ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

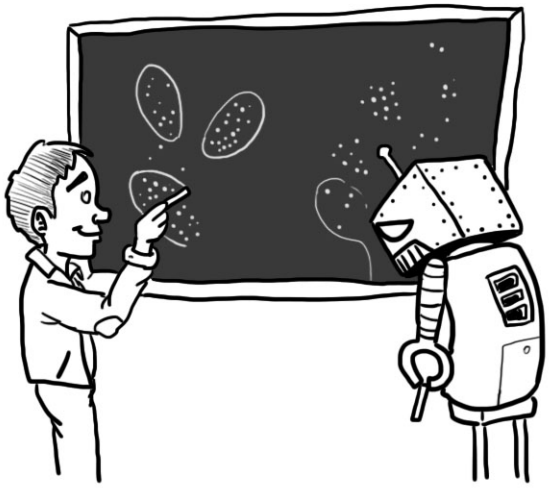
Prof. R. Wattenhofer

BA/MA/SA/Group/Lab:

Metric approximations and clustering

A metric space is, essentially, a set of points together with a rule for saying how far apart two such points are. For example, all points in $\mathbb{R}^d$ together with the Euclidean metric. Formally, a metric space is an ordered pair (M, d) where M is a set and d is a metric on M, i.e., a distance function $d : M \times M \mapsto \mathbb{R}$ such that for any $x, y, z \in M$, the following holds: **I.** $d(x, y) = 0 \Leftrightarrow x = y$, **II.** $d(x, y) = d(y, x)$, **III.** $d(x, z) \leq d(x, y) + d(y, z)$.

Metric spaces are widely used to model similarities, costs, etc. The third property, called triangle inequality, makes a difference in a bunch of combinatoric problems. For example, the traveling salesperson problem (TSP) is easier with the triangle inequality than without.



If the set M is finite, then this metric can be approximated by a distribution over tree metrics. This approximation is helpful in many applications because it allows us representing a metric by much less parameters. Besides, good properties of trees play crucial roles in algorithm design and analysis, e.g., in the online matching problem.

We like to extend this approximation to general metrics which map three (or more) points to a nonnegative "distance" among these points. The first step is to choose a reasonable axiom system for general metrics. The most important part here is to find a correct correspondence for the triangle inequality in high dimension. This step should not be too challenging, because there exist several general metrics proposed by mathematicians (to studied fixed-points theories).

The next step is to find some (hierarchical) clustering technique similar to one used in a previous approximation algorithm and show that there is a tree-like approximation of the metric you chose in the first step.

Requirements: Creativity and an interest in algorithm design. Being familiar with clustering techniques may be an advantage.

Interested? Please contact us for more details!

Contacts

- Yuyi Wang: yuyi.wang@tik.ee.ethz.ch, ETZ G94