

Extended Repository of Warped Distances and Additional Experimental Results Based on Generalized Dynamic Time Warping

I. ADDITIONAL DISCUSSION ON THE GDTW FRAMEWORK

Discussion of Practical Requirements for The Newly Designed Distances We examine our newly “warped distances” in light of practical desiderata for designing effective distance functions [1].

(1) User-centered. In data mining applications, the choice of similarity distance rests with the analysts and their perception of the results. We created the GDTW framework to enable analysts to use distances appropriate for their application domain which were previously not meaningful for comparing sequences with different lengths and alignments. Our experimental evaluation reveals that improvement in classification accuracy is indeed achieved by one of our new distances compared to the state-of-the-art DTW.

(2) Compactness. A new distance should be efficiently computable because its computation will likely be repeated many times, especially for large datasets. In Sec. 3.3 in the paper we show effective strategies for computing our new distances, as well as a general strategy toward achieving “compactness” of any new warped distance.

(3) Interpretability. Being able to know which distance provides the most influential insights into specific datasets is crucial for analysts. This interpretability of a distance mostly comes from the ability to explore domain specific datasets using diverse distances and comparing the results. Our framework gives analysts the opportunity to perform such comparative experiments and use the results in conjunction with their knowledge to decide the best similarity model for specific applications.

II. EXTENDED REPOSITORY OF WARPED DISTANCES

A. Warping the Cosine Distance

The Cosine distance, which measures the angles between two vectors, corresponds to the normalized Inner Product. In the literature, other names for this popular distance include: angular metric [4], Ochiai [4] and Carbo [10].

Given the Cosine distance d_{Cos} between two sequences X and Y defined as:

$$d_{Cos}(X, Y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}, \quad (1)$$

its **recursive expression** is:

$$f_{d_{Cos}}\left(\frac{a}{bc}, x_n, y_n\right) = \frac{a + x_n y_n}{\sqrt{b^2 + x_n^2} \sqrt{c^2 + y_n^2}} = \frac{a'}{b'c'}.$$

with a , b and c denoting respectively the sum of the product of the first $n-1$ coordinates, the sum of squares for the first $n-1$ values of x and the sum of squares for the first $n-1$ values of y . Lastly, its **dynamic programming recurrence** becomes:

$$\gamma(i, j) = \min \left\{ \begin{array}{l} \frac{a_1 + x_i y_j}{\sqrt{b_1^2 + x_i^2} \sqrt{c_1^2 + y_j^2}}, \\ \frac{a_2 + x_i y_j}{\sqrt{b_2^2 + x_i^2} \sqrt{c_2^2 + y_j^2}}, \\ \frac{a_3 + x_i y_j}{\sqrt{b_3^2 + x_i^2} \sqrt{c_3^2 + y_j^2}}, \end{array} \right.$$

$$\gamma(i-1, j-1) = \frac{a_1}{b_1 c_1}, \gamma(i-1, j) = \frac{a_2}{b_2 c_2}, \gamma(i, j-1) = \frac{a_3}{b_3 c_3}.$$

We showcase below how our Design Tool can assist in the implementation of this distance. Focusing on the right side of the equation, we see that there are three separate terms, the running sum of products and the running sum of individual squares of x and y respectively. The cache stores these three values.

- `init() = CosineCache()`
- `reduce() = new_a where {`
`new_a.sumSqXi = a.sumSqXi + Xi^2`
`new_a.sumSqYi = a.sumSqYi + Yi^2`
`new_a.sumXiYi = a.sumXiYi + Xi * Yi`
`}`
- `norm() =`
`a.sumXiYi / (sqrt(a.sumSqXi) * sqrt(a.sumSqYi))`

B. Warping Other Distances

As shown by the examples above, it is clear that all the distance measures in [2] can be warped by our GDTW methodology. These distances which are essential for classification, clustering and retrieval problems, are reviewed and categorized syntactically and semantically into eight families in [2]. Different distances in our case study correspond to different families described in this survey. Aside from the distances that we demonstrate, many other popular distances such as Jaccard, Dice and Pearson – which are based on similar arithmetic expressions – can be efficiently warped using our methodology.

III. FUTURE WORK

Studies leveraging variants of *GDTW* could now further contribute to solving a broad range of problems including

but not limited to classification, clustering, addressing singularities such as “over-warping”, etc. While our paper attempts to look beyond the competitive “apple polishing” task of improving the accuracy of time series classification [3], [5], it opens the avenue for important new research. That is, other studies leveraging variants of $GDTW$ could now further capitalize on these newly warped distances. Possible future studies on time series exploration include:

(1) If we could predict which version of $GDTW$ is most suitable on a *domain-by-domain* basis, we could assist analysts by providing this variant as the most suitable distance for their application.

One might achieve this by doing cross-validation on the training set; however this may be difficult to do robustly for time series domains, which typically have very small training sets.

(2) If we could predict which version of $GDTW$ is most likely to be correct on an *exemplar-by-exemplar* basis (much like a “gating network”), we could choose the class label of the most “confident” variant [6] for classification. This idea has been used in many settings, but to our knowledge it has not been explored in the time series context. The time considerations are inconsequential in most practical settings because all variants of $GDTW$ can exploit the lower bounding and early abandoning ideas in [11].

(3) Finally, the current state-of-the-art classifiers for time series are all Ensemble Classifiers [9]. The base methods in the Ensemble include existing variants of DTW, such as derivative DTW [8] and weighted DTW [7]. It is quite possible, according to our results, that increasing the diversity of the Ensemble with additional distances, such as our newly developed $GDTW_{MD}$, $GDTW_{Mink}$ or other of the distances that we can now warp, will further improve the accuracy of the state-of-the-art ensemble classifiers.

IV. ADDITIONAL EXPERIMENTAL EVALUATION

A. Experiment on Time Series Classification

A spreadsheet with the raw results is archived along with our code and the rest of the details of the experiments¹. More detailed pairwise comparisons using error-rate binary plots can be found in Fig. 1, 2 and 3. In addition to the summaries provided in Sec. 5.2.1, we include binary plots showing the comparisons $GDTW_{ED}$ vs $GDTW_{MD}$, $GDTW_{ED}$ vs $GDTW_{Mink}$ and $GDTW_{MD}$ vs $GDTW_{Mink}$ in Fig. 1, 2, and 3.

Fig. 1 is based on the error rates for $GDTW_{ED}$ and $GDTW_{MD}$ over the 85 datasets during the classification experiment. The plot indicates the areas where $GDTW_{ED}$ performs better than $GDTW_{MD}$, as well as the area where $GDTW_{MD}$ performs better than $GDTW_{ED}$. We note that although the performance seems very close, $GDTW_{MD}$ “wins” more times than $GDTW_{ED}$. Fig. 2 displays a pairwise

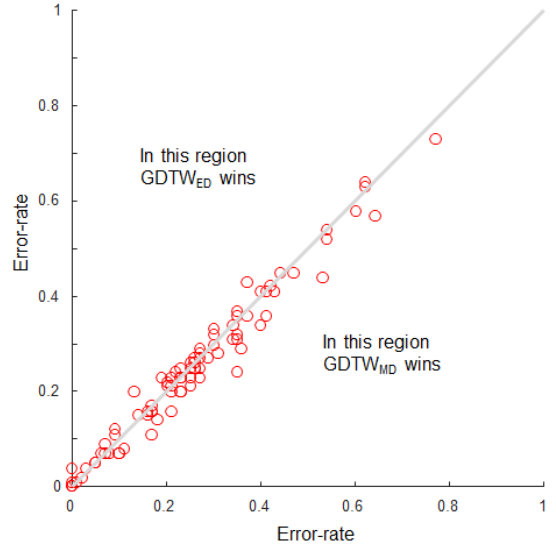


Fig. 1: Error-rate Binary Plot $GDTW_{ED}$ vs $GDTW_{MD}$

comparison of $GDTW_{ED}$ and $GDTW_{Mink}$ based on error rates. The area where $GDTW_{Mink}$ wins is much smaller than the area where $GDTW_{ED}$ wins, but it is important to note that there are 15 times where $GDTW_{Mink}$ has higher accuracy for classifying the time series in the 85 datasets. Fig. 3 offers the

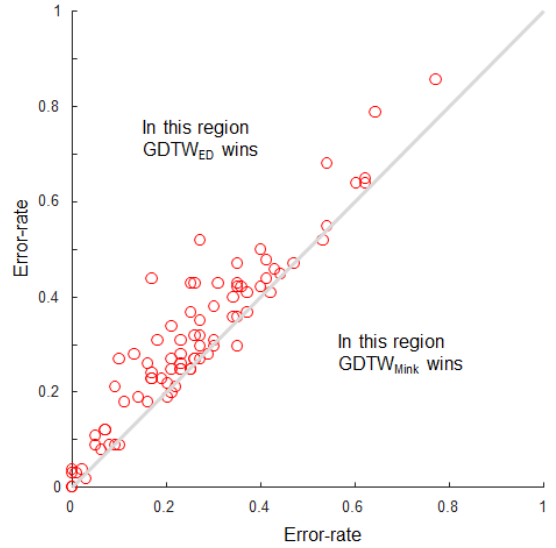


Fig. 2: Error-rate Binary Plot $GDTW_{ED}$ vs $GDTW_{Mink}$

pairwise comparison between $GDTW_{MD}$ and $GDTW_{Mink}$. We note again that the area where $GDTW_{Mink}$ wins is much smaller than the one where $GDTW_{MD}$ wins.

B. Experiment on Best Match Retrieval

Tables IV, V, and VI show the best match sequences in the ECG dataset using ED compared to classic DTW, then using Manhattan Distance (MD) compared to $GDTW_{MD}$, then using Chebyshev or Minkowski Distance (Mink) compared to $GDTW_{Mink}$.

¹<https://github.com/gdtw/GDTWSysstem>

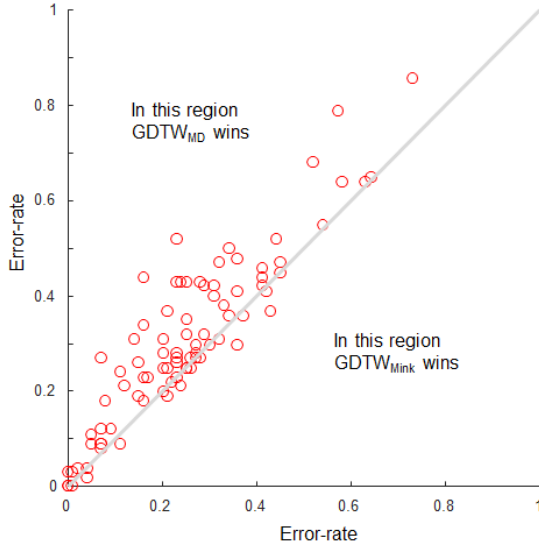


Fig. 3: Error-rate Binary Plot $GDTW_{MD}$ vs $GDTW_{Mink}$

We display in Tables I, II, and III the best match sequences in the ItalyPower dataset using ED compared to classic DTW, then using Manhattan Distance (MD) compared to $GDTW_{MD}$, then using Chebyshev or Minkowski Distance (Mink) compared to $GDTW_{Mink}$. The “best match” is displayed as the sequence “TS” followed by a number indicating the position of the time series in the dataset and the start and end points of the match sequence inside the square brackets.

Analyzing and summarizing the results, we note that the

TABLE I: Best match comparison ED vs DTW in ItalyPower

Query	ED	Match ED	DTW	Match DTW
TS0[0,23]	0.36	TS22[0,23]	0.22	TS53[1,23]
TS1[18,23]	0.06	TS37[18,23]	0.06	TS37[18,23]
TS2[0,11]	0.05	TS43[2,13]	0.05	TS43[2,13]
TS3[2,13]	0.09	TS31[2,13]	0.05	TS46[5,11]
TS4[2,13]	0.10	TS18[2,23]	0.35	TS0[3,14]
TS5[10,23]	0.07	TS45[10,23]	0.07	TS30[12,23]
TS6[5,23]	0.16	TS52[5,23]	0.12	TS36[5,23]
TS7[0,23]	0.08	TS29[0,23]	0.82	TS29[0,23]
TS8[0,23]	0.13	TS7[0,23]	0.11	TS56[0,23]
TS9[0,21]	0.20	TS31[0,21]	0.10	TS31[0,21]

TABLE II: Best match comparison MD vs $GDTW_{MD}$ in ItalyPower

Query	MD	Match MD	$GDTW_{MD}$	Match $GDTW_{MD}$
TS0[0,23]	0.90	TS22[0,23]	0.90	TS53[1,23]
TS1[18,23]	0.13	TS37[18,23]	0.47	TS36[0,23]
TS2[0,11]	0.13	TS43[2,13]	0.13	TS43[2,13]
TS3[2,13]	0.26	TS31[2,13]	0.16	TS46[5,12]
TS4[2,13]	0.26	TS43[2,23]	0.18	TS11[2,11]
TS5[10,23]	0.23	TS45[10,23]	0.22	TS30[12,23]
TS6[5,23]	0.52	TS36[5,23]	0.39	TS36[5,23]
TS7[0,23]	0.31	TS29[0,23]	0.30	TS29[0,23]
TS8[0,23]	0.49	TS7[0,23]	0.43	TS56[0,23]
TS9[0,21]	0.33	TS31[0,21]	0.32	TS31[0,21]

distance between the sample query sequence and the best match is minimal, as expected, when using warped distances instead of their counterpart point-wise versions. More detailed

TABLE III: Best match comparison Mink vs $GDTW_{Mink}$ in ItalyPower

Query	Mink	Match Mink	$GDTW_{Mink}$	Match $GDTW_{Mink}$
TS0[0,23]	0.16	TS22[0,23]	0.09	TS53[1,23]
TS1[18,23]	0.05	TS37[18,23]	0.13	TS37[18,23]
TS2[0,11]	0.13	TS43[2,13]	0.13	TS43[4,13]
TS3[2,13]	0.05	TS17[2,13]	0.04	TS46[5,12]
TS4[2,13]	0.07	TS18[2,13]	0.03	TS35[4,15]
TS5[10,23]	0.04	TS45[10,23]	0.04	TS45[12,23]
TS6[5,23]	0.07	TS33[5,23]	0.06	TS33[5,23]
TS7[0,23]	0.05	TS19[0,23]	0.03	TS60[0,23]
TS8[0,23]	0.07	TS7[0,23]	0.05	TS21[1,23]
TS9[0,21]	0.05	TS34[0,21]	0.05	TS17[0,23]

analysis reveals that using diverse distances leads to retrieving different best matches for the same sample sequence. In 40% of the scenarios, the ED and the classic DTW found the same best match. Also in 40% of the cases, MD and $GDTW_{MD}$ found the same best match. The best match was the same for Minkowski and $GDTW_{Mink}$ in only 20% of our trials. The results confirm that *warped distances are better suited than point-to-point distances to perform similarity searches*, except for the special situations when the sequences are aligned and have the same length. In that case, both versions (warped and point-wise) of a specific distance return the same result.

We provide an additional comparison of the three warped distances that are created with our framework. In 70% of the cases, the classic DTW and $GDTW_{MD}$ returned the same match, but in only 10% of the scenarios, the matches retrieved by the classic DTW and $GDTW_{Mink}$ were matched. 20% of the scenarios returned the same match for $GDTW_{MD}$ and $GDTW_{Mink}$. All three warped versions returned the same result only 10% of the time. This is another indication that the best match for a sample sequence is highly dependent on the chosen distance. In few cases the results produced by the warped distances are the same, which means these new distances can be useful in “discovering” best matches that otherwise would not be found. Bellow we offer the detailed results in retrieving the best match for the ECG dataset.

TABLE IV: Best match comparison ED vs classic DTW in ECG

Query	ED	Match ED	DTW	Match DTW
TS0[0,14]	0.15	TS39[0,14]	0.10	TS39[0,14]
TS1[17,47]	0.13	TS77[18,48]	0.11	TS198[16,50]
TS2[0,9]	0.07	TS140[9]	0.05	TS178[0,8]
TS3[6,52]	0.10	TS125[6,52]	0.07	TS188[5,40]
TS4[16,75]	0.19	TS91[18,77]	0.06	TS117[8,52]
TS5[0,30]	0.18	TS30[0,30]	0.13	TS30[0,30]
TS6[0,96]	0.24	TS46[0,96]	0.13	TS46[0,95]
TS7[2,15]	0.11	TS36[2,15]	0.08	TS131[27,96]
TS8[26,96]	0.14	TS151[26,96]	0.08	TS56[0,23]
TS9[91,95]	0.007	TS134[50,54]	0.004	TS21[72,75]

Table V shows the best match sequences found using Manhattan distance and its warped version $GDTW_{MD}$. In 10% of the scenarios, the best match found was the same for the Manhattan distance and its warped version.

Table VI showcases the results for finding the best match using Minkowski or Chebyshev distance and its warped version. In this case, the best match was never the same for the point-to-point distance and its warped version. In 50%

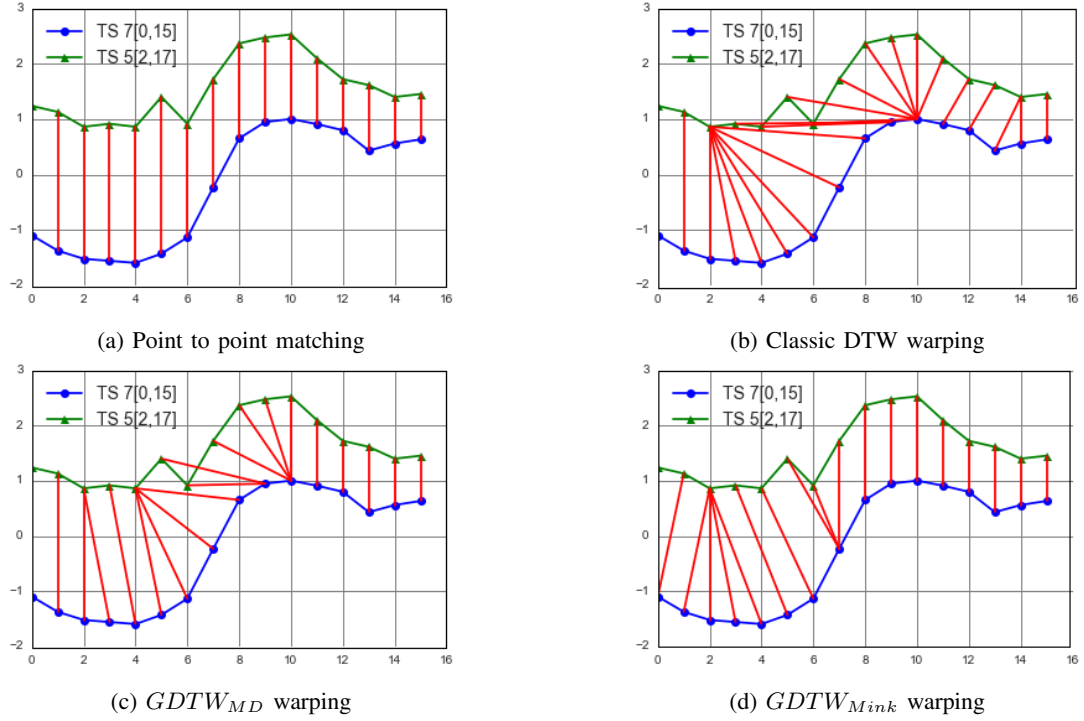


Fig. 4: Warping for a pair of sequences in ItalyPower.

TABLE V: Best match comparison MD vs $GDTW_{MD}$ in ECG

Query	MD	Match MD	$GDTW_{MD}$	Match $GDTW_{MD}$
TS0[0,14]	0.40	TS39[0,14]	0.33	TS39[0,14]
TS1[17,47]	0.59	TS153[18,48]	0.44	TS103[19,48]
TS2[0,9]	0.20	TS140[0,9]	0.14	TS178[0,8]
TS3[6,52]	0.57	TS125[6,52]	0.41	TS188[5,40]
TS4[16,75]	1.20	TS91[18,77]	0.43	TS117[8,52]
TS5[0,30]	0.75	TS30[0,30]	0.61	TS65[0,28]
TS6[0,96]	1.82	TS46[0,96]	1.14	TS46[0,95]
TS7[2,15]	0.36	TS36[2,15]	0.28	TS69[5,19]
TS8[26,96]	0.99	TS151[26,96]	0.58	TS131[12,96]
TS9[91,95]	0.01	TS24[64,68]	0.007	TS52[72,75]

TABLE VI: Best match comparison Mink vs $GDTW_{Mink}$ in ECG

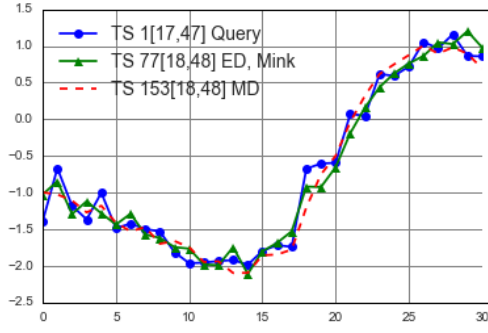
Query	Mink	Match Mink	$GDTW_{Mink}$	Match $GDTW_{Mink}$
TS0[0,14]	0.08	TS138[0,14]	0.06	TS39[1,14]
TS1[17,47]	0.06	TS 60[19,49]	0.04	TS77 [19,47]
TS2[0,9]	0.04	TS140[9]	0.02	TS178[0,8]
TS3[6,52]	0.04	TS125[6,52]	0.02	TS188[5,37]
TS4[16,75]	0.05	TS83[22,81]	0.01	TS91[18,66]
TS5[0,30]	0.07	TS65[0,30]	0.05	TS30 [0,24]
TS6[0,96]	0.08	TS46[0,96]	0.03	TS137[0,95]
TS7[2,15]	0.06	TS36[2,15]	0.04	TS65[5,17]
TS8[26,96]	0.04	TS151[26,96]	0.02	TS131[27,96]
TS9[91,95]	0.005	TS134[50,54]	0.002	TS52[72,75]

of the scenarios, the three point-wise distances returned the same best match, while only 10% the same best match was returned by the warped versions of these distances. In no scenario did all distances, point-wise and warped return the same match.

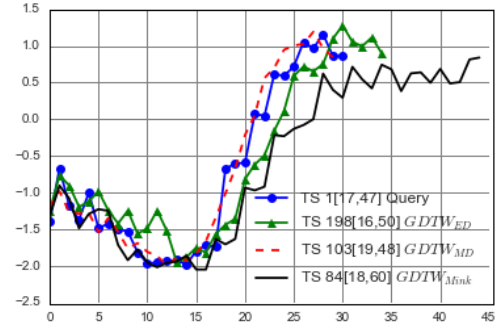
The results can vary significantly when using different point-to-point distances and their warped versions, as expected.

The point-wise distances can theoretically be *at most* as good as their warped versions. Their highest performance is achieved for sequences that are aligned in time. Below we offer the analysis that has been summarized in Table 1 in Sec. 5.2.1 in the paper. In 20% of the cases ED and DTW returned the same sequence as the best match. MD and $GDTW_{MD}$ also return the same match in 20% of the scenarios. Mink and $GDTW_{Mink}$ never returned the same result. We compare the results retrieved by the newly warped distances. In 10% of the cases DTW and $GDTW_{Mink}$ returned the same result, $GDTW_{MD}$ and $GDTW_{Mink}$ found the same match in 20% of the scenarios, while DTW and $GDTW_{MD}$ had the same result in 50% of the cases. The fact that the results are different in most cases attests to the usefulness of our distances in finding matches that otherwise might be overlooked.

In Fig. 5 we show a visual display of an example of the best match in ECG based on using the point-wise and their respectively warped counterpart distances. In this specific case, two of the point-wise distances returned the same best match (ED and Mink), while MD returned a different best match. Each warped distance returned a different match. If our distances would have all returned the same result or even the same result as their warped counterparts, then they would not be useful. It is the diversity of the results that proves the usefulness of these distances. We perform similar experiments on other datasets, but due to space constraints, we only display the results for two datasets. Extended results are available in our archive, along with our code.



(a) Best match using ED, MD, Mink



(b) Best match using DTW, $GDTW_{MD}$, $GDTW_{Mink}$

Fig. 5: Best matches in ECG retrieved with point-wise distances and their warped counterparts

C. Experiment on Warping Characteristics

Evaluating Cardinality of Warpings. Warpings indicate points of a sequence that are either matching or are being matched to more than one point of the other sequence. Points that are matched one-to-one are referred to as matchings. The warpings created by $GDTW_{MD}$ and $GDTW_{Mink}$ are fewer and more intuitive than the ones created by $GDTW_{ED}$, which is a similar conclusion with the conclusion of the experiments of [8]. This shows that indeed the classic DTW can “over-warp”, mainly due to the fact that it incorporates the ED as base distance. This knowledge can be useful to analysts who might choose to use distances that produce fewer warpings.

We find no warpings when using the point-wise distances, as expected, in Fig. 4a. The chart in (b) shows that the classic DTW (same as our $GDTW_{ED}$) warps 23 times. Examining the charts in (c) and (d), we see that $GDTW_{MD}$ warps 21 times for the same pair of sequences, while $GDTW_{Mink}$ warps only 17 times. It is observable on the charts that the warpings produced by this last distance are to points that are closer to each other and are therefore more intuitive. $GDTW_{Mink}$ is always smaller than the classic DTW and $GDTW_{MD}$, making it a better choice for analysts in application domains where only very small distances are considered to express similarity.

In summary, many possible alignments of the same pair of sequences arise, depending on the chosen distance. Analysts can choose the easier to interpret alignments for specific datasets by deciding what distance to use.

D. Studying Heart Arrhythmia Using $GDTW$ variants

Examples of such signals include ECG and heart rate recordings from patients at a high risk of sudden death and healthy control subjects. Other examples include fluctuations of hormone and other molecular biological signal messengers, transducers in neuroendocrine dynamics, multiparameter dynamics, and multiparameter recordings in sleep apnea and epilepsy.

We randomly selected records from the dataset, including that of the patient with record 107 and asked our cardiologist collaborators to find the best match for each of the samples

displayed in Figure 7. The cardiologists identified the ECG for the patient with record number 113, as having the closest heart rate, meaning average heart rate in beats per minute to the sample record number 107.

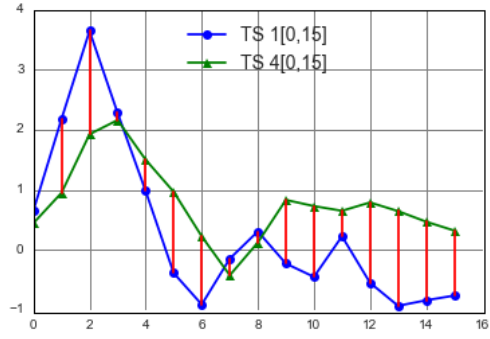
Independent of their findings, we retrieved the best match for the sample sequences by using each of our three $GDTW$ variants. The results are displayed in Table VII. We performed this best match experiment five times using different sample sequences of records as displayed in Figure 7 and arrived to the same conclusion based on comparing the answers provided by the cardiologists with the ones retrieved by our system.

TABLE VII: Best match record found by $GDTW$ variants

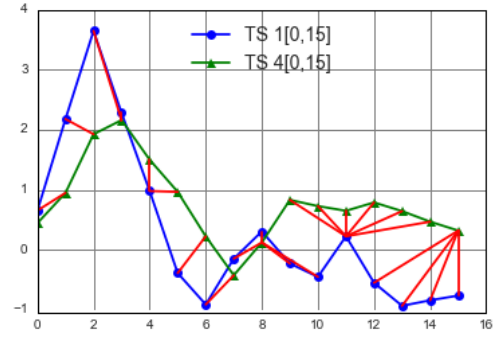
Sample	DTW	$GDTW_{MD}$	$GDTW_{Mink}$	Domain Experts
Sample 1	118	113	116	113
Sample 2	118	113	116	113
Sample 3	113	113	116	113
Sample 4	116	113	116	113
Sample 5	118	113	116	113

REFERENCES

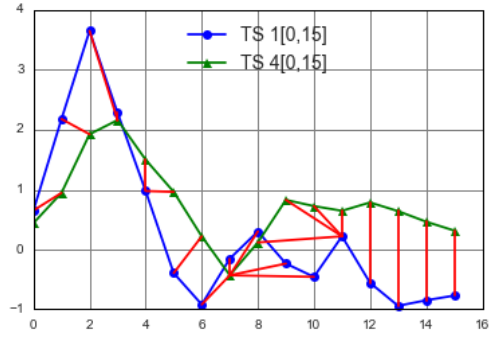
- [1] C. C. Aggarwal. Towards systematic design of distance functions for data mining applications. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 9–18. ACM, 2003.
- [2] S.-H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 2007.
- [3] L. Chen and R. Ng. On the marriage of lp-norms and edit distance. In *Proceedings of the Thirtieth international conference on VLDB-Volume 30*. VLDB Endowment, 2004.
- [4] M. Deza and E. Deza. *Dictionary of distances*. Elsevier, 2006.
- [5] Ding, P. Trajcevski, Scheuermann, X. Wang, and E. Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 2008.
- [6] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton. Adaptive mixtures of local experts. *Neural computation*, 3(1):79–87, 1991.
- [7] Y. Jeong, M. Jeong, and O. Omiaou. Weighted dynamic time warping for time series classification. *Pattern Recognition*, 2011.
- [8] E. Keogh and M. Pazzani. Derivative dynamic time warping. *SIAM*, 2001.
- [9] M. Långkvist, L. Karlsson, and A. Loutfi. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24, 2014.
- [10] V. Monev. Introduction to similarity searching in chemistry. *MATCH Commun. Math. Comput. Chem*, 2004.
- [11] T. Rakthanmanon, Campana, and et. al. Searching and mining trillions of time series subsequences under dynamic time warping. In *18th ACM SIGKDD*. ACM, 2012.



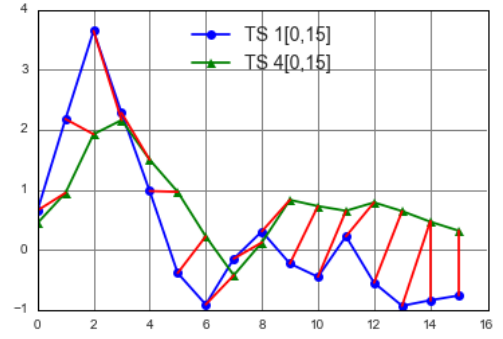
(a) Point-to-point matching



(b) Classic DTW warping

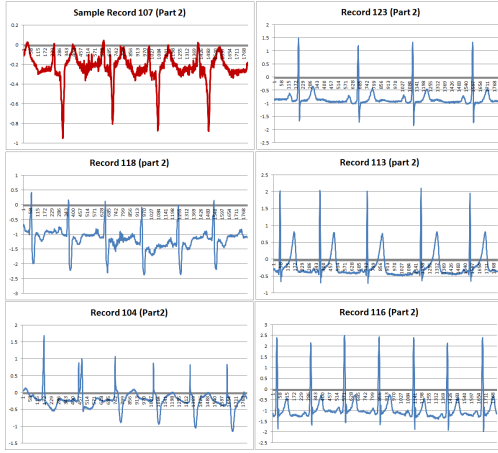


(c) $GDTW_{MD}$ warping

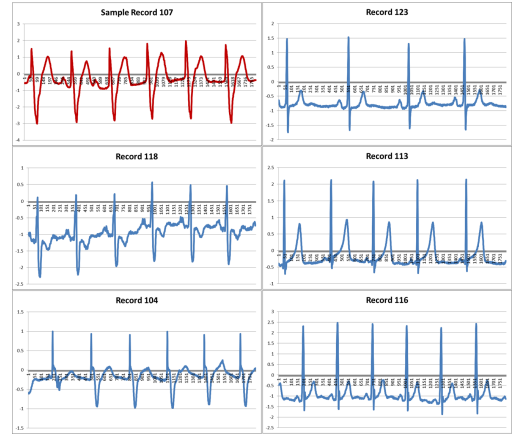


(d) $GDTW_{Mink}$ warping

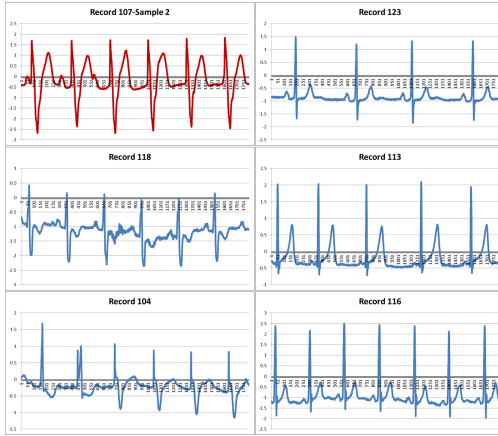
Fig. 6: Warping for a pair of sequences in ECG.



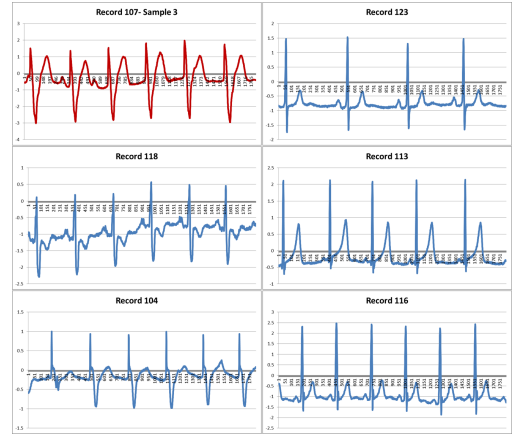
(a) Sample 1 of patient record 107



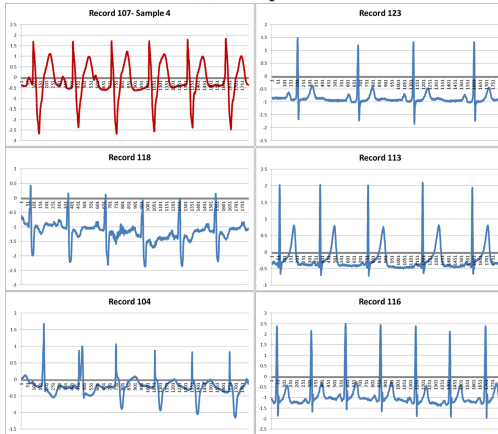
(b) Sample 2



(c) Sample 3



(d) Sample 4



(e) Sample 5

Fig. 7: Samples for best match retrieval for arrhythmia case study