Abstract

Matrix Profile (MP) has emerged as a new concept in the data mining community since 2016, which the advocating researchers believe has the potential to revolutionize time series data mining because of its generality, versatility, simplicity, and scalability.

Despite many its advantage, however, there is one inherent, fundamental problem with the current MP setting – the distance it uses is Euclidean, which can be brittle in measuring the similarity between time subsequences.

An ideal substitution for Euclidean distances is to use dynamic time warping (DTW). The researchers who proposed the MP have named this as an open question for future work. The challenge is that the time complexity for computing both MP and DTW are heavy themselves, making the combination of the two intimidatingly expensive.

Never before have the data mining community discussed the introduction of DTW into MP. In this project we make a first attempt to bridge this gap.

In particular, our focus is to adapt DTW's lower bound functions to the MP setting. We study and confirm the scalability of some versions of lower bound functions. After looking into the relationship between the entries of the MP, we show that certain lower bound functions, which were designed for calculating single DTW, can be calculated much easier in a large scale when dealing with multiple DTWs in the MP.

Besides, we resort to local search and randomized algorithms, for example the Metropolis algorithm, to exploit the pruning power of the lower bound functions. This further alleviates the computational burden.

We also talk about other directions that may facilitate the introduction of DTW into MP – to speed up the computation of individual DTWs, and to adapt the definition of DTW to the MP.