

ICT22-059 - Structured Data Learning with General Similarities

Abstract

In this project we will systematically investigate similarity-based machine learning with structured data such as strings, trees and graphs. While most off-the-shelf machine learning algorithms require data to be embedded in a (finite or infinite) dimensional inner product space, most intuitive notions of similarity for structured data by domain experts do not allow for such an embedding. Examples of such similarities are based on alignments, edit operations, or (graph) matching. Recent progress has allowed learning algorithms to use more general similarities which can be embedded in Krein space. While preliminary work shows the potential of this approach to learning with structured data, this possibility has never been systematically explored. Furthermore, even these approaches have no means for dealing naturally with asymmetric notions of similarity like the ones based on substructure relations. This project will close the described gaps by (i) designing and investigating general similarities for structured data, (ii) developing learning algorithms for general similarities, and (iii) applying combinations of these for concrete problems in cheminformatics. Progress in the design of RNA therapeutics, polyketide pharmaceuticals, and the prediction of mass spectra will have high impact on several areas of human society. Our approach promises higher predictive performance, more efficient learning, and better interpretability of the models by domain experts.

Scientific disciplines:

Machine learning (50%) | Theoretical computer science (30%) | Theoretical chemistry (20%)

Keywords:

machine learning, structured data, similarity

Principal Investigator: Nils Kriege
Institution: University of Vienna
Co-Principal Investigator(s): Thomas Gärtner (TU Wien)
Christoph Flamm (University of Vienna)



v.l.n.r. Nils Kriege; Christoph Flamm; Thomas Gärtner;

Status: Ongoing (01.05.2023 - 30.04.2027)

Further links to the persons involved and to the project can be found under

<https://www.gmbh.wwtf.at/funding/programmes/ict/ICT22-059/>