R

Frank Madrid

March 12, 2018

1 Introduction

During the STAMP, STOMP, SCRIMP, and SCRIMP++ presentation, it was suggested that calculating the Distance Profile of some sequence T of length n and a subsequence length of m is $\Omega(\frac{1}{m}n^2 \log n)$; however, I'm not terribly convinced. I believe the *Distance Profile Problem* is reducible to the *Matrix Multiplication Problem* for square matrices and since there are algorithms with a runtime complexity of $O(n^{1.68})$ for this secondary problem, I would believe the *Distance Profile Problem* to be solvable in $\mathcal{O}((n-m+1)^{1.68})$
Claim: The Distance Profile Problem is reducible to the Matrix Multiplication problem.

Proof. Let T be a sequence of length n. Consider the matrix $\mathbf{T}$ whose column vectors are the subsequences of T,

$$\mathbf{T} = \begin{bmatrix} T_{(1,m)} & T_{(2,m)} & T_{(3,m)} & \dots & T_{(n-m+1,m)} \end{bmatrix}^T$$

then

$$\mathbf{T} \times \mathbf{T}^T = \mathcal{T}$$

such that $\mathcal{T}_{(i,j)} = T_{(i,m)} \cdot T_{(j,m)}$ or the inner products of the subsequences of T. □

2 Definitions

Definition 1 (The Distance Profile Problem)**.** Given two sequences T and S, for each subsequence $T_{i,m} \in T$ find the subsequence $S_{j,m} \in S$ such that $Distance(T_{i,m}, S_{j,m})$ is minimized where $Distance$ is $EuclideanDistance$.

Definition 2 (Matrix Multiplication Problem)**.** Given two square matrices $A_{[n,n]}$ and $B_{[n,n]}$, find the product $A \times B$.

Definition 3. A **subsequence** $T_{i,m}$ of T is a contiguous proper subset of the values from T of length m starting from position i. Formally, $T_{i,m} = [t_i, t_{i+1}, \dots, t_{i+(m-1)}]$.

Definition 4. Let $T = [t_1, t_2, \ldots, t_n]$ and $S = [s_1, s_2, \ldots, s_n]$ be two sequences of length n, then the **Euclidean Distance** of T and S is defined as

$$\text{EuclideanDistance}(T, S) = \sqrt{\sum_{i=1}^{n} (t_i - s_i)^2}$$

3 Notes