Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Matthias Schubert
Julian Busch, Daniyal Kazempour

Munich, 24.11.2017

Big Data Management and Analytics
WS 2017/18

Tutorial 5: Stream Processing

Assignment 5-1 *Streaming*

Given the following terms:
Aggregation, Compression, Data Reduction, Histograms, Load Shedding, Microclusters, Sampling, Wavelets

(a) Explain each of the terms by providing a short definition.

(b) Illustrate how the terms are related to each other.

Assignment 5-2 *Discrete Wavelet Transformation (DWT)*

Given the following input sequence $S =$ (4,1,2,3,6,1,7,6)

(a) Perform a Haar Wavelet Transformation on S and determine the Wavelet coefficients

(b) Reconstruct the original sequence S using the Wavelet coefficients

(c) For a loss afflicted reconstruction we assume that -0.5 and 0.5 are close to 0. Sum up the resulting errors per residue to a total (linear) approximation error

Assignment 5-3 *Piecewise Aggregate Approximation (PAA)*

Given the following input sequence $S =$ (4,1,2,3,6,1,7,6)

(a) Compute the reduced representation of S using PAA (box size $M = 4$).

Hint: A PAA approximates a time series X of length N with a vector $\bar{X} = (\bar{x_1},...,\bar{x_M})$ of arbitrary length $M \leq N$, where for each $\bar{x_i}$ holds:

$$\bar{x_i} = \frac{M}{N} \sum_{j=\frac{N}{M}(i-1)+1}^{\frac{N}{M}i} x_j \tag{1}$$

(b) Convince yourself that PAA and DWT (using Haar Wavelets as basis functions!) are equivalent.

Assignment 5-4 *Reservoir Sampling*

Given a data stream of size N. Randomly select $k \leq N$ elements from the stream. Here k represents the size of the reservoir.

(a) Setting $k = 1, N = 2$. The first element is in the reservoir, the second is not. What is the probability of both elements to be in the reservoir?

(b) Setting $k = 1, N = 3$. What is now the probability for each of the elements to be in the reservoir?

(c) Setting $k = 1$. What is the probability for any given N?

(d) What is the probability for an arbitrary reservoir size k and an arbitrary stream size N?