Towards biomedical neurosymbolic AI: From semantic knowledge infrastructure to explainable predictions

Michel Dumontier*

* Maastrich University

Résumé

The increased availability of biomedical data, particularly in the public domain, offers the opportunity to better understand human health and to develop effective therapeutics for a wide range of unmet medical needs. However, data scientists remain stymied by the fact that data remain hard to find and to productively reuse because data and their metadata i) are wholly inaccessible, ii) are in non-standard or incompatible representations, iii) do not conform to community standards, and iv) have unclear or highly restricted terms and conditions that preclude legitimate reuse. These limitations require a rethink on data can be made machine and AI-ready – the key motivation behind the FAIR Guiding Principles. Concurrently, while recent efforts have explored the use of deep learning to fuse disparate data into predictive models for a wide range of biomedical applications, these models often fail even when the correct answer is already known, and fail to explain individual predictions in terms that data scientists can appreciate. These limitations suggest that new methods to produce practical artificial intelligence are still needed.

In this talk, I will discuss our work in (1) building an integrative knowledge infrastructure to prepare FAIR and « AI-ready » data and services along with (2) neurosymbolic AI methods to improve the quality of predictions and to generate plausible explanations. Attention is given to standards, platforms, and methods to wrangle knowledge into simple, but effective semantic (e.g. knowledge graphs and ontologies) and latent representations (i.e. embeddings), and to make these available into standards-compliant and discoverable interfaces that can be used in model building, validation, and explanation. Our work, and those of others in the field, creates a baseline for building trustworthy and easy to deploy AI models in biomedicine.

Biographie

Dr. Michel Dumontier is the Distinguished Professor of Data Science at Maastricht University, founder and executive director of the Institute of Data Science, co-founder of the Department of Advanced Computing Sciences, and co-founder of the influential FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles. His research focuses on computational learning and reasoning for drug discovery and personalized medicine. He is coordinator for the REALM project to build a regulatory sandbox for (AI) medical software devices, Principal

Investigator (PI) for the AIDAVA project to create an AI assistant in the curation of personal health knowledge graphs, PI for the NIH/NCATS Biomedical Data Translator, PI for the ARPA-H CHARM project, and a PI for the iCare4CVD IHI Project. He is the editor-in-chief for the journal Data Science and is internationally recognized for his contributions in bioinformatics, biomedical informatics, and semantic technologies.