Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Peer Kröger
Yifeng Lu

Knowledge Discovery in Databases II

SS 2017

Exercise 6: Sequential Data and Time Series

Exercise 6-1 Longest Common Subsequence for Time Series

Given the longest common subsequence distance for time series is defined as following:

Given two time series $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_m)$ and two threshold ϵ and δ,

$$LSC(X,Y) = \begin{cases} 0 & if\ n = 0 \vee m = 0 \\ LCS(start(X), start(Y)) + 1 & if\ match(last(X), last(Y)) \\ max(LCS(start(X), Y), LCS(X, start(Y))) & else \end{cases}$$

where the matching function is defined as:

$$match(x_i, y_j) = \begin{cases} true & if\ |x_i - y_j| \leq \epsilon \wedge |i - j| \leq \delta \\ false & else \end{cases}$$

Then the distance is:

$$D_{LCS}(X,Y) = 1 - \frac{LCS(X,Y)}{\min(n,m)}$$

(a) Given $X = (3, 5, 9, 2, 3, 6, 3)$ and $Y = (3, 4, 6, 10, 1, 3, 2, 7, 4)$, $\epsilon = 2, \delta = 2$, what is the $D_{LCS}(X,Y)$?

(b) Prove or disprove if the D_{LCS} is a metric distance.

Exercise 6-2 Discrete Fourier Transformation

Show that the original signal can be recovered by the inverse Discrete Fourier Transformation, i.e.: $f(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(k) \exp(\frac{2\pi jnk}{N})$.

Exercise 6-3 PAA / DWT(Optional)

Given two time series: $X = (6, -2, -7, -1, 1, -3, 6, 8)$, $Y = (1, 3, -8, -4, 5, -1, 2, 10)$,

(a) compute the L_1- and L_∞-distance.

(b) Compute the dimension reduction representations of X using: DWT with Haar-Wavelet and PAA with $M = 4$ boxes.