International Journal of Trade, Economics and Finance, Vol.1, No.4, December, 2010
2010-023X

Financial Data Representation and Similarity Model

Yongwei Ding, Xiaohu Yang, Alexsander J. Kavs and Juefeng Li

***Abstract*— Similarity metric is of fundamental importance for similarity matching and subsequence query in time series applications. Most existing approaches measure the similarity by calculating and aggregating the point-to-point distance, few of them take the segment trend duration into account. In this paper, upon analyzing the properties of financial time series, we define a time series notation which is more intuitive and expressive. Base on that, a new similarity model is proposed. Experiments on both real foreign currency exchange rate data and stock market data are performed. The result shows the effectiveness and good accuracy of our method. The similarity model is also proved to be segmentation algorithm independent thus can be combined with other segmentations for similarity query, pattern matching, classification, and clustering.**

***Index Terms*—financial data, piecewise linear representation, radian distance, segment duration, similarity, time series**

I. Introduction

In recent years, mining financial time series has been gaining more and more attentions. A time series is a sequence of real numbers, each number representing a value a time point [1]. Unlike transactional databases with discrete item, time series data are characterized by their numerical and continuous nature [2]. Time series are ubiquitous. In particular, daily indices of stock markets, foreign currency exchange rates, and daily net values of funds are typical examples of time series in financial domain. We are often more interested in finding or revealing the hidden pattern from time series rather than querying a value with given time point. For instance, we may want to find stocks that behave in approximately the same way (or the opposite way, for hedging); or stocks that rose for five consecutive days followed by a limit down opening the next day. In this type of queries, approximate matching is required. Before that, similarity or distance metric should be properly defined and measured.

Moreover, time series are often extremely huge, so the computation cost can be considerable large if we conduct the distance calculation on raw time series data directly. Thus, the time series data need to be compressed in a both effective and intuitive manner, enabling efficient indexing to support the distance calculation. For financial time series, two of the most representative characteristics are trend and duration. These is no difference between ordinary individual investors and professional investment institutions, they are essentially the same. What they care is the tendency of the market next, and how long would the current momentum last.

In this paper, we propose a new representation for time series, Radian-Duration-Time (RDT) based piecewise linear representation. Each segment can be represented as a 3-tuple with trend, duration and starting time encoded. More importantly, on basis of the RDT representation, Radian Distance is firstly introduced to measure the similarity for time series.

The rest of paper is organized as follows. In section 2, we discuss related work on distance measure and their drawbacks. Section 3 describes in detail the idea of RDT based time series representation. Our work is motivated by some other alternative distances as listed in section 4. The radian distance is defined and analyzed theoretically in section 5. Experimental results for financial dataset are presented in section 6. Section 7 provides the conclusions of this paper.

II. Related Work

In contrast to traditional database applications, users in financial domains typically search for patterns within the data that fit some approximate notion of what they want. They are not interested in the exact values in the time series as much as the overall shape of some subsequences [3]. Searching for the similarity or dissimilarity between sequences is normally measured by distance metric. The most famous measure in the area of time series is Euclidean distance. For the point-to-point similarity of two series, Euclidean Distance is widely adopted as the similarity distance function due to its simplicity. However, the Euclidean distance requires the time series to be equal interval, although linear interpolation can help achieve this after segmentation, but the computation cost for Euclidean distance calculation would be extraordinarily high if the time series are very long. Besides the potential poor performance, Euclidean distance has ambiguous concepts in certain scenario. To address this, authors in [4] introduce the pattern model presentation and series pattern distance (SPD) to measure the trend similarity of time series. SPD overcomes the problem of time series based on point distance. A shape-based discrete symbolic representation and

Manuscript received May 1, 2010. This work is supported by State Street Corporation.

Yongwei Ding is with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027 P.R. China (phone: 86-135-8803-7210, e-mail: ywding@zju.edu.cn).

Xiaohu Yang is with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027 P.R. China (e-mail: yangxh@zju.edu.cn).

Alexsander J. Kavs is with the State Street Corporation, Boston, MA 02111 USA (e-mail: ajkavs@statestreet.com)

Juefeng Li is with the State Street Technology (Zhejiang), Hangzhou, Zhejiang 310030 P.R. China (e-mail: lijuefeng@statestreet.com).

its corresponding distance measure was proposed in [5] by extending the concept of SPD. Both techniques allow fast calculation compare with the Euclidean distance, but they might probably fail to work at times as they cannot retain all major shape features by using finite and coarse grained divisions of the segment trend. An alternative definition of similarity based the included angle distance is proposed in [6], two sequences are similar if their distance is less than a predefined threshold. Although the method is invariant to translation and rotation, it considers the two nonparallel time series (e.g., one is a monotonic increasing straight line while the other is monotonic decreasing) as whole matching which are evidently not. The reason is that the segment represented by included angle does not contain any sub trend information.

The work presented in this paper lays in the intersection of two domains, namely, approximation and similarity measure. The basic idea of our method can be characterized procedurally in two steps, (i) reduce the dimensionality of raw time series data from *n* to **K** and represent the compressed data with a set of RDT based segments, the PLR_CRE segmentation algorithm that we first proposed in [7] would be adopted; (ii) align the turning points which are identified during segmentation, calculate the point-to-point similarity of two series, and aggregate these measures.

III. RADIAN-DURATION-TIME BASED REPRESENTATION OF TIME SERIES

A time series is a set of observations x_t, each one being recorded at a specified time *t* [8]. As with most computer science problems, representation of data is the key to efficient and effective solutions [9] if we want to further analyze or mine the data. The Radian-Duration-Time based representation for time series that we first present will be elaborated in this section.

A. Notation

For clarity we will refer to 'raw', unprocessed time series data of length n as time series, denoted as $\boldsymbol{S}$, and the compressed, piecewise linear representation of a time series as a sequence, $\boldsymbol{S'}$, length for K. We will use these notations throughout the paper.

B. RDT Representation of Time Series

Unlike pattern model representation, shape-based representation, and other common alternatives in research literature, more segment shape features could be retained by means of the 3-tuple property set of each segment, which are radian, duration, and starting time.

Definition 1: Segment radian is the radian of the included angle which is created by the segment (the straight line connecting two adjacent points) and the time axis. Fig. 1 illustrates this notation.

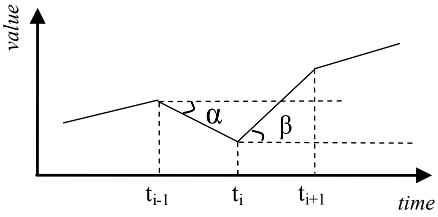

Figure 1. Segment radian of time series

The i[th] segment radian is calculated by the following equation:

$$rad_i = \arctan[(x_{i+1} - x_i)/\Delta t],\ i \in [1, n-1] \tag{1}$$

Theoretically, $rad_i \in (-\frac{\pi}{2}, \frac{\pi}{2})$, β in Fig. 1 is the radian for ith segment, β>0. Normally, the time series value would be normalized by mapping entries of the series from their natural range to values between 0 and 1, then $\Delta x \in [-1, 1]$. Generally, $\Delta t = 1$ if we assume the time series S is indexed by the natural numbers, therefore, $rad_i \in [-\frac{\pi}{4}, \frac{\pi}{4}]$.

With the PLR_CRE segmentation method we first proposed in [7], we can get the piecewise linear representation of the original time series as shown in (2),

$$S' = \{(rad_1, t_1), ..., (rad_i, t_i), ..., (rad_K, t_K)\} \tag{2}$$

Apparently, Δt of i[th] segment in (1) denotes the duration of current segment trend. Thus, we have the representation modified to (3).

$$S' \equiv \{R, D, T\} \tag{3}$$

$\boldsymbol{S'}$, containing **K** linear segments, is a 3-tuple of vectors of length **K**. The i[th] element of sequence S' is represented by the line starting from T_i with segment radian R_i, ending after D_i time interval, denoted as $s_i = (R_i, D_i, T_i)$. Fig. 2 depicts this notation.

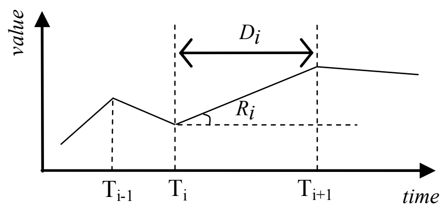

Figure 2. RDT based representation of segment

IV. DISTANCES OF TIME SERIES

In this section, the aforementioned three typical point-to-point similarity measures, Euclidean distance, pattern distance, shape distance, and included angle distance will be revisited.

A. Euclidean Distance (EucDist)

For any two points in 2-dimension space, $p=(x_1, y_1)$, and $q=(x_2, y_2)$, the Euclidean distance is defined as follows:

$$EucDist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

However, for time series data of length n, the Euclidean distance between two series $P=(p_1, p_2, ..., p_n)$, and $Q=(q_1, q_2, ..., q_n)$ is defined by the following equation:

$$EucDist = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

B. Pattern Distance (PatDist)

The time series mode m represents the single variation trend of a subsequence s. The mode is a ternary set {uptrend, sideway, downtrend}, denoted as M= {1, 0, -1}. $m \in M$. Pattern distance is the distance between two modes which are of equal duration.

$$PatDist(s_i, s_j) = |m_i - m_j|$$

Since the trend duration for each mode may vary, an Equal Pattern Number (EPN) process is required to align the starting and ending time of mode between sequence pair. Projection method is in common use for EPN.

Time series pattern distance indicates the degree of difference in terms of trend between two equal-length series.

$$PatDist(S_1, S_2) = \left[\sum_{i=1}^{N} \left(t_{ih} \times PatDist(s_{1i}, s_{2i})\right)\right] / t_N$$

C. Shape Distance (ShapeDist)

Authors in [5] proposed this method by extending the Pattern Distance. In fact, both distances share the same idea. It describes the shape of each segment with a 7-tuple mode. By importing the factor named "action strength change", A_{ih}, it improves the accuracy of similarity measurement and eliminates the influence of noises.

$$ShapeDist(M_1, M_2) = \left[\sum_{i=1}^{n} \left(t_{ih} \times |A_{1ih} - A_{2ih}| \times |M_{1i} - M_{2i}|\right)\right] / L$$

D. Included Angle Distance (IaDist)

Included angle distance of time series is the degree of difference of two EPN processed equal-length sequences which are represented by included angle set. Here the included angle, denoted as α, is constructed by two adjacent segments.

$$IaDist(S_1, S_2) = \left(\sum_{i=1}^{n-1} |\alpha_{1i} - \alpha_{2i}|\right) / [\pi(n-1)] \in [0,1]$$

V. RADIAN DISTANCE OF TIME SERIES

The goal of this work is to devise a method that effiectively measures the similarity of any two given sequences with higher accuracy and less computation cost. This section formally defines our similarity model and describes how to calculate the distance by extending the proven online piecewise linear segmentation algorithm, PLR_CRE.

A. Piecewise Linear Segmentation

Time series segmentation has gained much attention in last decade. Piecewise Linear Representation (PLR) [10][11][12][13] is preferred among various segmentation methods because of its great simplicity and intuitiveness. The basic idea is, firstly, identify the "key points" which certain behavior changes occur in time series, then, approximate the data by fitting the segments between successive "key points". Simply stated, the original data of length n is divided into K segments. As a dimensionality reduction technique, PLR makes the storage, transmission and computation of the data more efficient [9]. The segmentation method PLR_CRE that we first propose in [7] employs a novel online turning-point (TP) identification based on the cumulative radian error of data point. It's been proven to be an effective and efficient segmentation algorithm and can retain most of the local trend features while attaining high compression ratio in a fast and scalable manner. The PLR_CRE approach is extended in this paper to achieve the similarity measurement.

B. Turning point alignment

Given two equal-length time series S_1 and S_2, it is possible that the length of sequences S_1' and S_2', segmented with PLR_CRE, might differ from one to the other, meanwhile, each segment (or sub-trend) duration can in arbitrary length. In order to measure the point-to-point distance, a mutual projection method is used to align the sequences by projecting the turning points between S_1' and S_2'.

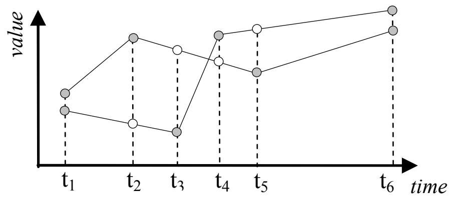

Figure 3. Turning points alignment

As illustrated in Fig. 3, the solid spots on each sequence represent turning points while the hollow circles represent Peering Points (PP) which are projected from corresponding turning points on the other sequence. The radian of any peering point equals to the radian of its last previous turning point. The aligned sequence is denoted as $\widetilde{S}$.

The pseudo code of the alignment algorithm, Piecewise Linear Turning Points Alignment (PLTPA) algorithm, is outlined in Fig. 4. Before conducting the PLTPA, we need to apply the PLR_CRE segmentation algorithm to the target time series.

$\widetilde{S}_1 = S_1', \widetilde{S}_2 = S_2'$
$last\_tp\_idx\_in\_S_1 = 1$
$last\_tp\_idx\_in\_S_2 = 1$
$for\; i = 2 : K$
$\quad if\; x_{1i}\; is\; TP\; and\; x_{2i}\; is\; not\; TP$
$\quad\quad x_{2i}.ppRad = x_{2(last\_tp\_idx\_in\_S_2)}.ppRad$
$\quad\quad\quad = x_{2(last\_tp\_idx\_in\_S_2)}.rad$
$\quad\quad \widetilde{S}_2 \cup \{(x_{2i}, t_i)\}$
$\quad\quad last\_tp\_idx\_in\_S_1 = i$
$\quad end\; if$
$\quad else\; if\; x_{1i}\; is\; not\; TP\; and\; x_{2i}\; is\; TP$
$\quad\quad x_{1i}.ppRad = x_{1(last\_tp\_idx\_in\_S_1)}.ppRad$
$\quad\quad\quad = x_{1(last\_tp\_idx\_in\_S_1)}.rad$
$\quad\quad \widetilde{S}_1 \cup \{(x_{1i}, t_i)\}$
$\quad\quad last\_tp\_idx\_in\_S_2 = i$
$\quad end\; if$
end

Figure 4. Piecewise Linear Turning Points Alignment (PLTPA) algorithm

C. Radian Distance (RadDist) of Time Series

As defined in (3), the formal piecewise linear representation of time series is a 3-tuple of vectors. For simplicity, we denoted the aligned sequences as shown in (4) as the segment radian play the key role in our similarity model, this is not a requirement of our approach, however it does simplify notation.

$$\widetilde{S}_1 = \{rad_1', rad_2' \ldots, rad_i', \ldots, rad_M'\},$$
$$\widetilde{S}_2 = \{rad_1'', rad_2'' \ldots, rad_i'', \ldots, rad_M''\} \qquad (4)$$

Let M be the dimensionality of the aligned sequence we wish to index (K≤M≤n). Thus, the radian distance of corresponding subsequences or segments in $\widetilde{S}_1$ and $\widetilde{S}_2$ is calculated as follows,

$$RadDist(s_i', s_i'') = \left| rad_i' - rad_i'' \right| \qquad (5)$$

The radian distance of original series can be obtained by calculating the *RadDist* of $\widetilde{S}_1$ and $\widetilde{S}_2$ which are the approximation of original series in (6). By aggregating the individual measures, we can get the overall distance measure.

$$RadDist(S_1, S_2) \approx RadDist(\widetilde{S}_1, \widetilde{S}_2)$$
$$= \left[af \times \sum_{i=1}^{M} \left(\Delta t_{rad_i} \times \left| rad_i' - rad_i'' \right| \right) \right] / (\pi \times t_M) \qquad (6)$$

Obviously, segments in different length should not be equal-weighted in distance measure. Long-duration segment has more impact to the whole shape of sequence than short-duration segment. To reflect this, in our model, the segment duration is taken into account, working as a weight either amplify or diminish the individual distance. Δt_{rad_i} is the trend duration of ith segment, namely, D_i in Fig. 2. Here *af* is another factor used to amplify the value of similarity, it's proportional to the size of time series.

According to the radian distance calculation formula, we know that the smaller the *RadDist* value is, the more similar the two series' shapes are. Radian distance can be used for time series whole matching and subsequence matching, by modifying the turning point alignment rule, the proposed similarity model can easily support time series scaling and measure the similarity of non-equal length sequences.

VI. Experimental Results

The purpose of the experiment performed in this section is to test the validity of the proposed similarity mode based on radian distance and to compare with aforementioned types of distance in section 3. It is carried out on real financial datasets which have more than 10 years' trading records.

To prove the effectiveness of our method and perform the comparison with other existing approaches, two different financial datasets are employed. One is the foreign currency exchange rate data [1] (daily exchange rate of eight international currencies that are expressed as number of units of the foreign currency per US dollar). The other is the stock market data[2] (daily indices of eight major stock markets worldwide). 4 currencies (Dutch guilder, German deutschmark, Swiss franc, and Japanese yen) are selected from the first dataset while 4 indices (FTSE100, SPCOMP, HNGKNGI, and FRCAC40) are picked up from the second stock market dataset. The selected date ranges are both last 2500 trade dates. PLR_CRE algorithm is used to compress the raw data to 100 data points.

The experimental results are summarized in table 3 and table 4. Table 1 shows the Euclidean Distance based benchmark for the uncompressed Foreign Exchange Rate data while table 2 gives the benchmark for Stock Market data. As shown in table 3, the recognition accuracy of our model is as good as Euclidean distance but with less computation cost. Although pattern distance, shape distance, and included angle distance succeed in finding the most similar pair, neither of them can identify the most dissimilar one accurately. Table 4 shows that, the pattern distance, shape distance, and included angle distance can determine neither the most similar nor the dissimilar pairs. The result of similarity measured by radian distance is exactly the same as Euclidean distance, even sorted by the distance value for Foreign Exchange Rate data.

TABLE I. DISTANCE BENCHMARK FOR FOREIGN EXCHANGE RATE DATA (COMPRESSION RATIO = 0%)

Distance	Series Pair					
	(S1,S2)	*(S1,S3)*	*(S1,S4)*	*(S2,S3)*	*(S2,S4)*	*(S3,S4)*
EucDist	**0.334**	3.861	10.323	3.886	**10.513**	9.943

S_1: Dutch guilder, S_2: German deutschmark, S_3: Swiss franc, S_4: Japanese yen

TABLE II. DISTANCE BENCHMARK FOR STOCK MARKET DATA (COMPRESSION RATIO = 0%)

Distance	Series Pair					
	(S1,S2)	*(S1,S3)*	*(S1,S4)*	*(S2,S3)*	*(S2,S4)*	*(S3,S4)*

[1] http://robjhyndman.com/TSDL/data/FVD2.dat
[2] http://robjhyndman.com/TSDL/data/FVD1.dat

Distance	Series Pair					
	(S_1,S_2)	*(S_1,S_3)*	*(S_1,S_4)*	*(S_2,S_3)*	*(S_2,S_4)*	*(S_3,S_4)*
EucDist	**2.913**	4.440	6.071	5.719	7.430	**8.633**

S_1: FTSE100, S_2: SPCOMP, S_3: HNGKNGI, S_4: FRCAC40

TABLE III. DISTANCE COMPARISON FOR FOREIGN EXCHANGE RATE DATA (COMPRESSION RATIO = 96%, TI = 1.0001)

Distance	Series Pair					
	(S_1,S_2)	*(S_1,S_3)*	*(S_1,S_4)*	*(S_2,S_3)*	*(S_2,S_4)*	*(S_3,S_4)*
PatDist	**0.014**	0.094	0.158	0.084	0.234	**0.243**
ShapeDist	**0.081**	0.152	**0.234**	0.219	0.196	0.200
IaDist	**0.004**	0.008	0.241	0.007	0.243	**0.248**
RadDist	**0.650**	1.079	1.450	1.283	**1.540**	1.401

S_1: Dutch guilder, S_2: German deutschmark, S_3: Swiss franc, S_4: Japanese yen

TABLE IV. DISTANCE COMPARISON FOR STOCK MARKET DATA (COMPRESSION RATIO = 96%, TI=1.0001)

Distance	Series Pair					
	(S_1,S_2)	*(S_1,S_3)*	*(S_1,S_4)*	*(S_2,S_3)*	*(S_2,S_4)*	*(S_3,S_4)*
PatDist	0.595	0.773	**0.188**	**0.788**	0.760	0.756
ShapeDist	0.099	**0.157**	0.141	**0.049**	0.129	0.149
IaDist	0.562	0.588	**0.489**	0.576	0.613	**0.635**
RadDist	**0.570**	0.721	0.978	0.654	1.082	**1.109**

S_1: FTSE100, S_2: SPCOMP, S_3: HNGKNGI, S_4: FRCAC40

In addition, to illustrate that the *RadDist* is segmentation algorithm independent, we conducted two more experiments by using PLR_PF [11] and PLR_PIP [13] to compress the time series respectively. Both PLR_PF and PLR_PIP are classical piecewise linear segmentation algorithms. Table 5 shows that, under same compression ration with PLR_PF, *RadDist* keeps good accuracy in identifying the most similar and dissimilar pairs. The same thing happens to PLR_PIP, as listed in table 6.

TABLE V. DISTANCE COMPARISON FOR STOCK MARKET DATA COMPRESSED BY PLR_PF (COMPRESSION RATIO = 96%)

Distance	Series Pair					
	(S_1,S_2)	*(S_1,S_3)*	*(S_1,S_4)*	*(S_2,S_3)*	*(S_2,S_4)*	*(S_3,S_4)*
PatDist	0.710	0.764	**0.543**	0.922	0.727	0.788
ShapeDist	0.047	0.083	0.069	**0.070**	0.086	0.084
IaDist	0.638	0.696	**0.646**	0.643	0.615	0.703
RadDist	**0.340**	0.463	0.575	0.418	0.739	**0.740**

S_1: FTSE100, S_2: SPCOMP, S_3: HNGKNGI, S_4: FRCAC40

TABLE VI. DISTANCE COMPARISON FOR STOCK MARKET DATA COMPRESSED BY PLR_PIP (COMPRESSION RATIO = 96%)

Distance	Series Pair					
	(S_1,S_2)	*(S_1,S_3)*	*(S_1,S_4)*	*(S_2,S_3)*	*(S_2,S_4)*	*(S_3,S_4)*
PatDist	0.299	0.552	**0.703**	0.530	0.631	0.936
ShapeDist	0.038	0.095	0.114	**0.056**	0.336	0.300
IaDist	0.588	0.691	**0.652**	0.572	0.558	0.687
RadDist	**0.273**	0.482	0.542	0.389	0.552	**0.659**

S_1: FTSE100, S_2: SPCOMP, S_3: HNGKNGI, S_4: FRCAC40

VII. CONCLUSION

In this paper, we introduce a new distance for similarity model based on the Radian-Duration-Time representation over financial time series. The RDT notation features good intuitiveness yet expresses the trend information of segment in a readable form. Recognition accuracy using the proposed measure is compared with similarity measures like Euclidean distance, pattern distance, shape distance, as well as included angle distance. Experimental results show that our approach has better performance over the rival measures. Apart from that, the proposed method is segmentation algorithm independent thus can be combined with other piecewise linear representations. Further research can proceed by incorporating indexing in similarity search and subsequence query.

REFERENCES

[1] Davood Rafiei, Alberto O. Mendelzon, "Similarity-based queries for time series," Proc. of ACM SIGMOD International Conference on Management of Data, May 1997, pp. 13-25.

[2] F. Chung, T. Fu, R. Luk, V. Ng, "Evolutionary time series segmentation for stock data mining," Proc. of the 2nd International Conference on Data Mining (ICDM 2002), IEEE Computer Society, Dec. 2002, pp. 83-90.

[3] Hagit Shatkay, Stanly B. Zdonik, "Approximate queries and representations for large data sequences," Proc. of the Twelfth International Conference on Data Engineering (ICDE1996), 26 Feb. -1 Mar. 1996, pp. 536-545.

[4] WANG Da, RONG Gang, "Pattern distance of time series," Journal of Zhejiang University, 2004, 38(7), pp. 795-798.

[5] DONG Xiaoli, GU Chengkui, WANG Zheng'ou, "Research on shape-based time series similarity measure," Journal of Electonics & Information Technology, 2007, 29(5).

[6] ZHANG Peng, LI Xueren, ZHANG Jianye, ZHANG Zonglin, "Included angle distance of time series and similarity search," Pattern Recognition and Artificial Intelligence, 2008, 21(6).

[7] Y. Ding, X. Yang, A. J. Kavs, and J. Li, "A novel piecewise linear segmentation for time series," Proc. of the 2nd International Conference on Computer and Automation Engineering (ICCAE 2010), Feb. 2010, pp. 52-55.

[8] Peter J. Brockwell and Richard A. Davis, Time Series: Theory and Methods, 2nd ed., Springer, 2009, pp. 1.

[9] E. Keogh, S. Chu, D. Hart, M. Pazzani, Segmentation Time Series: A Survery and Novel Approach, Data Mining in Time Series Databases, World Scientific Publishing Co., 2004, pp.1-22.

[10] V. Guralnik, J. Srivastava, "Event detection from time series data," Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 1999), ACM, Aug. 1999, pp. 33-42.

[11] K. Pratt, E. Fink, "Search for patterns in compressed time series," International Journal of Image and Graphics, World Scientific, 2000, 2(1), pp. 89-106.

[12] F. Chung, T. Fu, R. Luk, V. Ng, "Flexible time series pattern matching based on Perceptually Important Points," International Joint Conference on Artificial Intelligence (IJCAI 2001) Workshop on Learning from Temporal and Spatial Data, Aug. 2001, pp. 1-7.

[13] T. Fu, F. Chung, R. Luk, "Stock time series pattern matching: Template-based vs. rule-based approaching," Engineering Applications of Artificial Intelligence, 2007, 20(3), pp. 347-364.

Yongwei Ding, from P.R. China, received his B.S. degree from Zhejiang University in the year 2005 in the field of Software Engineering. He also received M.S. in Software Engineering from Zhejiang University in 2007. He is currently pursuing the Ph.D. degree in the College of Computer Science and Technology, Zhejiang University.
His area of research includes financial data processing and visual analysis.