Abstract

Similarity search in time-series subsequences is an important time series data mining task. Searching in time series subsequences for matches for a set of shapes is an extension of this task and is equally important. In this work we propose a simple but efficient approach for finding matches for a group of shapes or events in a given time series using a Nearest Neighbor approach. We provide various improvements of this approach including one using the GNAT data structure. We also propose a technique for finding similar shapes of *widely* varying temporal width. Both of these techniques for primitive shape matching allow us to more accurately and efficiently form an event representation of a time-series, leading in turn to finding complex events which are composites of primitive events. We demonstrate the robustness of our approaches in detecting complex shapes even in the presence of "don't care" symbols. We evaluate the success of our approach in detecting both primitive and complex shapes using a data set from the Fluid Dynamics domain. We also show a speedup of up to 5 times over a naïve nearest neighbor approach.