Data Representation and Pattern Recognition in Financial Time Series

Seunghye J. Wilson[1] and James E. Gentle[1]

[1]Department of Statistics, George Mason University, 4400 University Drive, Fairfax, VA 22030

Abstract
One major interest of financial time series analysis is to identify changepoints of trends and recognize patterns that can be used for classification and clustering of time series. Because of the large amounts of data, nonlinear relationship of the data elements and the presence of random noise, some method of data reduction is necessary. The data reduction, however, must preserve the important characteristics of the original data. Many representation methods in the time domain or frequency domain have been suggested to accomplish efficient extraction of information. These include, for example, piecewise linear approximation, symbolic representation, and discrete wavelet transformation (DWT). However, most of the existing methods do not take into consideration time information of trends and/or depend on user-defined parameters, for example the number of segments for piecewise approximation. We introduce piecewise band smoothing (PBS) for data representation based on linear regression using small sets of current data points. The proposed method is flexible and interpretable in the sense that it allows the acquisition and addition of new data points (online method) to detect meaningful trends and changepoints. Changepoints are confirmed once new data points stray far enough outside of the band, creating a reduced dataset of changepoints to utilize. Next, we define patterns from the reduced data which preserve trends and the length of a trends duration. Finally, a distance metric is suggested as a similarity measure to classify the present application example of classification.

Key Words: Data reduction, Financial time series, Piecewise linear approximation, Online method, Similarity measure, Classification

1. Introduction

A lot of large sized datasets, particularly financial datasets, are in the form of time series data or can be converted into temporal data. Although traditional time series models using statistical techniques, such as ARIMA and GARCH, are relatively simple to use and easy to understand data generating processes, the assumptions of these models are not likely realistic for massive time series datasets. As a result, data mining techniques for financial time series analyses have been explored and developed as alternatives to traditional statistical models.

In time series data mining, unlike general data mining, data representation or data preprocessing is required to reduce the dimensionality of data. Reduced data by data representation, however, must preserve the important characteristics of interest from the original data. While data representation methods in the time domain or frequency domain have been suggested for financial time series analysis, many of these methods do not take into consideration time information when the local characteristics of data, for example "up" or "down" linear trends, but rather only focus on approximating the original data based on user-defined parameters such as the number of segments.

In this paper we introduce new data representation methods, *Alternating Trends Smoothing* (ATS) and *Piecewise Band Smoothing* (PBS). They are piecewise linear approximation methods. These methods reduce the original time series data with continuous straight lines by identifying "up" or "down" trend changepoints. While the reduced data by ATS represent increasing and decreasing trends alternatively, the reduced data by PBS does not necessarily represent up/down alternating patterns since it can define trend "change" more flexible with several parameters. Next, we suggest methods of pattern definition on the reduced data by ATS or PBS in numerical and symbolic forms with appropriate similarity measures. Then, we show an application example of stock price data classification based on their visual similarities.

2. Methods

2.1 Data Representation

While the major purpose of data representation is data reduction, the reduced data must preserve interesting patterns and features of the original data as much as possible. In financial time series data, most interesting pattern and feature are probably increasing/decreasing patterns, and their changepoints. Our two new data representation methods, *alternating trends smoothing*(ATS) and *piecewise band smoothing*(PBS) represent lengthy and noisy time series data by detecting these trends and their changepoints.

2.1.1 Alternating Trends Smoothing (ATS)

Alternating trends smoothing, or ATS, is a piecewise linear smoothing method that represents the original time series by "up" and "down" straight lines alternatively. The algorithm is given in **Algorithm 1**.
The output of **Algorithm 1** for a time series $x_1, x_2, \cdots$ would be,

$$(b_1, c_1), (b_2, c_2), \cdots$$

where $b_1 = 1$, and b_i's and c_i's $(i = 1, 2, \cdots)$ are time indices and values at changepoints. ATS has a tuning parameter h in **Algorithm 1**, that is a "step size". The algorithm begins with examining h data points to identify the next changepoint beyond the current changepoint. Larger step size h tends to identify a fewer since the algorithm moves faster with more data points (larger h) to begin to identify the next changepoints (Figure 1). However, the distance between two changepoints can be smaller than the step size, even as small as 1 time unit. Notice that the ATS may overshoot the peaks and valleys because the identification of a changepoint is delayed untill the trend change is confirmed by a true trend (Figure 1, right), a peak between $t = 60$ and $t = 70$.

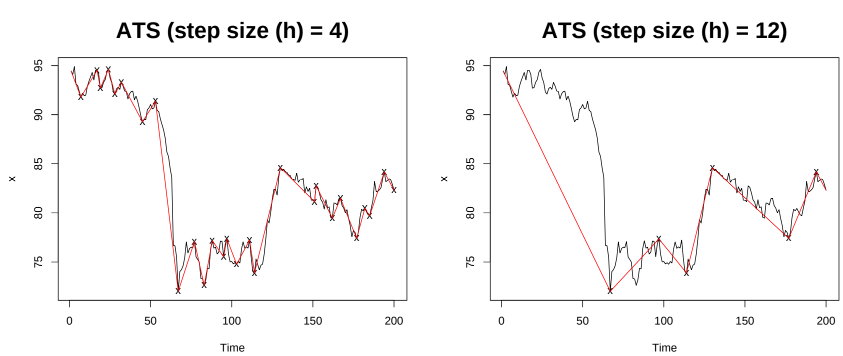

Figure 1: Data representation by ATS with $h = 4$ and $h = 12$

Algorithm 1 Alternating Trends Smoothing

Set $d \leftarrow 1$ (changepoint counter)
while more data in first time setp **do**
  for $i = 1, 2, \cdots, m$, where $m = h$ if h additional data available, or m is last data item **do**
    input x_i;
    $b_d \leftarrow 1; c_d \leftarrow x_1$
    Determine $j_+, j_-, x_{j_+}, x_{j_-}$ such that
      $x_{j_+} = \max(x_1, x_2, \cdots, x_h)$ and $x_{j_-} = \min(x_1, x_2, \cdots, x_h)$
    Set $s = (x_k - x_i)/(k - i)$ and $r = sign(s)$
    while $r = 0$ **do**
      Continue inputting more data; stop with error at end of data
    end while
  end for
end while
Set $j \leftarrow i$ (index of last datum in previous step); and set $d \leftarrow d + 1$
while more data **do**
  for $i = j + 1, j + 2, \cdots, j + m$ ($m = h$ if h additional data available, or $j + m$ is last data item) **do**
    Input x_i;
    while $sign(s) = r$ **do**
      Set $k \leftarrow \min(i + h, n)$ where n is the number of data points
      if k=i **then** break
      end if
      Set $s \leftarrow (x_k - x_j)/(k - j)$
      Set $j \leftarrow k$
    end while
    Determine j_+ such that $rx_{j_+} \leftarrow \max(rx_{j+1}, \cdots, rx_{j+m})$
    Set $b_d \leftarrow j_+$; and set $c_d \leftarrow x_{j_+}$
    Set $d \leftarrow d + 1$; set $j \leftarrow j_+$; and set $r \leftarrow -r$
  end for
  Set $b_d \leftarrow j_+$; and set $c_d \leftarrow x_{j_+}$
end while

2.1.2 *Piecewise Band Smoothing (PBS)*

While ATS identifies changepoints based on trend direction, *piecewise band smoothing* (PBS) considers not only the increasing/decreasing direction change but also the changes in the magnitude of trends in the same direction, thus it rarely overshoots the peaks and valleys. In the algorithm of PBS, there are some parameters to control the criteria of trend change, *initial window size* (w), *bandwidth* (B), and *change ratio* (R). *The initial window size*, w, is the number of data points to determine the current linear trend. Using the initial window size $(w > 3)$, the algorithm fits the linear model by least squared methods, and the coefficient, $\hat{\beta}_1$, of this fitted line in (1) is considered "current trend".

$$\hat{x_t} = \hat{\beta}_0 + \hat{\beta}_1 t, \quad t = 1, 2, \cdots, w \tag{1}$$

To meet continuity constraints of the approximated data, the linear models without intercept ($\hat{\beta}_0 = 0$) is used. After the current trend is fitted, the algorithm examines one point at a time to identify if the trend change occurred or not. *Bandwidth* (B) is the tolerable range of data fluctuation under the current trend line. That is, if the vertical distance from the trend line to x_j, $(j = w+1, \cdots)$ is greater than B, it is assumed that the trend change occurred at time $j-1$. Otherwise, the algorithm moves to the next point and examines its deviation. The larger bandwidth tends to identify a fewer number of changepoints (Figure 2).

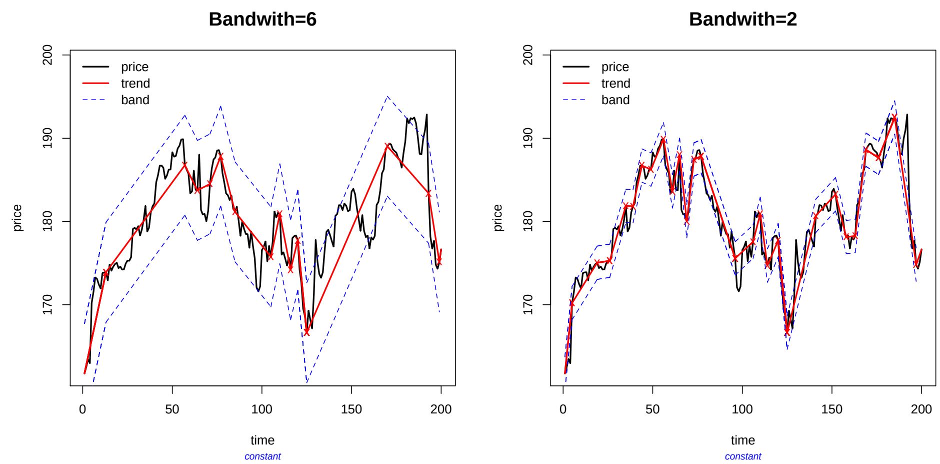

Figure 2: Data representation with bandwidth 6 (left) and 2 (right) - IBM daily closing price from January 13, 2012 to October 26, 2012

The *change ratio* (R) is the threshold ratio to determine "significance" of change in the magnitude of two successive trends. Sometimes after PBS identifies a changepoint, the next trend actually may not be significantly different or seem to be almost, since the algorithm identifies changepoints by examining one point at a time. Thus, the change ratio specifies the ratio of the magnitude between two successive trend lines that is thought of as being a significant change in trend. Specifically, for two successive trends β_i and β_{i+1} and $sign(\beta_i) = sign(\beta_{i+1})$, the change ratio R is defined by

$$R = \min\left(\frac{\beta_{i+1}}{\beta_i}\right) \text{ if } \left(\frac{\beta_{i+1}}{\beta_i}\right) > 1 \quad (2)$$

$$= \min\left(\frac{\beta_i}{\beta_{i+1}}\right) \text{ if } \left(\frac{\beta_{i+1}}{\beta_i}\right) < 1. \quad (3)$$

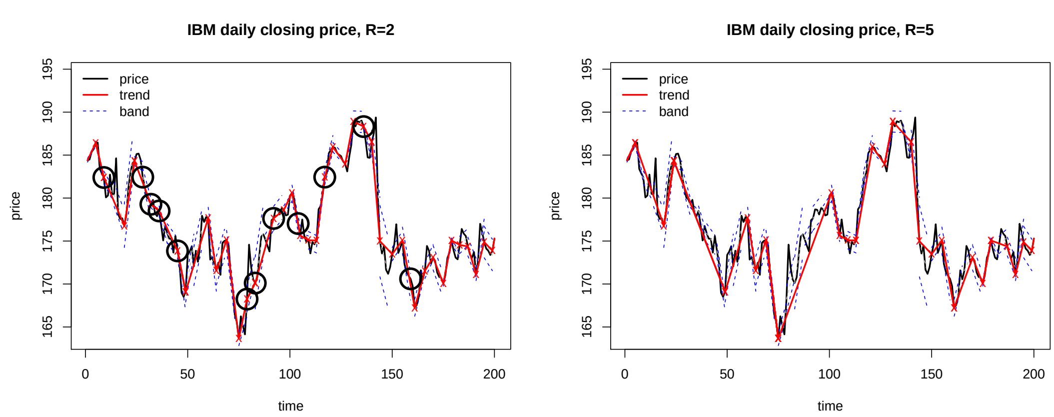

Figure 3: Data representation with the change ratio $R = 2$ (top) and $R = 5$ (bottom) - IBM daily closing price from March 27, 2012 to January 11, 2013

To determine parameters, we may use the *objective function* that consists of *loss* function and *regularization* to control the loss of information and roughness of the represented data.

$$\min \sum_{i=1}^{\infty} \sum_{k=1}^{n_i} \left\{ \frac{(x_{ik} - \hat{x}_{ik})^2}{n_i} + \lambda I(x_{ik} = x_{i1}) \right\} \tag{4}$$

where $\hat{x}_{ik}$ is the fitted value of k-th point and n_i is the number of data points in the i-th regime, and λ is a tuning parameter for regularization term. Note that unlike ATS, the length of trend regime in data reduced by PBS is always greater than the initial window size w.

2.2 Patterns

Once the original time series data is represented by linear lines, the subsequence of the local models can be *patterns* of interest. Let $\mu_i(t)$ be the i-th local model between $(i-1)$-th and i-th changepoints, then a particular sequence $\mu_i(t), \mu_{i+1}(t), \cdots, \mu_{i+r}(t)$ defines a pattern. There are two approaches to define a pattern. We may define patterns using a sequence of duplets $(s_1, l_1), (s_2, l_2) \cdots$, where s_i is the *linear trend* (slope value of the fitted line) and l_i is the length of the i-th trend regime respectively. Another way of defining a pattern on the reduced data is to us *symbolic* patterns. That is, transforming the real-values in the sequence of $(s_i, l_i),\ i = 1, 2, \cdots$ into the ordinal categorical values as seen in **Table 1** and **Table 2** based on their quantiles.

Table 1: Up and Down patterns with the lengths

$\mathbf{U_1}$ (C_1^+)	Up short length	$\mathbf{D_1}$ (C_1^-)	Down short length
$\mathbf{U_2}$ (C_2^+)	Up moderate length	$\mathbf{D_2}$ (C_2^-)	Down moderate length
$\mathbf{U_3}$ (C_3^+)	Up long length	$\mathbf{D_3}$ (C_3^-)	Down long length

Table 2: The magnitude of the trends

S (V_1)	Small in magnitude
M (V_2)	Medium in magnitude
L (V_2)	Large in magnitude

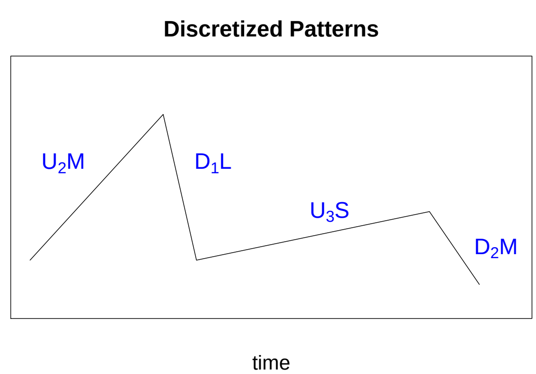

Figure 4: Discretized Pattern

The symbolic patterns also consist of two symbols: one for direction and length of trend, and the other for the magnitude of the trend. The magnitude of the trends are measured by the absolute value of the linear trends. Figure 4 illustrates an example of a sequence of symbolic patterns. $\mathbf{U_2M}$ and $\mathbf{D_1L}$ means "moderate length of increasing trend with medium magnitude" and "short length of decreasing trend with large magnitude" respectively.

2.3 Similarity Measures

2.3.1 Similarity Measures for numerical patterns

For the sequence of numerical patterns, that is a sequence of duplets that consists of the trends and the length of the trends (s_i, l_i), we transform the local functions to the piecewise constant functions, since the length of the trend l_i in a piece of numerical pattern (s_i, l_i) is actually a length between two successive changepoints c_{i-1} and c_i, each numerical pattern (s_i, l_i) can be transformed to a constant function,

$$f_i(t) = s_i I([c_{i-1}, c_i]),\ i = 1, 2, \cdots \tag{5}$$

In this fashion, the sequence of numerical patterns is transformed to a discontinuous constant function over time, and the distance of two time series can be measured from the different area between two discontinuous constant functions as follows:

$$\begin{aligned} \text{Distance}(T, Q) &= \left(\sum_{i=1}^{m+n} \int_{a_{i-1}}^{a_i} \left(f_T(t) - f_Q(t)\right)^2 dt \right)^{\frac{1}{2}} &\quad (6) \\ &= \left(\sum_{i=1}^{m+n} \left(f_T(a_i) - f_Q(a_i)\right)^2 (a_i - a_{i-1}) \right)^{\frac{1}{2}} &\quad (7) \end{aligned}$$

where f_T and f_Q are discontinuous constant functions of time series T and Q, and a_i's ($a_i \leq a_{i+1}$ for all i) are a sequence of combined changepoints of T and Q. Figure 5 illustrates the process of data representation, transformation to the numerical patterns and discontinuous constant functions, and calculating distance between two time series.

2.3.2 Similarity Measures for symbolic patterns

Distance measure for symbolic patterns can be calculated in a simpler way. As we categorize real-values in numerical patterns to a finite number of discrete values (symbols) according to the trend direction, the length of the trends, and the magnitude of the trends, we may assign integers with +/- signs on the symbols as seen in Figure 6. Thus, the distance between two sequences of symbolic patterns can be obtained by

$$\text{Discrete Pattern (DP) distance} = \sum_{i=1}^{m} \sqrt{(C_i^1 - C_i^2)^2 + (V_i^1 - V_i^2)^2} \tag{8}$$

where C_i's and V_i's are symbols in the symbolic pattern (C_i, V_i). Note that to measure *discrete pattern distance*, the length of two sequence of symbolic patterns must be the same since it is Euclidean distance base. One simple idea to meet this condition might to merge two adjacent trend regimes of the longer sequence based on some criteria, for example, the minimum increase in errors when merged.

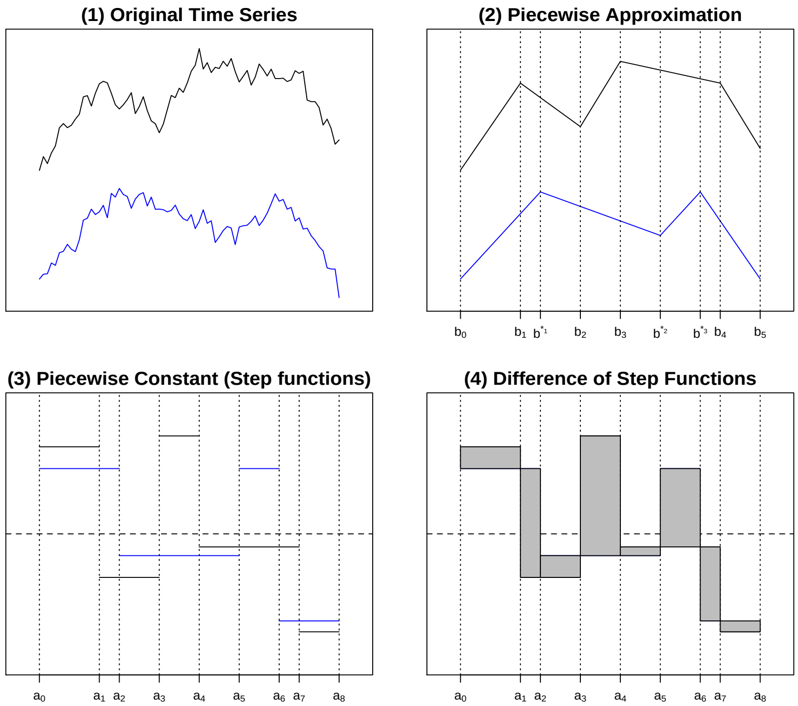

Figure 5: (1) Two raw data (2) Data representation by ATS or PBS (3) Transforming numerical pattern sequence to discontinuous constant functions (4) Distance measure by integration of difference between two constant functions

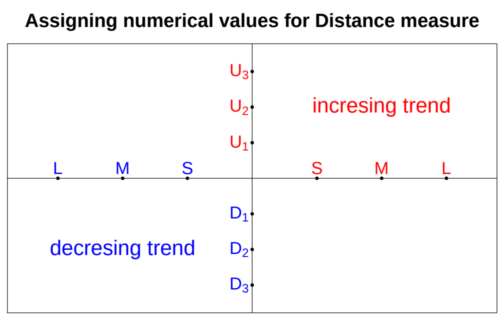

Figure 6: Distances for discretized patterns

3. Application Example

In this section, we demonstrate an example of classification for time series data using PBS and distance measure for numerical patterns. The purpose of this application example is not to evaluate the performance, but to illustrate PBS for classification and see if it would detect similar patterns in time series data.

- **K-Nearest Neighbors** (**K-NN**) - The K-nearest neighbors classifier uses K-nearest classified observations of a new observation x_0 to predict x_0's class. It assigns the most frequent class of K-nearest observations from x_0. We used $K = 1$ which is most widely used in time series data mining (Esling et. al, 2012) and distance measure for numerical pattern sequence in (6) and (7).

- **Datasets** - We collected four groups of stock price data from January 13, 2012 to January 12, 2015 based on their "visually similar" pattern determined by local features at specific time periods (Figure 7).
- **Choice of parameters** - Each time series is smoothed by *piecewise band smoothing* (PBS) with *change ratio*, $R = 3.5$. The *initial widow size* (w) and *bandwidth* (B) are selected from $S = \{w \mid 4 \leq w \leq 15,\ w \text{ is integer}\}$ and $B \in \{0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ which minimize the mean of squared error sum in (4).

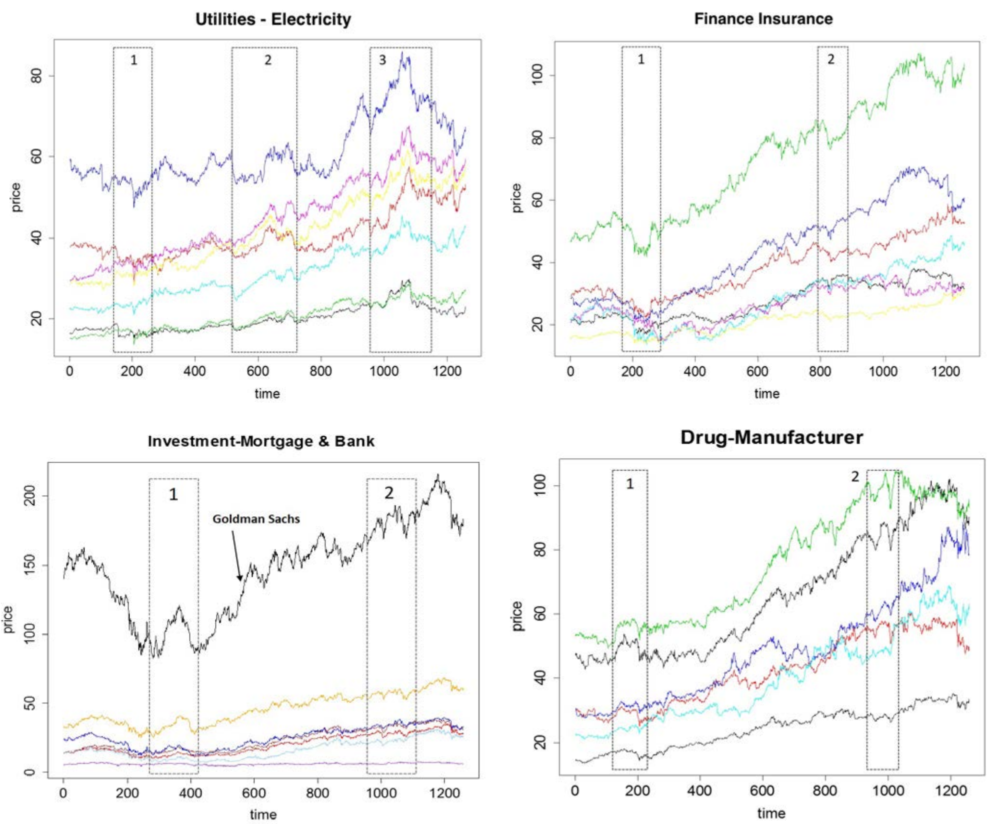

Figure 7: Four classes of stock prices from January 13, 2012 to January 12, 2015. They are classified based on visual similarity for specific time periods.

The results are shown in Table 3, Table 4, and Table 5. Table 3 is the result of 1-NN classification using raw data and Euclidean distance. Euclidean distance is rarely used for large size time series data, however, it might be a straightforward reference to compare mining performance for relatively short sequences of datasets. Table 4 and Table 5 are the result of classification on raw data and normalized data respectively, using PBS and the integral distance in (6) and (7). It is noticeable that normalization of datasets considerably improved the performance. However, zero error rate in Table 5 does not imply the methods used are perfect since the sample size is relatively small, we need more experiments with larger size samples to verify its performance.

Table 3: The result of 1- NN classification for raw data by Euclidean Distance

		Predicted				row
		U	**F**	**M**	**D**	total
Actual	**U**	7	0	0	0	7
	F	0	4	3	0	7
	M	0	0	7	0	7
	D	0	2	0	4	6
		total				**27**
Error Rate = 0.19						

Table 4: The result of 1-NN classification for raw data using PBS and the integral distance

		Predicted				row
		U	**F**	**M**	**D**	total
Actual	**U**	6	0	1	0	7
	F	1	3	3	0	7
	M	1	0	6	0	7
	D	0	0	1	5	6
		total				**27**
Error Rate = 0.26						

Table 5: The result of 1-NN classification for normalized data using PBS and the integral distance

		Predicted				row
		U	**F**	**M**	**D**	total
Actual	**U**	7	0	0	0	7
	F	0	7	0	0	7
	M	0	0	7	0	7
	D	0	0	0	6	6
		total				**27**
Error Rate = 0.00						

4. Conclusions and Future Work

In this research, we developed two new time series data representation methods, *alternating trends smoothing* (ATS) and *piecewise band smoothing* (PBS). These methods assume that a large time series consists of linear data generating processes, and thus the represented data by ATS or PBS are continuous straight lines with various lengths. While ATS has an tuning parameter "step size", PBS has three parameters *initial window size* (w), *bandwidth* (B), and *change ratio* (R) which controls the roughness of the reduced data. These parameters can be determined by the objective function that has two components, loss function and regularization. The fitted lines in trend regimes are used to define patterns in the form of real-values or discrete values. The numerical patterns can be transformed to discontinuous constant functions using line slope values to measure distance by integration of the difference of two functions. Thus, the lengths of two numerical pattern sequences do not have to be the same while those of two symbolic patterns must be since it is measured by Euclidean distance.

There are some possible modifications and extensions in piecewise band smoothing. In the algorithm of PBS, w data points used to fit the current trend line cannot be examined to identify changepoints, therefore it might miss true changepoints in the case that there is actually a trend changepoint among these w points, and the length of a trend regime in the reduced data is always equal to or greater than initial window size w. It may result in forcing the linear data generating process to be always have at least a certain number of length regardless of the true size of data generating process. This issue can be resolved by having an additional process to search a deviated point (changepoint) among the points in the initial window. In this step, the rule for decision of bandwidth should be considered together.

Also, these methods can be extended for correlated and multivariate time series data for future research. Although we assume that a large time series comprises many independent data generating processes, generally time series data are more likely correlated. Extended research for multivariate time series is also very important since it will give insight into the relational properties between sets of changepoint and patterns among the times series datasets.

REFERENCES

NYSE daily stock price data. http://finance.yahoo.com, accessed April, 2016.

Patterns in technical analysis. http://www.forexblog.org, accessed April, 2016

Chung, F.-L., Fu, T.-C., Luk, R., and 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* (2001), pp. 17.

Esling, P., and Agon, C., Time-series data mining, *ACM Computing Surveys(CSUR)* 45, 1, (2012), pp. 12.

Gentle, J. E. Mining for patterns in financial time series. In *JSM 2012 Proceedings (Alexandria, VA, 2012)*, American Statistical Association, pp. 29782988.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.

Keogh, E., Chakrabarti, K., Pazzani, M., and Mehrotra, S. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems 3*, 3 (2001), pp. 263286.

Lin, J., Keogh, E., Lonardi, S., and Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* (2003), ACM, pp. 211.

Shatkay, H., and Zdonik, S. B. Approximate queries and representations for large data sequences. In *Data Engineering, 1996. Proceedings of the Twelfth International Conference* on (1996), IEEE, pp. 536545.

Wilson, J. S., "Trend detection and Pattern Recognition in Financial Time Series", dissertation, George Mason university, (2016)