ICT22-059 - Structured Data Learning with General Similarities

Zusammenfassung

Techniken des maschinellen Lernens und der künstlichen Intelligenz gewinnen für der Wirtschaft und Wissenschaft zunehmend an Bedeutung. Liegen die Daten nicht in Form von Zahlen vor, sondern haben eine komplexe Struktur wie Proteine, RNA oder Moleküle, werden diese zunächst mit unterschiedlichen Verfahren in numerische Repräsentationen überführt, um in einem nächsten Schritt Techniken des maschinellen Lernens anwenden zu können. Die Interpretierbarkeit im Bezug auf die ursprüngliche Struktur der Daten geht hierbei in der Regel verloren. Für viele derartige Daten existieren jedoch auch von Experten entworfene Ähnlichkeitsmaße, die komplexes Domänenwissen beinhalten. Diese können für das maschinelle Lernen bis jetzt jedoch nur stark eingeschränkt genutzt werden. Ziel des Projekts ist es daher, (i) solche Ähnlichkeitsfunktionen zu analysieren und neue zu entwerfen, (ii) Lernverfahren zu entwickeln, die diese unterstützen, und (iii) mit diesen konkrete Probleme der Chemieinformatik zu lösen. Fortschritte beim Design von RNA Therapeutika, polyketid-basierten Pharmazeutika und der Vorhersage vom Tandem-Massenspektren wird einen großen Einfluss auf viele Bereiche der menschlichen Gesellschaft haben. Unser Ansatz verspricht genauere und effizientere Lernverfahren, die darüber hinaus eine bessere Interpretierbarkeit aufweisen.

Wissenschaftliche Disziplinen:
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: Laufend (01.05.2023 - 31.10.2027)

Weiterführende Links zu den beteiligten Personen und zum Projekt finden Sie unter
https://www.wwtf.at/funding/programmes/ict/ICT22-059/