

# A Recent-Biased Dimension Reduction Technique for Time Series Data

Yanchang Zhao, Chengqi Zhang and Shichao Zhang

Faculty of Information Technology, University of Technology, Sydney, Australia  
{yczhao, chengqi, zhangsc}@it.uts.edu.au

**Abstract.** There are many techniques developed for tackling time series and most of them consider every part of a sequence equally. In many applications, however, recent data can often be much more interesting and significant than old data. This paper defines new recent-biased measures for distance and energy, and proposes a recent-biased technique based on DWT for time series in which more recent data are considered more significant. With such a recent-biased technique, the dimension of time series can be reduced while effectively preserving the recent-biased energy. Our experiments have demonstrated the effectiveness of the proposed approach for handling time series.

## 1 Introduction

Analyzing time series is a challenging topic in the field of data modelling and mining. In many applications, such as stock market, one concerns more about the recent data than what happened long ago. Besides the global trend, the recent data are very important to judge the similarity between time series and more significant to predict and make decisions than the detail of old data. For example, for a stockbroker, the long-term (say, six years) trend of stock price and the detailed variances in the last month of a stock are important, but the variance in a certain month four years ago is of little significance. In such kind of scenarios, a mechanism which favors the recent is called for. Nevertheless, most of the techniques for time series give equal significance to all data in the series. In this paper, we design a recent-biased technique to tackle the above problem. With our method, recent data are given more significance and kept with finer resolution, while old data with coarser resolution. Weights for DWT coefficients are derived from a decaying function, and then the coefficients with the largest  $k$  weights are chosen as the representation of the time series. Our technique is different from SWAT [1] in that the largest  $k$  coefficients are kept with our method while only a single coefficient is maintained at each level with SWAT. Our technique is also different from the traditional method of keeping the largest  $k$  coefficients and the RAM-DS algorithm [8], because with our technique the largest  $k$  coefficients are obtained from the weights only and has nothing to do with specific time series. However, the subsets of largest coefficients are different for different time series with both the tradition method and RAM-DS. The same subset of coefficients for all time series is used in [6], but the subset of coefficients are obtained from all time series without weighting.

## 2 Related Work

In this section, related work on similarity measures, dimension reduction and recent-biased techniques for time series data will be introduced. Euclidian Distance and other  $L_p$ -norms are popular to measure the distance between time series. Another measure, DTW (Dynamic Time Warpping), is designed to handle time series with some time shifts.

Because time series is usually of very high dimension, the dimension has to be reduced to improve the efficiency of computation. Popular techniques include PCA, DFT (Discrete Fourier Transform), DWT (Discrete Wavelet Transform) [6], Landmark [7], major minima and maxima [4], PIP [3], PAA [5], etc. With DWT, a time series can be represented by a rough sketch by keeping only the first coefficients. Some researchers propose to use the largest coefficients to preserve the optimal amount of energy, or to choose the same subset of the coefficients for all time series for the ease of computing similarity [6].

As to recent-biased techniques, Bulut et al designed a structure named SWAT [1] to process queries over data streams that are biased towards the more recent values. A time weighting function is defined in [8] so that the old data values are with lower weights, and then more resources can be utilized to explore more recent data with finer granularities. Cohen et al uses decay functions to maintain time-decaying stream aggregates [2].

## 3 A Recent-Biased Technique for Dimension Reduction

Our idea comes from the observation that recent data are usually more important than ancient data. Considering the time series shown in Figure 1, which pair of  $\mathbf{S}_1\mathbf{S}_2$  and  $\mathbf{S}_1\mathbf{S}_3$  is more similar? In many applications, we care more about recent data than ancient data, then  $\mathbf{S}_1$  and  $\mathbf{S}_3$  is more similar than  $\mathbf{S}_1$  and  $\mathbf{S}_2$ , since the difference between  $\mathbf{S}_1$  and  $\mathbf{S}_3$  happened long ago. However, the  $L_p$ -norm distance between  $\mathbf{S}_1$  and  $\mathbf{S}_2$  is 1, while that between  $\mathbf{S}_1$  and  $\mathbf{S}_3$  is also 1. So the two pairs are of the same similarity according to  $L_p$ -norm distance. If using  $(\frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}, 1)$ , as the weights, the distance between  $\mathbf{S}_1$  and  $\mathbf{S}_2$  is 0.75, while that between  $\mathbf{S}_1$  and  $\mathbf{S}_3$  is 0.25. Therefore,  $\mathbf{S}_1$  is more similar to  $\mathbf{S}_3$  than to  $\mathbf{S}_2$  when a biased distance measure is used.

$\mathbf{S}_1$	0	0	0	0	0	0	0	0	0
$\mathbf{S}_2$	0	0	0	0	0	0	1	0	0
$\mathbf{S}_3$	0	1	0	0	0	0	0	0	0
Time Elapsed	7	6	5	4	3	2	1	0	0
	Time $\rightarrow$								

**Fig. 1.** An Example of Time Series

Since old data become less significant as time goes by, a bias can be given to recent data. A simple idea is to give larger weights to more recent data. Actually, decaying functions are used widely for processing time series data. For example, exponential decaying functions are used for time series to find recent frequent itemsets adaptively, and to explore temporal and support count granularities in data streams [8]. In addition to exponential functions, polynomial and ployexponential decaying functions are also considered to maintain time-decaying stream aggregates [2]. Similarly, we define a recent-biased function as

follows to help keep more recent data at finer scales. Recent-biased function  $\mathbf{B} = b(t), t \geq 0$  is a monotonously decreasing function with  $b(0) = 1$  and  $b(+\infty) = 0$ , where  $t$  is the time elapsed till now.

The most common used decay function is exponential function,  $b_t = d^t$ , where  $t$  is the time elapsed, and  $d$  is the decay factor,  $0 < d < 1$ ,  $\alpha > 0$ . Linear decay can be get with linear function  $b_t = \frac{n-t}{n}$ , where  $0 \leq t \leq n$  and  $n$  is the length of time series. Based on the above recent-biased function, the recent-biased  $L_p$ -norm distance with bias on recent is defined in the following.

**Definition 1 (Recent-Biased Distance & Energy).** *The recent-biased distance between time series  $\mathbf{S}$  and  $\mathbf{S}'$  is defined as*

$$Dist(\mathbf{S}, \mathbf{S}') = \frac{\|(\mathbf{S} - \mathbf{S}') \bullet \mathbf{B}\|}{\|\mathbf{B}\|} \quad (1)$$

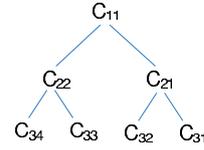
where ' $\bullet$ ' stands for the operator of inner product, ' $\|\cdot\|$ ' denotes  $L_p$ -norm, and  $\mathbf{B}$  is a recent-biased vector. If  $L_2$ -norm is used, the recent-biased energy of  $\mathbf{S}$  is defined as

$$E(\mathbf{S}) = \frac{\|\mathbf{S} \bullet \mathbf{B}\|^2}{\|\mathbf{B}\|^2} \quad (2)$$

### 3.1 Recent-Biased Dimension Reduction

In this paper, Haar wavelet transform is used because it is very simple and widely used and is of linear time complexity. For Discrete Wavelet Transform, there are two different ways for choosing coefficients, the first  $k$  or the largest  $k$  coefficients. If the first  $k$  coefficients are selected, the global trend and variation can be preserved. With the largest coefficients kept, the parts of large energy are preserved and the energy is better kept, which is better for compressing a single signal. Nevertheless, when dealing with multiple time series, more storage space is required to keep the positions of coefficients and the distance computations is more complex. To keep more detail for recent data and preserve the recent-biased energy, a recent-biased technique based on DWT is designed in the following. Instead of keeping the largest or the first coefficients, the largest recent-biased coefficients are kept, which are computed from the weights only.

The DWT coefficients for a time series with 8 values are shown in Figure 1. We can make the average to be zero by normalizing the time series, so the average  $C_0$  is zero and is not considered here. Suppose that the time series are normalized that the average is zero. The original time series is  $\mathbf{S} = (-C_{11} - C_{21} - C_{31}, -C_{11} - C_{21} + C_{31}, -C_{11} + C_{21} - C_{32}, -C_{11} + C_{21} + C_{32}, C_{11} - C_{22} - C_{33}, C_{11} - C_{22} + C_{33}, C_{11} + C_{22} - C_{34}, C_{11} + C_{22} + C_{34})$  from recent data to old data. Assume that  $\mathbf{B} = b(t), t \geq 0$  is the bias function, so the recent-biased energy of  $\mathbf{S}$  is



**Fig. 2.** Coefficients of DWT

$$\begin{aligned}
E(\mathbf{S}) &= \|\mathbf{B} \bullet \mathbf{S}\|^2 / \|\mathbf{B}\|^2 \\
&= \frac{1}{\|\mathbf{B}\|^2} ((-C_{11} - C_{21} - C_{31})^2 b_0^2 + (-C_{11} - C_{21} + C_{31})^2 b_1^2 \\
&\quad + (-C_{11} + C_{21} - C_{32})^2 b_2^2 + (-C_{11} + C_{21} + C_{32})^2 b_3^2 \\
&\quad + (C_{11} - C_{22} - C_{33})^2 b_4^2 + (C_{11} - C_{22} + C_{33})^2 b_5^2 \\
&\quad + (C_{11} + C_{22} - C_{34})^2 b_6^2 + (C_{11} + C_{22} + C_{34})^2 b_7^2) \\
&= \frac{1}{\|\mathbf{B}\|^2} (C_{11}^2 \sum_{i=0}^7 b_i^2 + C_{21}^2 \sum_{i=0}^3 b_i^2 + C_{22}^2 \sum_{i=4}^7 b_i^2 \\
&\quad + C_{31}^2 \sum_{i=0}^1 b_i^2 + C_{32}^2 \sum_{i=2}^3 b_i^2 + C_{33}^2 \sum_{i=4}^5 b_i^2 + C_{34}^2 \sum_{i=6}^7 b_i^2 \\
&\quad + 2C_{11}C_{21}(b_0^2 + b_1^2 - b_2^2 - b_3^2) + 2C_{11}C_{22}(-b_4^2 - b_5^2 + b_6^2 + b_7^2) \\
&\quad + 2C_{21}C_{31}(b_0^2 - b_1^2) + 2C_{21}C_{32}(-b_2^2 + b_3^2) \\
&\quad + 2C_{22}C_{33}(b_4^2 - b_5^2) + 2C_{22}C_{34}(-b_6^2 + b_7^2))
\end{aligned} \tag{3}$$

In order to preserve the energy as large as possible, the coefficients with largest weights will be kept, and other coefficients are set to zero. It is difficult to tell which  $C_{ij}$  is of the greatest importance from Formula (3). To make it easy, we only consider  $C_{ij}^2$  while ignoring those  $C_{i_1 j_1} C_{i_2 j_2}$ , where  $i_1 \neq i_2$  or  $j_1 \neq j_2$ . Then, the question becomes to choose  $C_i$  with the largest coefficients in  $C_{11}^2 \sum_{k=0}^7 b_k^2 + C_{21}^2 \sum_{k=0}^3 b_k^2 + C_{22}^2 \sum_{k=4}^7 b_k^2 + C_{31}^2 \sum_{k=0}^1 b_k^2 + C_{32}^2 \sum_{k=2}^3 b_k^2 + C_{33}^2 \sum_{k=4}^5 b_k^2 + C_{34}^2 \sum_{k=6}^7 b_k^2$ . The weight of  $C_{ij}$  is

$$w_{ij} = \sum_{k=(j-1) \cdot 2^{L-i+1}}^{j \cdot 2^{L-i+1} - 1} b_k^2, \quad 1 \leq i \leq L, 1 \leq j \leq 2^i, L = \lceil \log_2 n \rceil \tag{4}$$

where  $n$  is the length of time series. Since  $b_i$  is a recent-biased function, it is monotonously decreasing with the increase of  $i$ . Therefore,  $C_{11}$  is of the largest coefficient,  $\sum_{k=0}^7 b_k^2$ , and it is the first one to choose. The second one is  $C_{21}$ , whose weight is  $\sum_{k=0}^3 b_k^2$ . The third one will be  $C_{22}$  (or  $C_{31}$ ) if  $\sum_{k=4}^7 b_k^2$  is greater (or less) than  $\sum_{k=0}^1 b_k^2$ . Different bias functions will probably lead to different sets of the largest  $k$  coefficients. Given the specific bias function, the weight of DWT coefficients can be calculated immediately with Formula (4) and ready for use for all time series. With our method, only the weighted function is used to decide which coefficients to keep, so the same subset of coefficients are chosen for all time series, which is different from the traditional method of keeping different subsets of the largest coefficients for different time series.

The Euclidean distance between the extracted  $k$  coefficients is used as the similarity between time series and the recent-biased Euclidean distance and energy are calculated with the following formulae.

$$Dist(\mathbf{S}, \mathbf{S}') \approx \frac{(\sum_{ij} w_{ij} (C_{ij} - C'_{ij})^2)^{\frac{1}{2}}}{(\sum_{ij} w_{ij})^{\frac{1}{2}}} \tag{5}$$

$$E(\mathbf{S}) \approx \frac{\sum_{ij} w_{ij} C_{ij}^2}{\sum_{ij} w_{ij}} \tag{6}$$

where  $C_{ij}$  is in the first  $k$  coefficients. If the bias function is set to  $b_i = 1, i \geq 0$ , that is, there is no bias, then the weights for those coefficients in Figure 1 are 8, 4, 4, 2, 2, 2, 2 from top to bottom. So those coefficients at higher level will be chosen first, which is the same as traditional DWT with the first coefficients.

### 3.2 Complexity Analysis

Assume the length of time series is  $n$ , and there are  $m$  time series. The time complexity for computing weights from bias function are  $O(n)$  and the time complexity for getting the largest  $k$  weights is  $O(kn)$ . The time complexity of DWT for a time series is  $O(n)$ . Therefore, the total time complexity for processing  $m$  time series is  $O(n + kn + mn)$ , i.e.,  $O((m + k + 1)n)$ . As to space complexity, the space requirement for keeping the positions of the largest  $k$  coefficients is  $O(k)$ , and there are  $k$  coefficients for each time series, so the total space complexity is  $O(k + mk)$ , i.e.,  $O((m + 1)k)$ .

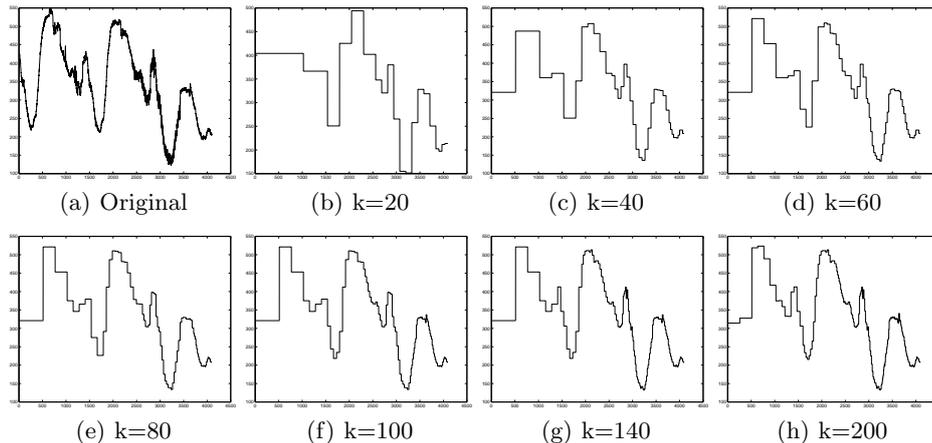
## 4 Experimental Results

Effectiveness of our algorithm for capturing recent details is shown in Figure 3. The original time series (see Figure 3a) is “leleccum” from Matlab, and the first 4096 values are kept. Linear bias function  $b(t) = \frac{n-t}{n}$  is used, where  $n=4096$ . The reconstructed times series after keeping the recent-biased largest  $k$  coefficients are shown in Figure 3b-h. These figures show clearly that the more recent data are preserved with more details while the older data kept with a coarser scale.

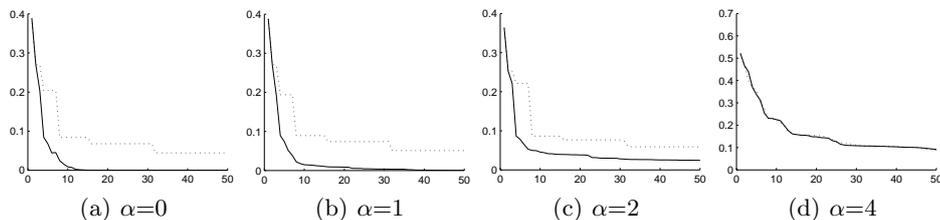
The Nasdaq indices from “Yahoo! Finance” (<http://finance.yahoo.com/>) is used to test the accuracy of our technique in experiments. The close prices of indices from Jun 1988 to Oct 2004 are chosen and each time series is composed of 4096 points. To evaluate the effectiveness of our technique, we design a criterion to measure the precision of approximation after dimension reduction. Assume that  $\mathbf{S}$  and  $\mathbf{S}'$  are respectively the original and reconstructed time series. The error of approximation between  $\mathbf{S}'$  and  $\mathbf{S}$  is defined as  $Err(\mathbf{S}', \mathbf{S}) = \frac{E(\mathbf{S}' - \mathbf{S})}{E(\mathbf{S})}$ , where  $E(\mathbf{S})$  is the recent-biased energy of  $\mathbf{S}$  defined in Formula (2). The experimental result for accuracy is shown in Figure 4. The horizontal axis stands for  $k$ , the number of coefficients kept, and the vertical axis stands for the error rate. The solid line denotes the error rate of recent-biased DWT, while the dotted denotes that of traditional DWT with the first coefficients. It is clear that the accuracy gets improved as more coefficients are kept. Exponential bias functions are used, and the decay factor is  $d = 1 - \frac{1}{1+10^\alpha}$ . From Figure 4a to 4d, the decay function becomes less biased on recent with the increase of  $\alpha$ . When the bias is large, higher accuracy can be achieved with our method than with traditional DWT with the first coefficients. When the bias is tiny (see Figure 4d), our method becomes nearly the same as traditional DWT with the first coefficients.

## 5 Conclusions

We have designed a recent-biased technique for time series, which gives greater weights to more recent data and also preserves more details of recent data. Our experiment shows that the recent-biased technique is very efficient and effective to handle time series. Our future work includes combining our recent-biased idea with DFT, PIP, PAA and other dimension reduction techniques for time series data, and extending dynamic time warping to a recent-biased measure.



**Fig. 3.** Effectiveness. The original time series is shown in (a), and the reconstructed times series by keeping the recent-biased largest  $k$  coefficients are shown in (b)-(h).



**Fig. 4.** Accuracy. The horizontal axis stands for the number of coefficients, and the vertical axis stands for the accuracy. The solid line denotes the error rate of recent-biased DWT, while the dotted denotes that of DWT with the first coefficients.

## References

1. A. Bulut and A. K. Singh: SWAT: Hierarchical Stream Summarization in Large Networks. Proc. of the 19th Int. Conf. on Data Engineering, Bangalore, India, 2003.
2. E. Cohen and M. Strauss: Maintaining time-decaying stream aggregates. Proc. of the 22nd ACM Symposium on Principles of Database Systems, 2003.
3. T. Fu, F. Chung, V. Ng and R. Luk: Pattern Discovery from Stock Time Series Using Self-Organizing Maps. Workshop Notes of KDD'01 . San Francisco, CA, USA, 2001.
4. E. Fink, K. B. Pratt, and H. S. Gandhi: Indexing of Time Series by Major Minima and Maxima. Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics, 2003.
5. E. Keogh, K. Chakrabarti, et al: Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowledge and Information Systems 3(3), 2000.
6. F. Mörchen: Time series feature extraction for data mining using DWT and DFT. In Technical Report No. 33, Philipps-University Marburg, 2003.
7. C.-S. Perng, H. Wang, et al: Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases. Proc. of the 16th ICDE, San Diego, CA, 2000.
8. W-G. Teng, M-S. Chen, and P.S. Yu: Using Wavelet-based Resource-aware Mining to Explore Temporal and Support Count Granularities in Data Streams. Proc. of the SIAM Int. Conf. on Data Mining, April 2004.