Proceedings of 2011 International Conference on System Science and Engineering, Macau, China - June 2011

Financial Time Series Segmentation Based On Turning Points

Jiangling Yin, Yain-Whar Si, Zhiguo Gong
Faculty of Science and Technology
University of Macau
{ma96589, fstasp, fstzgg}@umac.mo

Abstract—Segments extracted from financial time series are widely used in trend analysis as well as in predicting future tendency of the price movement. Recent approaches for time series segmentation often rely on an arbitrary threshold value and segments are generated at only one level. In this paper, we propose a novel time series segmentation method based on Turning Points which are extracted from the maximum or minimum points of the time series. The proposed segmentation method generates segments at different levels of details and achieves satisfactory results in preserving higher number of trends compared to an existing segmentation approach.

***Keywords*—financial time series, turning points, segmentation, trends.**

I. INTRODUCTION

Recent researches [1-5] analyze the movement of a stock based on a selected number of points from the time series. These points include *Important Points (IPs), Perceptually Important Points (PIPs)* and *Turning Points (TPs).* IPs [5] are local minimum and maximum points in a time series. IPs can be calculated after minor fluctuations are discarded with a threshold 'R' which is determined by the compression rate. Perceptually important points (PIPs) are used in identification of frequently appearing technical (analysis) patterns [2]. PIPs usually contains a few noticeable points and they are used for technical pattern matching in stock market applications [2].

Segments are usually used for representing financial time series. One frequently used segmentation method is Piecewise Linear Approximation (PLA) [6-10]. Specifically, PLA has been applied for pattern matching [11, 12] and predicting the trading points [13] in the stock market. In predicting stock movement, financial analysts not only consider the trend identified by the curve but also take into account certain points on the time series data. A segmentation method based on IPs is proposed in [5]. In this approach, original financial time series can be represented by a group of IPs for compression. However, some segmentation methods discard the points after generating the segments [9].

In this paper we propose a novel approach to segment the financial time series based on Turning Points. The main contribution of our work is that we allow the segmentation of the time series at different levels of details. Such segmentation allows top-down analysis on the time series in which highly visible trends are first identified and then more detailed segments are used in later stages. In addition, the proposed method can achieves satisfactory results in preserving higher number of trends compared to an existing method. Furthermore, our method does not depend on any arbitrary threshold which is used in PLA and IPs approaches.

This rest of the paper is organized into 5 sections. A review of recent work on Turning points and segmentation methods is given in Section 2. Proposed segmentation method based on Turning points is given in Section 3. In section 4, we detail the experimental results obtained from testing with price data from Hong Kong Stock Market. In Section 5, we summarize our ideas and future work.

II. RELATED WORK

In this section, we review recent work on Turning Points (TPs) and segmentation methods for financial time series.

A. Turning Point

A common method in predicting the movement of a stock is to use local minimum and maximum points from the historical price data. These local maximum and minimum points are often called Turning Points (TPs) since they indicate the change in the trend of the stock during a period. TPs are usually located near the top and the bottom of the financial time series during a period [3].

TPs are widely used in technical analysis since they contain more information than other points. TPs represent the trend of the stock change and they can be used to identify the beginning or end of a transaction period. We denote a point a_i on a time series as a Turning Point:

1. if the stock price ends the increasing trend at a_i and starts a decreasing period.
2. if the stock price ends the decreasing trend at a_i and starts an increasing period.

From mathematical point of view, we can define the TPs to be the maximum or minimum points during a certain period. In Figure 1, TPs are depicted in filled circle.

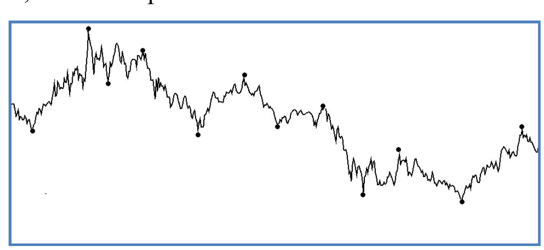


Figure 1. Turning Points on a time series

978-1-61284-471-8/11/$26.00 ©2011 IEEE

B. Segmentation Method of Financial Time Series

Segmentation is one of the important methods for preprocessing financial time series for further analysis. For instance, Piecewise Linear Approximation (PLA) has been applied for pattern matching [11, 12] and predicting the trading point [13]. In this section we review three common Piecewise methods for segmentation namely PLA, PIP, and IP.

PLA can be obtained through top down, bottom up, sliding window and B-spline wavelets approaches [9]. Segments are generated by applying certain actions such as partitioning (for top down) and merging (for bottom up) on the time series recursively until a predefined stopping criterion is met. The complexity of PLA is $O(n^2)$. One of the problems in PLA approach is the design of stopping condition (denoted as a threshold value [13]). The lengths of the segments extracted by PLA approach can increase when the higher threshold values are used. However, when PLA approach is applied to decompose historical data of different stocks, analysts may want to use different threshold values to match the underlying characteristics of the stocks. To alleviate this problem, several approaches are used to choose the right threshold value. For instance, in [13], Chang et al. employs GA algorithms to choose a valid threshold.

The perceptually important points (PIP) segmentation method is introduced in [2]. The process to identify the PIPs [2] is described in Algorithm 1. For a given the time series T, all the data points are processed by the PIP identification function. The first two PIPs will be the first and last points of T. The next PIP will be the point in T with the greatest distance to the first two PIPs. The process continues until all the points in T are added to a list (PIPList). To calculate the distance from the next PIP (x_2, y_2) to the two adjacent PIPs ((x_1, y_1) and (x_3, y_3), equation (1) is used to measure the distance.

$$\mathrm{Dis} = |\hat{y}_2 - y_2| = \left| y_1 + \frac{(y_3 - y_1) \times (x_2 - x_1)}{(x_3 - x_1)} - y_2 \right| \quad \mathbf{(1)}$$

ALGORITHM 1. PSEUDO CODE OF THE PIP IDENTIFICATION PROCESS FROM [2].

```
Function PIP_Identification (P)
   Input:   sequence P[1..m]
   Output:  PIPList L[1..m]
Begin
   Set L[1] = P[1], L[2] = P[m]
   Repeat until L[1..m] all filled
   Begin
     Select point P[j] with maximum distance to the
     adjacent points in PIPList (L[1] and L[2]
     initially)
     Append P[j] To L
   End
Return L
End
```

Important Points (IPs) [5] are local minimum and maximum points at which the change in the trend of the stock is bigger than a threshold R (the compression rate). For instance, to be a minimum IPs, the a_m should satisfy following formula [5]:

$$R \ge \left(\frac{a_i}{a_m}\right) and \left(\frac{a_j}{a_m}\right) \le R \text{ where } [i < m < j] \quad \mathbf{(2)}$$

According to [3], Turning Points (TPs) are usually situated near the top or bottom of the financial time series. Not all IPs extracted using equation (2) are considered as TPs since these points may not be near the top or bottom on the time series even although they indicate sharp changes in the trend.

Segmentation method based on Important Points is useful for compressing time series data since the principle of this method is to use a group of IPs to represent the original series [5]. The criterion for deciding a point to be important is based on the compression rate (R). Furthermore, the complexity of segmentation method for IPs is $O(n)$ which is much faster than PLA. In this method, a point in a time series is selected for producing a segment if the resulted R is sufficiently large. Otherwise the point will be discarded. Similar to PLA approach, a larger R value may create a small number of long segments and a smaller R may create a large number of short segments. Therefore, the choice of R can significantly influence the result of any future analysis. In this paper, we propose a novel approach to segment financial time series based on Turning points. In this approach, segments can be generated with varying level of granularity.

III. SEGMENTATION METHOD BASED ON TURNING POINTS

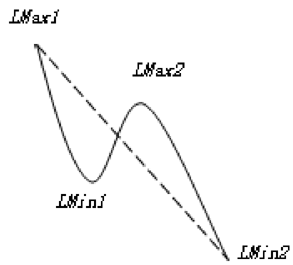
In this section, we describe a novel segmentation method based on Turning Points to decompose the time series at different levels of details. The key principle of our segmentation method is to properly divide the time series into different periods and to make sure that each segment extracted should have a single trend during that period. In our approach, Turning Points are identified iteratively to produce the corresponding segments. During the segmentation, some local Turning Points are discarded for combining small trends into big ones. Segmentation procedure is described as follows.

- First level: In this level, all local minimum and maximum points are marked based on following criteria. Let $f(x)$ be the height of a point x in a time series:
  - For point p_i and the neighboring points p_{i-1} *and* p_{i+1}, if $f(p_i) < f(p_{i-1})$ and $f(p_i) < f(p_{i+1})$ then p_i is a local minimum point.
  - For point p_i and the neighboring points p_{i-1} *and* p_{i+1}, if $f(p_i) > f(p_{i-1})$ and $f(p_i) > f(p_{i+1})$ then p_i is a local maximum point.
- Second level: In this level, some of the less significant points from previous step are filtered based on four strategies described in Figure 2. These points are filtered since they do not contribute to the shape of the overall trend of a segment.

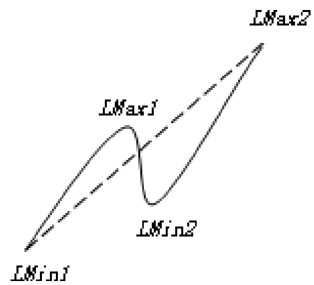
The algorithm for generating turning points is given in Algorithm 2. Above segmentation method can repeatedly performed to further reduce the level of details. For instance, in the third level, the steps from the second level can be again repeated to prune the remaining Turning Points. If further reduction in the level of detail is required, the steps from the second level can be again repeated to further reduce the points on the time series. Such iterative approach reduces the level of detail by progressively merging smaller trends into more significant trends.

FIGURE 2. STRATEGIES FOR ELIMINATING POINTS

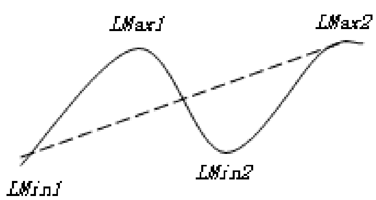
S1:
If *LMax1* > *LMax2* and *LMin1* > *LMin2* and (| *LMax2* – *LMin1*|< (|*LMax2* - *LMax1*|+ |*LMin1* - *LMin2*|)
Then
Eliminate *LMin1* and *LMax2*



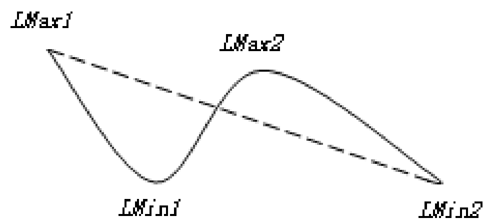
S2:
If LMin1 < LMin2 and LMax1 < LMax2 and (| LMax2 – LMin1| < (|LMax2 - LMax1| + |LMin1 - LMin2|)
Then
Eliminate LMax1 and LMin2



S3:
If LMax1 ≈ LMax2 and LMin1 ≈ LMin2
Then
Eliminate LMax1 and LMin2



S4:
If LMax1 ≈ LMax2 and LMin1 ≈ LMin2
Then
Eliminate LMin1 and LMax2



ALGORITHM 2. ALGORITHM FOR GENERATING TURNING POINTS

```
TURNING-POINTS// Function for obtaining the 1st TPs
Input: stock price (P), date
Output: TP1
for each point i do
if (P_i < P_i+1 And P_i < P_i-1) or (P_i > P_i+1 And P_i > P_i-1) //is Local-Minimum or Maximum
  Output(i)
return

TURNING-POINTS// Function for obtaining the higher level TPs
Input: Lower-Level TP index
Output: Higher-Level TP index
for each index i in Lower-Level TP do
if CONTAINS-POINT-IN-DOWNTREND(i) //refer S1 in Table 2
 Or CONTAINS-POINT-IN-UPTREND(i) // refer S2 in Table 2
 Or POINTS-IN-SAME-TREND(i)     // refer S3,4 in Table 2
  Output(i,i+3); i=i+3; //i+3 is the 3rd index in Lower-Level TPs after i
Else
  Output(i); i=i+1;    //i+1 is the next index in Lower-Level TPs after i
return

CONTAINS-POINT-IN-UPTREND(i)
if P_i < P_i+1 And P_i < P_i+2 And P_i+1 < P_i+3
 And P_i+2 < P_i+3 And |P_i+1 - P_i+2| < |P_i - P_i+2| + |P_i+1 - P_i+3|
  return true
else
  return false
```

```
CONTAINS-POINT-IN-DOWNTREND(i)
if Pi > Pi+1 And Pi > Pi+2 And Pi+1 > Pi+3
 And Pi+2 > Pi+3 And |Pi+2 - Pi+1| < |Pi - Pi+2| + |Pi+1 - Pi+3|
   return true
else
  return false

POINTS-IN-SAME-TREND(i)
if Pi ≈ Pi+2 And Pi+1 ≈ Pi+3
   return true
else
  return false
```

We extract a part of a time series to demonstrate how the algorithm works in Figure 3. For overall illustration, we select one of the stocks (HSBC 0005.HK) from Hong Kong Stock Market. The sample data is extracted from 2007-05-30 to 2009-12-30. In Figure 4, we compare the generated time series based on Turning Points at different levels of details. The compression ratio based on the number of points used at each level is summarized in Table 1.

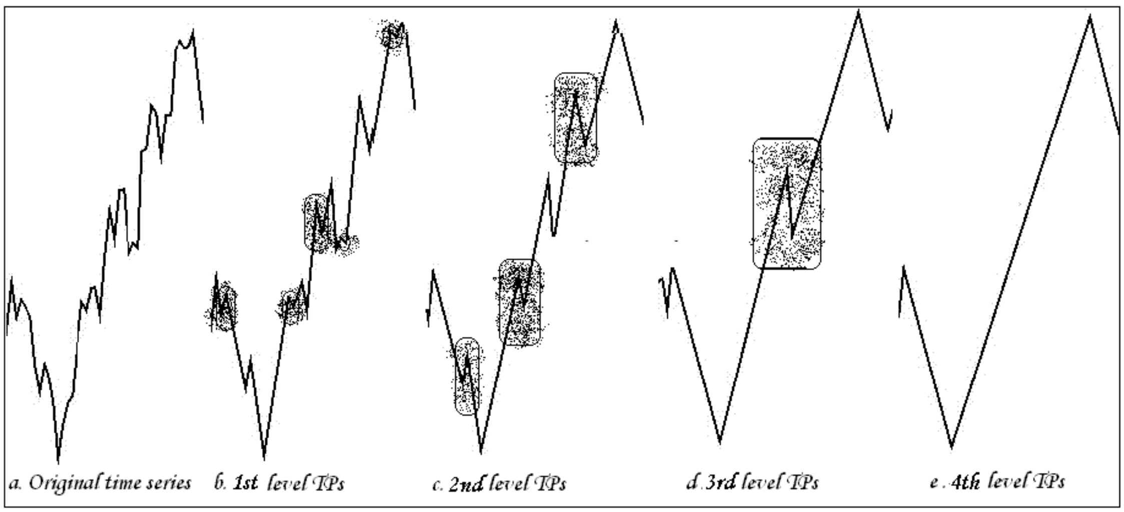


Figure. 3. (a) Original price data, (b) TPs after first level, (c) TPs after second level, (d) TPs after third level, and (e) TPs after fourth level. TPs in shaded areas are discarded.

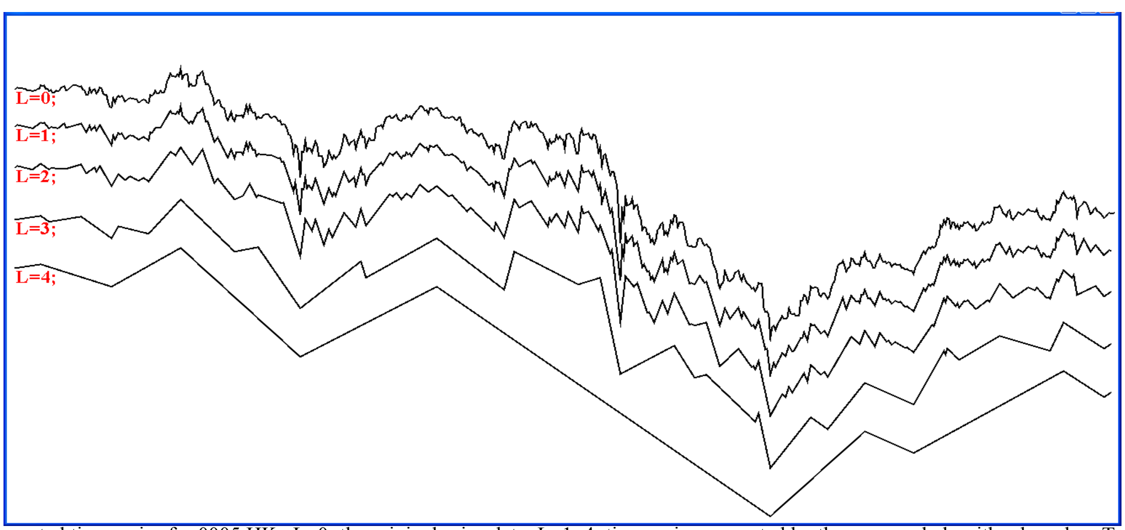


Figure. 4. Generated time series for 0005.HK. L=0: the original price data; L=1..4: time series generated by the proposed algorithm based on Turning Points

TABLE 1. THE NUMBER OF POINTS USED AT EACH LEVEL

Time Series	No. of points	Compression Ratio
Original	653	100
First level	311	48
Second level	164	25
Third level	38	5.8
Fourth level	12	1.8

IV. EXPERIMENTAL RESULTS

The experiment data set is taken from Hong Kong stock market. In our experiment, we evaluate curve fitting and trend preservation performance of the proposed algorithm in comparison with PIP. We choose PIP for comparison because PIP and TP are based on points selected from the original time series. Due to the space limitation, we select Hong Kong Hang Seng Index (HSI) from 2000-01-01 to 2010-5-30 for evaluation. In Figure 5, we plot the time series at different levels of details by using our approach and PIP.

We compare the error at different levels in Table 2. Error is defined as the sum of distance for every point between the original time series and the series generated by TP or PIP approach. The distance is calculated by using equation (3). In Table 3, we compare the number of up trends and down trends preserved by our approach and PIP approach at different levels.

$$\text{error} = \sum_{i=1}^{N} \left| \hat{y}_i - y_i \right| \quad \mathbf{(3)}$$

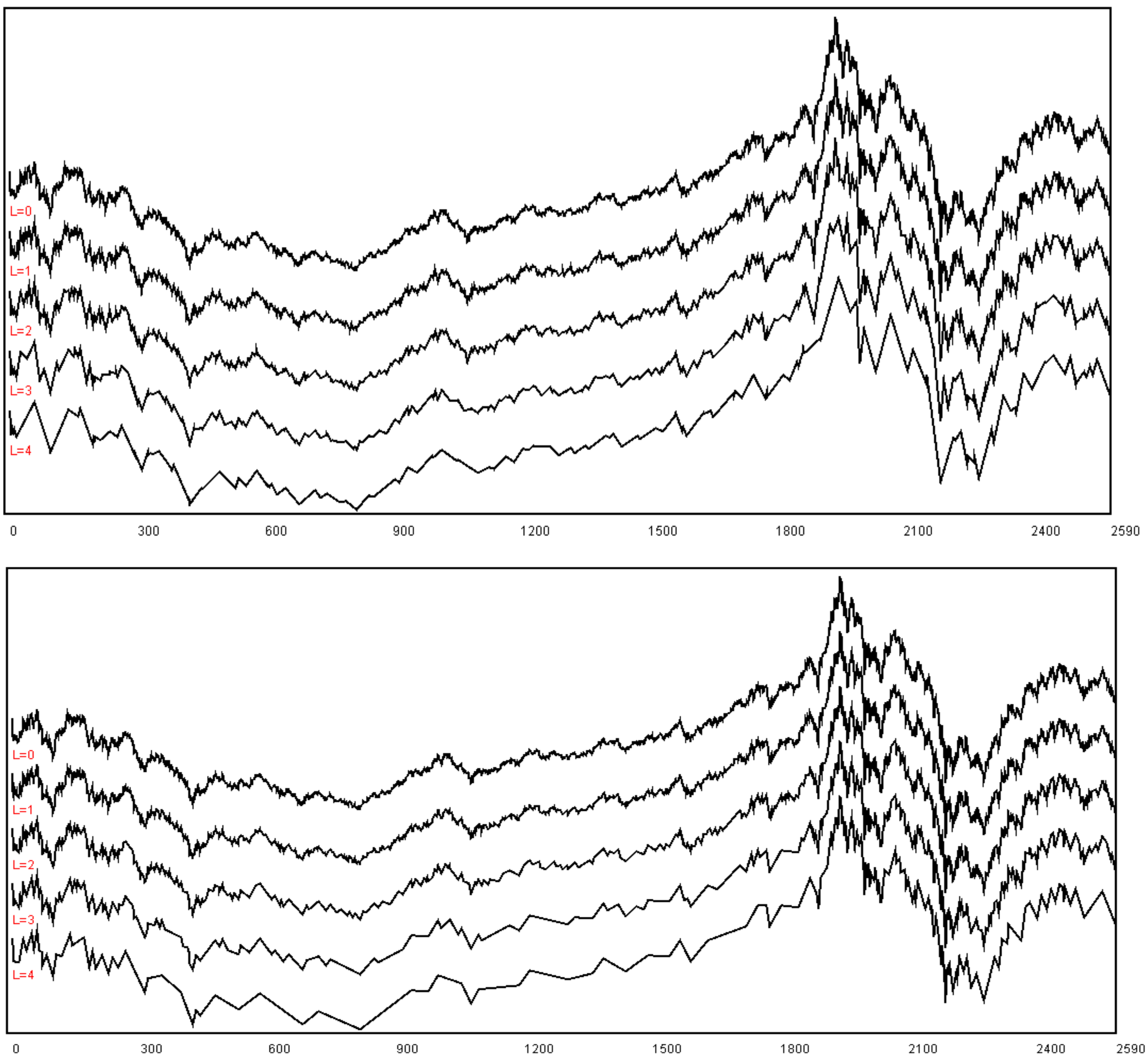


Figure 5. Time series generated by the proposed algorithm based on TP (top) and PIP approach (bottom) for HSI.

TABLE 2. ERROR COMPARISON AT EACH LEVEL

Level	No. of points selected	Error produced by TP	Error produced by PIP
Original	2590	0	0
1st level	1284	117	57
2nd level	611	228	184
3rd level	310	371	343
4th level	166	625	551

TABLE 3. TRENDS PRESERVATION COMPARISON AT EACH LEVEL

Level	No. of points selected	PIP-uptrend	TP-uptrend	PIP-downtrend	TP-downtrend
1st level	1284	419	636	420	636
2nd level	611	235	304	235	304
3rd level	310	129	154	128	154
4th level	166	70	82	69	82

By comparing the errors and trends produced by the two approaches, we find that PIP can produce less error than TP, but TP can preserve more trends than PIP. These results make sense since PIP approach is designed to keep the overall shape of the time series whereas the objective of our approach is designed to extract as many trends as possible from the time series. Financial analysts may use the extracted trends for building patterns for predicting stock movement.

In addition, TPs proposed in this paper are more sensitive to the movement of the stock than PIPs since PIPs tend to maintain the shape of the curve while TPs can be used to identify the change in the trend of the curve.

V. CONCLUSION

In this paper, we propose a novel segmentation method based on Turning Points. The time series generated by the proposed methods maintain the shape of original trends. Furthermore, our proposed algorithm can generate segments at different levels of details and preserve higher number of trends compared to an existing approach. Such capability is extremely useful for analyzing stock data in top-down fashion. As for the future work, we are planning to use the generated segments for predicting stock movement directions. We are also planning to extract similar segments from the historical price data for training a neural network model.

REFERENCES

[1] F.-l. Chung, T.-c. Fu, R. Luk, and V. Ng, “Evolutionary time series segmentation for stock data mining,” in Proceedings of the 2002 IEEE International Conference on Data Mining, 2002, pp. 83-90.

[2] T.-c. Fu, F.-l. Chung, R. Luk, and C.-m. Ng, “Representing financial time series based on data point importance,” *Eng. Appl. Artif. Intell.,* vol. 21, no. 2, pp. 277-300, 2008.

[3] D. Bao, and Z. Yang, “Intelligent stock trading system by turning point confirming and probabilistic reasoning,” *Expert Systems with Applications,* vol. 34, no. 1, pp. 620-627, 2008.

[4] D. Bao, “A generalized model for financial time series representation and prediction,” *Applied Intelligence,* vol. 29, no. 1, pp. 1-11, 2008.

[5] K. B. Pratt, “Locating Patterns In Discrete Time-Series,” Master Thesis, Department of Computer Science and Engineering, University of South Florida, Florida, 2001.

[6] E. J. Keogh, and M. J. Pazzani, “Relevance feedback retrieval of time series data,” in Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, California, United States, 1999, pp. 183-190.

[7] C.-S. Li, P. S. Yu, and V. Castelli, “MALM: a framework for mining sequence database at multiple abstraction levels,” in Proceedings of the 7th International Conference on Information and knowledge Management, Bethesda, Maryland, United States, 1998, pp. 267-272.

[8] P. Sanghyun, “Efficient searches for similar subsequences of different lengths in sequence databases,” in Proceedings of the 16th International Conference on Data Engineering, San Diego, CA, United States, 2000, pp. 23-23.

[9] E. J. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, United States, 2001, pp. 289 -296.

[10] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting time series: a survey and novel approach", in *Data Mining in Time Series Databases*, M. Last, A. Kandel, and H. Bunke Eds, World Scientific , pp. 1-22, 1993.

[11] Z. Zhang, J. Jiang, X. Liu, W. C. Lau, H. Wang, S. Wang, X. Song, and D. Xu, "Pattern recognition in stock data based on a new segmentation algorithm," in Lecture Notes in Computer Science, pp. 520-525, Springer Berlin / Heidelberg, 2010.

[12] H. Wu, B. Salzberg, and D. Zhang, “Online event-driven subsequence matching over financial data streams,” in Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, Paris, France, 2004, pp. 23-34.

[13] P.-C. Chang, C.-Y. Fan, and C.-H. Liu, “Integrating a piecewise linear representation method and a neural network model for stock trading points prediction,” *Trans. Sys. Man Cyber Part C,* vol. 39, no. 1, pp. 80-92, 2009.