ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed
Computing

Prof. R. Wattenhofer

Algorithm Learning

Efficient Graph Traversal Learning from Paths Travelled and Untravelled

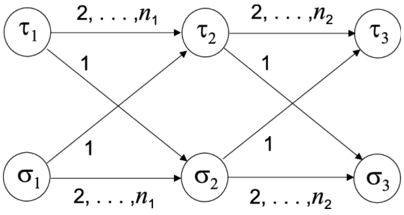
Choose your favourite graph data sample where all nodes are labelled according to some scheme. Given a starting point or and a set of endpoints from that graph, how can we identify the traversal query (algorithm) that was used to reach the endpoints?

The general problem is quite hard. So, make the assumption that every step in the sought traversal is decided only by the labels of the neighbours of the current node. This simplifies the problem greatly and makes it very similar to the problem of regular expression learning, though in this case we benefit implicitly from the additional information coming from all the paths that we *did not take* to reach the endpoints.

Why is this interesting and what is our goal?

Graph (network) data is ubiquitous and often very complex. Unlike relational database data (that can too be viewed as graph data), it usually does not benefit from a fixed, explicit structure and is therefore somewhat more difficult to work with.

Our goal is to design an algorithm capable of learning simpler types of graph traversals (namely graph finite automaton traversals) from $\langle startpoints, endpoints \rangle$ pairs on a fixed graph. A brute, inefficient solution is not too hard to come up with, but we can certainly do better. If we show that it is possible to do this learning efficiently, we will have created an undoubtedly powerful and useful analysis approach for data scientists, not to mention the advancement of frontiers of algorithm learning.



Who is this for? Bachelor's or master's students interested in diving into the above-mentioned problem from either the probabilistic or optimisation perspective. Some familiarity with graphs, graph terminology, or finite automata will make for a speedier start.

Interested? Please reach out to us for more details.

Point of contact: Peter Belcak, email pbelcak@ethz.ch, ETZ G63.