Research Article

Piecewise Trend Approximation: A Ratio-Based Time Series Representation

Jingpei Dan,[1] Weiren Shi,[2] Fangyan Dong,[3] and Kaoru Hirota[3]

[1] College of Computer Science, Chongqing University, Chongqing 400044, China
[2] School of Automation, Chongqing University, Chongqing 400044, China
[3] Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259 Nagatsuta, Midoriku, Yokohama 226-8502, Japan

Correspondence should be addressed to Jingpei Dan; danjingpei@cqu.edu.cn

A time series representation, piecewise trend approximation (PTA), is proposed to improve efficiency of time series data mining in high dimensional large databases. PTA represents time series in concise form while retaining main trends in original time series; the dimensionality of original data is therefore reduced, and the key features are maintained. Different from the representations that based on original data space, PTA transforms original data space into the feature space of ratio between any two consecutive data points in original time series, of which sign and magnitude indicate changing direction and degree of local trend, respectively. Based on the ratio-based feature space, segmentation is performed such that each two conjoint segments have different trends, and then the piecewise segments are approximated by the ratios between the first and last points within the segments. To validate the proposed PTA, it is compared with classical time series representations PAA and APCA on two classical datasets by applying the commonly used K-NN classification algorithm. For ControlChart dataset, PTA outperforms them by 3.55% and 2.33% higher classification accuracy and 8.94% and 7.07% higher for Mixed-BagShapes dataset, respectively. It is indicated that the proposed PTA is effective for high dimensional time series data mining.

1. Introduction

Time series representation is one of the key issues in time series data mining, since the suitable choice of representation greatly affects the ease and efficiency of time series data mining. To address the high dimensionality issue in real-world time series data, a great number of time series representations by applying dimensionality reduction have been proposed.

Dimensionality reduction methods help to compare time series efficiently by modeling time series into a more compact form, whereas significant information about main trends in a time series, which are essential to effective similarity search, may be lost. To support accurate and fast similarity detection in time series, a number of special requirements that should be satisfied by any representation model are summarized as follows [1].

(i) *Time Warping-Awareness*. Time series should be modeled into a form that can be naturally mapped to the time domain. This will make it feasible to benefit from using dynamic time warping (DTW) that can compare time series with local time shifting and different lengths for similarity detection.

(ii) *Low Complexity*. Due to the high dimensionality of time series data, modeling time series should be performed maintaining a reasonably low complexity, which is possibly linear with the series length.

(iii) *Sensitivity to Relevant Features*. It is clearly desirable that time series approximation is able to preserve as much information in the original series as possible. For this purpose, approximating a time series should be accomplished in such a way that it tailors itself to the local features of the series, in order to capture the important trends of the series.

(iv) *Absence of Parameters*. Most representation models and dimensionality reduction methods require the

user to specify some input parameters, for example, the number of coefficients or symbols. However, prior domain knowledge is often unavailable, and the sensitivity to input parameters can seriously affect the accuracy of the representation model or dimensionality reduction method.

From an empirical viewpoint, it has been recently observed that there is no absolute winner among the time series representations in every application domain. Therefore, it is critical for time series representation to keep features that are important for corresponding application domains. The sensitivity to features can be considered according to three main subrequirements for the segments detected in an individual time series: (a) segments may have different lengths, (b) any segment represents different slopes (trends) of a subsequence of data points, and (c) segments capture the series trends [1].

Slopes [2] and derivative estimation [1] are adopted to denote trend of time series commonly in the literature. Due to the property of tangent function that is used to calculate slopes, it is difficult to distinguish two trends when the degrees of angles are close to $\pm 90°$ by using slope to represent trend. In derivative time series segment approximation (DSA) representation [1], original time series is firstly transformed into the first derivative estimations of the points, and segmentation and approximation are based on the derivative estimations of time series. It has been observed that relative variations, the ratios between any two consecutive data points in a given time series, are suitable for representing trend in time series [3]. The magnitude of ratio reflects the variation degree of trend and the sign of ratio represents the changing direction of trend naturally. Based on ratio-based time series, a time series representation, piecewise trend approximation (PTA), is proposed, which retains the important feature of main trends of original time series concisely by dimensionality reduction. In contrast to the conventional representations based on raw time series data, the proposed PTA representation is based on local trends of raw time series data. That is to say, the raw data is firstly transformed into local trends (ratios), segmentation that separates time series into segments of different trends is then performed based on the ratios, and each segment is finally approximated by the ratios between the first and the last data points in the segment.

PTA is able to satisfy the first three requirements mentioned earlier.

(i) PTA representations can be compared by using DTW directly.

(ii) The ratio-based feature generation allows for representing a time series by focusing on the characteristic trends in the series.

(iii) Computational complexity for PTA is linear with the length of series, and the dimensionality of PTA is adaptive with the identified trends of the series.

To validate the proposed PTA, the performance of PTA for time series classification is compared to conventional representations. The experiments are based on two classical datasets by applying K-nearest neighbor (K-NN) classification method. The comparative experimental results show that PTA outperforms conventional representations in classification accuracy.

In Section 2, the time series representations with respect to different dimensionality reduce, techniques are reviewed. PTA representation is proposed in Section 3, and the experiments to validate the proposed PTA for time series classification are illustrated in Section 4.

2. Time Series Representations

To reduce dimensionality of a time series, a piecewise discontinuous function or low-order continuous function is usually applied to approximate it into a compact form. This study focuses on the first dimensionality reduction method, and the time series representations based on piecewise discontinuous functions are reviewed as follows.

The piecewise approximation-based representations include discrete wavelet transform (DWT) [4, 5], swinging door (SD) [6], Piecewise Linear Approximation (PLA) [7, 8], piecewise aggregate approximation (PAA) [9–11], adaptive piecewise constant approximation (APCA) [12], symbolic aggregate approximation (SAX) [13], and derivative time series segment approximation (DSA) [1].

Using DWT, a time series is represented in terms of a finite length, fast decaying, oscillating, and discretely sampled wave form (mother wavelet), which is scaled and translated in order to create an orthonormal wavelet basis. Each function in the wavelet basis is related to a real coefficient; the original series is reconstructed by computing the weighted sum of all the functions in the basis, using the corresponding coefficient as weight. The Haar basis [14] is the most widely used in wavelet transformation. The DWT representation of a time series of length n consists in identifying n wavelet coefficients, whereas a dimensionality reduction is achieved by maintaining only the first p coefficients (with $p > n$).

SD is a data compression technique that belongs to the family of piecewise linear trending functions. SD has been compared to wavelet compression. The SD algorithm employs a heuristic to decide whether a value is to be stored within the segment being grown or it is to be the beginning of a new segment. Given a pivot point, which indicates the beginning of a segment, two lines (the "doors") are drawn from it to envelop all the points up to the next one to be considered. The envelop has the form of a triangle according to a parameter that specifies the initial amplitude of the lines. The setup of this parameter has impact on the data compression level.

In the PLA method, a time series is represented by a piecewise linear function, that is, a set of line segments. Several methods have been proposed to recognize PLA segments (e.g., [7, 8]).

PAA transforms a time series of n points in a new one composed by p segments (with $p > n$), each of which is of size equal to n/p and is represented by the mean value of the data points falling within the segment.

Like PAA, APCA approximates a time series by a sequence of segments, each one represented by the mean value of its data points. A major difference from PAA is that APCA

can identify segments of variable length. Also, the APCA algorithm is able to produce high quality approximations of a time series by resorting to solutions adopted in the wavelet domain.

In SAX method, dimensionality of original time series is first reduced by applying PAA, then the PAA coefficients are quantized, and finally each quantization level is represented by a symbol so that SAX is a symbolic representation of time series.

The DSA representation is based on the derivative version of the original time series. DSA entails derivative estimation, segmentation, and segment modeling to map a time series into a different value domain which allows for maintaining information on the significant features of the original series in a dense and concise way.

For representing a time series of n points, it can be performed in $O(n)$ by using DWT, SD, (the fastest version of) PLA, PAA, SAX, and DSA, whereas the complexity of APCA is $O(n \log(n))$.

There are some other kinds of time series representations applying continuous polynomial functions to approximate time series, include Singular Value Decomposition (SVD) [15, 16], Discrete Fourier Transforms (DFT) [17, 18], splines, nonlinear regression, and Chebyshev polynomials [19, 20], of which the details are kindly referred to the references.

In contrast to conventional representations based on raw data, a time series representation based on ratios between any two consecutive data points in a given time series is proposed by applying piecewise segment approximation to reduce dimensionality in Section 3.

3. PTA: Piecewise Trend Approximation

Given a time series $Y = \{(y_1, t_1), \ldots, (y_n, t_n)\}$, where y_i is a real numeric value and t_i is the timestamp, it can be represented as a PTA representation

$$T' = \left\{ \left(R_1, R_{t_1}\right), \ldots, \left(R_m, R_{t_m}\right) \right\}, \quad m \leq n, \ n \in N, \quad (1)$$

where R_{t_i} is the right end point of the ith segment, R_i $(1 < i \leq m)$ is the ratio between $R_{t_{i-1}}$ and R_{t_i} in the ith segment, and R_1 is the ratio between the first point t_i and R_{t_1}. The length of the ith segment can be calculated as $R_{t_i} - R_{t_{i-1}}$.

PTA approximates a time series by applying a piecewise discontinuous function to reduce dimensionality. The algorithm of PTA consists of three main steps:

(1) local trend transformation: the original time series is transformed into a new series where the values of data points are ratios between any two consecutive data points in original series;

(2) segmentation: the transformed local trend series is divided into variable-length segments such that two conjunctive segments represent different trends;

(3) segment approximation: each segment is represented by the ratios between the first and last data points within the segment, which indicates the characteristic of trend.

3.1. Local Trend Transform.
Given a time series $Y = \{(y_1, t_1), \ldots, (y_n, t_n)\}$, $n \in N$, a new series $T = \{(r_1, t_2), \ldots, (r_n, t_n)\}$ is achieved from Y by local trend transform, where r_i is the value of ratio between (y_{i-1}, t_{i-1}), $i = 2, \ldots, n$.

Ratios between each two consecutive data points in Y are calculated according to the equation by justifying (1) as follows:

$$r_i = \frac{y_i - y_{i-1}}{y_{i-1}}, \quad i = 2, \ldots, n. \quad (2)$$

T is indeed a feature space of local trends mapped from the original data space with one dimension reduced. Although slope is often used to represents trend in the literature, it is difficult to distinguish two trends when the degrees of angles are close to $\pm 90°$ due to the property of tangent function which is used to calculate slopes. Ratio, however, is more suitable for representing trend because the magnitude of ratio reflects the variation degree of trend and the sign of ratio represent the changing direction of trend naturally. Although T is one dimension reduced, it is not enough for many real-world applications. Hence, T will be compressed by the next two steps into a more concise form.

3.2. Segmentation.
Given a time series $Y = \{(y_1, t_1), \ldots, (y_n, t_n)\}$, $n \in N$, Y is segmented into $S = \{S_1, \ldots, S_m\}$ $(m \leq n, m \in N)$, where S_k $(k = 1, \ldots, m)$ is a subsequence of Y, which is decided by key points that certain behavior changes occur in Y. In PTA, segmentation is based on the local trend series T of original series Y. That is to say, the sequence $T = \{(r_1, t_2), \ldots, (r_n, t_n)\}$ is divided into the sequence $S = \{S_1, \ldots, S_m\}$ $(m \leq n, m \in N)$, which is composed of m variable-length segments $S_k = \{(r_{k,1}, t_{k,1}), \ldots, (r_{k,j}, t_{k,j})\}$ $(k = 1, \ldots, m)\}$. Each two consecutive segments represent different trends. Since the segmentation in PTA is based on the ratios by local trend transform, of which signs represent trend directions, the main idea for segmentation is to separate T by finding out the first point such that the sign of it is different from those of the previous points. Assume that ε denotes the threshold of the ratios and $\text{sign}(r_i)$ denotes the sign of r_i in T, the sequence S_k is identified as a segment if and only if $\text{sign}(r_{k,1}) = \text{sign}(r_{k,2}) = \cdots = \text{sign}(r_{k,j})$, $|r_{k,j}| > \varepsilon$, and $\text{sign}(r_{k,j}) = \text{sign}(r_{k+1,1})$, $k = 1, \ldots, m$.

Accordingly, the raw data Y is segmented as $S' = \{S'_1, \ldots, S'_m\}$, $S'_k = \{(y_{k,l}, t_{k,l}), \ldots, (y_{k,j}, t_{k,j})\}$, $(k = 1, \ldots, m, m \leq n, m \in N)$.

This segmentation aggregates the data points having the same changing directions so that the subsequences represent fluctuations in raw data intuitively. Thus, the reduced dimensionality is adaptive to the trend fluctuations and no parameter is needed.

3.3. Segment Approximation.
To approximate the segments $S' = \{S'_1, \ldots, S'_m\}$, $m \leq n, m \in N$, the ratio between the first and last point within each segment is calculated to represent the main trend information of any segment. Finally, the PTA representation $T' = \{(R_1, R_{t_1}), \ldots, (R_m, R_{t_m})\}$, $m \leq n, n \in N$, is yielded such that

$$R_k = \frac{y_{k,j} - y_{k,l}}{y_{k,l}}, \quad k = 1, \ldots, m,$$

$$R_{t_k} = t_{k,j}. \tag{3}$$

The PTA representation maintains the important feature of trend variations in a concise form, while the computation complexity of it is linear with the length n of the sequence, that is, $O(n)$. In addition, since the length of PTA representation is determined by the fluctuations in original time series, similarities between PTA representations can be compared by applying dynamic time warping.

3.4. Distance Measure. To compare two time series data in similarity search tasks, various distance measures have been introduced. By far the most common distance measure for time series is the Euclidean distance [21, 22]. Given two time series X and Y of the same length n, the Euclidean distance between them is defined as

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

In PTA representation, original time series is segmented according to the change of local trend, and the length of the transformed PTA representation is thus adaptive with the trend variations in original time series. The Euclidean distance is limited to compare time series of equivalent length, and thus it cannot be applied to time series similarity search on PTA directly.

To address the limitation of Euclidean distance, dynamic time warping (DTW) has been proposed to evaluate the similarity of variable-length time series [23]. Unlike Euclidean distance, DTW allows elastic shifting of a sequence to provide a better match with another sequence; hence, it can handle time series with local shifting and different lengths. Therefore, DTW can be directly applied to measure similarity of time series in PTA form.

4. Experiments on Time Series Classification

Classification of time series has attracted much interest from the data mining community [24–26]. To validate the performance of the proposed PTA representation for similarity search in time series data, we design a classification experiment based on two classical datasets ControlChart and Mixed-BagShapes [27] by applying the most common classification algorithm, K-nearest neighbor (K-NN) classification. ControlChart is a synthetic dataset of six classes: normal, cyclic, increasing trend, decreasing trend, upward shift, and downward shift. Each class contains 100 instances. Figure 1 shows that representative sample instances in each class of ControlChart dataset. Mixed-BagShapes contains time series derived from 160 shapes with nine classes of objects, including bone, cup, device, fork, glass, hand, pencil, rabbit, and tool. The sample instances from each class of Mixed-Bag-Shape are shown in Figure 2.

The proposed PTA is compared to two classical representations, PAA and APCA, which are introduced in Section 2.

The K-NN classification algorithm is briefly reviewed in Section 4.1, data preprocessing is introduced in Section 4.2, and the experimental results are illustrated in Section 4.3.

4.1. K-Nearest Neighbor (K-NN) Classification. K-NN is one of the most widely used instance-based learning methods [28]. Given a set of n training examples, upon receiving a new instance to predict, the K-NN classifier will identify K-nearest neighboring training examples of the new instance and then assign the class label holding by the most number of neighbors to the new instance [29]. To classify time series data, it is straightforward to investigate the ability of time series representations for similarity search by applying K-NN algorithm since time series can be compared to the others as instances in K-NN.

4.2. Data Preprocessing. In order to reduce the noise in the data, original time series are usually preprocessed by smoothing techniques in time series data mining. It is essential to make data amenable to further data mining tasks by denoising. In PTA, it is necessary to denoise time series data before local trend transformation to avoid that the main trends are undistinguished from noise. Thus, smoothing is applied to denoise raw data before local trend transformation in PTA.

Commonly used smoothing techniques are moving average models including simple moving average, weighted moving average, and exponential moving average. In our experiments, exponential smoothing is applied to preprocess original data to reduce noise. Given a time series $X = \{x_t\}$ ($t = 0, \ldots, n$), the output $S = \{s_t\}$ ($t = 0, \ldots, n$) of the exponential smoothing algorithm is defined as

$$s_1 = x_0,$$

$$s_t = \alpha x_{t-1} + (1 - \alpha) s_{t-1}, \quad t > 1, \tag{5}$$

where α is the smoothing factor and $0 < \alpha < 1$.

4.3. Experimental Results of Time Series Classification. The most commonly used K-NN algorithm is utilized to facilitate independent confirmation of the proposed PTA representation. Concerning with the neighborhood size K in K-NN algorithm, the simple yet very competitive 1-NN algorithm is adopted in this experiment, that is, K-NN with K equal to 1. The parameter of sliding window in PAA representation and the threshold in PTA need to be predefined. The number of segments for PAA is decided by the sliding window while those of PTA and APCA are adaptive with fluctuations in original data. To compare the representations effectively, the parameters are tried several times such that the compressions (i.e., number of segments) of the representations are equal or at least very close. Classification accuracy is defined as

$$\text{accuracy} = 1 - E, \tag{6}$$

where E is the error rate.

The comparative results on ControlChart and Mixed-BagShapes by using leaving-one-out cross-validation are shown in Table 1. The results are the best results of each representation by trials of different parameters. For ControlChart,
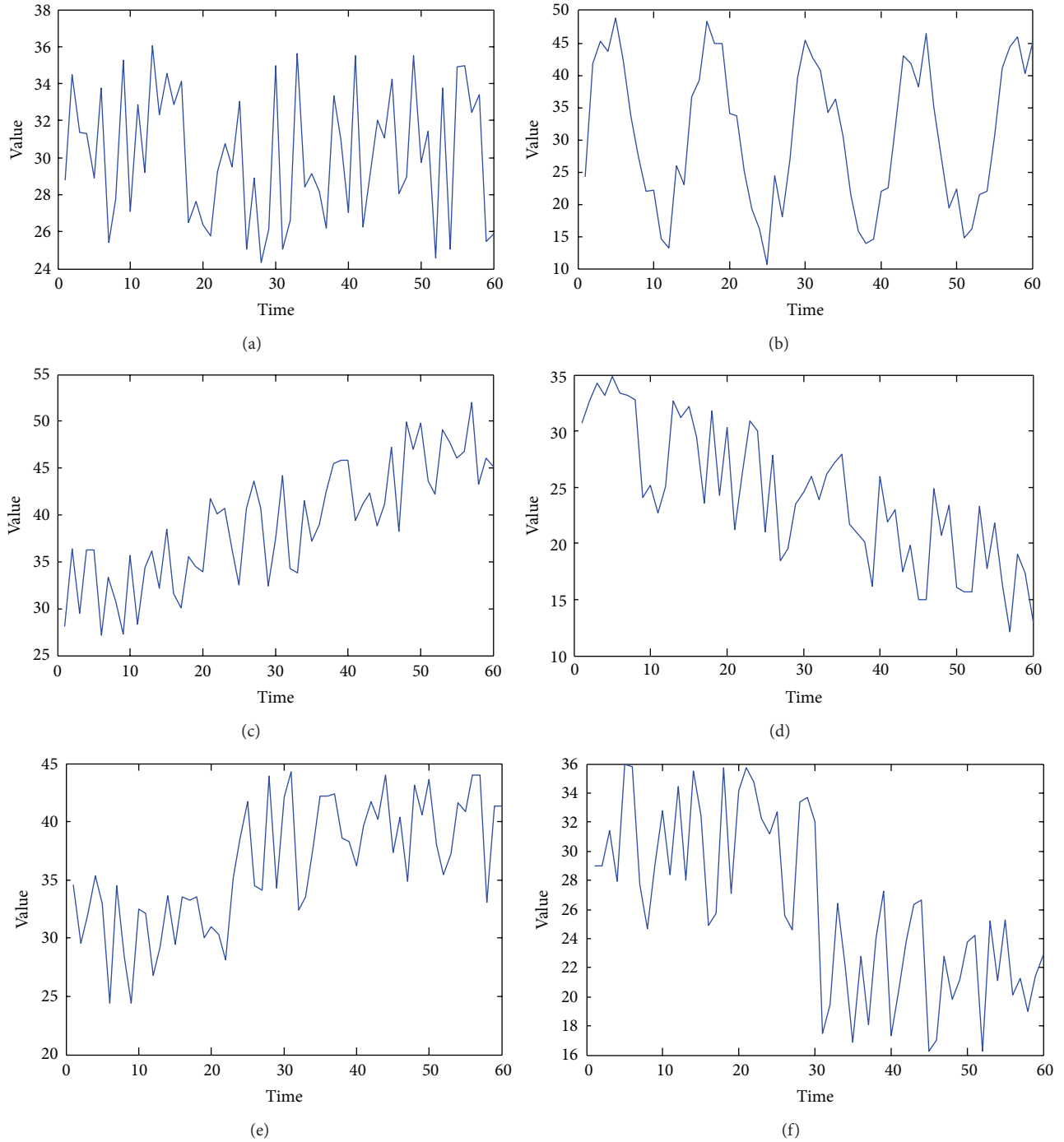
Figure 1: Sample instances from each class in ControlChart dataset: (a) normal; (b) cyclic; (c) increasing; (d) decreasing; (e) upward shift; (f) downward shift.

the proposed PTA outperforms PAA and APCA by 3.55% and 2.33% higher classification accuracy, respectively. For Mixed-BagShapes, PTA yields 8.94% and 7.07% improvement in classification accuracy compared with PAA and APCA, respectively. It is shown that the PTA outperforms the competitive representations by higher classification accuracy, which indicates that PTA is effective for time series classification by representing original data concisely with retaining important feature of trend variation.

5. Conclusions

In order to improve efficiency of time series data mining in high dimensional large-size databases, a time series representation piecewise trend approximation (PTA) is proposed to represent original time series into a concise form while retaining important feature of trend variations. Different from the representations based on original data space, PTA transforms original data space into the feature space of ratio
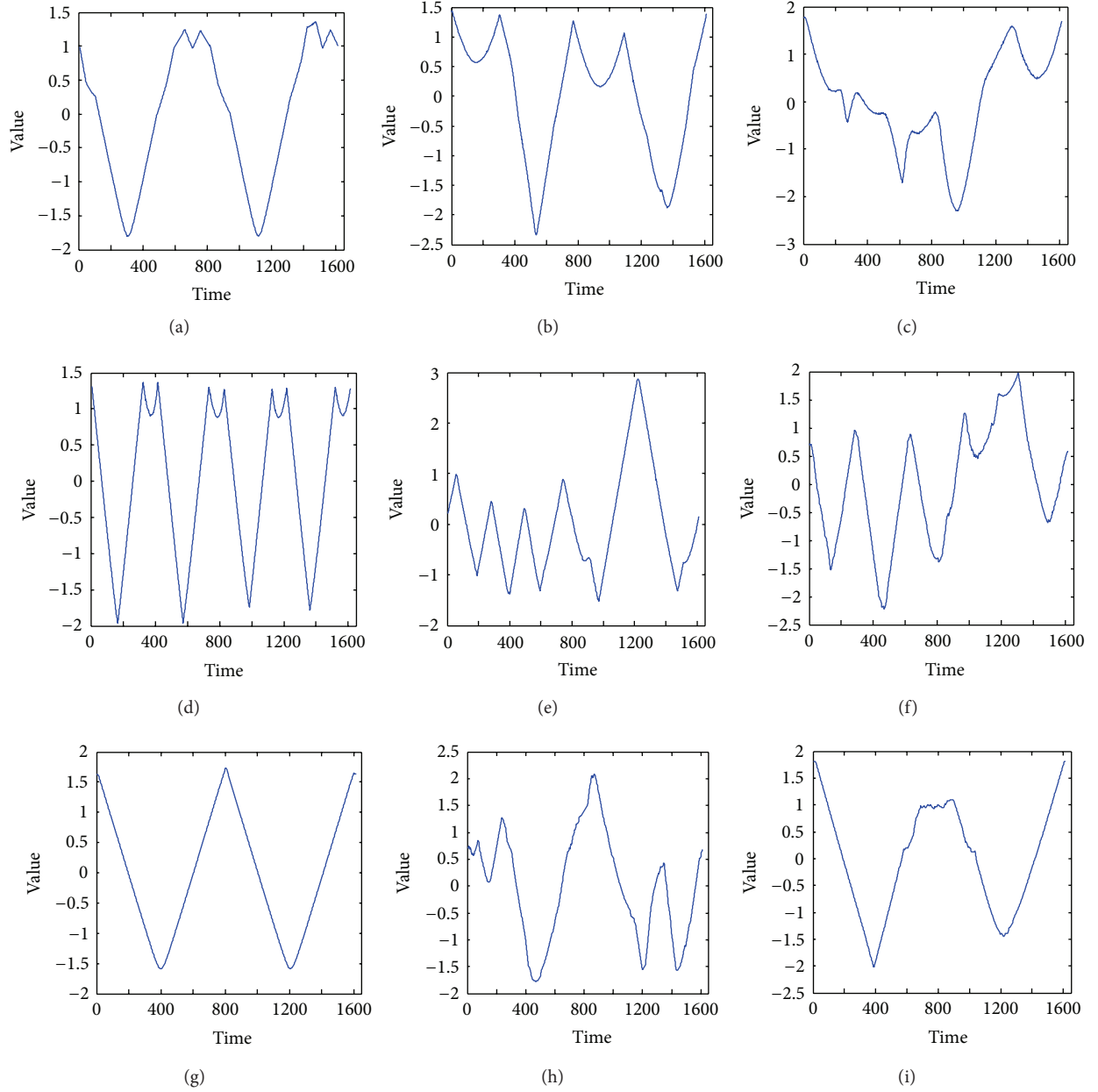
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

FIGURE 2: Sample instances from each class in Mixed-BagShapes dataset: (a) bone; (b) cup; (c) device; (d) fork; (e) glass; (f) hand; (g) pencil; (h) rabbit; (i) tool.

TABLE 1: Comparative results of classification accuracy for ControlChart and Mixed-BagShapes.

	PAA	APCA	Proposed PTA
ControlChart	0.952	0.964	0.987
Mixed-BagShapes	0.876	0.894	0.962

between any two consecutive data points in original time series, of which sign and magnitude indicate changing direction and degree of local trend, respectively. Based on the ratio-based feature space, segmentation is performed such that each two conjoint segments have different trends, and then the piecewise segments are approximated by the ratios between the first and last pints within the segments; dimensionality is, hence, reduced while keeping important feature of main trends in original data.

Based on two classical datasets, ControlChart and Mixed-BagShapes, by applying the commonly used time series classification algorithm K-NN, PTA is compared with classical PAA and APCA representations using DTW distance measure. The results for ControlChart show that PTA yields 3.55% and 2.33% improvements in classification accuracy compared to PAA and APCA, respectively. For Mixed-BagShapes, PTA outperforms PAA and APCA by 8.94% and 7.07% improvement, respectively. The time complexity of PTA algorithm is

linear with the length of original time series. The efficiency of time series data mining is, hence, enhanced by applying PTA representation. The applications of PTA in time series clustering, indexing, and other similarity search tasks will be validated and a symbolic time representation derived from PTA can be further developed.

Acknowledgment

References

[1] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco, "A time series representation model for accurate and fast similarity detection," *Pattern Recognition*, vol. 42, no. 11, pp. 2998–3014, 2009.

[2] Y. Ding, X. Yang, A. J. Kavs, and J. Li, "A novel piecewise linear segmentation for time series," in *Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE '10)*, pp. 52–55, February 2010.

[3] K. Huarng and T. H. K. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 2, pp. 328–340, 2006.

[4] K. Chan and A. Fu, "Efficient time series matching by wavelets," in *Proceedings of the International Conference on Data Engineering (ICDE '99)*, pp. 126–133, 1999.

[5] Y. L. Wu, D. Agrawal et al., "A comparison of DFT and DWT based similarity search in time-series databases," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '00)*, pp. 488–495, 2000.

[6] E. H. Bristol, "Swinging door trending: adaptive trending recording," in *Proceedings of the ISA National Conference*, pp. 749–753, 1990.

[7] T. Pavlidis and S. L. Horowitz, "Segmentation of plane curves," *IEEE Transactions on Computers*, vol. 23, no. 8, pp. 860–870, 1974.

[8] E. Keogh and M. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '98)*, pp. 239–241, 1998.

[9] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and Information Systems*, vol. 3, pp. 263–286, 2001.

[10] E. Keogh and M. Pazzani, "Scaling up dynamic time warping for data mining applications," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '00)*, pp. 285–289, 2000.

[11] X. Feng, C. Cheng, L. Changling, and S. Huihe, "An improved process data compression algorithm," in *Proceedings of the International Conference on Intelligent Control and Automation*, pp. 2190–2193, 2002.

[12] K. Chakrabarti, E. Keogh, S. Mehrotra, and M. Pazzani, "Locally adaptive dimensionality reduction for indexing large time series databases," *ACM Transactions on Database Systems*, vol. 27, no. 2, pp. 188–228, 2002.

[13] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (DMKD '03)*, pp. 2–11, 2003.

[14] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms, a Primer*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1997.

[15] F. Korn, H. V. Jagadish, and C. Faloutsos, "Efficiently supporting Ad Hoc queries in large datasets of time sequences," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 289–300, 1997.

[16] K. V. Kanth, D. Agrawal, and A. Singh, "Dimensionality reduction for similarity searching in dynamic databases," in *Proceedingss of the ACM SIGMOD International Conference on Management of Data*, pp. 166–176, 1998.

[17] D. Rafiei and A. Mendelzon, "Similarity-based queries for time series data," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 13–25, 1997.

[18] D. Rafiei and A. Mendelzon, "Efficient retrieval of similar time sequences using DFT," in *Proceedings of the International Conference on Foundations of Data Organization and Algorithms (FODO '98)*, pp. 249–257, 1998.

[19] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2003.

[20] Y. Cai and R. Ng, "Indexing spatio-temporal trajectories with Chebyshev polynomials," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 599–610, 2004.

[21] G. Reinert, S. Schbath, and M. S. Waterman, "Probabilistic and statistical properties of words: an overview," *Journal of Computational Biology*, vol. 7, no. 1-2, pp. 1–46, 2000.

[22] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: a survey and empirical demonstration," *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349–371, 2003.

[23] D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the Workshop on Knowledge Discovery in Databases, at the 12th International Conference on Artificial Intelligence*, pp. 359–370, 1994.

[24] J. McQueen, "Some methods for classification and analysis of multivariate observation," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

[25] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-based classification of time-series data," in *Information Processing and Technology*, pp. 49–61, Nova Science Publishers, Commack, NY, USA, 2001.

[26] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proceedings of the 4th SIAM International Conference on Data Mining (SDM '04)*, pp. 11–22, April 2004.

[27] E. Keogh and T. Folias, "The UCR time series data mining archive," 2002, http://www.cs.ucr.edu/~eamonn/time_series_data/.

[28] T. M. Mitchell, ,*Machine Learning*, Computer Sciences Series, McGraw-Hill, New York, NY, USA, 1997.

[29] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transaction on Information Theory*, vol. 13, pp. 21–27, 1967.