Dynamic Time Warping Application for Financial Pattern Recognition

Yangqi Li, Taihua Hu

Abstract

Close price pattern is one of the widely applied technical indicators in market operations and trading. However, datasets or programming packages including calculation or labeling of close price patterns barely exist in public. With understanding of difficulties and flexibility of identifying a typical close price pattern, we developed a pattern recognition algorithm specially for close price patterns using Featured-based Dynamic Time Warping and Pattern Rule. The algorithm exhibits identified sequences with close resemblance of target financial patterns while incorporating some variations tolerable in human sense.

Keywords: Technical Analysis, Financial Pattern Recognition, Feature-based DTW, Pattern Rule

1. Introduction

Financial chart pattern is one type of technical analysis that predicts the future dynamics of financial market based on empirical evidence of historical sample paths that forms into certain shapes. It is mainly comprised of two types, candlestick pattern and close price pattern. Candlestick pattern is generally a short sequence of OHLC data that shares qualitative relations with each other. For example, Three Black Crows is three subsequent red candles with decreasing open, high, low and close price. Morningstar is one candle with merging open and close price but high price higher than low price. Close price pattern is a time series of close price that resembles some geometric shapes with no specified length. A typical example is head-shoulder-top. It is a time sequence with three peaks in which the second peak should be substantially higher than the first and last peaks. Many instruments of technical analysis, including candle data patterns, are widely applied by major investment software, except financial chart pattern detention, especially variable length sequence close price pattern detention. We hereby present a close price recognition algorithm, an auto-detention mechanism that searches and flags chart patterns with desirable characteristics on shape in close price path. This algorithm supports locating patterns with variable length and noisy sequence with outstanding characteristics satisfying the criteria of pattern. Hereinafter, we use close price pattern and financial pattern interchangeably.

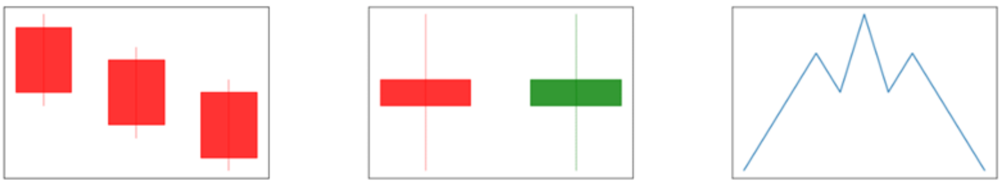


Figure 3.1-1 Three Black Crows, Morningstar and Head Shoulder Top

2. Literature Review

Among academic literature that targets pattern recognition, there is one alternative set of algorithms that fit an approximate pattern from raw, noisy sequences. Such is called segmentation method, including perceptually important points (PIP), piecewise aggregate approximation (PAA), piecewise linear approximation (PLA) and turning points (TP). It aims to reduce the number of data points in the time series (Wan et al., 2015) to extract an overall shape to represent the tendency of it. The advantage of segmentation is that it can draw a shape from sequences with variable length. But segmentation does not solve the pattern recognition. Financial patterns are time series with clear, fixed shape insightful for trading strategy construction. For every input sequence segmentation returns a shape, with which one still needs to design another detention scheme to document the approximate shape to existing public-recognized chart patterns. Besides, financial pattern recognition requires determinant capturing on shape in order to serve patterns as momentum signal, reversal signal, oscillation signals, etc. Conformity or approximation does not meet the criteria for investment purpose.

Another common pattern matching method is called Symbolic Aggregate approXimation (SAX). It transforms a numerical time series into a sequence of symbols according to which quantile a value sits in from a standard Gaussian distribution. Number of quantiles used to divide the distribution equals the number of characters used to represent the whole sequence. The distance measure between each character is the difference of quantile endpoint in which the corresponding value lies. One merit of SAX is that it supports fuzzy search as a range of value is represented by the same character, so that different data points laying in the same range has zero distance. However, despite its complexity in implementation, support of noise at each data point independently does not help capture the general shape of sequence. Instead, pure random noise sometimes distorts the characteristics on some points, thus distorts the general shape of the sequence, therefore such sequence is undesirable to locate.

After all the research and methodology evaluations, we adopt a pattern matching algorithm commonly applied beyond financial field, Dynamic Time Warping.

3. Model Framework

3.1. Dynamic Time Warping Distance Measure

Dynamic Time Warping, short for DTW, is designed to measure the distance between two time series. Given two time series $\mathrm{T} = \{t_1, t_2, \dots, t_n\}$ and $\mathrm{S} = \{s_1, s_2, \dots, s_m\}$, DTW seeks a path with minimum distance in a $n \times m$ matrix:

$$D = \begin{bmatrix} d(t_1, s_1) & d(t_1, s_2) & \cdots & d(t_1, s_m) \\ d(t_2, s_1) & d(t_2, s_2) & & \\ \vdots & & \ddots & \\ d(t_1, s_m) & & & d(t_n, s_m) \end{bmatrix}$$

where $D(i,j) = d(t_i, s_j)$ stands for ith row, jth column entry of matrix D and $d(t_i, s_j)$ is the distance measure between two points t_i and s_j . One may refer such path as warping path $\mathrm{W} = \{\mathrm{w}_1, \mathrm{w}_2, \dots, w_k, \dots, \mathrm{w}_\mathrm{K}\}$, in which $max(n,m) < K < m + n - 1$. $w_k = D(i,j)$ such that it minimizes the following function, which is namely DTW measure:

$$\mathrm{DTW}(T, S) = \min\left(\sqrt{\sum_{k=1}^{K} w_k}\right)$$

In walking the warping path, several constraints must be met. Suppose $w_k = d(t_i, s_j)$ and $w_{k-1} = d(t_{i'}, s_{j'})$ with $i, i' \leq n$ and $j, j' \leq m$:

1. **Boundary condition.** $w_1 = d(t_1, s_1)$ and $w_K = d(t_n, s_m)$
2. **Continuity.** $i - i' \leq 1$ and $j - j' \leq 1$
3. **Monotonicity.** $i - i' \geq 0$ and $j - j' \geq 0$

DTW warping path is generally computed with Dynamic Programming, with which a cumulative distance matrix C of the same dimension as D is created to store ith row, jth column entry, $C(i,j)$, with the following value:

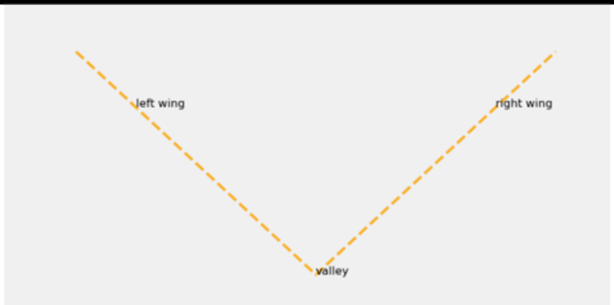
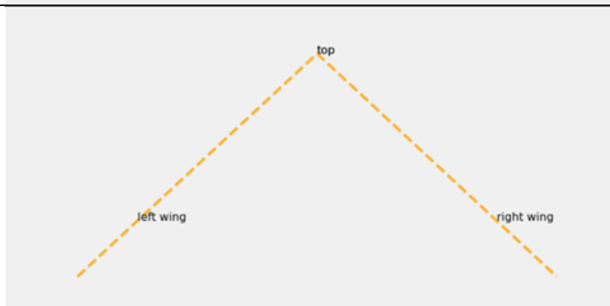
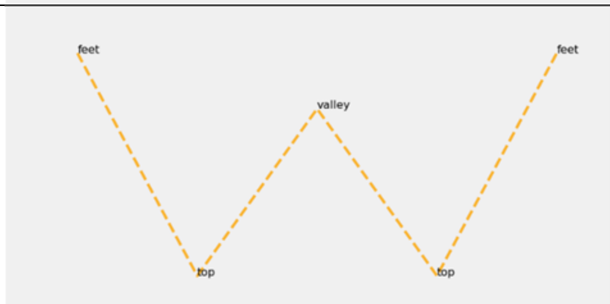
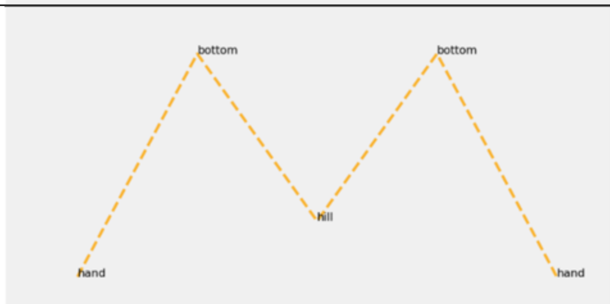
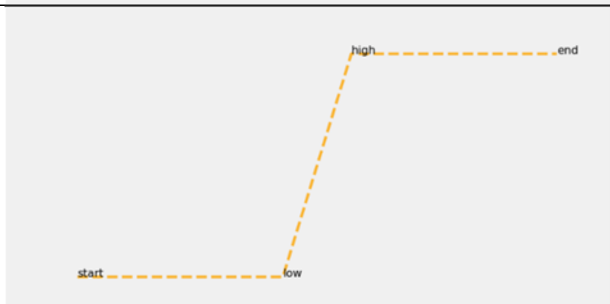
$$C(i,j) = d(t_i, s_j) + min\{C(i-1, j-1), C(i-1, j), C(i, j-1)\}$$

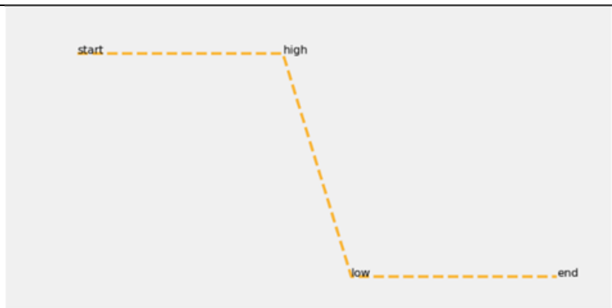
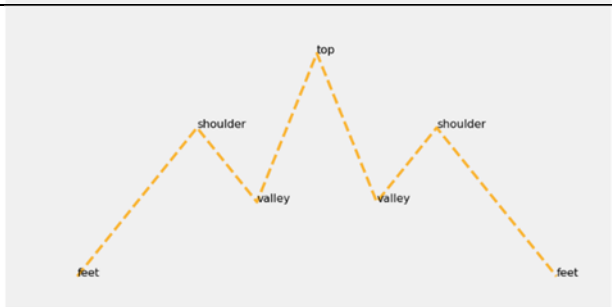
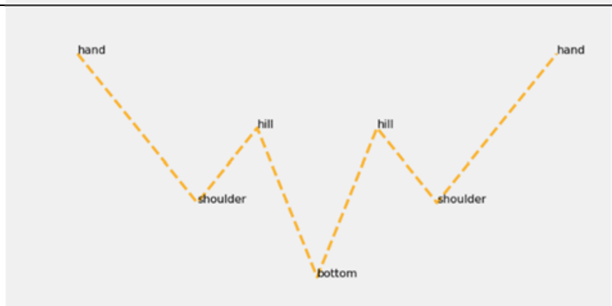
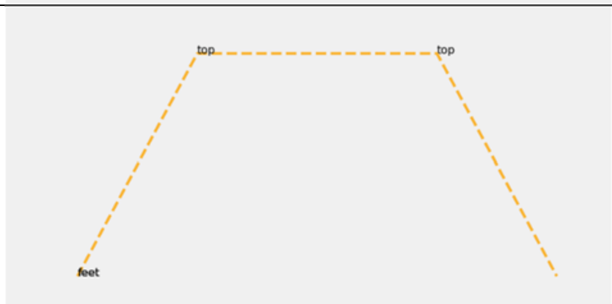
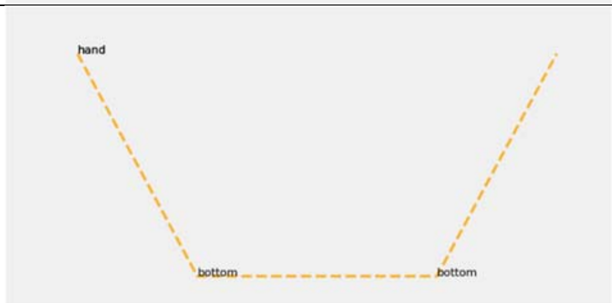
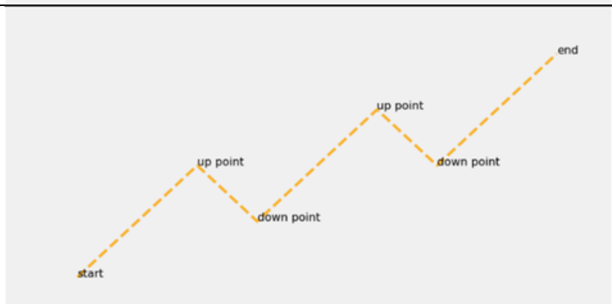
If $C(i-1, j-1)$, $C(i-1, j)$, or $C(i, j-1)$ is adopted as minimum, such single operation is called match, insertion, or deletion.

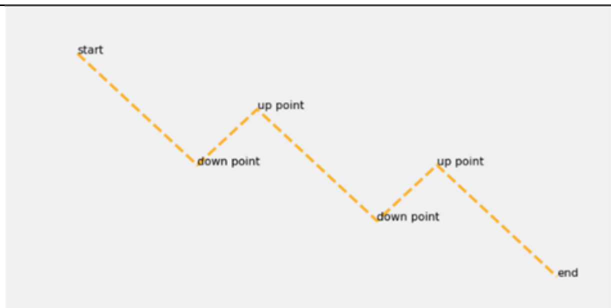
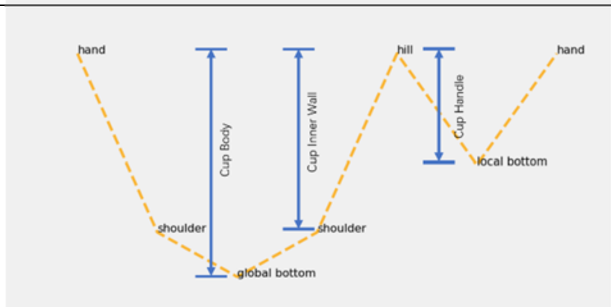
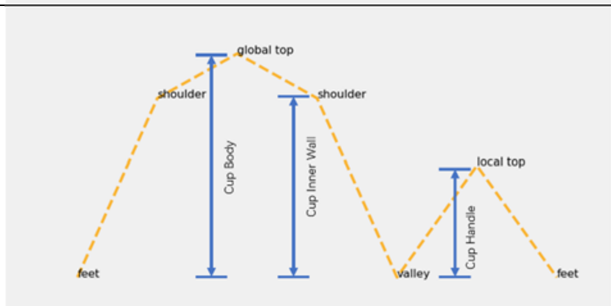
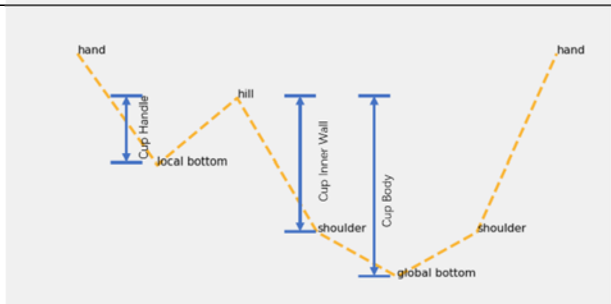
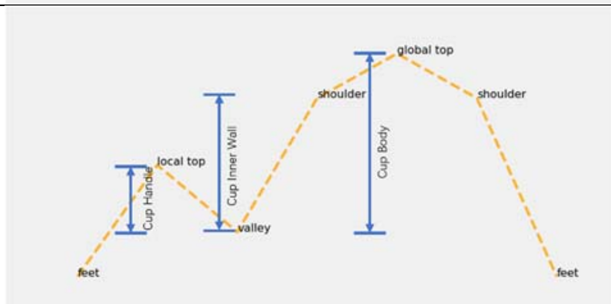
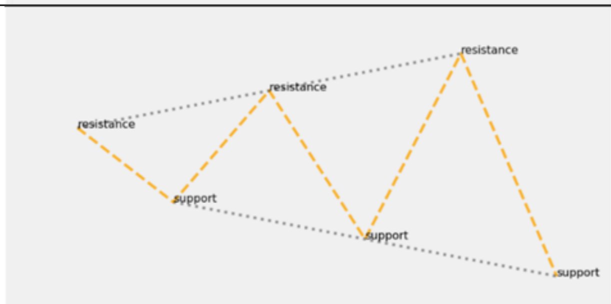
3.2. Model I: Baseline DTW Model

In order to fit DTW into financial pattern recognition, one must specify desirable patterns to search. We designed a series of dummy time series of financial patterns as templates and constructed a numerical array such that its shape matches the standard financial pattern in definition. In the parenthesis of each pattern name are acronyms of formal pattern name used in implementation with programming scripts.

A majority of patterns are inspired by works from Yuqing Wan and Yain-Whar Si. (Wan et al., 2017)

Pattern Name	**Chart Pattern Shape**
Single Valley (VUp)	
Single Hill (VDn)	
Double Valley (W)	
Double Hill (M)	
Step Up (StUp)	

<table>
<tr><td>Step Down (StDn)</td><td>
</td></tr>
<tr><td>Head Shoulder Top (HST)</td><td>
</td></tr>
<tr><td>Head Shoulder Bottom (HSB)</td><td>
</td></tr>
<tr><td>Flat Top (FT)</td><td>
</td></tr>
<tr><td>Flat Bottom (FB)</td><td>
</td></tr>
<tr><td>Up Channel (UC)</td><td>
</td></tr>
</table>

Down Channel (DC)	
Cup with Handle (CwH)	
Cup with Handle Inverted (CwHI)	
Bump and Run Reversal Bottom (BRRB)	
Bump and Run Reversal Top (BRRT)	
Broading Bottom (BrdB)	

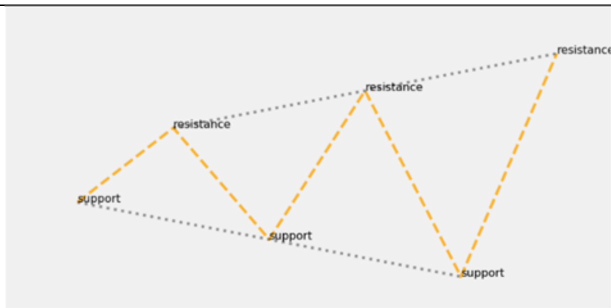
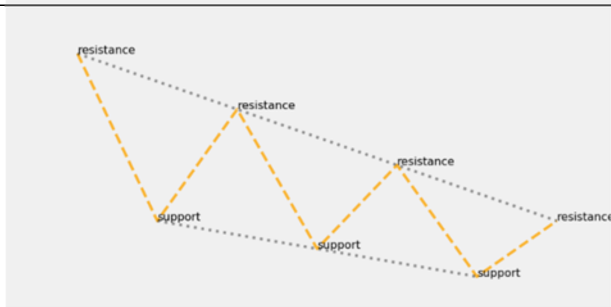
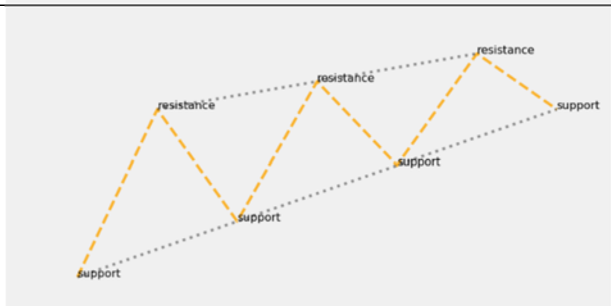
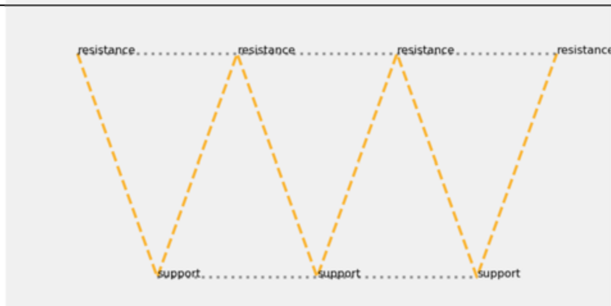
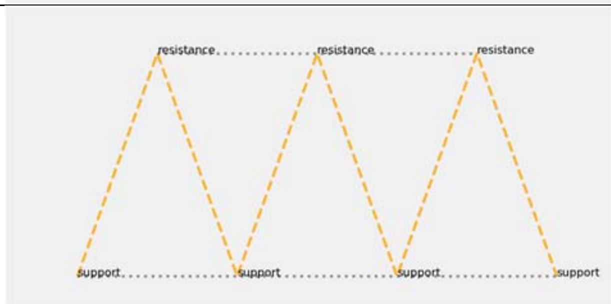
<table>
<tr><td>Broading Top (BrdT)</td><td>
</td></tr>
<tr><td>Wedge Falling (WegF)</td><td>
</td></tr>
<tr><td>Wedge Rising (WegR)</td><td>
</td></tr>
<tr><td>Rectangular Bottom (RecB)</td><td>
</td></tr>
<tr><td>Rectangular Top (RecT)</td><td>
</td></tr>
</table>

Table 3.2-1 Chart Pattern and Template Shape with Characteristics

First, we specified a fixed length of target sequence of close price. It is the time series that we want to search whether it has the shape of a desired financial pattern. Note that the target sequence does not have to be the same length of template thanks to DTW matching, therefore we can search sequences of different length in a certain range from template length. Before entering the target sequence and template into DTW computation, we applied a min-max scaling to both time series respectively in order to put them in the same range. DTW distance computed in that way will thus be unambiguous and consistent with shape resemblance. For example, after min-max scaling, it is possible to have DTW of 0 with Euclidean distance measure, meaning that two candidate time series are identical. Moreover, with min-max scaling, it is definite that the smaller the DTW distance, the more resemble the shapes of two time series. In a way, min-max scaling is necessary. Standardization, i.e. mean-variance scaling, is not considered even though some literature claims it may reserve the shape of time series better, because standardization does not grant two time series in the same range.

$$Min - Max\ Scaling: \tilde{x} = \frac{x - X_{min}}{X_{max} - X_{min}}\ \forall x \in X$$

After computing DTW from min-max scaled time series, we need a filter to decide how small the DTW distance should be to identify that target sequence as pattern. Since the comparison of slices of historical close price with template forms a distribution of DTW distance, we can initiate a threshold from the distribution to filter small DTW distances and their corresponding target sequences. Specifically, we build a quantile scheme and a deviation scheme. For quantile scheme, if a DTW computed from comparison with a template is smaller than a specified quantile, for example, 5 percent quantile, the target sequence corresponding to that DTW is labelled pattern with template shape. For deviation scheme, if a DTW is smaller than a multiple of standard deviation subtracted from mean, that corresponding target sequence is considered of template shape. In the following equation, ~ stands for resemblance in shape or "is similar to".

$$Threshold_{quantile} = q_\alpha$$
$$Filter_{quantile}: \mathrm{T} \sim \mathrm{S}\ \forall\ DTW_{(T,S)} < q_\alpha$$
$$Threshold_{deviation} = \mu - C\sigma$$
$$Filter_{deviation}: \mathrm{T} \sim \mathrm{S}\ \forall\ DTW_{(T,S)} < \mu - C\sigma$$

Since we used rolling window in slicing the target pattern, it is of high probability that some filtered targeted sequences will be highly overlapped. For example, one target might be just one-step away from the other with only one data point different. We designed two schemes of checking overlap to tackle it. With merge scheme, filtered target sequences with overlap ratio up to a certain degree, e.g. 0.67, will be merged into a large target sequence. With drop scheme, only the first target sequence is kept and the rest overlapped sequences will be dropped. We did not apply a multi-step-ahead rolling window, as we were concerned that multiple steps might miss some fine-shaped target pattern. Note that merge/drop scheme is only applied in exhibition, because captured overlapped sequence in a series is not a noise, but rather shows a process of sequence moving into and out of template pattern. Withholding those observations is beneficial for some deep learning model, e.g., ConvLSTM.

The initial implementation of DTW to extract patterns brings some valuable insights for further improvements. For instance, DTW appreciates certain degree of noise in the time series, therefore it can capture variations of template shapes. However, too much noise would inherently break the global shape, exceeding noise tolerance of attributing a variation of time series into one shape from a human perspective. In fact, the output from baseline DTW sometimes exhibits that the captured sequence does not resemble template because noise is added to some critical characteristic points and thereby lose the overall form. Considering this and thanks to the warping path of DTW algorithm, we were successful in constructing pattern rules for each template.

3.3. Model II: DTW Model with Pattern Rules

Our requirement on financial pattern is strict in terms of characteristics of shape. So, we restricted the patterns found by DTW with preset pattern rules. Pattern rule is a collection of limitations of characteristics of a shape in a time series, such as global maximum, global minimum and trend of time series. Applying pattern rule is appropriate for financial pattern detention because it is essentially data mining, where overfitting or fine tuning is appreciated and desired. Pattern rule is distinct for each template, specifying the relations between each characteristic point or the trend of some parts of the time series. In implementing the pattern rule in scripts, it does not strictly follow the descriptions as there is generally an easy approach to conduct with geometric relations in min-max scaled sequence. A concrete description can be seen below:

Pattern Name	Rule

Single Valley	➢ Valley should be the lowest ➢ Two hands (endpoints) should be roughly equal ➢ Left wing should have spikes only in the same direction of valley ➢ Right wing should have spikes only in the same direction of valley
Single Hill	➢ Hill should be the highest ➢ Two feet (endpoints) should be roughly equal ➢ Left wing should have spikes only in the same direction of hill ➢ Right wing should have spikes only in the same direction of hill
Double Valley	➢ Two hands (endpoints) should be higher than tops and valley ➢ Hill should not be too low compared with the vertical distances of hands and bottom ➢ Two valleys should be roughly equal ➢ Two hands should be roughly equal ➢ Two hands should be the highest
Double Hill	➢ Two feet should be lower than tops and valley ➢ Valley should not be too shallow compared with the height of tops ➢ Two tops should be roughly equal ➢ Two feet should be roughly equal ➢ Two tops should be the highest
Step Up	➢ Jump should be greater than 30% of the lower side ➢ Lower side should not differ too much ➢ Higher side should not differ too much
Step Down	➢ Jump should be greater than 30% of the lower side ➢ Lower side should not differ too much ➢ Higher side should not differ too much
Head Shoulder Top	➢ Two feet should be relatively equal ➢ Two shoulders should be relatively equal ➢ Two valleys should be relatively equal ➢ Top should be considerably high compared with shoulders ➢ Valleys should be low compared with shoulder ➢ Valleys should be considerably high compared with feet ➢ All points except top should be strictly lower than top
Head Shoulder Bottom	➢ Two hands should be relatively equal ➢ Two shoulders should be relatively equal ➢ Two hills should be relatively equal ➢ Hills should be high compared with shoulder ➢ Hills should be considerably low compared with hands ➢ Bottom should be fairly small compared with shoulders ➢ All points except top should be strictly higher than bottom
Flat Top	➢ Two feet should be relatively equal ➢ Top should be flat ➢ Top should exceed feet to a certain degree ➢ All points except top should be strictly lower than top ➢ Recognized tops must be consecutive and have at least 2 points
Flat Bottom	➢ Two hands should be relatively equal

	➢ Bottom should be flat ➢ Hand should exceed bottom to a certain degree ➢ All points except bottom should be strictly higher than bottom ➢ Recognized bottoms must be consecutive and have at least 2 points
Up Channel	➢ First upper changepoint (up point) is lower than second upper changepoint ➢ First lower changepoint (down point) is lower than second downward changepoint ➢ First upper changepoint is higher than first lower changepoint ➢ Second upper changepoint is higher than second lower changepoint ➢ Starting point is lower than the first lower point ➢ Ending point is higher than the second upper point
Down Channel	➢ First upper changepoint is higher than second upper changepoint ➢ First lower changepoint is higher than second lower changepoint ➢ First upper changepoint is lower than first lower changepoint ➢ Second upper changepoint is lower than second lower changepoint ➢ Starting point is higher than the first lower point ➢ Ending point is lower than the second upper point
Cup with Handle	➢ Two hands should be relatively equal ➢ Two shoulders should be relatively equal ➢ Hands should not be too high for hill ➢ Cup handle should not neither be too high nor low compared with cup body ➢ Local bottom should be considerably high for shoulders ➢ Cup inner wall shall be steep ➢ All points except itself should be strictly higher than global bottom
Cup with Handle Inverted	➢ Two feet should be relatively equal ➢ Two shoulders should be relatively equal ➢ Feet should be approximately equivalent to valley ➢ Cup handle should not neither be too high nor low compared with cup body ➢ Cup handle should be lower than shoulders ➢ Cup inner wall shall be steep ➢ All points except itself should be strictly lower than global top
Bump and Run Reversal Bottom	➢ Two hands should be relatively equal ➢ Two shoulders should be relatively equal ➢ Hands should not be too high for hill ➢ Hills should be lower than hands ➢ Cup handle should not either be too high nor low compared with cup body ➢ Local bottom should be considerably high for shoulders ➢ Cup inner wall shall be steep ➢ All points except top should be strictly higher than global bottom
Bump and Run Reversal Top	➢ Two feet should be relatively equal

	➢ Two shoulders should be relatively equal ➢ Feet should not be too low compared with valley ➢ Feet should be lower than hands ➢ Cup handle should not either be too high nor low compared with cup body ➢ Cup handle should be lower than shoulders ➢ Cup inner wall shall be steep ➢ All points except top should be strictly lower than global top
Broading Bottom	➢ Resistance level slope (RLS) is positive ➢ Support level slope (SLS) is negative ➢ Initial trend is considerably downward
Broading Top	➢ Resistance level slope is positive ➢ Support level slope is negative ➢ Initial trend is considerably upward
Wedge Falling	➢ Resistance level slope is negative ➢ Support level slope is negative ➢ Initial trend is considerably downward ➢ RLS is larger than SLS in absolute term
Wedge Rising	➢ Resistance level slope is positive ➢ Support level slope is positive ➢ Initial trend is considerably upward ➢ SLS is larger than RLS in absolute term
Rectangular Bottom	➢ Resistance level slope should have slight or no difference from 0 ➢ Support level slope should have slight or no difference from 0 ➢ Initial trend is considerably downward ➢ RLS and SLS is equivalent in absolute term
Rectangular Top	➢ Resistance level slope should have slight or no difference from 0 ➢ Support level slope should have slight or no difference from 0 ➢ Initial trend is considerably upward ➢ RLS and SLS is equivalent in absolute term

Table 3.3-1 Pattern Rule

We provide two patterns with two typical types of restrictions to illustrate the exact application of pattern rule. One is restrictions on changepoints, typical shown in reversal patterns, such as head shoulder, cup handle, etc. For example, in head shoulder bottom, we need the bottom to be considerably lower than shoulders. In implementation, we enforce that filtered target sequences by quantiles or deviations should have its minimum shoulder value larger than 1.2 times the bottom value. Bottom and shoulders of each frame is realized by DTW warping path. Our head shoulder bottom template takes a form from numerical array of length 9, in which bottom is located at index 4 (index starting from 0) and shoulder is located at index 2 and 6. According to the figure 3.3-1 with 12-point-long target sequence below, template data point at index 2, denoted as t_2, matches target data point at index 3, denoted as s_3. Then the algorithm determines that s_3 is recognized by DTW as left shoulder. Similarly, t_4 matches s_5 and s_6, then s_5 and s_6 are the top of target sequence. t_6 matches s_8, meaning s_8 is the right shoulder of target sequence, at least in DTW regards. According to our pattern rule number 4 of head shoulder bottom, bottom should be considerably small compared with shoulders. In our implementation, maximum of two shoulders should be larger than 1.2 times bottom. Down to our notation, it should be $\min(s_3, s_8) > 1.2 \cdot \min(s_5, s_6)$. Target sequence that does not satisfy this condition will be thrown out of head shoulder bottom pattern, even though its DTW is lower than the threshold. Please note that in exhibition graphs below, red lines only represent matchings between points from two arrays, whose distance is not necessary distance measure in DTW, as points in sequences are 1D and temporal distance is negligible, or even devastating to consider in pattern matching.

Figure 3.3-1 A Ruled-in Head Shoulder Bottom Target Sequence by Pattern Rule
(deviation_rule: 1 means pattern rule imposed onto sequences filtered by 1 deviation)

Another is restrictions on trend, typical shown in oscillation patterns and trend patterns, such as rectangular patterns, broading patterns, wedges, etc. Though trend is mostly reflected upon changepoints, the rule of restriction takes a different form other than positional relations. As a typical instance, Wedge Rising can be seem as comprised of two increasing levels, support level and resistance level. Support level is formed by index 0, index 2, index 4 and index 6 of template, so support level in target sequences is formed by points matching index 0, index 2, index 4 and index 6 of template. Similarly, resistance level in target sequences is formed by points matching index 1, index 3, index 5 and index 7 of template. One of the Wedge Rising pattern rule regulates that support level slope, short for SLS, is larger than resistance level slope, short for RLS, in absolute term. In implementation, the slopes are computed from linear regression of candidate data points. SLS minus RLS needs to be larger than 0.2 times maximum of SLS and RLS to satisfy this pattern rule. Target sequence not meeting the criterion is not considered Wedge Rising. The target sequence ought to satisfy all criteria in Wedge Rising to be elected as such pattern, so does criteria in other template patterns.

Figure 3.3-2 A Ruled-in Wedge Rising Target Sequence by Pattern Rule

It is worth noticing that DTW algorithm and pattern rules complement each other. According to the plots elected by DTW and by pattern rule respectively, plots chosen by DTW conform to the general shape of template, but sometimes would have a few points off, causing mismatch of partial characteristic points. As one can see in the left part of Figure 3.3-3, the shape of target sequence resembles Bump and

Run Reversal template except for the last few data points. Pattern rules rigorously formulate the relations of shape characteristics, but they ignore the overall form. In the right part of Figure 3.3-3, the characteristic points, such as hill, hands, local bottom and global bottom, etc., conforms to the pattern rules. But a few points in the later part make the general shape closer to Double Valley than to Bump and Run Reversal Bottom. To benefit from both methodologies, we apply pattern rules on target sequences filtered by DTW threshold. In Figure 3.3-4, the general shape fits a Bump and Run Reversal, and characteristic points follows the pattern rule. Therefore, the target sequence has a fine shape resembling the Bump and Run Reversal.

Figure 3.3-3 Target Sequence Flagged as Bump and Run Reversal by Deviation Threshold (Left) and Pattern Rule (Right)

Figure 3.3-4 Target Sequence Flagged as Bump and Run Reversal by Pattern Rule Applied to Deviation Threshold Filtered Sequences

Applying DTW solely and DTW with pattern rule filter do not search satisfying patterns that resembles templates all the time. It is mainly due to the fact that a time series of close price is recognized as pattern because of its shape. However, DTW matches data points from two time series according to its numerical value. Although the time series are transformed with min-max scaler, one feature from DTW that one data point in one time series would match multiple points in the other would sometimes distort the desirable shape mandated by the shape of templates. There are some modifications on DTW to limit the number of points in one sequence matched by one point of the other, such as Sakoe-Chiba band or Itakura parallelogram (Cassisi et al., 2012), which essentially constraint the warping path not to stray away from the diagonal. However, such multi-matching feature is somewhat helpful in finding target patterns that have characteristics scattered over the time frame and recognizing multiple points as one characteristic. Besides, constrained DTW restricts the length of time frame from straying too far from length of templates, since DTW matches all points in both time series. Considering limitation in constrained DTW and its weak recognition in shape, we experimented two variations of DTW that transformed the target sequences to search for shapes.

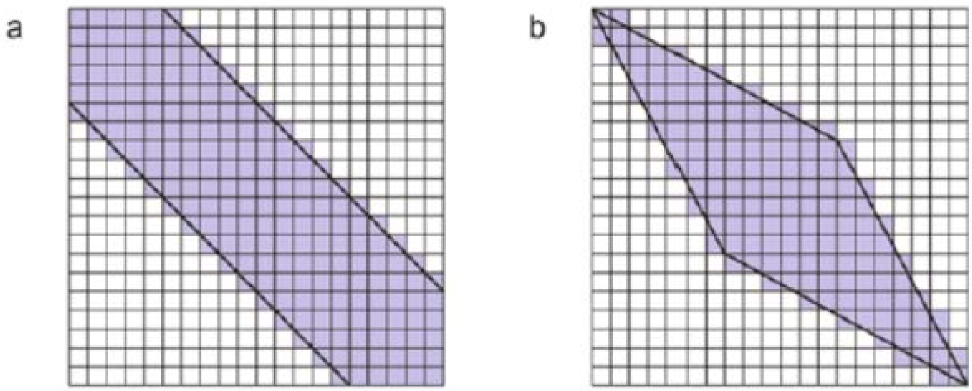


Figure 3.3-5 Sakoe-Chiba Band (a) and Itakura Parallelogram (b) Shown in Distance Matrix

3.4. Model III: DTW with Feature-based Objective Function

In better addressing shape in pattern matching to search sequence with fine shape, we worked around different DTW variations. One is applying derivative of sequence into DTW. Derivative DTW is designed to match the points in sequence according to its derivative rather than the ground base value. First order derivative calculates slope, a fine representation of shape. Second order derivative calculates curvature. Summation of ground base DTW, first order DTW and second order DTW weighs both numerical value, slope and curvature of time series, thus forming a Taylor-like DTW computation. However, the application of Taylor DTW is incompatible with application of pattern rule, as it introduces a dilemma critical to rule implementation, i.e. which warp path should be used in implementing pattern rule, warp path of ground base value, first-order derivative or second-order derivative. It is also infeasible to incorporate all three warp path and compute an average warp path or the like. Eventually, we aborted the implementation of Taylor DTW. A formula is provided below as a reference of this abandoned methodology.

$$Taylor\ DTW_{T,S} = DTW_{T,S} + DTW_{T',S'} + DTW_{T'',S''}$$

Our remedy lies at Feature based DTW, a change to the loss function of distance measure within DTW. (Xie et al., 2010) It computes global feature and local feature of sequences and adds them up as distance measure between two data points. For each point, it first computes a local vector consisting of two values, current point minus previous point and current point minus next point. It better reflects the local trend than the first derivative of the point. Denoting $R = \{r_1, r_2, \dots, r_T\}$ and $Q = \{q_1, q_2, \dots, q_S\}$, we can write the following formula:

$$f_{local} = (r_t - r_{t-1}, r_t - r_{t+1})$$

While local feature can address the shape like first derivative can, it does not address the position of data point in whole sequence. The value of data point can serve that purpose; However, it is not on the same scale with local feature. So global feature is computed in the same fashion of local feature with a slightly different term to subtract. In the equation below, T and S denote the final index of sequences.

$$f_{global} = \left(r_t - \sum_{i=1}^{t-1} r_i/(t-1)\, , r_t - \sum_{i=t+1}^{T} r_i/(T-t) \right)$$

Our new distance measure is summing local distance and global distance:

$$dist(r_t, q_s) = dist_{local}(r_t, q_s) + dist_{global}(r_t, q_s)$$

in which distance measure on vectors of local and global feature has two methods of computation. The first method is elementwise:

$$\mathrm{dist}_{local}(r_t, q_s) = \left|\left(f_{local}(r_t)\right)_1 - \left(f_{local}(q_s)\right)_1\right| + \left|\left(f_{local}(r_t)\right)_2 - \left(f_{local}(q_s)\right)_2\right|$$
$$\mathrm{dist}_{global}(r_t, q_s) = \left|\left(f_{global}(r_t)\right)_1 - \left(f_{global}(q_s)\right)_1\right| + \left|\left(f_{global}(r_t)\right)_2 - \left(f_{global}(q_s)\right)_2\right|$$

The second method is norm of vector:

$$\mathrm{dist}_{local}(r_t, q_s) = \left|f_{local}(r_t) - f_{local}(q_s)\right|$$
$$\mathrm{dist}_{global}(r_t, q_s) = \left|f_{global}(r_t) - f_{global}(q_s)\right|$$

In our implementation we write both methods but choose the second method as default because it is a standard vector operation. Eventually we carry out DTW with $dist(r_t, q_s)$ as our new pointwise distance measure.

One may notice that using feature-based measure would cut off the beginning point and ending point of sequence, because their features cannot be computed. It results in shortening of template sequence so our first implementation on Feature-based DTW did not harvest too many sequences with fine shape as short template does not fix the shape of target sequence well enough because of multi-matching. To cope with it we elongate each template by 2 and mark length of effective target sequence as length of original target sequence subtracted by 2. For example, we use elongated template to search target sequence with length of 12 and note that only 10 points are in the warping path of Feature-based DTW. Template Elongation is specific for different types of pattern. If the pattern is oscillation pattern, we add one point each at beginning and ending point according to its preceding and following cycle of period. If the pattern is reversal pattern or trend pattern, one point each is added according to the trend of beginning and ending point.

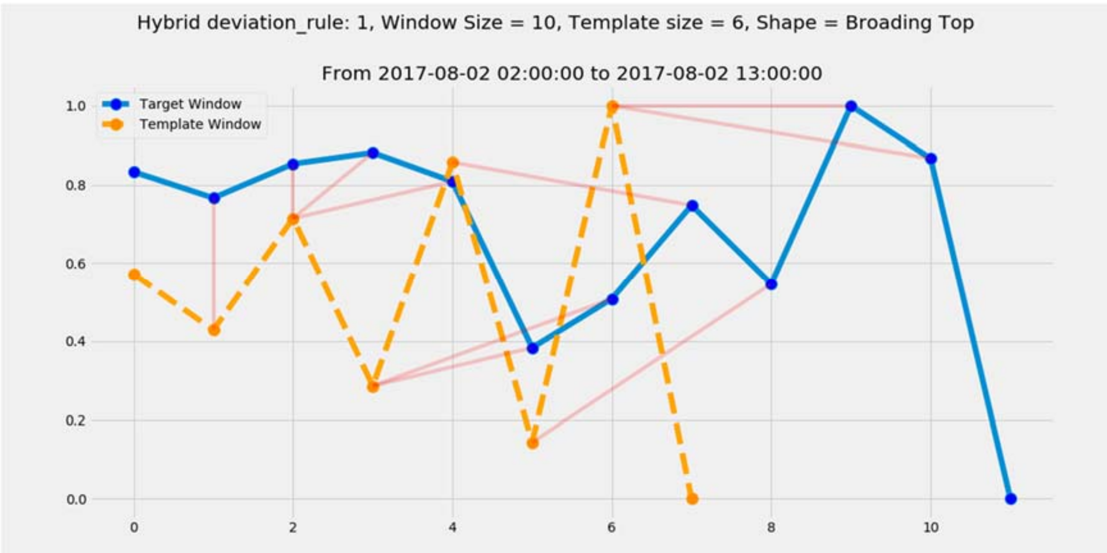
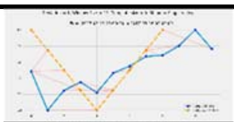
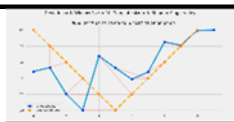
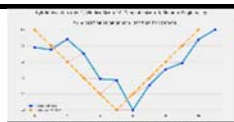
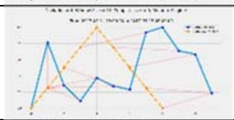
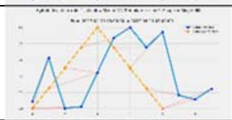
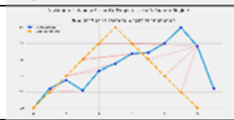
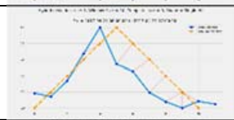
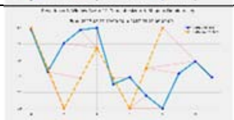
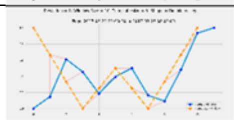
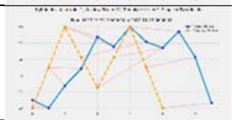
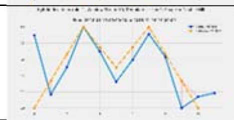
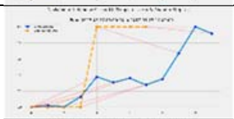
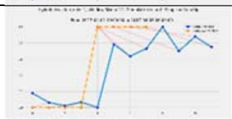
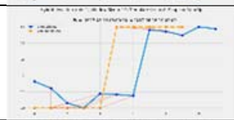
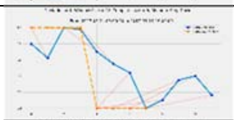
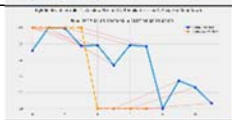
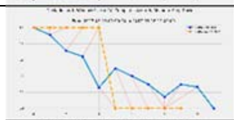
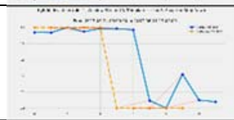
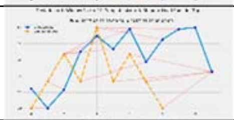
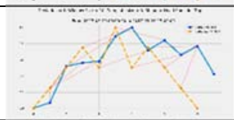
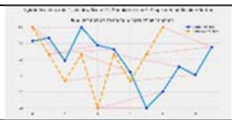
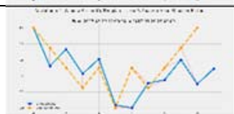
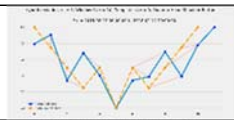
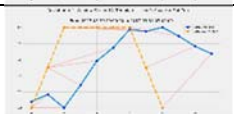
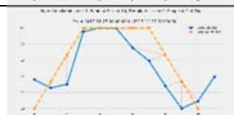
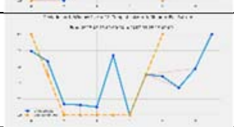
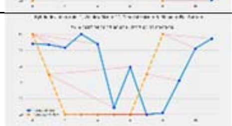
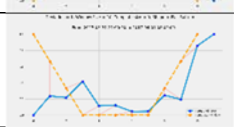


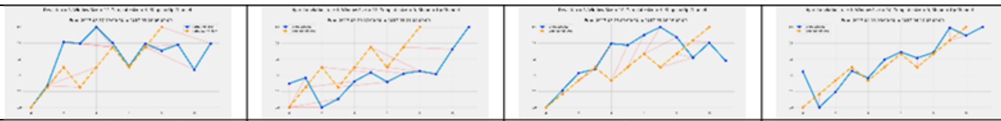
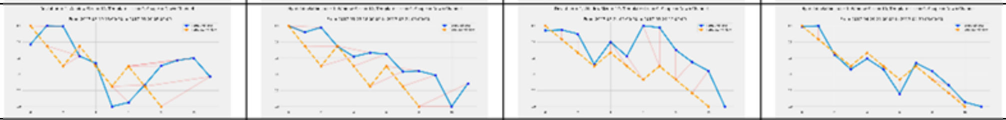
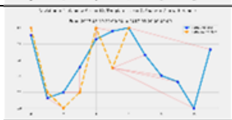
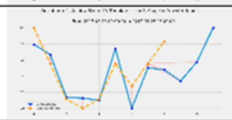
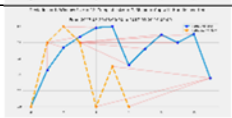
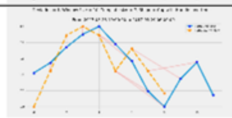
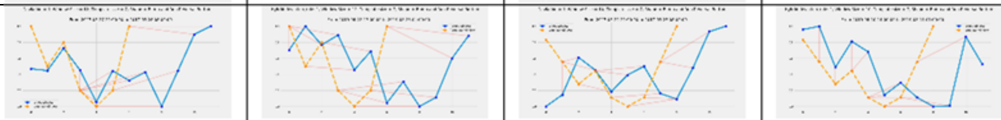
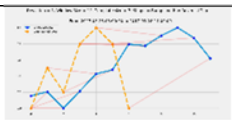
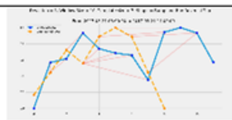
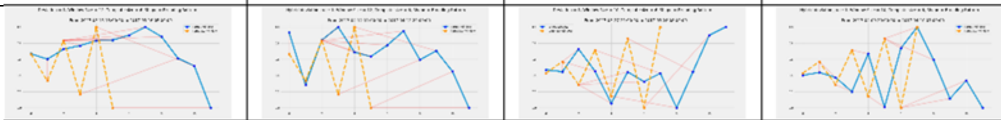
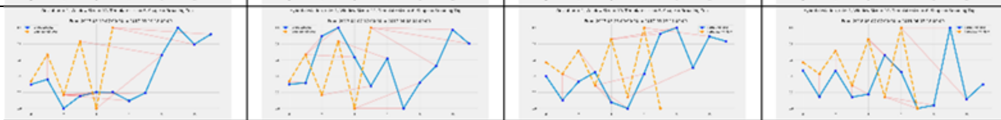
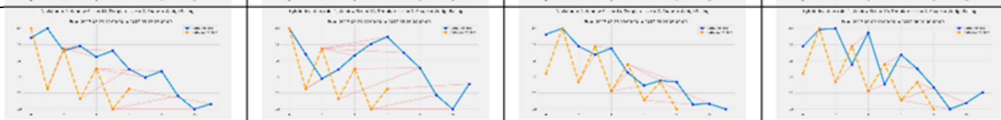
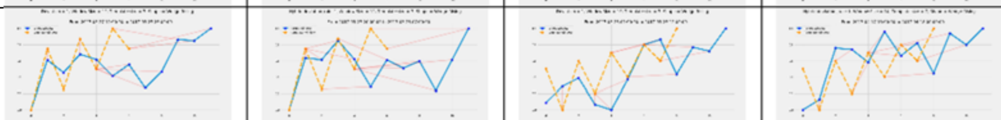
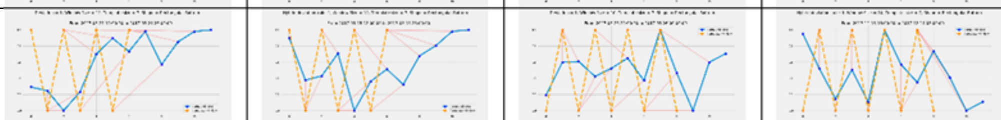
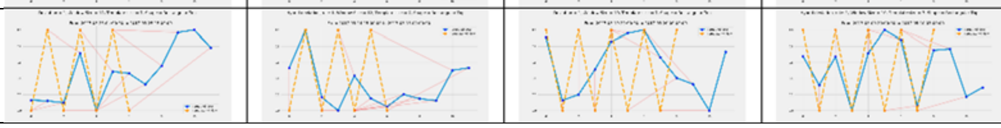
Figure 3.4-1 An Illustration of Feature-based DTW Warping Path. Note that the beginning and ending point of both template and target sequence are not matched.

Applying Feature-based DTW solely already guarantees closeness of target sequence to template no matter in shape or ground base value. Another notable advantage over other DTW variations, such as Taylor DTW is that Taylor DTW needs a considerable allocation of weights on each Derivative DTW, including ground base DTW. It is due to that ground base DTW does not have the same scale with first derivative DTW and second derivative DTW. Technically, to tune three components in Taylor DTW under the same scale even requires weights to be different for each target sequence in order not to let one of the three values overweighs other two values. (Górecki et al., 2013) The result is that executing an unbiased Taylor requires more expensive computational effort than executing Feature-based DTW. Our pattern searching methodology rests at pattern rule application after Feature-based DTW for warranty of substantial characteristic point relations.

4. Empirical Analysis

In this section, we exhibit results from each pattern recognition method in the form of target sequence captured as pattern. In the exhibition, we will show 22 patterns with four methods, baseline DTW, DTW plus Pattern Rule, Feature-based DTW, and Feature based DTW with Pattern Rule. The improvement with each methodology innovated from DTW is demonstrated in comparison of two pairs: DTW versus DTW plus Pattern Rule and DTW versus Feature-based DTW. Feature-based DTW plus Pattern Rule is presented as the ultimate version. Our experiment was carried out in crypto pair BTC/USDT with 1-hour frequency from 2017/03/14 12:00 UTC to 2019/09/17 4:00 UTC. Please note that even Feature-based takes beginning point and ending point into computation, the warping path construction does not consider the beginning point and ending point. Thus, some distortion of the general shape happened at the beginning point and ending point is not necessarily at the responsibility of FBDTW algorithm. Theoretically, FBDTW would tolerate those distortions if the local/global feature at the other side outweighs local/global feature at beginning/ending side, at which local feature should be the same value of global feature. All target sequences are selected with deviation method of 1 as threshold, i.e. sequence with DTW smaller than 1 standard deviation less the mean of DTW.

Pattern Name	**DTW**	**DTW + Rule**	**FBDTW**	**FBDTW + Rule**
Single Valley				
Single Hill				
Double Valley				
Double Hill				
Step Up				
Step Down				
Head Shoulder Top		N/A in trial crypto pair		N/A in trial crypto pair
Head Shoulder Bottom				
Flat Top				
Flat Bottom				

Up Channel				
Down Channel				
Cup with Handle		N/A in trial crypto pair		
Cup with Handle Inverted		N/A in trial crypto pair		N/A in trial crypto pair
Bump and Run Reversal Bottom				
Bump and Run Reversal Top		N/A in trial crypto pair		N/A in trial crypto pair
Broading Bottom				
Broading Top				
Wedge Falling				
Wedge Rising				
Rectangular Bottom				
Rectangular Top				

(Reversal Pattern: Single Valley, Single Hill, Double Valley, Double Hill, Head Shoulder Top, Head Shoulder Bottom, Flat Top, Flat Bottom, Cup with Handle, Cup with Handle Inverted, Bump and Run Reversal Bottom, Bump and Run Reversal Top) Reversal Pattern comprises a majority of financial patterns. There are simple patterns formed with just two monotonic opposite trends and there are complicated patterns with multiple zigzags during reversal. For simple patterns, DTW captures the general shape of the whole sequence quite well. DTW and FBDTW achieve similar results for simple patterns such as Single Hill, Single Valley, Double Hill and Double Valley because their partial trend is relatively stable and their changepoints are not many to recognize. FBDTW generally keeps a better general shape than DTW does. But both pure DTW distance-based measures tend to tolerate distortion from one of two points because DTW returns an average of pointwise distances in the warping path. This is where pattern rule would help restrict distortions, contributing to a captured sequence with almost all points necessary comforting to the outlook of template.

Flat Top and Flat Bottom are also simple reversal patterns by definition. But DTW performs poorly in capturing sequences satisfying the characteristics of them. The main reason lies at its rigorous requirement of flat parts. Because DTW accepts distortion at a few points, DTW-captured sequences always have imperfection at flat parts. With FBDTW flat features are mandated to a certain degree, but still distortions at one or two points are still present on the trend, in which we call left/right wings according to the location of trend. So FBDTW with pattern rule constraining the points at wings turns out as the last resort.

In complicated patterns is the FBDTW most effective. In our searched patterns, complicated patterns not only have more than two turning points, but also each turning point has ordinal relationship to one another. Turning points are changepoints where trend is reversed. For example, Head Shoulder Bottom has five turning points and the third points should be the lowest, the second and fourth turning points should be the two highest points among the five points and relatively equal, and the first and fifth turning points should also be relatively equal. For a pattern with such detailed characteristics, baseline DTW is vulnerable, frequently compromising inferiors as template pattern. For instance, DTW would mark sequence resembling Double Valley or even Triple Valley as Head Shoulder Bottom. With FBDTW, the characteristics points tend to follow the ordinal order regulated by template. But provided the inherent tolerance of minor pointwise distortions of all DTW variations, pattern rule is the ultimate safeguard for turning point relations.

(Trend Pattern: Down Channel, Up Channel, Step Up, Step Down) Trend Pattern is easy for traders to manipulate capital allocation and an easier pattern to detect than reversal pattern. It is mainly due to its monotonicity and few characteristics to capture. Either for DTW or FBDTW, the uptrend/downtrend is guaranteed in captured sequences. The performance of each methodology remains the same as in reversal pattern: FBDTW better captures the shape, and Pattern Rule keeps the characteristics well-behaved. This observation is typically seen in Up Channel and Down Channel, as they have three turning points not violating the overall monotonic trend. In FBDTW the turning points are more apparent and with Pattern Rule the turning points forms an upward/downward support-resistance level more clearly, thus fit the template shape more closely than general DTW and DTW with Rule. In Step Up and Step Down, the huge increase/decrease is emphasized by all methods, since otherwise the pointwise distance would be extremely large around the steep, too large to even out by average. Still, relatively flat lower/upper end is secured by FBDTW and Pattern Rule, in which FBDTW seems to outperform Pattern Rule in maintaining flat end. One cause is that we made a relatively slack rule for Step Up and Step Down at ends as the steep should be the most outstanding feature rather than flat ends. Secondly, ends are more of a partial shape than of a few characteristic points. As is previously stated, FBDTW is good at maintaining fine shapes locally and globally, thus prevails Pattern Rule. Empirically, Step Up/Down may be the only patterns that FBDTW achieves more marginal improvements than Pattern Rule does.

(Oscillation Pattern: Broadening Bottom, Broadening Top, Rectangular Bottom, Rectangular Top) Oscillation Pattern is one of the reasons we seek the DTW variations. In this category, Broadening shapes can be called magnifying variance patterns and Rectangular shapes can be called constant variance pattern. DTW-captured sequences are considerably ambiguous for us to assert a property of sequence variance. To Take Broadening Bottom as an example, sequences captured by DTW seems to have magnifying variance as time elapses, but no turning points. In fact, it is common across all sequences from DTW. Posing Pattern Rule makes the sequences slightly better but does not solve the fundamental problem. In our example of DTW-Rule-captured sequence, it has obvious turnings points, however the trend between turning points does not exhibit magnifying momentum sometimes. Therefore, the overall variance looks like increasing with time, but there are always some parts with seemingly shrinking variance. The underlying causes are also inconclusive at the beginning. We first mainly attributed such outcome to multi-matching, that one point in the template matches three or more points in the target sequence. It also reminds us that the length of target sequence should not be too distant from length of template. But multi-matching phenomenon brings more insights in volatile patterns. Multi-matching occurs from multiple points with similar values cluster around one location. For one point in template that has sought a point in the cluster into the warping path, it is natural for it to link the neighbors for minimizing pointwise distance. What DTW neglects is the relation of neighboring points, with which points form a sequence and a shape. Point in the global location is certainly one key, but its location related to other plots a partial form in the whole sequence. Missing one partial shape may be tolerable in some patterns where the global shape matters, but not in broadening patterns that requires a magnification of oscillation in the progress of walking the sequence. To put weights on global and local shape, adopting FBDTW thus becomes a natural call. Capturing rectangular patterns follows the same rationale, except constant-variance sequence is innately rare in crypto markets. Using either deviation or quantile filter with thresholds uniform with other patterns will usually include some noisy pictures in the selection. Even so, FBDTW is capable of selecting one of the best few shapes from the candidate pool.

(Oscillation-Trend Pattern: Wedge Falling, Wedge Rising) Hybrid patterns contain features of other single-featured patterns. Named after its wedge-like shape pointed at either an upward or downward direction, wedge pattern has an overall upward/downward trend and oscillates with shrinking variance during the process. The resolution in searching this pattern is indeed FBDTW, as the problem of baseline DTW deteriorates in wedge patterns. Neither DTW- nor DTW-Rule captured sequences exhibit clear shrinking variance nor explicit turning points conforming to narrowing support-resistance levels. With FBDTW, turning points are not obscure and forming a narrowing channel upward/downward. With Pattern Rule posed onto FBDTW, the variance is decreasing more aggressively in walking the sequence, assuring a finer wedge shape than with plain FBDTW.

5. Conclusion and Future Work

Our conclusion, based on theoretical soundness and empirical analysis, is that Feature-based DTW and Pattern Rule is effective in financial pattern recognition, especially for patterns with variable length and complicated shapes. Besides recognition of general shape and characteristic points of the sequences resembling certain close price patterns. DTW method could include some slight mutations of financial patterns present in real-life cases. Our implementation on applying DTW into close price financial patterns recognition is therefore deemed successful.

Though DTW distance measure is a powerful instrument in analyzing sequence similarity and is a numeric value. DTW value is arbitrary to quantify. Indeed, the smaller DTW value is, the more similar two sequences are. However, the relation is somehow qualitative. Our next step is to transform DTW, which is ranging from 0 to $+\infty$, to a closed interval range, such as probability measure, so that this metrics is easy to interpret. Furthermore, we will build a prediction scheme on future DTW and Pattern Rule with contemporary information with neural networks techniques incorporating features other than close price, to understand other factors contributing to pattern formation, such as volume.

Reference

Cassisi, Carmelo & Montalto, Placido & Aliotta, Marco & Cannata, Andrea & Pulvirenti, Alfredo. (2012). Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining. 10.5772/49941.

Górecki, Tomasz & Łuczak, Maciej. (2013). Using derivatives in time series classification. Data Mining and Knowledge Discovery. 26. 310-331. 10.1007/s10618-012-0251-4.

Górecki, Tomasz & Łuczak, Maciej. (2014). First and Second Derivatives in Time Series Classification Using DTW. Communications in Statistics: Simulation and Computation. 43. 10.1080/03610918.2013.775296.

Wan, Yuqing & Gong, Xueyuan & Si, Yain Whar. (2015). Effect of segmentation on financial time series pattern matching. Applied Soft Computing. 38. 10.1016/j.asoc.2015.10.012.

Wan, Yuqing & Si, Yain Whar. (2017). A formal approach to chart patterns classification in financial time series. Information Sciences. 411. 10.1016/j.ins.2017.05.028.

Xie, Y., & Wiltgen, B. (2010). Adaptive Feature Based Dynamic Time Warping. IJCSNS International Journal of Computer Science and Network Security, 10(1), 10.