Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping

Liyue Fan

Here I listed a few points for discussion or revisit later. The goal is to propose a hybrid system for privacy preserving nearest neighbor search on time series data. The idea is to perturb the lower bound with differential privacy mechanism for higher accuracy. For candidate subsequences, the exact DTW distance can be computed by a SMC protocol for security purpose. The proposed system can be evaluated against the baseline LPA algorithm and FAST algorithm.

Discussions:

1. Suppose both the time series and the query sequence are normalized, what is the sensitivity of LB_Keough(Q,C)? (Hint: read (Ada Wai-chee Fu, 2005) for detailed documentation of the lower bound)

2. With the first question answered, what is the sensitivity of evaluating LB_Keough over all subsequences \( C_i = \{x_i x_{i+1} ... x_{i+k}\}, \ i = 1, ... n-k, \) where n is the length of X and k is the length of Q?

3. How to protect the LB_Keough(Q,C)? Laplace Mechanism or Exponential Mechanism?

4. Most of existing encryption systems deal with integer input, e.g. Paillier, given normalized time series data, what secure protocol can be used to compute DTW(Q,C)?


Bibliography