Internship proposition: Fairness in multi-stage decision making

Keywords: machine learning, fairness, ethics of algorithms

Lab: Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Eric Gaussier)
Team in the lab: POLARIS (head: Arnaud Legrand)

Advisor: Patrick Loiseau (Univ. Grenoble Alpes, LIG)
patrick.loiseau@univ-grenoble-alpes.fr, http://www.eurecom.fr/~loiseau/

General presentation of the topic:

Machine learning algorithms have become omnipresent in everyday life; they are used in increasingly many areas such as advertising, but also hiring or justice. As learning algorithms become more ubiquitous, there is a growing concern about how they affect humans in possibly inconspicuous ways. In particular, it was recently observed that many algorithms discriminate (even unintentionally) users based on features (such as race or gender) which are deemed unacceptable. This started a very intense debate on the fairness of machine learning and a high interest in designing fair learning algorithms.

Internship program and objectives:

In this internship, we will study the case of multi-stage decision making for classification or candidates selection. In this setting, we have a set of candidates. In a first stage, the algorithm selects a subset of the candidates based on some features. In a second stage, for the candidates preselected, the algorithm gets access to additional features to make the final selection. Such multi-stage decision making is used for instance in most hiring processes. The goal of the internship will be to investigate fairness in multi-stage decision making. The student will start by investigating optimal algorithms for multi-stage decision making and their fairness properties according to different definitions (there exist multiple definitions of fairness, see [1-5]). Then, he/she will design modified algorithms that respect fairness constraints and investigate whether, for certain definitions, it is impossible to do so.

Expected abilities of the student:

Strong background in probability and machine learning

References:

[1] M. B. Zafar, I. Valera, M. Gomez-Rodriguez, K. P. Gummadi, A. Weller. From Parity to Preference-based Notions of Fairness in Classification. In NIPS, 2017.
[2] M. Joseph, M. Kearns, J. Morgenstern, A. Roth. Fairness in Learning: Classic and Contextual Bandits. In NIPS, 2016.
[3] M. Hardt, E. Price, N. Srebro. Equality of Opportunity in Supervised Learning. In NIPS, 2016.
[4] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork. Learning Fair Representations. In ICML, 2013.
[5] M. B. Zafar, I. Valera, M. Gomez-Rodriguez, K. P. Gummadi. Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment. In WWW, 2017.