

Alignment of Shifted Signals in Time Series

Recep Furkan Koçyiğit, Ahmet Elbir
Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü
furkankocyiigitfk@gmail.com, aelbir@gmail.com

Özetçe —Zaman düzleminde olan sinyallerin hizalanması, geniş uygulama alanı sebebiyle önemli sinyal işleme problemleri arasında yer almaktadır. Bu çalışmada ön işleme teknikleri kullanılarak zaman düzleminde olan sinyallerin gürültüsü azaltılmış, taban hizası düzeltilmiş ve tepe değeri belirginleştirilmiştir. Elde edilen yeni sinyaller dinamik zaman eğrilmesi, polinomal zaman eğrilmesi veya korelasyon eniyilemeli eğrilme yöntemleri kullanılarak hizalanmıştır ve elde edilen sonuçların performans karşılaştırılması yapılmıştır. Çalışma kapsamında GTZAN veri tabanı kullanılmış ve en yüksek başarı PTW yöntemiyle elde edilmiştir.

Anahtar Kelimeler—Dinamik Zaman Eğrilmesi, Korelasyon Eniyilemeli Eğrilme, Parametrik Zaman Eğrilmesi, Sinyallerin Hizalanması.

Abstract—Alignment of signals in the time plane is one of the major signal processing problems due to its wide range of applications. In this study, by using preprocessing techniques, the noise of the signals in the time plane was reduced, the base alignment was corrected and the peak value was determined. The new signals obtained were aligned using dynamic time warp, polynomial time warp or correlation optimized warp methods and the results were compared. The GTZAN database was used in the study and the highest success was achieved with PTW method.

Keywords—Dynamic Time Warping, Correlation Optimized Warping, Parametric Time Warping, Alignment of Signals

I. INTRODUCTION

Important scientific studies are being carried out to learn the structure of viruses such as HIV and HCV that threaten the health of millions of people and to produce drugs that will in effect the functioning of these viruses that negatively affect human health and also to diagnose from the heart. It is necessary to analyze the data of the virus RNA used in these studies and the sound signals formed in the heart. In order to analyze time series quickly and accurately, signals should be aligned with some techniques. Through this study, experts working on these issues will be able to better analyze the data.

Analysis and alignment of time series have an important role in meaningful extraction from biological data. In this study, dynamic programming based Dynamic Time Warp-DTW approach has been examined to solve the problem [1] [2] [3] [4] [5] [6] [7] [8] [9]. In addition, in many scientific studies, the most commonly used Correlation Optimized Warping-COW and Parametric Time Warping have been emphasized [10] [11] [12] [13]. DTW-based approaches are widely used to align time series. The classical DTW algorithm used for the first time alignment of audio signals has been modified with different approaches to solve this problem more effectively.

To compare performance results, all warping methods has been applied on the GTZAN dataset. In the second section, methods used for preprocessing and warping has been elaborated. Experimental results has been summarized and discussed in the third section and the last section includes conclusion.

II. METHODS

A. Preprocessing

Smoothing is a method that minimizes noise or outliers in a numerical data series. With this method, some outliers such as erroneous peaks in the raw data are eliminated. Although there are many softening techniques in the literature, three different techniques have been performed in this study.

- **Smoothing with Arithmetic Mean:** This smoothing technique is often used in mathematics as the sum of N numbers divided by N number is applied. The new values obtained as a result of the arithmetic mean in this region reduce the outliers in that region.
- **Smoothing with Weighted Arithmetic Mean:** A weighted arithmetic mean is preferred, especially for better detection of peaks in signals with peaks.
- **Smoothing by Gauss Density Function:** In this smoothing technique, it is aimed to determine coefficients by using probabilistic distribution instead of giving certain coefficients and to make smoothing process with these coefficients. In this way, a more successful smoothing occurs and the disturbing effects found in the original data can be eliminated.

Resolution Enhancement: The peak highlighting or sharpening step can be used to highlight the peaks in the raw data. After the signal smoothing process, some peaks in the original signal may be disturbed and the peak values may be lost. Peak valuation techniques can be used to avoid such distortions. It is generally done using second and fourth order derivatives. Firstly, the second or fourth order derivative of the raw data is taken and a new time series is obtained. This multiplied signal is then multiplied by a constant k to remove the raw data, resulting in a new signal whose peaks are more pronounced [2] [3] [13] [14].

Baseline Adjustment: Changes in the base level of the signals in the time plane may occur. In order to perform signal alignment successfully, the signals to be aligned must have the same base level. However, the change in base level prevents the accurate determination of the height value of the peaks of the raw data at hand. By adjusting the level

of the base level, the peaks of the raw data are made more pronounced and it is easier to accurately determine the level of these peaks. Linear interpolation was used to perform base level correction[2] [3] [14]. In order to perform base level correction, the raw data must be divided into pieces of a certain length and the smallest value in each piece must be found. From these values, a new series is formed by linear interpolation.

B. Warping

Dynamic Time Warping:Dynamic Time Warping-DTW algorithm aims to find the highest possible match between two time series. In two time series, it spins non-linearly by aligning the regions where the distance is less and achieving the minimum distance in this sequence. DTW is performed by skewing the time axes until the maximum possible similarity is found between the two sequences. Furthermore, many modifications may be applied to this method to increase time or alignment success. These; time and match penalties, receiving the first order derivative of the signal, can be given as different distance functions [2] [3] [4] [5] [9] [15].

For the alignment of the two signals with the dynamic time warp algorithm, $X = (x_1, x_2, \dots, x_N)$, $Y = (y_1, y_2, \dots, y_M)$ and $N = M$ are assumed. Firstly, the distance value of each element in these two series from each signal to the cost matrix of $N * M$ dimension is placed by a distance finding method. Every $M[i][j]$ equal to distance between $X[i]$ and $Y[j]$. Since the numerical value of the distance between two points will affect the success of the alignment process, this value is calculated with different distance functions.

Correlation Optimized Warping: The second method for signal alignment is the correlation optimized warping method. This method is developed by Nielsen and his friends. For alignment of chromatogram data, is technically very similar to dynamic time warp [9] [10]. Basically, the minimum distance is found in the DTW method and the maximum similarity or correlation value is used in the COW method. However, this similarity value is not calculated between two points, but is calculated between the sub-parts of the signal [9] [10]. The COW method is basically implemented with two parameters. The first is the length of the part m used in the division of the raw signal into small pieces. Thus, both signals are divided into n pieces and the length of each piece is m . In the alignment step, the limit values of the points with the most similarity between these parts are determined and warped. The slack (s) parameter shows the variation between the boundary values at the stage of finding the most similarity of the mutually divided segments. The sub-segments are reconstructed between the points with the highest valence similarity among all segments of the signal to determine which part to match with which part. In this method proposed by Nielsen and his friends, Pearson Correlation Coefficient was used as the similarity value[9] [10].

Parametric Time Warping: One of the commonly used methods to eliminate the distortion in the time axis is the parametric time warp method. This method was developed by Eiler and his friends to align chromatogram data,

which can run faster than other methods using dynamic programming, and allocates less memory [9] [12]. Basically, the interpolation is performed using a quadratic function to align the two signals, and the signal is aligned. [15].

The two signals with distortion in the time axis are defined as time varying by the notation $X_i = X(t_i)$ and $Y_i = Y(t_i)$. The skew function, which will allow the alignment of these two signals, is determined as $W(t_i)$ [9] [12] [15]. In the parametric time warp method, the X signal that is specified as a reference is aligned to the Y signal by the $X(W(t_i))$ operation by means of the skew function [9] [12] [15].

III. EXPERIMENTAL RESULTS

This section summarizes the results according to the alignment performance criteria as a result of applying the warping methods used in signal alignment to different data sets. Different methods to solve the time shift problem in raw data and the criteria that determine which parameter will work better have been tried in the sample data sets and summarized in the tables below. Correlation and root mean square(RMS) values of unaligned sample data set are given.

Test ID	Correlation	RMS
1	0.67	335.5
2	0.62	12008.35
3	0.82	9955.37
4	0.45	10044.82
5	0.52	8236.2
6	0.61	6900.37
7	0.66	4713.27

Table 1 Some Numerical Values of Raw Datas

After the data set is aligned with the DTW method, the correlation, RMS, the percentage of matching of the peaks of the aligned signal and the calculation times are shown.

Test ID	Correlation	RMS	PMS	Time
1	0.99	8159537.93	72.62	3.84
2	0.76	5599247.48	19.0	4.15
3	0.9	13330998.6	30.83	4.02
4	0.84	4888473.11	28.78	4.05
5	0.57	119449441.94	26.49	4.8
6	0.62	10890938.45	22.86	4.0
7	0.71	21786624.45	18.85	3.94

Table 2 Numerical Values of Alignment with DTW Method

When the numerical data of the unaligned and signals that aligned with DTW method are compared, the correlation value between the new signals obtained after alignment is increased. Peak matching percentage was found to be 72.62 percent. During these operations, the gap penalty was selected as 3 and the time penalty as 0.

When the numerical data of the unaligned and signals that aligned with COW method are compared, the correlation value between the new signals obtained after

Test ID	Correlation	RMS	PMS	Time
1	0.96	17083118.63	62.92	3.65
2	0.62	7476053.03	23.3	6.09
3	0.85	18709877.25	34.16	6.16
4	0.59	11029956.78	26.53	6.22
5	0.49	125763345.12	22.95	7.42
6	0.46	13172644.47	18.08	6.22
7	0.75	21829318.66	29.02	6.11

Table 3 Numerical Values of Alignment with COW Method

alignment is increased. Peak matching percentage was found to be 62.82 percent. During these operations, the segment length was selected as 10 and the slack as 3.

Test ID	Correlation	RMS	PMS	Time
1	1.00	814.13	74.44	0.09
2	0.64	11141.39	17.75	0.44
3	0.8	6487.67	18.45	0.41
4	0.75	2008.92	22.4	0.27
5	0.52	7372.79	20.77	0.44
6	0.67	5801.7	18.64	0.14
7	0.88	8450.08	17.86	0.42

Table 4 Numerical Values of Alignment with PTW Method

When the numerical data of the unaligned and signals that aligned with PTW method are compared, the correlation value between the new signals obtained after alignment is increased. Peak matching percentage was found to be 74.14 percent. During these operations, the gap penalty was selected as 3 and the warping function as x . This method works faster than other methods and has more peak matching percentages.

IV. CONCLUSION

As can be seen from all the results summarized in the previous section, the highest correlation and peak mapping percentage values after signal alignment were calculated very close to each other on average in all methods. PTW method stands out compared to other methods in terms of signal alignment in a short time. However, the PTW method did not achieve the desired success when there was base alignment or noise in the signal. In these cases, DTW and COW methods yielded more successful results. In order to adjust the base level within the signal, the recommended base line adjustment method is insufficient. Derivative-based alignment used in the DTW method, cost matrix calculation using different distance functions, time penalty and gap penalty modifications positively affected the alignment results both in terms of accuracy and time. Because the calculation complexity of the correlation process used in the COW method is high, the calculation time is longer in cases where the number of data is higher. For this reason, cosine similarity or different similarity functions can be used in future studies.

REFERENCES

- [1] F. Karabiber, J. L. McGinnis, O. V. Favorov, and K. M. Weeks, "Qushape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis," *Rna*, vol. 19, no. 1, pp. 63–73, 2013.
- [2] F. Karabiber, "An automated signal alignment algorithm based on dynamic time warping for capillary electrophoresis data," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 21, no. 3, pp. 851–863, 2013.
- [3] F. Karabiber and M. Balcilar, "A signal alingment method based on dtw with new modification," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2013, pp. 1–4.
- [4] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proceedings of the 2001 SIAM international conference on data mining*. SIAM, 2001, pp. 1–11.
- [5] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.
- [6] V. Niennattrakul and C. A. Ratanamahatana, "On clustering multimedia time series data using k-means and dynamic time warping," in *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*. IEEE, 2007, pp. 733–738.
- [7] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [8] P. Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, vol. 855, no. 1-23, p. 40, 2008.
- [9] T. G. Bloemberg, J. Gerretzen, A. Lunshof, R. Wehrens, and L. M. Buydens, "Warping methods for spectroscopic and chromatographic signal alignment: a tutorial," *Analytica chimica acta*, vol. 781, pp. 14–32, 2013.
- [10] N.-P. V. Nielsen, J. M. Carstensen, and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping," *Journal of chromatography A*, vol. 805, no. 1-2, pp. 17–35, 1998.
- [11] G. Tomasi, F. Van Den Berg, and C. Andersson, "Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 5, pp. 231–241, 2004.
- [12] A. Van Nederkassel, M. Daszykowski, P. Eilers, and Y. Vander Heyden, "A comparison of three algorithms for chromatograms alignment," *Journal of Chromatography A*, vol. 1118, no. 2, pp. 199–210, 2006.
- [13] B. Írmikci, A. Elbir, and F. Karabiber, "Jtimewarp: A software for aligning biological signals using warping methods," 2013.
- [14] F. Karabiber, "A peak alignment algorithm with novel improvements in application to electropherogram analysis," *Journal of bioinformatics and computational biology*, vol. 11, no. 05, p. 1350011, 2013.
- [15] P. H. Eilers, "Parametric time warping," *Analytical chemistry*, vol. 76, no. 2, pp. 404–411, 2004.