the same as that in the template pattern. In these methods, segmentation is a necessary pre-processing step to decrease the number of points in the time series. Note that segmentation can affect both TB and RB pattern matching results, as the similarity measure is based on the segmented time series. To decrease the number of data points in the original time series, segmentation methods have commonly been used as a pre-processing step in time series analyses. These segmentation methods include the perceptually important points (PIP) [6],[5], piecewise aggregate approximation (PAA) [15], piecewise linear approximation (PLA) [14] and turning point (TP) methods [19]. Fu et al. [8] compare three variations of the PIP method. They consider the vertical distance PIP (PIP-VD) to be the best choice. Wan et al. [20] report that rule-based approach with PIP has a higher accuracy of recognizing Head-and-Shoulders Tops than the same approach based on other segmentation methods. Segmentation has also been applied in other domains. Image segmentation is the process of simplifying the representation of an image. Fuzzy c-means algorithms (FCM) have been widely applied in image clustering and image segmentation. To make the standard FCM more robust to image noise and outliers, Zheng et al. [27] introduce a new generalized hierarchical fuzzy c-means (GHFCM) for image segmentation. GHFCM consists of two new algorithm, generative FCM and hierarchical FCM. For vehicle classification, Wen et al. [21] propose a rapid learning algorithm to accelerate the training process of AdaBoost. A Haar-like feature extraction method is used to represent a vehicle's edges and structures.

The TB approach measures the similarity between the query pattern template and the segmented time series by calculating their point-to-point amplitude distance and temporal distance. The RB approach uses certain defined rules to identify the segmented time series. Setting rules is subjective and therefore can affect the pattern matching results. Although TB and RB can be used to rapidly calculate the similarity between the segmented time series and the standard template, these approaches lack the flexibility in identifying the variability of the patterns. Specifically, the TB approach relies on fixed templates for pattern matching. The rules defined in RB approach can also be highly subjective and strict.

The ED and DTW pattern matching approaches are distanced-based approaches without segmentation. The ED approach is a simple similarity measure, and the

DTW [1] approach is a popular similarity measure with many variations. Note that the TB approach mentioned in the preceding paragraph is a distanced-based pattern matching approach. These three approaches calculate the ED between the query pattern and time series in different ways. The TB approach calculates the amplitude and temporal distance between the segment points in a time series and the points on the template pattern. The ED and DTW approaches calculate the distance between all of the data points in a time series and the points on the enlarged (or re-scaled) template pattern. A threshold is needed when using distance-based methods. When the calculated distance is less than the threshold, the times series is accepted as a pattern. DTW can recognize more variations than ED since it is capable of calculating the similarity of time series that are out of phase. DTW can also recognize the length variation among input time series. Although DTW allows length variation in similarity calculation, the degree of difference should be kept minimum.

Labelled sequences or time series can also be classified using a SVM, which is a supervised learning model with associated learning algorithms that can analyse data and recognise patterns. The SVM is a non-probabilistic binary linear classifier that assigns a new input sequence into one category or the other. A SVM model maps a sequence from a low dimensionality feature space into a high dimensionality feature space and finds the maximum-margin hyper-plane to separate the two categories. A kernel function is used to map the low dimensionality feature space to a high dimensionality feature space. Choosing a kernel function and speeding up the computation of the kernel matrix comprise two challenges in the study of the SVM [23]. LIBSVM [4] is a popular open source machine learning library that implements the sequential minimal optimisation (SMO) algorithm for kernel SVMs that support classification and regression. SVM is a supervised approach and it can recognize more variations in the input patterns. SVM can be used to find the maximal-margin hyperplane in the high dimensional space to separate two kind of patterns. When SVM is used in separating two similar patterns, their positions in the high dimensional space are often found to be located at a short distance away. For the same reason, the distance between two similar patterns are so similar that TB, ED and DTW approaches cannot distinguish them in a good or satisfactory way. In contrast to SVM approach, the similarity measurement from HSMM is not based on the distance calculation. Therefore, a HSMM can better distinguish similar patterns and recognize more variations from the input sequences.

A SVM classifier was also applied in steganography. Least significant bit (LSB) matching is a steganography method that embeds messages to the pixel values of the cover image. Xia et al. [22] reveal that the histogram of the difference between pixel gray values will be smoothed by the stego bits introduced by LSB matching. Extracted features from the differences between nonadjacent pixels are used to train a SVM classifier. In [18], Schölkopf et al. introduced a variant of SVM, called v-SVM, for both regression estimation and pattern recognition. In [11], an incremental support vector learning for ordinal regression (ISVOR) algorithm is proposed. Numerical experiments on datasets show that ISVOR can converge to the optimal solution in a finite number of steps. ISVOR is also faster than the existing batch and incremental SVOR algorithms. Gu et al. [10] propose a robust regularization path algorithm for v-support vector classification (v-SvcRPath) based on lower upper decomposition with partial pivoting. The robust v-SvcRPath can avoid exceptions and handle singularities in the key matrix. In addition, it is more efficient than original v-SvcPath. A structural minimax probability machine (SMPM) which utilize the advantages of generative and discriminative approaches is proposed by Gu et al. [12]. Typical generative approaches include hidden Markov models, while typical discriminative approaches include support vector machine. SMPM can be used to calculate more reasonable decision hyperplanes. In addition, SMPM can be interpreted as a large margin classifier and can be transformed to SVM under certain special conditions.

A hidden Markov model (HMM) is a doubly stochastic process [24]. The underlying stochastic process is a discrete time finite-state homogeneous Markov chain. An HMM has a hidden state sequence and an observable sequence of observations that are influenced by the hidden states. As such, it is referred to as a double process. A HMM is characterised by the number of states in the model, the number of distinct observation symbols per state, the state transition probability distribution and the observation symbol probability distribution in a state [17]. The HMM is limited in applications such as speech recognition, as its state duration is geometrically distributed and each state in the HMM corresponds to one observation.

The HSMM, an extension of the HMM, is intended to overcome the limitations of the HMM in some applications. The HSMM allows each state to have a variable duration. Each state corresponds to a number of observations. The HSMM is also known as an explicit duration HMM, a variable-duration HMM, a HMM with explicit duration, a hidden semi-Markov model, a generalised HMM, a segmental HMM and a segmen-

Table 1 A comparison of the pattern matching approaches.

	Distance-based	Segment	Training
TB	√	√	×
RB	×	√	×
ED	√	×	×
DTW	√	×	×
SVM	×	×	√
HSMM-CP	×	×	√

tal model [24]. The HSMM has been applied to many different areas. Homes et al. [13] present an HSMM theory using a linear trajectory description characterised by slope and mid-point parameters in the application of speech recognition. Ge et al. [9] propose a general HSMM-based approach to detect specific waveform patterns in time series and compare its performance with the DTW approach. The proposed HSMM-based approach can find a specific pattern in a manufacturing process accurately, a feat that the DTW approach cannot accomplish. Kim et al. [16] use a HSMM to characterise waveform shape and add random effects to capture the variation of the waveform shape. In addition, a training method known as the expectation conditional maximisation either (ECME), which provides faster convergence than a standard EM procedure, is proposed to train the parameters of the HSMM with random effect. Kim et al. also compare the performance of the random-effect HSMM with the simple HSMM, DTW and ED approaches with bubble-probe interaction data and ECG data. They score the input sequences and compare the number of true positive sequences in the top 10 and top 20 waveforms selected by each algorithm. The experimental results show that an HSMM with a random effect performs better in the segmentation and recognition of waveforms. Kim et al. also use ECME to train a HSMM with random effect. ECME is a modification of EM applied to find a set of satisfying parameters. In contrast to these approaches, a novel way to train a HSMM is proposed in this article. The proposed approach is highly effective for chart patterns matching because it employs the stored positions of the points from the generated positive sequence. Besides, the extracted split points of each state and the parameters for each state are used to build the model. Table 1 presents a brief comparison of the pattern matching approaches.

3 Classification of chart patterns

Bulkowski introduces 53 chart patterns [2] in his book and details the features of each pattern and their relationships to stock trends. In Table 2, we classify the chart patterns from [2] into five categories in terms of their shapes. We number the five categories as C1 to C5. Patterns with fluctuations are divided into two categories: multiple-versions (C1) and single-version (C2) patterns. A single-version pattern has a certain number of highs or lows, and a multiple-versions pattern has an uncertain number of highs or lows and therefore various versions. However, all of these versions should conform to the shape features of the pattern. For patterns with curves (C3), the shapes form various curves due to the price decline/increase, which results in a gentle rounding turn akin to a half-moon or U/inverted U. Points (i.e., the closing price of the stock) can be used to represent the three pattern categories. Candlesticks are used to represent patterns with spikes (C4) and gaps (C5). A daily candlestick is a price stick that consists of the opening, highest and lowest prices in a day. Some daily candlesticks also plot the closing price or both the opening and closing prices. Patterns with spikes have two remarkably longer downward/upward spikes than other downward/upward spikes around them. In these cases, the highest/lowest prices of two candlesticks are much higher/lower than those of the candlesticks around them. To some extent, A gap forms when a candlestick's highest/lowest price is lower/higher than the lowest/highest price of the candlesticks around it. We select two patterns each from C1, C2 and C3 and one pattern each from C4 and C5 for our experiments. Figure 1 shows the templates for the selected six patterns (i.e., Head-and-Shoulders Tops; Triple Tops; Cup with Handle; Double Tops, Eve and Eve; Wedges, Rising; Triangles, Ascending) defined with points are shown in Figure 1. For C4 and C5, we define rules for Horn Bottom (HB) and Island Long (IL). We also define rules for the six selected patterns from C1, C2 and C3. All of the rules are designed in accordance with the RB pattern matching method, which is described in Section 4.

As shown in Figure 1(a), Head-and-Shoulders Tops (H&S-T) have a three-peak formation with a centre peak taller than the others. The two shoulders appear at about the same price level and the distance from the shoulders to the head is approximately the same. In Figure 1(b), Triple Tops (Trip-T)) have three highs at about the same price and are well separated and distinct. In Figure 1(c) Cup with Handle (CWH) is a U-shaped cup with a handle on the right side. As shown in Figure 1(d), both eve peaks of Double Tops, Eve and Eve (DT-E&E) appear rounded and wide and are not made of a single, narrow price spike.

In a Wedge Rising (Wed-R) pattern, an upward price spiral is bounded by two intersecting, up-sloping trend lines. Both trend lines must have upward slopes and eventually intersect with the bottom trend line,

Table 2 Classification of chart patterns.

C1: Patterns with fluctuations (multiple versions)
Broadening Bottoms
Broadening Formations, Right-Angled and Ascending
Broadening Formations, Right-Angled and Descending
Broadening Tops
Broadening Wedges, Ascending
Broadening Wedges, Descending
Flags
Flags, High and Tight
Head-and-Shoulders Bottoms, Complex
Head-and-Shoulders Tops, Complex
Pennants
Rectangle Bottoms
Rectangle Tops
Triangles, Ascending*
Triangles, Descending
Triangles, Symmetrical
Wedges, Falling
Wedges, Rising*
Diamond Bottoms
Diamond Tops
C2: Patterns with fluctuations (single version)
Double Bottoms, Adam and Adam
Double Tops, Adam and Adam
Head-and-Shoulders Bottoms
Head-and-Shoulders Tops*
Measured Move Down
Measured Move Up
Three Falling Peaks
Three Rising Valleys
Triple Bottoms
Triple Tops*
C3: Patterns with curves
Bump-and-Run Reversal Bottoms
Bump-and-Run Reversal Tops
Cup with Handle*
Cup with Handle, Inverted
Double Bottoms, Adam and Eve
Double Bottoms, Eve and Adam
Double Bottoms, Eve and Eve
Double Tops, Adam and Eve
Double Tops, Eve and Adam
Double Tops, Eve and Eve*
Rounding Bottoms
Rounding Tops
Scallops, Ascending
Scallops, Ascending and Inverted
Scallops, Descending
Scallops, Descending and Inverted
C4: Patterns with spikes
Horn Bottoms*
Horn Tops
Pipe Bottoms
Pipe Tops
C5: Patterns with gaps
Gaps
Island Reversals
Islands, Long*

producing a steeper slope than that at the top. A wedge that has at least five touches (three on one side and two on the other) is a reliable pattern in prediction. The Wed-R pattern has several variations, and we show four variants in Figures 1(e)-(h), respectively. Each of these variants has two upward intersecting trend lines and five touches on Figures 1(e) and (f) and six touches on Figures 1(g) and (h).

In Triangle Ascending (Tria-A) patterns, a horizontal top trend line and up-sloping bottom trend line form a triangle shape. In these patterns, there are at least two high touches on the horizontal line and at least two low touches on the up-sloping trend. For illustration purposes, we present four variations of Tria-A patterns in Figures 1(i)-(l), respectively.

The HB pattern in Figure 1(m) has two downward spikes separated by a one-week duration. In this pattern, at least two spikes should be longer than similar spikes over the previous year. There are no downward spikes that come near the length of the horn spikes looking back over the months. The IL pattern from Figure 1(n) is formed by several continuous sticks that are separated by the two gaps as a whole, resembling an island in a long time series represented by candlesticks.

4 Review of current pattern matching approaches

Five pattern matching approaches are reviewed in this section. These approaches include the TB, RB, ED, DTW and SVM approaches.

4.1 Template-Based Pattern Matching

The TB pattern matching approach [8] measures the similarity between the defined pattern templates and the segmented sequences, which have the same number of points as the templates, by calculating their point-to-point amplitude distance (AD) and temporal distance (TD). AD is defined as

$$AD(SP,Q) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(sp_k - q_k)^2} \tag{1}$$

TD is defined as

$$TD(SP,Q) = \sqrt{\frac{1}{n-1}\sum_{k=2}^{n}(sp_k^t - q_k^t)^2} \tag{2}$$

The similarity measure can be defined as

$$D(SP,Q) = w1 \times AD(SP,Q) + (1-w1) \times TD(SP,Q)$$

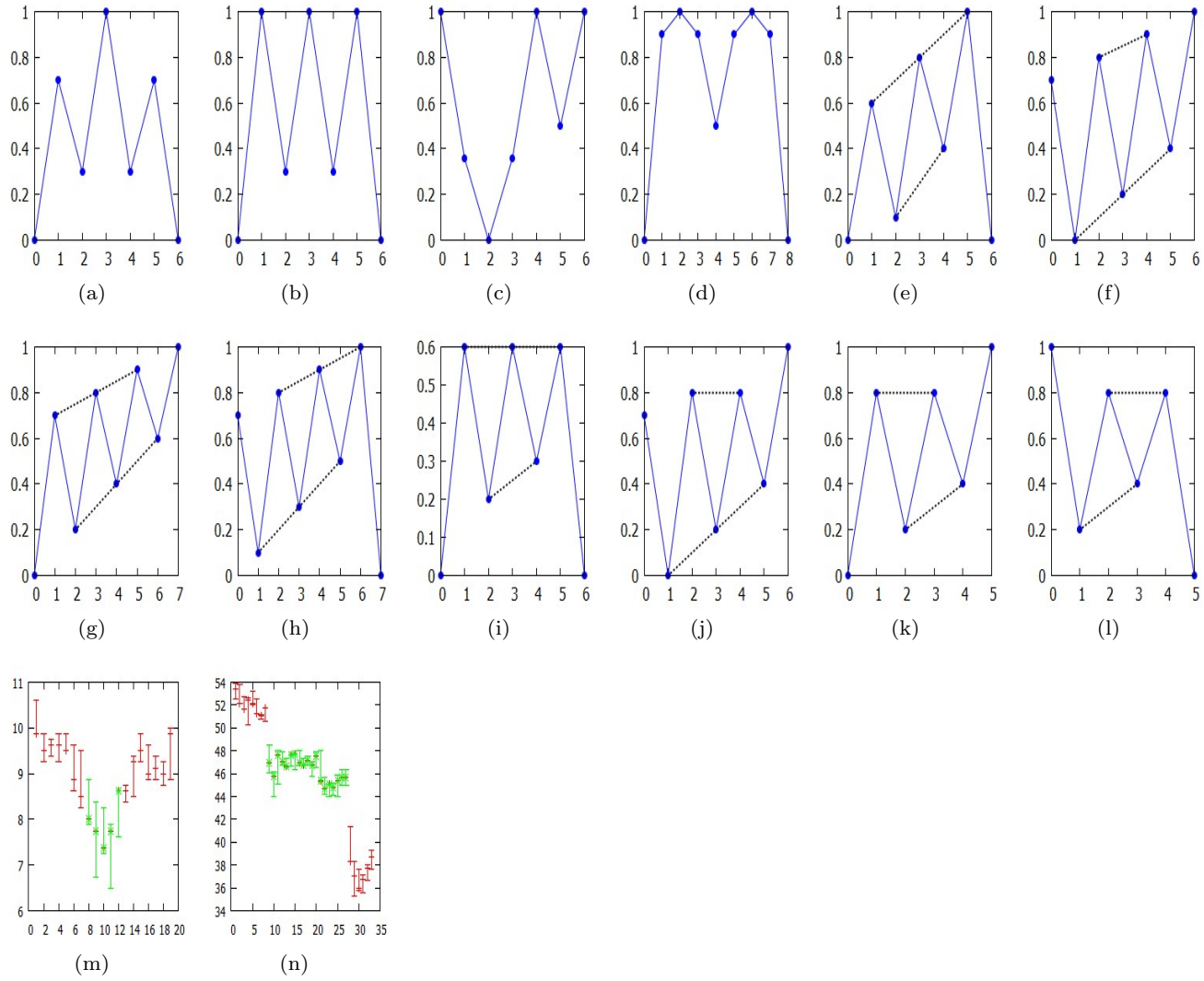


Fig. 1 Pattern templates including (a) Head-and-Shoulders Tops; (b) Triple Tops; (c) Cup with Handle; (d) Double Tops, Eve and Eve; (e)-(h) Wedges, Rising; (i)-(l) Triangles, Ascending; (m) Horn Bottom (green); (n) Island Long (green).

(3)

where Q denotes the query pattern template and SP denotes the segmented sequence. $w1$ is the weight required to balance AD and TD. sp_k and q_k denote the points in SP and the pattern template, respectively. sp_k^t and q_k^t denote the time coordinates of the points sp_k and q_k . Similar to the weights configuration described by Fu et al. [8], we set $w1 = 0.5$ for our experiment. When $D(SP, Q)$ for a sequence was less than a threshold, we accepted the sequence as a pattern.

4.2 Rule-based Pattern Matching

The RB pattern matching approach [8] uses predefined rules to identify patterns. A sequence is recognised as a matching pattern if its segmented sequence complies with the rules of a given pattern. The rules for the six selected patterns are listed in Tables 3-8, respectively. Note that these rules are designed according to the descriptions of the shape features given in [2]. Depending on analyst preference, adjustments can be made to restrict or relax the rules for pattern recognition. Only the RB approach is used to match the HB and IL patterns, which are represented by candlesticks and cannot be represented by lines and points.

Many researchers have examined the well-known H&S-T pattern. We combine the descriptions from [2] and [25] to define a set of rules in Table 3 for this pattern.In Table 3, $sp_k(k = 1, 2, 3, 4, 5, 6, 7)$ denotes the seven points of the H&S-T pattern and $\min(sp_m, sp_n)$ denotes the minimum value of the points. Table 4 presents the

Table 3 Rules for the Head-and-Shoulders Tops pattern

$sp_4 > sp_2$ and sp_6	$sp_2 > sp_1$ and sp_3
$sp_6 > sp_5$ and sp_7	$sp_2 \geq 0.5(sp_6 + sp_5)$
$sp_6 \geq 0.5(sp_2 + sp_3)$	$sp_7 \leq sp_5$
diff(sp_4^t,sp_2^t)<2.5*diff(sp_6^t,sp_4^t)	
diff(sp_6^t,sp_4^t)<2.5*diff(sp_4^t,sp_2^t)	

Table 4 Rules for the Triple Tops pattern

$sp_2 > sp_1$ and sp_3	$sp_4 > sp_3$ and sp_5
$sp_6 > sp_5$ and sp_7	diff(sp_3,sp_5)<15%
diff(sp_2,sp_4)<15%	diff(sp_4,sp_6)<15%
$sp_7 \leq \min(sp_3, sp_5)$	

rules for the Trip-T pattern.In Table 4, diff(sp_m, sp_n) denotes the difference between the two points sp_m and sp_n. Table 5 presents the rules for the CWH pattern. In Table 5, L_{mn} denotes the line decided by two points sp_m and sp_n, dis(sp_6, L_{15}) denotes the distance between sp_6 and L_{15}, dis(min(sp_2, sp_3, sp_4,L_{15}))denotes the distance between min(sp_2, sp_3, sp_4) and L_{15}. Except the rules in Table 5, sp_1, sp_2, sp_3, sp_4 and sp_5 form a second-order polynomial, the determination coefficient of which is greater than or equal to 0.8. Table 6 presents the rules for the DT-E&E pattern. In Table 6, max(sp_m, sp_n) denotes the maximum value of points. Table 7 presents the rules for the Wed-R pattern from Figure 1(e). Table 8 presents the rules for the Tria-A pattern from Figure 1(i).

We define the following rules for the HB pattern, which is formed by five weekly candlesticks:

- diff (lowest price of stick 2, lowest price of stick 4) $\leq 6\%$
- Lowest price of stick 2 <lowest prices of stick 1, stick 3 and stick 5
- Lowest price of stick 4 <lowest prices of stick 1, stick 3 and stick 5
- The two spikes of stick 2 and stick 4 should be longer than the average length of similar spikes over the previous year (about 48 weeks in weekly data)

An IL pattern is formed by several continuous daily candlesticks separated by two gaps. Two gaps separating a series of continuous candlesticks should be located initially to identify the pattern. Gaps appear when the previous days daily highest price is below the current days lowest price or when the previous days lowest price is above the current days highest price. Gaps should be at least \$1 wide and the pattern should be shorter than 4 months (about 120 days in daily data).

Table 5 Rules for the Cup with Handle pattern

$sp_1 > sp_2$, sp_3 and sp_4
$sp_6 < sp_5$and sp_7
dis(sp_6, L_{15}) $\leq$ 0.5dis(min(sp_2, sp_3, sp_4,L_{15}))
$sp_5 > sp_2$, sp_3 and sp_4
diff(sp_1,sp_5)<6%

Table 6 Rules for the Double Tops, Eve and Eve pattern

$sp_1 < sp_2$,sp_3 and sp_4
$sp_5 < sp_6$,sp_7 and sp_8
diff(max(sp_2, sp_3, sp_4), max(sp_6,sp_7,sp_8)) <15%
diff(max(sp_2,sp_3,sp_4,sp_6,sp_7,sp_8),sp_5)$\geq$ 0.1sp_5
$sp_5 < sp_2$, sp_3 and sp_4
$sp_9 < sp_6$, sp_7 and sp_8
$sp_9 \leq sp_5$

Table 7 Rules for Wedge Rising pattern from Figure 1(e)

$sp_2 > sp_1$ and sp_3	$sp_4 > sp_3$ and sp_5
$sp_6 > sp_5$ and sp_7	$sp_5 > sp_3$
$sp_6 > sp_4 > sp_2$	The slope of L_{35}>The slope of L_{46}

Table 8 Rules for Triangle Ascending pattern from Figure 1(i)

$sp_2 > sp_1$ and sp_3	$sp_4 > sp_3$ and sp_5
$sp_6 > sp_5$ and sp_7	$sp_3 < sp_5$
diff(sp_2,sp_4)<6%	diff(sp_4,sp_6)<6%

4.3 Pattern Matching Based on Euclidean Distance

The ED approach measures the similarity between two sequences by calculating the point-to-point ED between them. We enlarge the template pattern to the same length as the tested sequence in the experiments. The ED between the two sequences $X(x_1, ..., x_n)$ and $Y(y_1, ..., y_n)$ can be represented by the following equation.

$$ED(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{4}$$

The two sequences are said to be similar if the ED is less than a given threshold.

4.4 Dynamic Time Warping

The DTW [1] approach measures the similarity between time series of different lengths. DTW can match similar sequences that are out of phase. In our experiments, we enlarged the template pattern to the same size as the tested sequence. Given two sequences $X(x_1, ..., x_n)$ and $Y(y_1, ..., y_m)$, an n-by-m matrix M was constructed. The elements $d(x_i, y_j)$ in the matrix represent the ED of the points x_i and y_j. The warping path $W = w_1, w_2, ...,$

$w_k, ..., w_K (max(m,n) \leq K < m+n-1)$ is a neighbouring set of elements in M. The warping path follows three constraints, i.e., boundary conditions, continuity and monotonicity. The boundary conditions indicate that $w_1 = (x_1, y_1)$ and $w_K = (x_n, y_m)$. Continuity means that given $w_k = (a, b)$,$w_{k-1} = (a', b')$, where $a - a' \leq 1$ and $b - b' \leq 1$. Monotonicity indicates that $a - a' \geq 0$ and $b - b' \geq 0$. The optimal warping path $DTW(x, y)$ can be defined as follows:

$$DTW(x,y) = min\sqrt{\sum_{k=1}^{k=K} w_k} \tag{5}$$

The optimal warping path $DTW(x, y)$, which minimises the warping cost, is calculated by dynamic programming. The cumulative distance $\gamma(i, j)$ is defined as the distance $d(x_i, y_j)$, which is found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma(i,j) = d(x_i, y_j) + min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{6}$$

When $\gamma(m, n)$ is less than a certain threshold, the two sequences are said to be similar.

4.5 Pattern Matching with a Support Vector Machine

According to the SVM pattern matching approach, a set of time series labelled as positive and negative cases are used for training. Using the trained SVM model, a sequence can be classified as either a positive or negative pattern. In our experiment, we used LIBSVM [4], a popular open-source machine learning library for the SVM. We input a set of training data (i.e., the positive and negative time series of a pattern) and used the C-Support Vector Classification (C-SVC) in the LIBSVM to build a model to classify the given pattern. A training vector $x_i \in R^n, i = 1, , l$in two classes, and an indicator vector $y \in R^l$ such that $y_i \in$ {-1,1}. Each time series is a vector. The positive time series are labelled as 1 and the negative time series are labelled as -1. C-SVC solves the following optimisation problem:

$$\underset{w,b,\xi}{min} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{7}$$

Subject to $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, ..., l.$ where $\phi(x_i)$ maps x_i into a higher dimensional space and C comprises the regularisation parameters.

5 The Hidden Semi-Markov Model for Chart Patterns (HSMM-CP)

In this section, we describe the proposed HSMM for chart patterns, the Viterbi algorithm for pattern detection and the process of training an HSMM.

5.1 Overview of the HSMM-CP approach

The HSMM is a parametric learning model and EM [7] is a traditional approach to training the model. EM chooses the parameters by iteratively calculating the re-estimation formulas to maximise the sum probability of each time series in the training dataset. The initialisation of EM parameters affects the training results, and the more parameters an HSMM has, the more complex the implementation of EM. Figure 2(a) depicts the overall procedure for training an HSMM using EM. The first step in the EM process is to initialise the HSMM parameters. Using these parameters, the sum probability of all of the positive time series in the training dataset are calculated. Next, the parameters for the HSMM are updated using re-estimation formulas. The sum probability calculation and re-estimation process repeats until the sum probability of the sequence in a training dataset converges.

In this paper, we propose a training process that is tailored for recognising chart patterns effectively. The proposed approach, which is known as HSMM-CP, is different from the traditional EM. In our approach, we iteratively calculate the parameters so that the HSMM-CP can correctly classify the training dataset up to a given accuracy. Figure 2(b) depicts the overall training process for the HSMM-CP. First, we generate a positive sequence of the template chart pattern to calculate the parameters of an HSMM. Next, the model is used to classify the syntactic training dataset. All of these processes are repeated until the classifications reach a predefined accuracy threshold. HSMM-CP has the following main advantages.

- In the HSMM-CP, the parameters do not require initialisation. As a positive sequence of the template is used in each iteration to calculate the parameters, our approach is more suitable for cases in which the template patterns are known in advance.
- It is more convenient and straightforward to find a set of parameters without the use of re-estimation formulas from EM. In the HSMM-CP, by using the stored positions of the points from the generated positive sequence, the split points of each state can be extracted and the parameters for each state can be calculated to build a model.

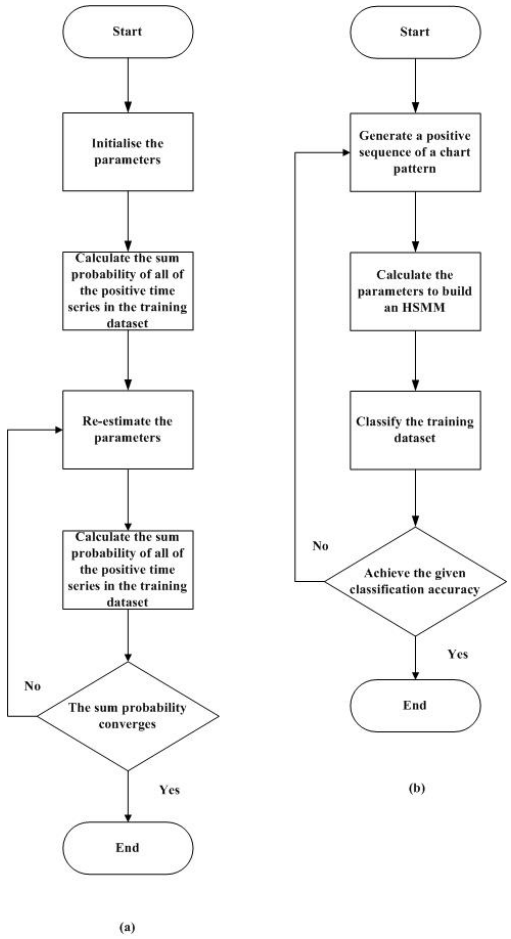


Fig. 2 (a) The EM process of an HSMM. (b) The training process of the HSMM-CP.

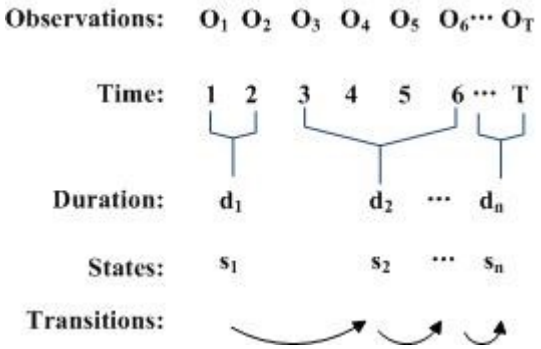


Fig. 3 A general HSMM, in which one observation corresponds to one time point. Each hidden state can emit a sequence of observations. States have variable durations as modelled by the duration distribution. The transition matrix describes transitions from one state to another.

5.2 The Hidden Semi-Markov Model (HSMM)

The general HSMM explained in Figure 3 is defined without specific assumptions on the state transitions and duration and observation distributions[24]. An HSMM has the following elements.

a) The number of state N is defined according to the shape of the query chart pattern. A line segment is a state in the template pattern, which is defined by the analyst before the matching process. For example, the number of state of the H&S-T chart pattern shown in Figure 1(a) is 6. Each state is represented by a number, and the end state for H&S-T is 6.

b) The state transition matrix A [3] is assumed to be a N×N left-right matrix. In this matrix, a_{ij} denotes the probability of being in state j at time $t+1$ from state i at time t. In a left-right model, $a_{ij} = 0$ for $j \leq i$ and $a_{NN} = 1$ for the last state. For instance, the transition matrix of H&S-T is a 6×6 left-right matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Furthermore, the initial probabilities π have the following property:

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{8}$$

The state sequence must begin at state 1 and end at state N.

c) The output probability of the observations are described by the Gaussian distribution, which is the most common and easily analysed continuous distribution. The probability density function (PDF) of the Gaussian distribution can be defined as follows:

$$p(x \mid \mu, \sigma^2) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-(x-\mu)^2}{2\sigma^2}) \tag{9}$$

Although the PDF $p(x)$ is not the probability of an observation with a value of x, it is proportional to the probability that x lies in a small interval centre on x. Therefore, the output probability of an observation is calculated by its PDF, i.e., equation 9.

We use the synthetic chart pattern H&S-T presented in Figure 4 to illustrate and discuss the HSMM pattern matching process. The pattern is first divided into six parts (each part is represented by a line) and therefore has six states. y in equation 10 is the observation of duration d (i.e., the number of points) generated by a state k and is presented by a linear regression function of time t added with noise e.

$$y = a_k + b_k t + e_k \tag{10}$$

where a_k is the intercept, b_k is the slope of the linear regression function and e_k is the mean square error in

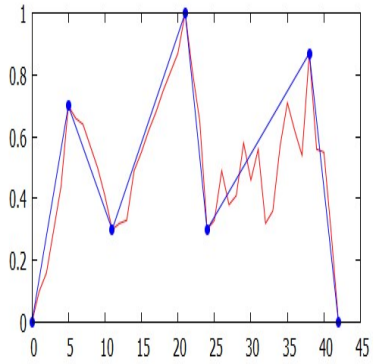


Fig. 4 Six line segments (blue) of a synthetic H&S-T pattern (red).

the k state. The output probability of an observation y_i in state k, denoted by $c_k(y_i) = p(y_i \mid \mu_k, \sigma_k^2)$, is calculated by the probability density function (i.e., equation 9) of state k. μ_k is the linear regression function of time t in state k. In other words, $\mu_k = a_k + b_k t$. σ_k^2 is defined as the mean square error (MSE) of $\hat{y}^{'}$ (the value of the linear regression function) as fitted to the original data y, i.e.,$\sigma_k^2 = \frac{(y-y^{'})^2}{d}$. There are d Gaussian distributions with different μ and the same σ^2 in a state. Note that the time t is a variable although the intercept and the slope are fixed in a state.

The duration of a state is indicated by the number of data points in that state. Instead of modelling the number of points in a state, we model the length of a state (i.e., the length of a line segment in the state) using a Poisson distribution:

$$p(l \mid \lambda_k) = \begin{cases} 1, & d = 1 \\ \frac{\lambda_k^l}{l!} exp(-\lambda_k), & d > 1 \end{cases} \tag{11}$$

where l is the length of a sequence of observations calculated by the ED of the start and end points of the sequence of observations and λ_k is the length of a state (i.e., the length of the line segment of state k). Any two line segments of the same length may have different numbers of points and may also have the same $p(l \mid \lambda_k)$.

The probability of a sequence of observations of d points generated by state k is calculated as

$$p(y \mid \mu_k, \sigma_k^2, \lambda_k) = p(l \mid \lambda_k) \prod_{i=1}^{i=d} p(y_i \mid \mu_k, \sigma_k^2) \tag{12}$$

Therefore, an HSMM has a set of elements $N, A, \mu, \sigma, \lambda$ and building an HSMM for a particular chart pattern requires deriving these parameters by training the model on the dataset of that pattern. Note that the parameters N and A are known when we define a chart pattern template. Given a new observation sequence and the HSMM, the Viterbi algorithm can be used to choose a corresponding state sequence that best explains the observation sequence. If the end state of the hidden state sequence of a new observation sequence is the same state as the end state of the defined state sequence of the query chart pattern, then the new observation sequence is accepted as a chart pattern.

5.3 The Viterbi Algorithm

The Viterbi algorithm [17] can be used to find the optimal state sequence associated with the given observation sequence. The optimal state sequence is the state sequence that maximises $p(y \mid \mu_k, \sigma_k^2, \lambda_k)$ (given in equation 12). To find a single best state sequence $S = (s_1, s_2, ..., s_T)$ for the given observation sequence $y = (y_1, y_2, ..., y_T)$, we first define a quantity $\delta_t(i)$ to explain the Viterbi algorithm.

At time t in state i, $\delta_t(i) = \max\limits_{s_1, s_2, ... s_{t-1}} P[s_1 s_2 ... s_t = i, y_1 y_2 ... yt \mid \mu_k, \sigma_k^2, \lambda_k]$ denotes the maximum probability to generate a sequence of observations $(y_1, y_2, ...y_t)$ along a single state path $(s_1, s_2, ...s_{t-1})$. By induction, we can calculate the following:

$$\delta_t(i) = \max_d [\max_j \delta_{t-d}(j) a_{ji} p(l \mid \lambda_i, d) \prod_{s=t-d+1}^{s=t} c_i(y_s)] \tag{13}$$

For example, we calculate $\delta_t(i)$ as follows:
$\delta_1(i) = \pi_i p(l \mid \lambda_i, d = 1) c_i(y_1)$
$\delta_2(i) = \max\limits_d [\pi_i p(l \mid \lambda_i, d = 2) \prod_{s=1}^{s=2} c_i(y_s), \max\limits_j [\delta_1(j) a_{ji} p(l \mid \lambda_i, d = 1) c_i(y_2)]]$
$\delta_3(i) = \max\limits_d [\pi_i p(l \mid \lambda_i, d = 3) \prod_{s=1}^{s=3} c_i(y_s), \max\limits_j [\delta_1(j) a_{ji} p(l \mid \lambda_i, d = 2) \prod_{s=1}^{s=2} c_i(y_s)], \max\limits_j [\delta_2(j) a_{ji} p(l \mid \lambda_i, d = 1) c_i(y_3)]]$
$\delta_4(i) = ...$
and so on until $\delta_T(i)$ is calculated. If the state that maximises $\delta_T(i)$ is the same as the end state of the chart pattern that we defined beforehand, then the sequence of observations is identified as a chart pattern.

One of the important questions here is how to to choose the parameters so that the Viterbi Algorithm can accurately detect a chart pattern. EM[7] is a traditional approach to choosing parameters that maximise the sum probability of all of the sequences in a training dataset.

We train the HSMM in a simple and effective way for predefined chart patterns. The training dataset is made up of randomly generated positive and negative examples. After we define the template chart pattern with points (shown in Figure 1), we adopt the algorithm from [8] and [26] to generate the positive and negative examples of the chart pattern (i.e., variations

of the chart pattern with a given length). We use the synthetic data-generating algorithm from [8] and [26]to add noise and change the positions of the original data points in the template chart pattern. We then iteratively generate a positive example of the chart pattern and record the latest positions of the defined points in the template chart pattern after they are modified by the synthetic data-generating algorithm. These points separate the sequence into line segments, each of which is a state. We then calculate the intercept (a), slope (b), length (λ) and MSE (σ^2) of each state of the positive example. Given the calculated parameters and number of defined states of a chart pattern template, we classify the training dataset using the Viterbi algorithm. If the classification result is satisfactory, we record the set of parameters as the final result. These parameters are now ready for use in classifying patterns in both synthetic and real datasets. This process of training an HSMM for chart patterns is described in Algorithm 1.

Algorithm 1 Training process of HSMM-CP.

Input: a training dataset TDS, a chart pattern template PT and the end state ES of PT
Output: the parameters (a, b, λ, σ^2)
  Accuracy=0
  while Accuracy<0.98 **do**
    Generate a synthetic positive example E=(y_1,...,y_T) of length T of the chart pattern template PT
    Calculate the parameters (a, b, λ, σ^2) of E
    for each sequence SQ in TDS **do**
      Input the calculated parameters into the Viterbi algorithm to find the best state sequence BSS=(s_1,...,s_T) of SQ
      if s_T=ES **then**
        SQ is recognised as a positive pattern of PT
      end if
    end for
    Calculate accuracy
  end while

6 Experiments

In this section, we describe the experimental results when HSMM-CP is compared with the TB, RB, ED, DTW, SVM pattern matching approaches based on synthetic and real datasets. We adopted the PIP segmentation method [6] in our experiment to decrease the number of points in a time series in the TB and RB approaches.

6.1 Experiments on a synthetic dataset

In terms of the synthetic dataset, we selected the H&S-T, Trip-T, Wed-R, Tria-A, CWH and DT-E&E from the three categories described in Section 3 as the example patterns for the experiments.

Distance-based pattern matching approaches such as the TB, ED and DTW approaches calculate the distance (i.e., similarity) between a sequence and a template pattern. If the distance is less than a predefined threshold, then the sequence is accepted as a positive pattern. To set a threshold for a pattern, we calculated the distances from the template pattern to the synthetic positive examples and the synthetic negative examples for samples with varying lengths. We attempted to set a threshold that well separated these two kinds of distance. Note that the TB approach must segment the original time series to decrease the number of points of the time series to that of the template pattern. For any chart pattern template in Figure 1, the number of points on the template was well known in advance. According to equation 3, the distance between a pair of points does not depend on the length of the time series. Therefore, we defined a fixed threshold for the TB approach in our experiments. However, for the ED and DTW approaches, the length of the template had to be the same as that in the input time series, and the distance calculated by equations 4 or 5 could grow when the length of the input time series increased. Therefore, we defined the threshold for the ED and DTW approaches as a linear regression function of length. The threshold calculation procedure for our experiments is detailed in Appendix A. Note that different patterns may have different thresholds.

We adopted the methods in [8] and [26] to generate synthetic data. There are three steps in the generation process of synthetic data: time scaling, time warping and noise adding. Time scaling alters the length of a pattern to generate different sub-sequences with different lengths. Time warping changes the positions of important points and therefore makes the subsequence looked warped compared with the pattern. Noise adding alters the value of each point to generate sub-sequences with more small fluctuations. The pseudo codes are given in Algorithms 2, 3 and 4. Note that for a given chart pattern, we stored the modified positions of the points that were originally on the defined template of the chart pattern after time scaling, time warping and noise adding to calculate the parameters for the HSMM. When generating negative cases, we revised the noise adding code by changing a parameter in Algorithm 4. To generate a negative case of a template chart pattern, we changed the range of parameter M in Algorithm 4 from [0, 0.3] to [0, 3] after time scaling and time warping the template pattern.

We used nine variables (P, Q, L,η, θ,α, β, γ, ε) to describe the experimental settings for a chart pattern P.

Table 9 Experimental settings for a pattern P.

	Training Dataset	Segmentation Points	Threshold
TB	None	η	θ
RB	None	η	None
ED	None	None	$\alpha L + \beta$
DTW	None	None	$\gamma L + \varepsilon$
SVM	P-Q-L-train	None	None
HSMM-CP	P-Q-L-train	None	None

Algorithm 2 Time scaling. Time scaling adds a number of points to expand the template pattern. .

Input: a pattern P (length of P is n) and a number m (m is the number of points generated)
  for each two adjacent Pi and Pi+1 **do**
    X = (m-n)/(n-1)
    Insert X points
  end for

Algorithm 3 Time warping. Time warping changes the positions of the points that were on the template pattern before time scaling.

Input: a pattern P
  for each critical point Pi **do**
    Change the position of Pi between Pi-1 and Pi+1 randomly
  end for

Algorithm 4 Noise adding. After time scaling and time warping, the value of each point on the original template pattern is changed at random to make more variations.

Input: a pattern P
  for each point Pi **do**
    Generate a probability R randomly
    if R <threshold(0.5 is used in this paper) **then**
      Generate a random value M (M $\in$ [0, 0.3])
      Pi = Pi + M (Pi+1 - Pi)
    end if
  end for

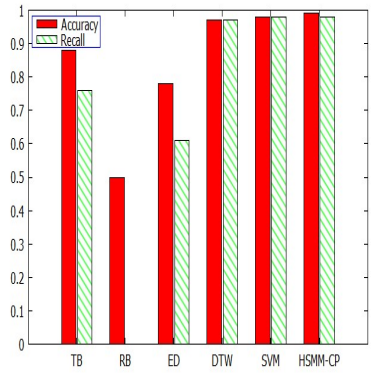
Q denotes the number of synthetic time series (50 positive examples and 50 negative examples) with a length of L. Recall that the TB and RB approaches require segmentation as pre-processing steps to decrease the number of points in the time series. After this step, the numbers of points in the sub-sequence and chart pattern template become the same. η denotes the number of points in a chart pattern template. θ is the threshold of the TB approach. The thresholds of the ED and DTW approaches are linear functions of the length of the time series. The threshold of the ED approach for a chart pattern P is presented by $\alpha L + \beta$, where α is the slope and β is the intercept of the linear function. The threshold of the DTW approach for a chart pattern P is presented by $\gamma L + \varepsilon$, where γ is the slope and ε is the intercept of the linear function. Table 9 summarises the experimental settings. For our experiment involving a chart pattern P, all six of the pattern matching approaches used the same test dataset containing Q time series (50 positive cases and 50 negative cases) with a length of L. The SVM and HSMM-CP approaches used the same training dataset containing Q time series (50 positive cases and 50 negative cases) with a length of L for a chart pattern P. In Table 9, None means that the corresponding pattern matching approach did not require that attribute. For the TB and RB approaches, the number of segmentation points was the same as the number of points in the chart pattern template denoted by η. A threshold was needed to measure the similarity in the distance-related TB, ED and DTW approaches. All of the experimental settings for the synthetic data are given in Appendix B.

6.1.1 C1: Patterns with fluctuations (multiple versions)

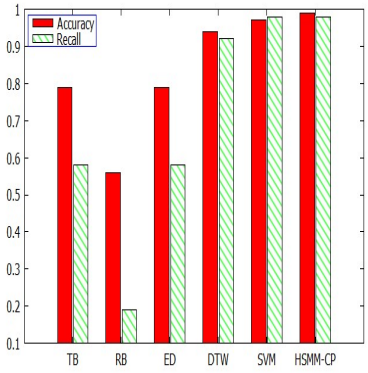
The Wed-R and Tria-A patterns contained several up and down trends and many variations. We chose the four variations shown in Figures 1(e)-(h) and Figure 1(i)-(l) for the experiments. Figures 5(a) and (b) show the accuracy and recall of the combination of four versions for Wed-R and Tria-A, respectively. As shown in Figure 5(a), the DTW, SVM and HSMM-CP approaches had similarly high levels of accuracy and recall in the Wed-R experiment. As shown in Figure 5(b), the HSMM-CP approach had the highest accuracy in the Tria-A experiment. The recall of the RB approach was zero for Wed-R and the lowest for Tria-A. The low result of the RB approach was caused by the strict rules defined for the template.

6.1.2 C2: Patterns with fluctuations (single version)

The H&S-T and Trip-T patterns are similar patterns and both contain fluctuations. Figures 6(a) and (b) depict the accuracy and recall of different pattern matching methods for the H&S-T and Trip-T patterns, respectively. From Figure 6(b), we can observe that RB achieves low recall for Trip-T pattern. The reason behind the low recall is that the rules employed by RB are



(a)

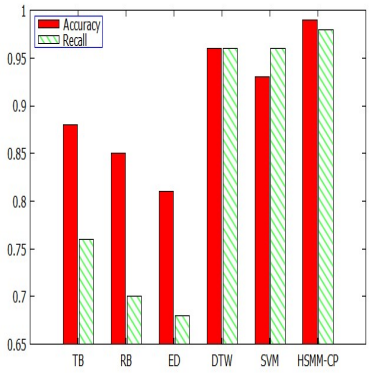


(b)

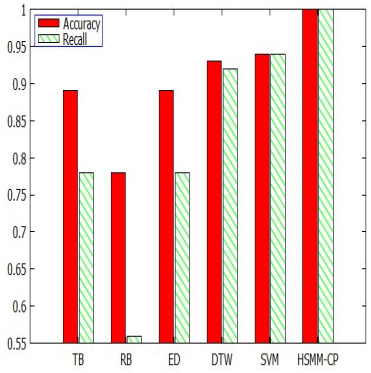
Fig. 5 (a) The accuracy and recall of different pattern matching methods for the four variations of Wed-R. (b) The accuracy and recall of different pattern matching methods for the four variations of Tria-A.

too strict when comparing the similarity of the three highs from the Trip-T pattern. We can also observe that TB achieves low recall for the Trip-T pattern. The low recall of TB is due to the threshold setting which is prone to reject the negative patterns. Therefore TB has a high rate of true negative but low recall. The HSMM-CP approach has the highest accuracy and recall for both patterns.

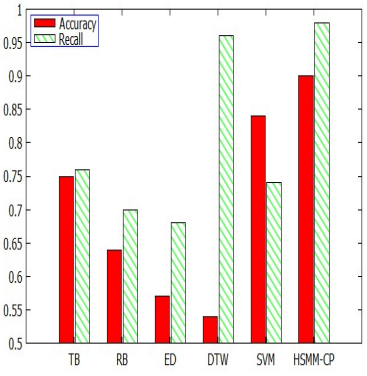
As the shape of the H&S-T and Trip-T patterns are similar, we conducted an additional experiment on the ability of the six pattern matching approaches to distinguish these two patterns. We designed the H&S-T pattern as a positive example and the Trip-T pattern as a negative example. Figure 6(c) shows the experimental results. From the experiment result, we can observe that TB and RB has a higher accuracy than ED and DTW. The HSMM-CP approach had the highest accuracy and recall, and we can conclude that it was able to distinguish these two similar patterns more effectively than the other five approaches. TB pattern matching method cannot accurately separate these two patterns since the templates of these two patterns are similar in shape. ED and DTW methods also have low accuracies in distinguishing these two patterns because the distance calculation of these two patterns are similar. For the SVM method, the positions of two similar patterns are situated too close in the high dimensional space and cannot be easily separated by SVM.



(a)

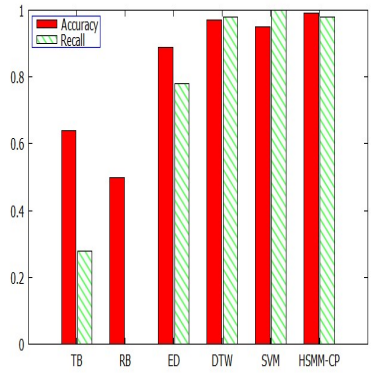


(b)

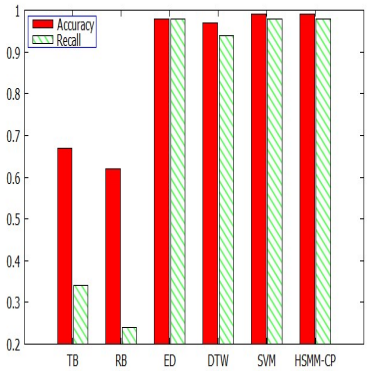


(c)

Fig. 6 (a) The accuracy and recall of different pattern matching methods for the H&S-T pattern. (b) The accuracy and recall of different pattern matching methods for the Trip-T pattern. (c) The accuracy and recall of different pattern matching methods for distinguishing the H&S-T and Trip-T patterns.



(a)



(b)

Fig. 7 (a) The accuracy and recall of different pattern matching methods for the CWH pattern. (b) The accuracy and recall of different pattern matching methods for the DT-E&E pattern.

6.1.3 C3: Patterns with curves

The CWH and DT-E&E patterns contain curves. As shown in Figure 7(a), SVM approach has a higher recall than the HSMM-CP and the HSMM-CP has a higher accuracy than the SVM approach for the CWH pattern. As shown in Figure 7(b), the HSMM-CP and SVM approach had the same accuracy and recall for the DT-E&E pattern. RB method achieves low recall when matching both patterns since RB relies on the rigid rules defined in advance. Likewise, the low recall in TB method reflects the fact that any bias in setting the threshold could significantly effect the pattern matching outcome.

6.1.4 Overall comparison

Figure 8 shows the overall accuracy and recall of different pattern matching approaches for patterns (H&S-T, Trip-T, Wed-R, Tria-A, CWH and DT-E&E). The results reveal that HSMM-CP has the highest average accuracy and recall among all methods.

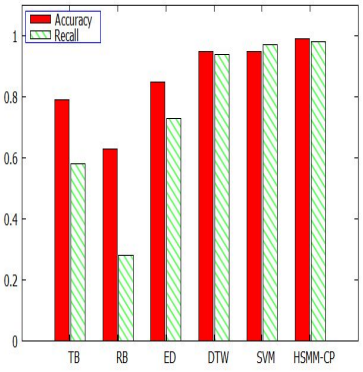


Fig. 8 The overall accuracy and recall of different pattern matching methods for the six chart patterns.

6.2 Experiments on a real dataset

For the real dataset, we included HB and IL in addition to the preceding six patterns. As the HB and IL approaches were represented by candlesticks, we used only the RB approach to detect them in a real dataset. For the experiments involving a real dataset, a sliding window was used and shifted one data point at a time to obtain a sub-sequence. We applied the PIP segmentation method [6] to segment the sub-sequence.

For the experiment involving a real dataset, we used the historical prices of the HANG SENG INDEX (HSI) from 1 January 2003 to 31 December 2012, which contained 2,506 points; the NYSE AMEX COMPOSITE INDEX (NYSE) from 1 February 2004 to 31 December 2014, which contained 2,769 points; and the Dow Jones Industrial Average (DJI) from 1 February 2004 to 31 December 2014, which contained 2,769 points. For this experiment, the redundant patterns within a +1 and -1 difference in position were eliminated. The number of patterns we found in the HSI, NYSE and DJI are shown in Tables 10, 11 and 12, respectively. The window sizes for the H&S-T, Trip-T, CWH and DT-E&E patterns were assigned to 43, 115, 121 and 121, respectively. The window size for the Wed-R-1 and Wed-R-2 patterns (from Figures 1(e) and (f), respectively) was assigned to 67 and that of the Wed-R-3 and Wed-R-4 patterns (Figures 1(g) and (h), respectively) was assigned to 71. The window size for the Tira-A-1 and Tria-A-2 patterns (from Figures 1-(i) and (j), respectively) was 115 and that for the Tria-A-3 and Tira-A-4 patterns (from Figures 1(k) and (l), respectively) was 116.

According to the experimental results, the TB and RB approaches found the least number of patterns out of the approaches for all three of the stock indexes. The rules set for the RB approach might have been too strict, leading to a lower number of patterns found com-

Table 10 The number of patterns found by different pattern matching approaches in the HSI.

	H&S-T	Trip-T	CWH	DT-E&E	Wed-R	Tria-A	Total
TB	2	1	1	3	9	5	21
RB	4	2	1	1	14	4	26
ED	21	8	6	8	70	27	140
DTW	22	6	8	10	56	29	131
SVM	24	7	8	5	18	27	89
HSMM-CP	17	6	11	6	64	37	141

Table 11 The number of patterns found by different pattern matching approaches in the NYSE.

	H&S-T	Trip-T	CWH	DT-E&E	Wed-R	Tria-A	Total
TB	2	0	6	4	8	7	27
RB	3	0	0	1	17	8	29
ED	28	9	7	9	74	33	160
DTW	28	9	10	10	75	31	163
SVM	27	8	9	7	59	35	145
HSMM-CP	22	7	13	7	76	47	172

Table 12 The number of patterns found by different pattern matching approaches in the DJI.

	H&S-T	Trip-T	CWH	DT-E&E	Wed-R	Tria-A	Total
TB	0	1	4	3	5	7	20
RB	3	0	0	2	12	9	26
ED	29	13	9	11	76	39	177
DTW	25	11	8	11	75	37	167
SVM	24	7	9	4	57	37	138
HSMM-CP	28	8	14	13	89	48	200

pared with other methods. Furthermore, segmentation might have affected the TB and RB pattern matching approaches. The TB approach measured similarity by calculating the point-to-point temporal and amplitude distance between the template and segmented subsequence. The RB and TB approaches were inflexible in detecting the variations of a pattern. Most of the patterns identified by the TB and RB approaches were also recognised by other approaches. Furthermore, 59 of the 68 patterns found by the TB approach in the three stock indexes were also found by other methods. Likewise, 66 of the 81 patterns identified by the RB approach were the same as those found by the other pattern matching approaches.

For illustrative purposes, Figure 9 shows some of the patterns found by the ED, DTW, SVM and HSMM-CP approaches. Note that these patterns were often recognised as negative patterns by the TB or RB approaches in some cases. Figure 9(a) shows a sub-sequence from 2/2/2012 to 30/3/2012 in the HSI, which the RB, ED, DTW, SVM and HSMM-CP approaches recognised as a positive H&S-T pattern. However, the TB approach recognised it as a negative case. Figure 9(b) shows a sub-sequence from 28/1/2009 to 21/7/2009 in the DJI, which the TB, ED, DTW, SVM and HSMM-CP approaches recognised as a positive CWH pattern and which the RB approach identified as a negative pattern. Figure 9(c) is a sub-sequence from 9/12/2009 to 25/5/2010 in the NYSE, which the ED, DTW, SVM and HSMM-CP approaches recognised as a positive Trip-T pattern. In this case, both the RB and TB approaches failed to recognise it as a Trip-T pattern.

A majority of patterns found using the ED approach differed from the patterns found using the DTW approach. The numbers of patterns found using the ED and DTW approaches were similar. These two approaches measured the similarity between the enlarged template and non-segmented time series in a different way. The ED approach calculated the point-to-point ED and the DTW approach mapped a point from the template to one or more points from the sub-sequence to find the optimal warping path. Figure 10(a) shows a sub-sequence from 23/1/2006to 14/7/2006 in the DJI, which the DTW approach identified as a DT-E&E pattern. Figure 10(b) shows a sub-sequence from 2/11/2004 to 3/2/2005 in the HSI, which the DTW approach identified as a Wed-R pattern. Figure 10(c) shows a sub-sequence from 5/5/2014 to 23/10/2014 in the DJI, which the ED approach identified as a DT-E&E pattern. Figure 10(d) shows a sub-sequence from 23/4/2003 to 30/7/2003 in the HSI, which the ED approach identified as a Wed-R pattern.

Two hundred and ninety-eight of the three hundred and seventy-two patterns found using the SVM approach in the three stock indexes were identical to the patterns found using the other approaches. Figure 11(a)

shows a sub-sequence from 29/9/2003 to 27/11/2003 in the HSI, which the SVM approach identified as an H&S-T pattern. However, none of the other five approaches recognised this sub-sequence as a pattern. We found the highest number of patterns in the three stocks using the HSMM-CP. The majority of the H&S-T, Trip-T, DT-E&E and CWH patterns found using the HSMM-CP could also be found using other approaches. In contrast, most of the Wed-R and Tria-A patterns found by HSMM-CP were not recognised as patterns by the other approaches. Figure 11(b) is a sub-sequence from 3/9/2003 to 12/12/2003 in the HSI, which the HSMM-CP approach recognised as a Wed-R pattern. Figure 11(c) shows a sub-sequence from 20/11/2008 to 8/5/2009 in the NYSE, which the HSMM-CP identified as a Tria-A pattern.

The HB and IL patterns are represented by candlesticks. Using the RB approach, we detected the HB pattern in the weekly historical price of Tenet Healthcare Corp. (THC) from 1 January 1993 to 31 July 1995 and the IL pattern in the daily historical price of Abbot Laboratories (ABT) from 6 April 1983 to 9 March 1987. We adopted these datasets from [2] to illustrate the matching of the HB and IL patterns with the RB approach. Figures 12 and 13show the pattern matching results. We found that the RB approach could detect the HB and IL patterns exactly the same as the examples given in [2].

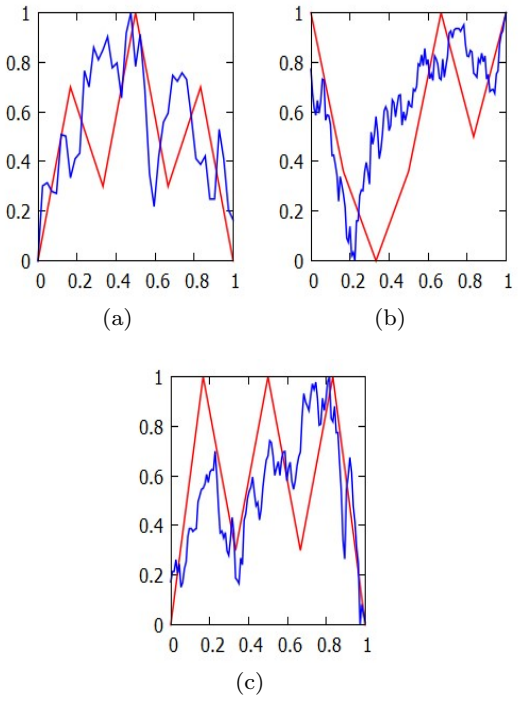


Fig. 9 (a) (Blue) A sub-sequence from 2/2/2012 to 30/3/2012 in HSI is identified as an H&S-T pattern; (Red) normalised H&S-T template. (b) (Blue) A sub-sequence from 28/1/2009 to 21/7/2009 in the DJI is identified as a CWH pattern; (Red) normalised CWH template. (c) (Blue) A sub-sequence from 9/12/2009 to 25/5/2010 in the NYSE is identified as a Trip-T pattern; (Red) normalised Trip-T template.

6.3 Efficiency of pattern matching approaches

In this section, we compare the efficiency of HSMM-CP with different pattern matching approaches. All the algorithms are coded in JAVA programming language on a computer with Intel(R) Core(TM) i7-4790 CPU 3.60GHz processor, 8 GB RAM, and 64-bit Microsoft Windows 7 operation system. For the experiment, we create synthetic datasets for the pattern H&S-T(Figure 1(a)). The lengths of the segments (window size) are set to 43, 85 and 127. In total, we create six synthetic datasets. Half of the datasets are used for training. The remaining sets are used for testing. Each dataset has 100 synthetic time series with 50 positive cases and 50 negative cases. SVM and HSMM-CP are trained by using the same datasets. The training accuracy is set to 0.97 for HSMM-CP. Table 13 shows the time cost of training for SVM and HSMM-CP with different length.

Table 14 shows the time required in testing for all the pattern matching methods for different lengths. The unit of time is millisecond. As shown in Table 13 and Table 14, for HSMM-CP, training and testing time increase rapidly when the length of the segments is increased. However, for SVM, training time is almost unchanged when the length of the segments is increased. We can also observe that the training and testing time SVM is much less than HSMM-CP. The testing time for DTW also increases with segments' length. We find that all the five approaches are more efficient (in time) than HSMM-CP.

Table 13 Training time of SVM and HSMM-CP for H&S-T synthetic datasets with length 43, 85 and 127.

	SVM	HSMM-CP
43	31	13463
85	46	110047
127	47	502607

We also compare the time cost of different pattern matching approaches for H&S-T pattern on a real dataset, which contains 2,506 points of the Hang Seng Index (HSI) from 1 January 2003 to 31 December 2012. A sliding window was used and shifted one data point at a time to obtain a sub-sequence. The window sizes used for the H&S-T pattern for the experiment are 43, 85 and 127. The comparison on time cost of different pattern matching approaches are shown in Table 15. The unit of time in Table 15 is millisecond.

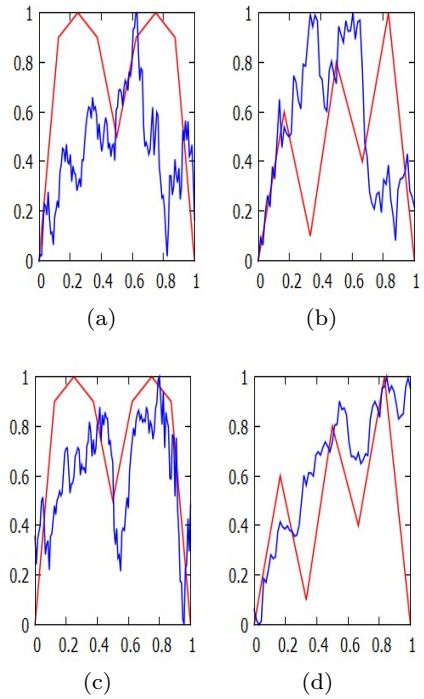


Fig. 10 (a) (Blue) A sub-sequence from 23/1/2006to 2006/7/14 in the DJI identified using the DTW approach; (Red) a normalised DT-E&E template. (b) (Blue) A sub-sequence from 2/11/2004 to 3/2/2005 in the HSI identified using the DTW approach; (Red) a normalised Wed-R template. (c) (Blue) A sub-sequence from 5/5/2014 to 23/10/2014 in the DJI identified using the ED approach; (Red) a normalised DT-E&E template. (d) (Blue) A sub-sequence from 23/4/2003 to 30/7/2003 in the HSI identified using the ED approach; (Red) a normalised Wed-R template.

Table 14 Testing time of TB, RB, ED, DTW, SVM and HSMM-CP for H&S-T synthetic datasets with length 43, 85 and 127.

	TB	RB	ED	DTW	SVM	HSMM-CP
43	16	15	15	32	16	1419
85	16	16	15	78	16	7363
127	32	43	15	141	16	28605

Table 15 Matching time of H&S-T pattern by TB, RB, ED, DTW, SVM and HSMM-CP methods from the HSI real dataset with window size of 43, 85 and 127.

	TB	RB	ED	DTW	SVM	HSMM-CP
43	47	47	47	359	94	31153
85	78	94	63	1248	140	177170
127	156	125	78	2636	188	670817

7 Conclusion

Financial analysts commonly use chart patterns to predict the future statuses of stocks. Locating chart patterns in financial time series is one of the most important steps for predicting stock market trends. In this paper, we present a novel approach for training

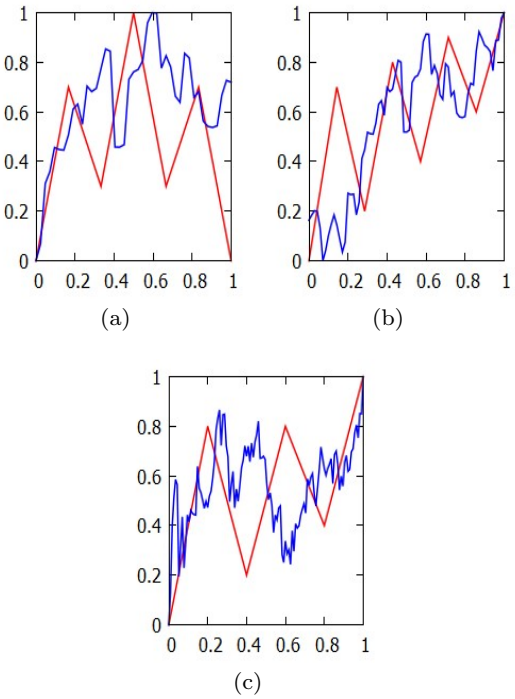


Fig. 11 (a) (Blue) A sub-sequence from 29/9/2003 to 27/11/2003 in the HSI identified by the SVM approach; (Red) normalised H&S-T template. (b) (Blue) A sub-sequence from 3/9/2003 to 12/12/2003 in the HSI identified by the HSMM-CP approach; (Red) normalised Wed-R template. (c) (Blue) A sub-sequence from 20/11/2008 to 8/5/2009 in the NYSE identified by the HSMM-CP approach; (Red) normalised Tria-A template.

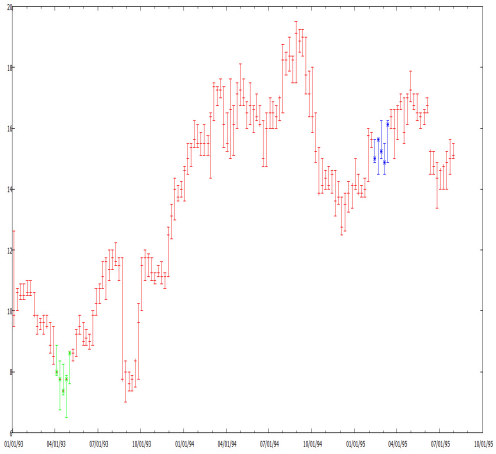


Fig. 12 Two HBs (green and blue) found using the RB approach on the THC stock (Red).

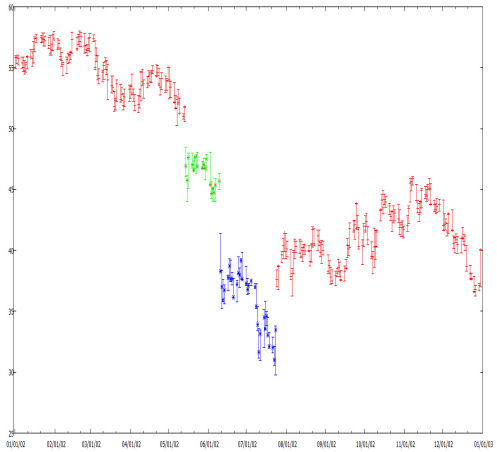


Fig. 13 Two ILs (green and blue) found using the RB approach on the ABT stock (Red).

and applying an HSMM for chart pattern matching. In this approach, we redesign the process of training an HSMM to recognise chart patterns with predefined templates. The proposed approach not only simplifies the traditional way of training an HSMM but also achieves good results in locating query templates from inputted time series. We evaluated the performance of a proposed method known as the HSMM-CP against other pattern matching approaches such as the RB, TB, ED, DTW and SVM approaches. We tested all of these approaches on both synthetic and real data from stock markets. For these experiments, we also designed the rules and templates for a set of representative patterns selected from 53 chart patterns. Experiments involving a synthetic dataset showed that the HSMM-CP had a higher level of accuracy and recall than the other approaches. Furthermore, experiments involving a real dataset showed that the HSMM-CP could detect more Wed-R and Tria-A patterns than the other approaches.

In our extensive experiments, we observed that segmentation could affect the pattern matching results of the TB and RB approaches. In particular, the rules designed for matching in the RB approach were subjective and could significantly influence the results. However, for HB and IL, which were represented by candlesticks, we found that the RB approach was a preferred method for detection. The TB approach measured similarity by calculating the temporal and amplitude distance between a fixed template and the segmented time series. We also found that the TB and RB approaches were inflexible at detecting variations in a given pattern. In addition, the TB, ED and DTW approaches were distance-based approaches, and a distance threshold was required for a decision to be made. Therefore, these approaches were susceptible to biases when the thresholds were designed. According to the TB approach, the numbers of points from the input time series had to be reduced via a segmentation process. In contrast, according to the ED and DTW approaches, the numbers of points in the pattern template had to be increased so that they equalled the queried time series. As the SVM and HSMM-CP approaches were trained with the positive and negative examples, they performed better in learning and recognising chart pattern variations.

As for the future work, we are planning to extend the HSMM-CP approach so that it can be used to recognise event patterns [2] from financial time series. We are also planning to extend the proposed model for pattern classification problems in health informatics and real-time control systems.

8 Acknowledgement

This research was funded by the Research Committee of University of Macau, grant MYRG2015-00054-FST and MYRG2016-00148-FST.

References

1. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: KDD workshop, vol. 10, pp. 359–370. Seattle, WA (1994)
2. Bulkowski, T.N.: Encyclopedia of chart patterns, 2nd edn. John Wiley & Sons (2011)
3. Cao, H., Jin, H., Wu, S., Ibrahim, S.: Petri net based grid workflow verification and optimization. The Journal of Supercomputing **66**(3), 1215–1230 (2013)
4. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) **2**(3), 27 (2011)
5. Chen, C.H., Tseng, V.S., Yu, H.H., Hong, T.P.: Time series pattern discovery by a PIP-based evolutionary approach. Soft Computing **17**(9), 1699–1710 (2013)
6. Chung, F.L., Fu, T.C., Luk, R., Ng, V.: Flexible time series pattern matching based on perceptually important points. In: International joint conference on artificial intelligence workshop on learning from temporal and spatial data, pp. 1–7 (2001)
7. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological) **39**(1), 1–38 (1977)
8. Fu, T.c., Chung, F.l., Luk, R., Ng, C.m.: Stock time series pattern matching: Template-based vs. rule-based approaches. Engineering Applications of Artificial Intelligence **20**(3), 347–364 (2007)
9. Ge, X., Smyth, P.: Deformable Markov Model Templates for Time-Series Pattern Matching. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 81–90. ACM (2000)
10. Gu, B., Sheng, V.S.: A robust regularization path algorithm for ν-support vector classification (2016)

11. Gu, B., Sheng, V.S., Tay, K.Y., Romano, W., Li, S.: Incremental support vector learning for ordinal regression. IEEE Transactions on Neural networks and learning systems **26**(7), 1403–1416 (2015)
12. Gu, B., Sun, X., Sheng, V.S.: Structural minimax probability machine (2016)
13. Holmes, W.J., Russell, M.J.: Probabilistic-trajectory segmental HMMs. Computer Speech & Language **13**(1), 3–37 (1999)
14. Keogh, E., Chu, S., Hart, D., Pazzani, M.: An online algorithm for segmenting time series. In: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on, pp. 289–296. IEEE (2001)
15. Keogh, E.J., Pazzani, M.J.: A simple dimensionality reduction technique for fast similarity search in large time series databases. In: Knowledge Discovery and Data Mining. Current Issues and New Applications, pp. 122–133. Springer (2000)
16. Kim, S., Smyth, P.: Segmental Hidden Markov Models with Random Effects for Waveform Modeling. The Journal of Machine Learning Research **7**, 945–969 (2006)
17. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE **77**(2), 257–286 (1989)
18. Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural computation **12**(5), 1207–1245 (2000)
19. Si, Y.W., Yin, J.: OBST-based segmentation approach to financial time series. Engineering Applications of Artificial Intelligence **26**(10), 2581–2596 (2013)
20. Wan, Y., Gong, X., Si, Y.W.: Effect of segmentation on financial time series pattern matching. Applied Soft Computing **38**, 346–359 (2016)
21. Wen, X., Shao, L., Xue, Y., Fang, W.: A rapid learning algorithm for vehicle classification. Information Sciences **295**, 395–406 (2015)
22. Xia, Z., Wang, X., Sun, X., Liu, Q., Xiong, N.: Steganalysis of lsb matching using differences between nonadjacent pixels. Multimedia Tools and Applications **75**(4), 1947–1962 (2016)
23. Xing, Z., Pei, J., Keogh, E.: A brief survey on sequence classification. ACM SIGKDD Explorations Newsletter **12**(1), 40–48 (2010)
24. Yu, S.Z.: Hidden semi-markov models. Artificial Intelligence **174**(2), 215–243 (2010)
25. Zapranis, A., Samolada, E.: Can Neural Networks Learn the "Head and Shoulders" Technical Analysis Price Pattern? Towards a Methodology for Testing the Efficient Market Hypothesis. In: Artificial Neural Networks–ICANN 2007, pp. 516–526. Springer (2007)
26. Zhang, Z., Jiang, J., Liu, X., Lau, R., Wang, H., Zhang, R.: A real time hybrid pattern matching scheme for stock time series. In: Proceedings of the Twenty-First Australasian Conference on Database Technologies-Volume 104, pp. 161–170. Australian Computer Society, Inc. (2010)
27. Zheng, Y., Jeon, B., Xu, D., Wu, Q., Zhang, H.: Image segmentation by generalized hierarchical fuzzy c-means algorithm. Journal of Intelligent & Fuzzy Systems **28**(2), 961–973 (2015)

9 Appendix A: How to set thresholds for the TB, ED and DTW approaches

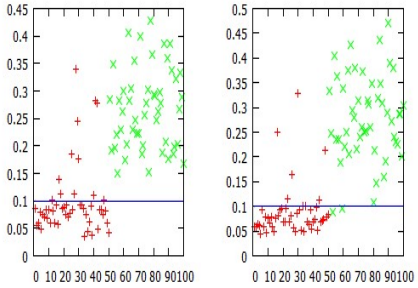
We use the H&S-T pattern as an example to illustrate how we set the thresholds in the experiment. We began by generating four datasets containing one hundred time series (the top fifty were H&S-T positive time series and the bottom fifty were randomly generated negative time series) with different lengths of 19, 43, 85 and 127. We used the TB, ED and DTW approaches, respectively, to calculate the similarities between each time series in the four datasets. The top 50 were positive cases and the distances had to be smaller than those in the bottom 50. As shown in Figures 14(a)-(d), the TB approach had a fixed threshold as the length of the time series increased. The threshold of the H&S-T pattern for TB θ=0.1. As shown in Figures 15(a)-(d), the threshold of the ED approach increased with the length of the time series. As shown in Figures 16 (a)-(d), the threshold of the DTW approach increased with the length of the time series. We modelled the threshold of the ED and DTW approach by a linear function of length. In Figures 15(a)-(d), the thresholds for the ED approach are 20, 40, 85 and 143 for lengths of 19, 43, 85 and 127, respectively. We regressed the threshold as a linear function of length, where the slope α=1.1417 and the intercept β=-6.2079. For the DTW approach, in Figures 16(a)-(d), the thresholds are 6, 10, 22 and 34 and correspond to lengths of 19, 43, 85 and 127. In the regression linear function of length, the slope γ=0.2649 and the intercept ε=-0.1457.

10 Appendix B:Experimental settings for a synthetic dataset

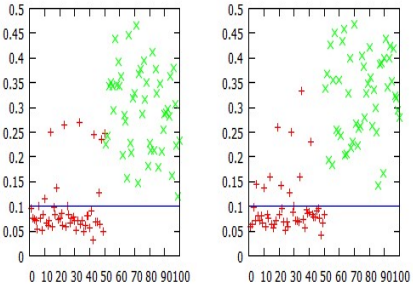
The experiment settings are shown in Table 10

Table 16 In the experiment conducted to distinguish H&S and Trip-T, the setting was (Distinguish, 100, 115, 7, 0.1 1.1417, -6.2079, 0.2649, -0.1457). Distinguish was a dataset containing 50 H&S-T and 50 Trip-T time series. As we designed the H&S-T patterns as positive cases, the threshold settings for the TB, ED and DTW approaches matched those in the H&S-T pattern recognition experiment.

P	Q	L	η	θ	α	β	γ	ε
Wed-R-1	100	115	7	0.1	0.8245	5.7715	0.2439	1.0408
Wed-R-2	100	115	7	0.1	0.9389	-0.5614	0.1578	3.4382
Wed-R-3	100	113	8	0.1	1.3247	-5.6701	0.2789	4.6973
Wed-R-4	100	113	8	0.1	0.8108	-3.9757	0.1893	1.8067
Tria-A-1	100	115	7	0.1	1.1905	-2.7998	0.2406	6.0177
Tria-A-2	100	115	7	0.1	0.8249	-0.7587	0.1274	6.0249
Tria-A-3	100	116	6	0.1	0.9888	-9.5791	0.2273	-0.332
Tira-A-4	100	116	6	0.1	1.0886	1.705	0.1803	3.7287
H&S-T	100	115	7	0.1	1.1417	-6.2079	0.2649	-0.1457
Trip-T	100	115	7	0.1	1.4494	-7.7872	0.2733	0.7797
Distinguish	100	115	7	0.1	1.1417	-6.2079	0.2649	-0.1457
CWH	100	115	7	0.16	0.6364	-4.345	0.1898	1.9956
DT-E&E	100	121	9	0.2	1.3708	-21.117	0.5646	-8.6917

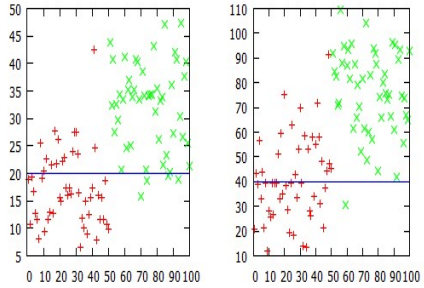


(a) (b)

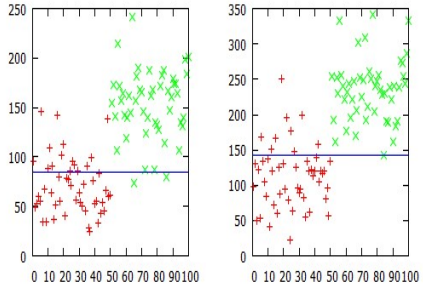


(c) (d)

Fig. 14 Similarity calculation results using the TB approach for the four datasets of varying sub-sequence length. The y-axis denotes the similarity and the x-axis denotes the case identification. The red crosses are the similarities calculated for the 50 positive cases and the green crosses are the similarities calculated for the 50 negative cases. The blue lines are the thresholds, which were found to be constant.



(a) (b)



(c) (d)

Fig. 15 Similarity calculation results using the ED approach on the four datasets of varying sub-sequence length. The y-axis denotes the similarity and the x-axis denotes the case identification. The red crosses are the similarities calculated for the 50 positive cases and the green crosses are the similarities calculated for the 50 negative cases. The blue lines are the thresholds, which increased when the length of the sub-sequence increased.

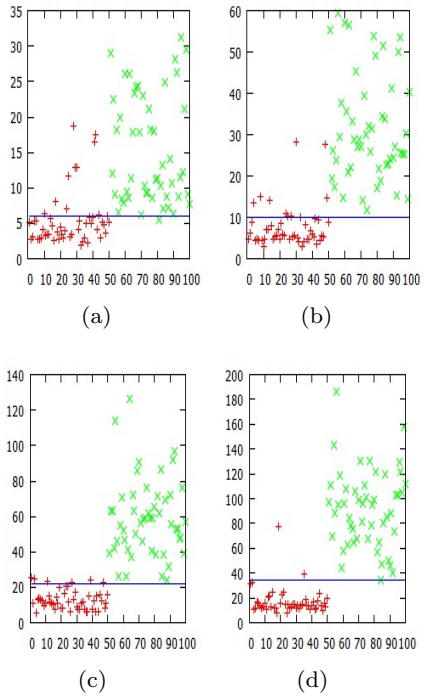


Fig. 16 Similarity calculation results using the DTW approach on the four datasets of varying sub-sequence length. The y-axis denotes the similarity and the x-axis denotes the case identification. The red crosses are the similarities calculated for the 50 positive cases and the green crosses are the similarities calculated for the 50 negative cases. The blue lines are the thresholds, which increased when the length of the sub-sequence increased.