# Enhancing DWT for Recent-Biased Dimension Reduction of 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.** In many applications, old data in time series become less important as time elapses, which is a big challenge to traditional techniques for dimension reduction. To improve Discrete Wavelet Transform (DWT) for effective dimension reduction in this kind of applications, a new method, *largest-latest-DWT*, is designed by keeping the largest $k$ coefficients out of the latest $w$ coefficients at each level of DWT transform. Its efficiency and effectiveness is demonstrated by our experiments.
**Keywords:** Time series, dimension reduction

## 1 Introduction

The "curse of dimensionality" is a notorious problem in the field of data mining, because it not only causes the explosion of search space, but also questions the meaningfulness of looking for the nearest neighbor. Since time series data are usually of very high dimensionality, techniques of dimension reduction are often applied to time series before storing them. Popular techniques are SVD (Singular Value Decomposition), DFT (Discrete Fourier Transform) and DWT (Discrete Wavelet Transform) [5], and some methods proposed recently are landmarks [7], major minima and maxima [3], PIP [2], PAA [4], SWAT [1] and online amnesic approximation [6]. DWT is widely used for compression and dimension reduction, and some researchers proposed 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 [5]. Bulut and Singh designed SWAT algorithm to process queries over data streams that are biased towards the more recent values [1]. Their method uses Haar wavelets and maintains a single coefficient at each level.

In many applications, such as stock market and weather forecast, old data become less important as time elapses. Recent data are more important to judge the similarity between time series and predict the future trend than the details of old data. In such applications, a mechanism which favors the recent is called for. Nevertheless, most of the techniques for time series, except SWAT [1] and online amnesic approximation [6], treats every part equally.

In this paper, we will present a technique named *largest-latest-DWT* to tackle the above problem. By decomposing time series with DWT, the latest $w$ coefficients at each level are put together to build a candidate set, which keeps the latest details. Then the largest $k$ coefficients out of the candidates are used to

represent the original time series, which captures the largest fluctuations. Our experiments show that the proposed method is more accurate than both traditional DWT and SWAT [1]. This work is different from our previous paper [9] in that the paper [9] focuses on preserving recent-biased energy and the same subset of coefficients are kept for all time series, because the set of coefficients to keep are computed from a given recent-biased function only . However, this paper will present a method designed for recent-biased dimension reduction for specific time series, which is more effective than using the same subset of coefficients for all time series. What makes it different from our another paper [8] in that this paper achieves recent-biased dimension reduction by keeping the largest ones out of the latest coefficients, while the paper [8] tackles the problem with segmentation. Moreover, this work focused on improving DWT for recent-biased representation with various decay functions, while the paper [8] focuses on a generalized framework which can make most existing dimension-reduction methods ready for online recent-biased dimension reduction.

## 2 Enhancing DWT for Recent-Biased Dimension Reduction

Our idea comes from the observation that recent data are usually more important than old data in many applications. For example, there are three time series as follows: **A**: 7 -1 8 10 12 2 -5 -1; **B**: 4 2 5 13 <u>12 2 -5 -1</u>; **C**: <u>7 -1 8 10</u> 11 3 -2 -4 (Figure 1). For time series **A** and **B**, the second halves of them are the same, while the first halves different. On the contrary, the first halves of **A** and **C** are the same, while the second halves different. Out of **B** and **C**, which one is more similar to **A**? If we care more about recent data than old data, then **B** is more similar to **A** than **C** to **A**. However, the Euclidean distance between **A** and **B** is 6 , while that between **A** and **C** is 4.5, which indicates that **C** is more similar to **A** than **B** to **A**! For the above problem, bias should be given to recent data, so a technique of *largest-latest-DWT* will be designed in the following.

There are two traditional DWTs: one is DWT with the first $k$ coefficients and the other is DWT with the largest $k$ coefficients. The latter is more effective for preserving the energy of time series. To apply DWT to those scenarios where old data become less important as time goes by, a new method of selecting coefficients is designed as follows. The latest $w$ coefficients at each level of DWT transform are selected and put together to build a candidate set. All coefficients are selected if there are less than $w$ coefficients at a certain level. Then the largest $k$ coefficients from the candidate set are used to represent the whole time series. The first selection of the latest $w$ coefficients at each level helps to preserve more details of recent data and less details of old data, which reduces dimension in a recent-biased way. Then the second selection of the largest $k$ coefficients out of the candidate set keeps the optimal energy of the time series. The DWT coefficients is organized in a hierarchy and the coefficients at level $l$ are labeled as $C_{l1}, C_{l2}, C_{l3}, ...$ from present to past. Then, the largest $k$ coefficients will be selected from the candidate set $\bigcup_{l=1..L}\{C_{l1}, C_{l2}, ..., C_{lm}\}$, where $m = \min\{w, 2^l\}$
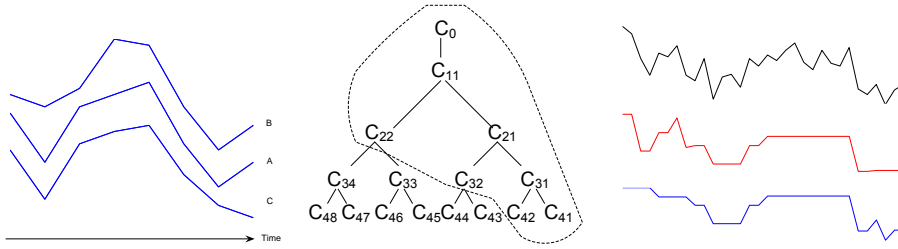
**Fig. 1. B** is more similar to **A** than **C** to **A** in recent-biased view.

**Fig. 2.** The candidate set of coefficients when the latest two coefficients are used at each level.

**Fig. 3.** A time series and the reconstructed ones with traditional DWT and our method (from top to bottom).

and $L$ is the number of levels. Figure 2 shows an example of a 4-level hierarchy of coefficients where $w$ is set to two. The coefficients circumscribed by the dotted line are candidates from which the largest $k$ coefficients will be selected. For DWT transformation, Haar wavelet transform is used in this paper because it is very simple, widely used and of linear time complexity.

An example is shown in Figure 3. The curves from top to bottom shows respectively the original time series, the reconstructed ones with traditional DWT and the proposed method. The largest 10 coefficients are kept for both transforms. For our method, the candidate set is composed of at most 6 coefficients from each level, and then the largest 10 coefficients are selected from those candidates. The figure clearly shows that our method captures more recent details.

## 3 Experimental Results

The proposed method is implemented with Matlab 7 and compared with DWT and SWAT [1]. The accuracy is computed as $1 - \frac{\|(S'-S)\bullet E\|}{\|S\bullet E\|}$, where $S$ and $S'$ are respectively the original/reconstructed time series and $E$ is a decay function [8].

The first dataset used is the control chart time series from UCI KDD Archive (http://kdd.ics.uci.edu/). It contains 600 examples of control charts and each example is composed of 60 values. We run the proposed method and traditional DWT on the dataset and the average accuracy is shown in Figure 4. The horizontal axis stands for $w$, the number of candidate coefficients at each level, while the vertical axis denotes accuracy. The horizontal dotted line shows the accuracy of traditional DWT by keeping the largest $k$ coefficients. Because the accuracies of traditional DWT with different decay functions are nearly the same on this dataset, only one line in each subfigure is used to show its accuracy. The other lines show the accuracy of our method with linear (dash-dot line), exponential (solid line) and window (dash line) decay functions. Since SWAT uses only one coefficient (the latest one) at each level [1], it is the same as our method with $w$=1. As we can see from the figure, SWAT ($w$=1) is always of the lowest accuracy and the accuracy is much improved with our method. However, the accuracy

is lower than that of traditional DWT when $w$ is small. The reason lies in that the number of candidate coefficients is less than $k$ when $w$ is small. For example, when $w=2$, there are only 11 candidate coefficients. Therefore, only 11 coefficients are used with the proposed method even when $k$ is set to 40, so it is not surprising that traditional DWT is of higher accuracy than our method when $w=2$, as shown in Figure 4a. When the same number of coefficients are used for both methods, the proposed method is of higher accuracy than traditional DWT in most cases. When $w$ is too large (e.g., when $w > 25$ in Figure 4a), the proposed method degrades to traditional DWT, because almost all coefficients become candidates.

The other dataset used is US monthly sales of petroleum and related products from Jan 1971 to Dec 1991 (http://www.robhyndman.info/forecasting/goto data.htm). There are 4 time series in the dataset which show the monthly sales of chemicals, coal, petrol and vehicles, and each time series is of 252 values. The largest 10 coefficients are kept for each time series and the average accuracy is given in Figure 5. The horizontal axis stands for $w$, while the vertical axis denotes accuracy. The dotted line shows the accuracy of traditional DWT, and the solid lines show the results of our method with (a)linear, (b)exponential and (c)window decay functions. The result also shows that our method is of higher accuracy than traditional DWT.

Another experiment is conducted on control chart data to test the relationship between the bias of decay function and $w$. The decay functions used are exponential function with bases as 0.99, 0.97, 0.95, 0.9 and 0.8 (Figure 6a). The smaller is the base, the function is more biased on recent data. The experimental result is shown in Figure 6b. In most cases, with the increase of $w$, the accuracy increases sharply first and then decreases slowly. The larger is the bias on recent data, the smaller $w$ is enough to get high accuracy, that is, the less coefficients are needed at each level of DWT transform. However, when the bias is very little (eg. when the base is 0.99), more coefficients are necessary to achieve high accuracy. Therefore, the value of $w$ should be chosen according to the required bias on recent data in specific applications.

## 4    Conclusions

We have designed a technique for efficiently representing time series data in applications where old data becomes less important as time goes by. The latest $w$ coefficients at each level of DWT transformation are put together to build a candidate set, then the largest $k$ coefficients in the candidate set are used to represent the whole time series. Our experiment shows that the proposed technique is more accurate than traditional DWT and SWAT.
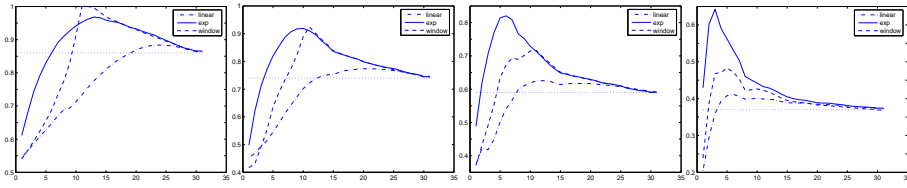
**Fig. 4.** Experimental results on control chart time series. The figures from left to right are the results with the largest (a) 40, (b) 30, (c) 20 and (d) 10 coefficients, respectively.
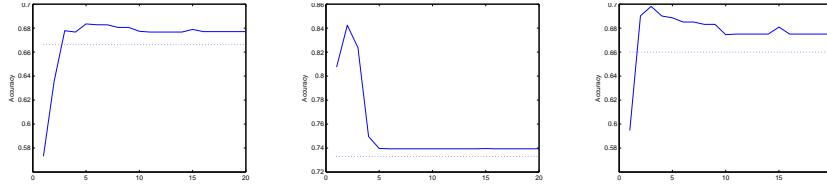


**Fig. 5.** Experimental results on petrol time series. The figures from left to right are respectively the results with (a) linear, (b) exponential and (c) window decay functions.
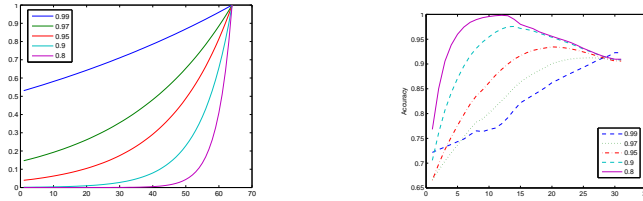


**Fig. 6.** Experimental results on various exponential decay functions. (a) Exponential decay functions, (b) Relationship between bias and $w$.

# References

1. A. Bulut, A.K. Singh: SWAT: hierarchical stream summarization in large networks. Proc. of ICDE'03.
2. T.-C. Fu, F.-L. Chung, V. Ng, R. Luk: Pattern discovery from stock time series using self-organizing maps. KDD'01 Workshop on Temporal Data Mining.
3. E. Fink, K. B. Pratt, H. S. Gandhi: Indexing of time series by major minima and maxima. Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics, 2003.
4. E. Keogh, K. Chakrabarti, et al.: Dimensionality reduction for fast similarity search in large time series databases. Knowl. and Info. Systems, 3(3):263-286, 2000.
5. F. Mörchen: Time series feature extraction for data mining using DWT and DFT. Tech. Report No. 33, Dept. of Maths and CS, Philipps-University Marburg, 2003.
6. T. Palpanas, M. Vlachos, E. Keogh, D. Gunopulos, W. Truppel: Online amnesic approximation of streaming time series. Proc. of ICDE'04.
7. C.-S. Perng, H. Wang, S. R. Zhang, D. S. Parker: Landmarks: a new model for similarity-based pattern querying in time series databases. Proc. of ICDE'00.
8. Y. Zhao, S. Zhang: Generalized dimension-reduction framework for recent-biased time series analysis. IEEE Trans. on Knowl. and Data Eng., 18(2):231-244, 2006.
9. Y. Zhao, C. Zhang, S. Zhang: A recent-biased dimension reduction technique for time series data. Proc. of PAKDD'05.