

# Chapter 1

## Segmented Dynamic Time Warping: A Comparative and Application Study

**Ruizhe Ma**

*Georgia State University, USA*

**Azim Ahmadzadeh**

*Georgia State University, USA*

**Soukaina Filali Boubrahimi**

*Georgia State University, USA*

**Rafal Angryk**

*Georgia State University, USA*

### ABSTRACT

*Initially used in speech recognition, the dynamic time warping algorithm (DTW) has regained popularity with the widespread use of time series data. While demonstrating good performance, this elastic measure has two significant drawbacks: high computational costs and the possibility of pathological warping paths. Due to the balance between performance and the tightness of restrictions, the effects of many improvement techniques are either limited in effect or use accuracy as a trade-off. In this chapter, the authors discuss segmented-DTW (segDTW) and its applications. The intuition behind significant features is first established. Then considering the variability of different datasets, the relationship between specific global feature selection parameters, feature numbers, and performance are demonstrated. Other than the improvement in computational speed and scalability, another advantage of segDTW is that while it can be a stand-alone heuristic, it can also be easily combined with other DTW improvement methods.*

DOI: 10.4018/978-1-5225-8446-9.ch001

## INTRODUCTION

Different data and applications usually require different distance measures. Generally speaking, distance measures can be categorized as either a lock-step measure or an elastic measure. Lock-step measure commonly refers to  $L_p$ -norms, meaning that the  $i$ -th element from one sequence is always mapped to the  $i$ -th element in the compared sequence. Since the mapping between the two sequences is fixed, lock-step measures are sensitive to noise and misalignments (Ranacher et al., 2014). The elastic measure, on the other hand, allows one-to-many, or even one-to-none mappings (Wang et al., 2013). A suitable distance measure should be chosen based on the data and task at hand.

With the improvement of data storage capability and processing power, more and more data are now being collected and used in the form of time series. Time series data is widely applied in a variety of domains, such as voice recognition, the stock market, solar activities, medical research, and many other scientific and engineering fields where measurements in the temporal sense are important. Time series data enjoys its popularity because it records details that can be overlooked by the summarization of data. At the same time, small temporal discrepancies and unequal length are very common in time series sequences. This length difference means lock-step distance measures are not as effective as elastic measures when identifying similarities in time series data (Ranacher et al., 2014).

A widely used elastic measure is the Dynamic Time Warping (DTW) algorithm (Serra et al., 2014). The warping path, or shortest global route, for DTW is identified based on the computation and comparisons of several options at each step. While this contributes to the general success of the DTW algorithm's performance with time series data, the large-scale computations make DTW a very time-consuming algorithm. Another common issue to run into when using DTW is the occurrence of pathological warping path, which happens when the warping path tries to compensate value differences by making wild and improbable mappings, as shown in Figure 1 (a). When the path is found in a greedy manner and never readjusted, the optimal warping path could be permanently missed. The root cause of pathological warping paths is that the decision at each step does not take significant global similarities into consideration; in other words, the bigger picture is neglected in standard DTW. In application, we can often observe the warping path between two time series does not match the expected intuitive mapping. Previously, it has been found that segmentation of time series based on feature peaks can alleviate this problem (Ma et al., 2018). Global similarity can be identified by adding a layer of peak identification, which acts as a form of restraints, then the focus on local similarity is found by mapping the points and computing the distances of each corresponding segments.

Figure 1 shows the mapping comparison between the standard DTW algorithm and an example of segDTW with peaks. Here the standard DTW is showing signs of pathological warpings, which does not match our intuitive expectations. In contrast, by identifying two peaks and segmenting the two time series into three paired segments, the mapping provides more intuitive results. Theoretically, any feature extracted from the data that could meaningfully segment time series can be used. When features are identified and paired, not only do we obtain a more intuitive mapping, but also makes it possible and exceptionally easy to parallelize DTW computation within a time series sequence. This segmentation makes our method scalable and sets segDTW apart from other DTW improvement heuristics.

Achieving speedup by segmenting one computation into several pieces of simultaneous computations is to be expected. Therefore, in this chapter, we establish the meaning and importance of intuitive segmentation, by comparing peak and valley segmentation against equal-window segmentation. This

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/segmented-dynamic-time-warping/230681](http://www.igi-global.com/chapter/segmented-dynamic-time-warping/230681)

## Related Content

---

### Semantic Web Rule Languages for Geospatial Ontologies

Philip D. Smart, Alia Abdelmoty and Baher A. El-Geresy (2009). *Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches* (pp. 149-169).

[www.irma-international.org/chapter/semantic-web-rule-languages-geospatial/35858](http://www.irma-international.org/chapter/semantic-web-rule-languages-geospatial/35858)

### An Interactive Viewpoint on the Role of UML

Dina Goldin, David Keil and Peter Wegner (2001). *Unified Modeling Language: Systems Analysis, Design and Development Issues* (pp. 250-264).

[www.irma-international.org/chapter/interactive-viewpoint-role-uml/30582](http://www.irma-international.org/chapter/interactive-viewpoint-role-uml/30582)

### Rapid Pattern-Oriented Scenario-Based Testing for Embedded Systems

Wei-Tek Tsai, Ray Paul, Lian Yu and Xiao Wei (2005). *Software Evolution with UML and XML* (pp. 222-262).

[www.irma-international.org/chapter/rapid-pattern-oriented-scenario-based/29615](http://www.irma-international.org/chapter/rapid-pattern-oriented-scenario-based/29615)

### XML Data Management in Object Relational Database Systems

Zhen Hua Liu, Anguel Novoselsky and Vikas Arora (2010). *Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies* (pp. 18-46).

[www.irma-international.org/chapter/xml-data-management-object-relational/41498](http://www.irma-international.org/chapter/xml-data-management-object-relational/41498)

### Efficient Storage and Parallel Query of Massive XML Data in Hadoop

Wei Yan (2019). *Emerging Technologies and Applications in Data Processing and Management* (pp. 242-262).

[www.irma-international.org/chapter/efficient-storage-and-parallel-query-of-massive-xml-data-in-hadoop/230692](http://www.irma-international.org/chapter/efficient-storage-and-parallel-query-of-massive-xml-data-in-hadoop/230692)