Lower-Bounding of Dynamic Time Warping Distances for Multivariate Time Series

Toni M. Rath and R. Manmatha*
Multi-Media Indexing and Retrieval Group
Center for Intelligent Information Retrieval
University of Massachusetts
Amherst, MA 01002

Abstract

A tight lower-bounding measure for dynamic time warping (DTW) distances for univariate time series was introduced in [Keogh 2002] and a proof for its lower-bounding property was presented. Here we extend these findings to allow lower-bounding of DTW distances for multivariate time series.

1. Introduction

Dynamic Time Warping (DTW) is an algorithm that aligns and compares two time series by calculating a matching error for them. Generally, the error measures obtained with DTW are more intuitive than other measures, e.g. Euclidean distance. Furthermore, the compared time series do not have to have equal lengths. For an introduction to DTW we refer the reader to standard literature, such as [Sankoff & Kruskal 1983].

One drawback of DTW is its high computational cost ($O(mn)$ for comparing sequences of length m and n). This makes the comparison of a large number of time series very expensive, if not impossible for a fixed amount of computational resources. The concept of lower-bounding measures can reduce the cost of many tasks that rely on DTW.

For example, the search for a time series in a data base that is most similar to a provided template can be found more efficiently than using DTW to compare the template to every series in the data base (i.e. sequential scan). A lower-bounding measure is cheap and approximates, but underestimates the actual cost determined by DTW. It can be used to avoid comparing the template to a time series with DTW when the lower-bounding estimate indicates that the time series is a worse match than the current best match (see [Faloutsos 1999] or [Keogh 2002] for a more detailed explanation).

E. Keogh presented a lower-bounding measure for comparing univariate time series, i.e. series where a scalar is recorded at every sampling point. He showed that his lower-bounding measure is tighter than previously proposed measures [Yi et al. 1998, Kim et al. 2001] for a great number of data sets. We have extended his findings to multivariate time series, which record a constant number of scalars per sampling point. Here we present our extension and prove its lower bounding property.

The presentation of the lower-bounding scheme closely follows the notation of the original work [Keogh 2002]. We refer the reader to that publication and to [Faloutsos 1999] for an introduction to the idea of lower-bounding. Without this theoretical background, the presentation in this work will be difficult to follow.

2. Multivariate Time Series Extension

In [Keogh 2002] the time series

$$\begin{aligned} C &= c1, c2, \ldots, c_i, \ldots, c_m \quad \text{and} \\ Q &= q1, q2, \ldots, q_j, \ldots, q_n \end{aligned}$$

are univariate. That is, the elements of Q and C are scalar: $c_i, q_j \in R$. In this work, Q and C are multivariate time series: $c_i, q_j \in R^l$, where l is an integer constant ≥ 1. As a notational convention, when referring to the p-th dimension of a time series element we use an extra subscript, e.g. $c_{i,p}$. For $l = 1$ the lower-bounding measure presented here reduces to that in [Keogh 2002], which we refer to as *LB_Keogh*.

2.1. Local Distance Measure

In order to align time series, a distance measure $d(\cdot, \cdot)$, which allows the similarity assessment of positions in two time series, must be defined. We call this the *local* distance

*This work was supported in part by the Center for Intelligent Information Retrieval and in part by the National Science Foundation under grant number IIS-9909073. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor.

measure, as opposed to the global "distance" between the two time series. *LB_Keogh* uses

$$d(c_i, q_j) = (c_i - q_j)^2$$

for this distance. For comparing multivariate time series components, we use

$$d(c_i, q_j) = \sum_{p=1}^{l} (c_{i,p} - q_{j,p})^2 \qquad (1)$$

in this work. Note that in this distance measure, which is the square of the Euclidean distance, each dimension of c_i and q_j contributes an equal amount to the total error. In practice, the ranges of all dimensions of C and Q are normalized to make them comparable.

Using the local distance measure, the DTW algorithm finds a warping path

$$W = w_1, w_2, \dots, w_k, \dots, w_K, \quad w_k = (i, j)_k,$$

which aligns corresponding locations (i.e. indices i and j) in the two time series C and Q. The length K of all warping paths W satisfies

$$max(m, n) \leq K \leq m + n - 1 \qquad (2)$$

The path W that DTW recovers is the one with minimum accumulated cost, i.e.

$$DTW(Q, C) = min_W \left\{ \sqrt{\sum_{k=1}^{K(W)} d(w_k)} \right\}, \qquad (3)$$

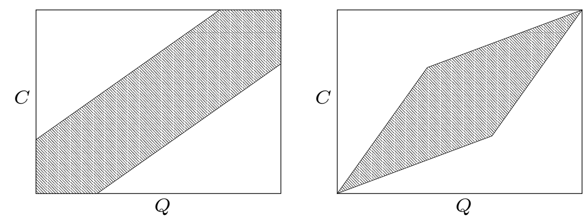
where $d(w_k) = d(c_i, q_j)$ (corresponding i and j at position k in warping path). K is a function of W, since warping paths can vary in length.

2.2. Lower-Bounding Measure

LB_Keogh exploits the fact that most DTW applications use global path constraints when comparing time series: i and j in $w_k = (i, j)$ are constrained to $j - r \leq i \leq j + r$, where r depends on the kind of path constraint that is used (e.g. Sakoe-Chiba band [Sakoe & Chiba 1978] or Itakura parallelogram [Itakura 1975]; see Figure 1).

Using this fact, two time series U and L (for *upper* and *lower*) can be constructed, such that they define an envelope that the time series Q must lie in, regardless of how much it is skewed under all possible warping paths that are allowed under the global path constraint. In the multivariate case, the original definition

$$\begin{aligned} u_i &= max(q_{i-r} : q_{i+r}) && (4) \\ l_i &= min(q_{i-r} : q_{i+r}) && (5) \end{aligned}$$



(a) Sakoe-Chiba band; (b) Itakura parallelogram;

Figure 1: The two most popular global path constraints used in DTW implementations. The index pairs $(i, j)_k$ are constrained to lie in the shaded regions of the cost matrices.

becomes

$$\begin{aligned} u_i &= (u_{i,1}, u_{i,2}, \dots, u_{i,l}) && (6) \\ l_i &= (l_{i,1}, l_{i,2}, \dots, l_{i,l}) && (7) \end{aligned}$$

where

$$\begin{aligned} u_{i,p} &= max(q_{i-r,p} : q_{i+r,p}) && (8) \\ l_{i,p} &= min(q_{i-r,p} : q_{i+r,p}). && (9) \end{aligned}$$

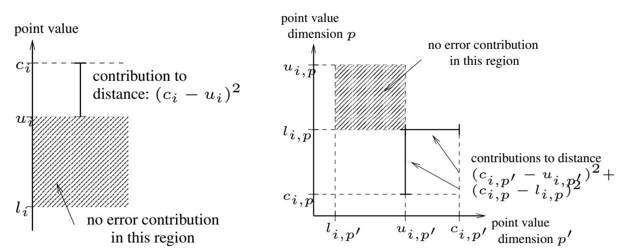
This definition leads to an inequality which we will use in the proof of the lower-bounding measure:

$$\forall i \, \forall p : \quad l_{i,p} \leq q_{i,p} \leq u_{i,p}.$$

Using the new definition of L and U, we can replace the original lower-bounding function *LB_Keogh* with the multivariate lower-bounding measure

$$LB\_MV(Q, C) = \qquad (10)$$

$$= \sqrt{\sum_{i=1}^{n} \sum_{p=1}^{l} \begin{cases} (c_{i,p} - u_{i,p})^2 & \text{if } c_{i,p} > u_{i,p} \\ (c_{i,p} - l_{i,p})^2 & \text{if } c_{i,p} < l_{i,p} \\ 0 & \text{otherwise} \end{cases}} \qquad (11)$$



(a) univariate case (b) multivariate case (2 dimensions shown)

Figure 2: Contributions to lower-bounding distance measure in uni- and multivariate cases.

Figure 2 shows how distance contributions are counted in the uni- and multivariate case of the lower-bounding measure.

2.3. Lower-Bounding Proof

Following the line of thought in E. Keogh's work, we now prove this

Proposition: For any two sequences Q and C of the same length n, for any global constraint on the warping path of the form $j - r \leq i \leq j + r$, the following inequality holds: $LB\_MV(Q, C) \leq DTW(Q, C)$.

Proof: We need to prove

$$\sqrt{\sum_{i=1}^{n}\sum_{p=1}^{l}\begin{cases}(c_{i,p}-u_{i,p})^2 & \text{if } c_{i,p} > u_{i,p}\\ (c_{i,p}-l_{i,p})^2 & \text{if } c_{i,p} < l_{i,p}\\ 0 & \text{otherwise}\end{cases}} \leq \sqrt{\sum_{k=1}^{K} d(w_k)},$$

where $W = w_1, w_2, \ldots, w_k, \ldots, w_K$ is the minimum-cost warping path with length K. Squaring both sides yields

$$\sum_{i=1}^{n}\sum_{p=1}^{l}\begin{cases}(c_{i,p}-u_{i,p})^2 & \text{if } c_{i,p} > u_{i,p}\\ (c_{i,p}-l_{i,p})^2 & \text{if } c_{i,p} < l_{i,p}\\ 0 & \text{otherwise}\end{cases} \leq \sum_{k=1}^{K} d(w_k).$$

We can prove this inequality by showing that for every summation term on the left-hand side there exists an equal or greater term on the right-hand side. Since the length K of the warping path is greater than or equal to n (see equation 2), every term of the summation $\sum_{i=1}^{n} \ldots$ on the left-hand side of the above equation can be matched with a greater or equal term of the summation $\sum_{k=1}^{K} \ldots$ on the right-hand side. Specifically, for a given index i on the left-hand side, we pick k on the right-hand side, such that $w_k = (i, j)$ for some j. An index k with this property is guaranteed to exist because of local continuity constraints. Terms w_k on the right-hand side are not matched more than once, since i is different for every matched term on the left-hand side.

We have left to show

$$\sum_{p=1}^{l}\begin{cases}(c_{i,p}-u_{i,p})^2 & \text{if } c_{i,p} > u_{i,p}\\ (c_{i,p}-l_{i,p})^2 & \text{if } c_{i,p} < l_{i,p}\\ 0 & \text{otherwise}\end{cases} \leq d(w_k).$$

Expanding the right-hand side yields

$$\sum_{p=1}^{l}\begin{cases}(c_{i,p}-u_{i,p})^2 & \text{if } c_{i,p} > u_{i,p}\\ (c_{i,p}-l_{i,p})^2 & \text{if } c_{i,p} < l_{i,p}\\ 0 & \text{otherwise}\end{cases} \leq \sum_{p=1}^{l}(c_{i,p}-q_{j,p})^2,$$

which we can prove by showing that every summation term on the left is less than or equal to the corresponding term on the right. We have three cases:

Case $c_{i,p} > u_{i,p}$:

$$(c_{i,p}-u_{i,p})^2 \leq (c_{i,p}-q_{j,p})^2 \tag{12}$$

We can take the square root of both sides, because the terms in parentheses are positive: the left-hand side is obvious, because of the case $(c_{i,p} > u_{i,p})$ we are treating.

By definition, our global path constraint guarantees $j - r \leq i \leq j + r$, from which we can deduce $i - r \leq j \leq i + r$. Using the definition of $u_{i,p} = max(q_{i-r,p} : q_{i+r,p})$, we get $q_{j,p} \leq u_{i,p}$. Since $u_{i,p} < c_{i,p}$ (definition of case), $c_{i,p} - q_{j,p}$ is positive.

Hence, we get

$$c_{i,p} - u_{i,p} \leq c_{i,p} - q_{j,p} \tag{13}$$
$$-u_{i,p} \leq -q_{j,p} \tag{14}$$
$$q_{j,p} \leq u_{i,p} \tag{15}$$

which is true.

Case $c_{i,p} < l_{i,p}$:

This proof is straightforward with an argument similar to the above.

Case $l_{i,p} \leq c_{i,p} \leq u_{i,p}$:

Trivially we have

$$0 \leq (c_{i,p}-q_{j,p})^2,$$

where the right-hand side is obviously non-negative. ■

Extension to Minkowski metrics: we would like to point out that this lower-bounding approach is not limited to the Euclidean metric, but can easily be extended to arbitrary Minkowski (p-) metrics (modify equations 1, 3 and 11).

3. Conclusion

We have presented and proven an extension of E. Keoghs tight lower-bounding measure to multivariate time series. This measure can form the basis for DTW indexing approaches as shown in [Keogh 2002]. Furthermore, sequential scans of time series data bases can be sped up tremendously, depending on the nature of the data collection.

References

[Faloutsos 1999] C. Faloutsos: *Multimedia IR: Indexing and Searching*. In: Modern Information Retrieval, R. Baeza-Yates and B. Ribeiro-Neto; Addison-Wesley, Reading, MA, 1999.

[Itakura 1975] F. Itakura: *Minimum Prediction Residual Principle Applied to Speech Recognition*. IEEE Transactions on Acoustics, Speech, and Signal Processing **23** (1975) 67-72.

[Keogh 2002] E. Keogh: *Exact Indexing of Dynamic Time Warping*. In: Proc. of the 28th Very Large Databases Conf. (VLDB), Hong Kong, China, August 20-23, 2002, pp. 406-417.

[Kim et al. 2001] S.-W. Kim, S. Park and W. W. Chu: *An Index-Based Approach for Similarity Search Supporting Time Warping in Large Sequence Databases*. In: Proc. of the 17th Int'l Conf. on Data Engineering, Heidelberg, Germany, April 2-6, 2001, pp. 607-614.

[Sakoe & Chiba 1978] H. Sakoe and S. Chiba: *Dynamic-Programming Algorithm Optimization for Spoken Word Recognition*. IEEE Transactions on Acoustics, Speech, and Signal Processing **26** (1978) 43-49.

[Sankoff & Kruskal 1983] D. Sankoff and J. B. Kruskal: *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Addison-Wesley, Reading, MA, 1983.

[Yi et al. 1998] B.-K. Yi, H. V. Jagadish and C. Faloutsos: *Efficient Retrieval of Similar Time Sequences Under Time Warping*. In: Proc. of the 14th Int'l Conf. on Data Engineering, Orlando, FL, February 23-27, 1998, pp. 201-208.