Machine explanation – Interpretation of Machine Learning Models for Medicine and Bioinformatics

Aim and scope

There is a recent shift towards human-centred AI, where analytical methods meet with human expertise to verify that decision support systems are doing what they do right, and also to validate that they're doing the right thing. In both cases, it is critical to look inside the machine. Machine learning systems are often treated as black-boxes, which makes it difficult to apply them in safety-related applications and may create serious difficulties given the right to explanation built into the new general data protection regulations.

The aim of the session is to report original research and case studies where models are explained and verified using clinical or bioinformatics knowledge. This may involve machine learning methods that are interpretable by design, links made between databased methods and knowledge-based systems, or integration of structure finding algorithms into medical decision making for instance in the form of Bayesian Belief Networks.

Topics of interest include, but are not limited to:

- Explanation of machine learning algorithms
- Integration of databased with knowledge-based methods
- Learning representations from multi-modal bioinformatics data
- Interpretability of deep learning models
- NLP for knowledge discovery and model validation in medicine and bioinformatics
- Reinforcement Learning for the optimization of medical treatments
- Explanation of models learning structured data in bioinformatics and chemistry
- Unsupervised learning and visualisation of high-dimensional data
- Prototype-based approaches

Session chairs

Ian H. Jarman, Liverpool John Moores University, UK
Alfredo Vellido, Universitat Politècnica de Catalunya, Barcelona, Spain
José D. Martìn-Guerrero, University of Valencia, Spain
Davide Bacciu, Università of Pisa, Italy