

**Non-overlapping Subsequence Matching of Stream Synopses**

**Abstract:**

In this paper, we propose SUbsequence Matching framework with cell MERgence (SUMMER) for online subsequence matching between histogram-based stream synopsis structures under the dynamic time warping distance. Given a query synopsis pattern, SUMMER continuously identifies all the matching subsequences for a stream as the bins are generated. To effectively reduce the computation time, we design a Weighted Dynamic Time Warping (WDTW) algorithm, which computes the warping distance directly between two histogram-based synopses. Furthermore, a Stack-based Overlapping Filter Algorithm (SOFA) is provided to remove the overlapping subsequences to avoid the redundant information. Finally, we design an optional refinement module to relax the subsequence range limit and improve the matching accuracy. Our experiments on real datasets show that the proposed method significantly speeds up the pattern matching without compromising the accuracy required when compared with other approaches.

**Existing System:**

Unfortunately, a general subsequence matching algorithm on histogram-based synopsis is yet to be developed. The original DTW computation is designed to handle time series with regular time intervals while each bin of a histogram-based synopsis stream may have various time spans. Of course we may convert each histogram bin to consecutive data points of the same value at fixed intervals and use the original DTW computation to compute the warping distance. However, such a method is obviously not efficient in a streaming environment since it is contrary to the original intent of reducing the size of data to be processed by constructing synopsis. Although there exist algorithms for subsequence matching on synopsis streams of equal-width bins such as, applying them directly on synopsis streams of arbitrary-width bins causes either accuracy loss or efficiency degradation. This motivates us to design a general subsequence matching algorithm that can deal with histogram-based synopsis streams of arbitrarywidth bins.

**Proposed System:**

Our goal is to design an online algorithm that can identify all non-overlapped subsequences satisfying a given distance threshold with minimized delays. In an online streaming environment, some applications require to continuously monitor whether there are similar instances in the evolving stream. In such applications, we may prefer to make range queries for subsequence matching rather than to use top one search. The range query aims to find all similar but nonoverlapping subsequences to the query pattern exceeding a given distance threshold. The non-overlapping requirement is a useful criteria for removing redundant subsequences because the overlapped subsequences usually indicate the same instance with little time shift. Note that the nonoverlapping examination may introduce the extra delay in reporting qualified subsequences. In light of this concern, the proposed algorithm should report these subsequences within an acceptable delay.