Anticipatory DTW for Efficient Similarity Search in Time Series Databases

Ira Assent
Hardy Kremer
Marc Wichterich
Thomas Seidl
Ralph Krieger

• RWTH Aachen University, Germany
○ Aalborg University, Denmark

VLDB 2009
Lyon, France
Overview

1. Introduction
2. Dynamic Time Warping
3. Anticipatory pruning
4. Experiments
5. Conclusion
Time series similarity search

Time series
- Sequence of time related values
- Stock data, sensor data, EEG measurements, climate data, ...

Similarity search
Find time series with similar patterns over time
Dynamic Time Warping (DTW)

- Most widely used distance functions: Euclidean distance and Dynamic Time Warping
- Dynamic Time Warping allows scaling and stretching for better alignment, but computationally costly

Figure: Euclidean distance (left) and Dynamic Time Warping (right)
DTW definition

k-band DTW

$$DTW([s_1, \ldots, s_n], [t_1, \ldots, t_m]) =$$

$$dist_{band}(s_n, t_m) + \min \left\{ DTW([s_1, \ldots, s_{n-1}], [t_1, \ldots, t_{m-1}]), \right.$$  

$$DTW([s_1, \ldots, s_n], [t_1, \ldots, t_{m-1}]),$$  

$$DTW([s_1, \ldots, s_{n-1}], [t_1, \ldots, t_m])$$  

$$\right\}$$

with

$$dist_{band}(s_i, t_j) = \begin{cases} 
    dist(s_i, t_j) & |i - \left\lfloor \frac{i \cdot n}{m} \right\rfloor| \leq k \\
    \infty & \text{else} 
\end{cases}$$

$$DTW(\emptyset, \emptyset) = 0,$$  

$$DTW(x, \emptyset) = \infty,$$  

$$DTW(\emptyset, y) = \infty$$
DTW Computation

\[ \text{dist}(s_i, t_j) + \min\{c_{i-1,j-1}, c_{i,j-1}, c_{i-1,j}\} \]

**Figure:** Cumulative warping matrix
Existing DTW algorithms

- DTW is computationally expensive
- Many approaches use multistep filter-and-refine architecture
- If filter lower bounds DTW $\Rightarrow$ lossless
- Different filters have been proposed and achieve substantial speed-ups

![Multistep filter-and-refine architecture](image)

**Figure:** Multistep filter-and-refine architecture
DTW properties for speed-up

**DTW is incremental.**

For any cumulative DTW matrix $C = [c_{i,j}]$, the column minima are monotonically non-decreasing: $\min_{i=1,\ldots,n}\{c_{i,x}\} \leq \min_{i=1,\ldots,n}\{c_{i,y}\}$ for $x < y$.

→ Existing approach: early stopping / early abandon
  - compute DTW cumulative matrix
  - after each filled column (band), check threshold for pruning
  - Not as tight as it could be
⇒ new **anticipatory pruning**

![Diagram of DTW algorithm](image)
Anticipatory pruning

- Multistep with anticipatory pruning
- Benefit from work already done in filter step
- Low overhead, substantial speed-up
Anticipatory pruning:

- As in early stopping, compute DTW incrementally
- Additionally: re-use filter information to anticipate full DTW (no additional computation necessary!)
- Requires: filter for remainder of time series
  - We characterize a class of filters to construct such anticipation: piecewise
  - DTW property: reversible
A piecewise lower bounding filter for the DTW distance is a set $f = \{f_0, ..., f_m\}$ with the following property:

$$
\begin{align*}
  j = 0 &: f_j(s, t) = 0 \\
  \forall j > 0 &: f_j(s, t) \leq \min_{(i,j) \in \text{band}_j} \text{DTW}([s_1, ..., s_i], [t_1, ..., t_j])
\end{align*}
$$

Piecewise is the property of the filter that complements the incremental nature of DTW.
DTW is reversible

For any two time series \([s_1, ..., s_n]\) and \([t_1, ..., t_m]\) their DTW distance is the same as for the reversed time series:

\[
DTW([s_1, ..., s_n],[t_1, ..., t_m]) = DTW([s_n, ..., s_1],[t_m, ..., t_1])
\]

Reversible is the DTW property that allows alteration of the time series direction between filter and DTW.
Anticipatory pruning

Anticipatory Pruning Distance.

Given two time series $s$ and $t$ of length $n$ and $m$, a cumulative distance matrix $C = [c_{i,j}]$, a piecewise lower bounding filter $f$ for reversed time series, the $j^{th}$ anticipatory pruning is

$$AP_j(s, t) := \min_{i=1,\ldots,n} \{c_{i,j}\} + f_{m-j}(s^-, t^-).$$
Example

Assent et al. Anticipatory DTW for Efficient Similarity Search in Time Series Databases 14 / 21
Anticipatory pruning is lossless

**Theorem: Anticipatory Pruning lower bounds DTW.**

Anticipatory pruning between two time series $s, t$ with respect to a bandwidth $k$ and a lower bounding filter $f$ lower bounds the DTW:

$$AP_j(s, t) \leq DTW(s, t) \quad \forall j \in \{1, \ldots, n\}$$

**Proof sketch**

- Anticipatory pruning: series of partial DTW paths plus lower bound estimate of the remainder from the previous filter step

(1) For each possible step $r, 1 \leq r \leq n$, column minima lower bound the true path

(2) DTW is reversible

(3) Combine (1),(2): lower bound of DTW

Lower bounding means: lossless pruning, i.e. speed-up + no loss of accuracy
Existing piecewise lower bounds

- Linearization $LB_{Keogh}$: Euclidean distance to envelopes (upper, lower bound of segments) [Keogh, VLDB 2002; Zhu/Shasha, SIGMOD 2003]
- Corner boundaries $LB_{Hybrid}$: piecewise corner-like shapes in the warping matrix through which every warping path has to pass [Zhou/Wong, ICDE 2007]
- Path Approximation $FTW$ (Fast search method for Dynamic Time Warping): go from coarser DTW (less costly) to finer as needed: [Sakurai/Yoshikawa/Faloutsos, PODS 2005]

→ are all piecewise lower bounds as required for anticipatory pruning (details in paper)
Experimental setup

Synthetic and real world data sets

- **SignLanguage:** 1,400 multivariate (11 attributes) of sign language finger tracking data[^1], length 64 to 512
- **TRECVid:** 650 to 2,000 benchmark data[^2]
- **NEWSVid:** 2,000 to 8,000 TV news recorded at 30 fps (20 attributes), length 64 to 2048
- Random walk RW1/RW2: zero normalized time series of length 512 and of cardinality 10,000 (1 to 50 attributes)

[^2]: Smeaton/Over/Kraaij, Evaluation campaigns and TRECVid, MIR 2006
Figure: Relative improvement (#calc.) for varying number of attributes on RW2
Experiments 2

Figure: Efficiency improvement (#calc.) for varying DTW bandwidths on NEWSVid
Figure: (log. scale) Absolute improvement (average query time), reduction, RW2
Conclusion

Anticipatory pruning

- Speed up DTW (Dynamic Time Warping)
  - Widely used distance function for time series similarity search
- Our novel anticipatory pruning makes best use of
  - A family of multistep filter-and-refine approaches
    - Compute an estimated overall DTW distance from already available filter information: series of lower bounds of the DTW

- Experiments demonstrate substantially reduced runtime
- AP can be flexibly combined with existing and future DTW lower bounds
- AP is orthogonal to speed-up via indexing etc.
Conclusion

Anticipatory pruning

- Speed up DTW (Dynamic Time Warping)
  - Widely used distance function for time series similarity search
- Our novel anticipatory pruning makes best use of
  - A family of multistep filter-and-refine approaches
    - Compute an estimated overall DTW distance from already available filter information: series of lower bounds of the DTW

- Experiments demonstrate substantially reduced runtime
- AP can be flexibly combined with existing and future DTW lower bounds
- AP is orthogonal to speed-up via indexing etc.

Thank you for your attention.

Questions?