

Michèle Sebag

Personal data

- Senior Researcher CNRS, deputy director Lab. of Computer Science, Univ. Paris-Sud
- Ecole Normale Supérieure (Maths 1975), Engineer Thalès (1980-1985), Consulting Engineer, PhD Computer Science (Univ. Paris-Dauphine, 1990), HdR Computer Science (Univ. Paris-Sud, 1997).
- Research Fellow CNRS Numerical Engineering (Ecole Polytechnique, 1991-2001), Senior Researcher Machine Learning (Univ. Paris-Sud, 2002-2018);

Research Interests

- Machine Learning: Deep Learning, Reinforcement Learning, Causal Modelling
- ML applications to human resources (hiring, economic indicators)
- Stochastic optimization

Publications

- 3 edited books, 12 chapters
- 20 publications in international journals
- 90 publications international conferences.

Graduate and post-graduate supervision

- Advisor of 19 PhDs (Olivier Bousquet, 1999-2002; Jérémie Mary, 2002-2005; Nicolas Baskiotis, 2005-2008; Riad Akrouf, 2011-2014);

Awards and Honors

- European AI Fellow, 2010
- Member of Académie of Technologies, 2017
- Test of time award, GECCO 2008-2018.

Editorial Boards and Program committees

- Machine Learning Journal (since 2001)
- Head of Steering Committee for European Machine Learning and Knowledge Discovery (2015-)
- Senior Area Chair IJCAI 2016, Area Chair NIPS 2017-2018; ICML 2016-2018
- President of the French Association for AI (2000-2007)

Invited talks (in the last 3 years)

- Le travail en mouvement, Colloque de Cerisy, 2018. *Qualité de la vie au travail et santé économique des entreprises : étude des causalités*
- JST Program on Big Data Applications, Tokyo, Sept. 2017. *Algorithm Recommendation*
- Lorentz Center, University of Leiden, Netherlands, September 2017. Space Weather: A Multidisciplinary Approach. *Causal Modeling with Generative Neural Networks*.
- 50th CREST Open Workshop, London, Jan. 2017. *Stochastic Gradient Descent: Going As Fast As Possible But Not Faster*.
- Deep Learning and Artificial Intelligence, Franco-Japanese Symposium, Tokyo, Oct. 2016. *Deep Learning without Hot Air*
- Constructive Machine Learning Workshop at ICML 2015, New York; AutoML Workshop at ICML 2015, New York.

5 key publications (in the last 3 last years)

- [1] Mustafa Misir and Michèle Sebag, *Alors: An algorithm recommender system*, Artif. Intell. **244** (2017), 291–314.
- [2] Thomas Schmitt and Francois Gonard and Phillippe Caillou and Michèle Sebag, *Language Modelling for Collaborative Filtering: Application to Job Applicant Matching*, 29th IEEE International Conference on Tools with Artificial Intelligence, ICTAI, 2017, pp. 1226–1233.
- [3] Yoann Isaac and Quentin Barthelemy and Cédric Gouy-Pailler and Michèle Sebag and Jamal Atif, *Multi-dimensional signal approximation with sparse structured priors using split Bregman iterations*, Signal Processing **130** (2017), 389–402.
- [4] Xiangliang Zhang and Cyril Furtlehner and Cécile Germain-Renaud and Michèle Sebag, *Data Stream Clustering With Affinity Propagation*, IEEE Trans. Knowl. Data Eng. **26** (2014), no. 7, 1644–1656.
- [5] Riad Akrouf and Marc Schoenauer and Michèle Sebag and Jean-Christophe Souplet, *Programming by Feedback*, Proceedings of the 31th International Conference on Machine Learning, ICML, 2014, pp. 1503–1511.

Research project

The project aims to advance the study (assessment, measure and enforcement) of the under-specified criteria at the core of “Fair AI”. In the domain of collaborative hiring for instance, it is likely that the available data reflect the implicit preferences of recruiters and candidates, ranging from geographic preferences to gender-related prejudices. Fair AI must avoid encapsulating unwanted biases in the models learned from the gathered data, thereby perpetuating them.

Several questions will be investigated. A first question is how to measure whether and to which extent a dataset is corrupted by a known bias (e.g., color, gender, status, -related). Counterfactual reasoning (what-if, the gender, age, origin,.. was different) will be deployed to construct shadow distributions, and distribution-based and/or adversarial learning will be achieved to determine how the actual data differ from the counterfactual data. This process aims both to measure the bias intensity, and “whiten” the data w.r.t. this bias. A second question is whether unwanted biases can be automatically distinguished from licit biases (e.g. personal preferences). A third question regards the definition of a methodology, enabling to probe an existing system and certify its fairness w.r.t. a given criterion.