SCIENTIFIC SEMINAR

TRACTABLE EXPLAINING

Friday 5th February 2021 – 3:00 p.m CET

With Martin COOPER - IRIT

https://univ-toulouse-fr.zoom.us/j/93887363532?pwd=R3MzR3ppTXhJSWF6L1hMTjgxZUhNdz09

ID meeting : 938 8736 3532
Code : 796903

ABSTRACT:
Explaining decisions is at the heart of explainable AI. We investigate the computational complexity of providing a formally-correct and minimal explanation of a decision taken by a classifier. In the case of threshold (i.e. score-based) classifiers, we show that a complexity dichotomy follows from the complexity dichotomy for languages of cost functions.In particular, submodular classifiers allow tractable explanation of positive decisions, but not negative decisions (assuming P not equal to NP). This is an example of the possible asymmetry between the complexity of explaining positive and negative decisions of a particular classifier. Nevertheless, there are large families of classifiers for which explaining both positive and negative decisions is tractable, such as monotone or linear classifiers. We extend tractable cases to constrained classifiers (when there are constraints on the possible input vectors) and to the search for contrastive rather than abductive explanations. Indeed, we show that tractable classes coincide for abductive and contrastive explanations in the constrained or unconstrained settings.

BIO:

Martin Cooper holds a post of professor of Computer Science at the University of Toulouse 3 and has spent more than 30 years in AI research.

ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE — Université Fédérale Toulouse Midi-Pyrénées