

Fairness: Knowing better what is good for you?

SUPERVISORS: Michèle Sebag, Philippe Caillou, Bruno Crépon, Solal Nathan

Michele.Sebag@lri.fr, Philippe.Caillou@lri.fr, Bruno.Crepon@ensae.fr, solal.nathan@ens-paris-saclay.fr

LAB: TAU – CNRS – INRIA – LISN, U. Paris-Saclay

ABSTRACT:

The VADORE (*Valorisation des Données de Recherche d'Emploi*) project, gathering MLers, economists and experts from the French public employment service, aims to recommend job offers best suited to job seekers by leveraging the wealth of proprietary data (signed contracts and applications) in France Travail [1].

Algorithms for human resources are considered to be high risk according to the EU AI act. Indeed, the data show biases; for instance women might trade a wage loss for a decrease in the distance between the job location and their home. The danger is the algorithm to carve in stone the biases present in the data. We have shown that algorithmic recommendations are no more biased than the actual applications of the job seekers [2].

Can we do better? This question raises thorny issues at the crossroad of algorithmic fairness, ethics and information theory. Two such questions are: i) determining the desired biases at the population level; ii) acting at the individual recommendation level.

The internship will tackle both questions through counterfactual reasoning [3]. Basically, the VADORE algorithm learns a model $s(x, y)$ expressing how job y is desirable for job seeker x based on contracts and applications. This model will be enriched through auxiliary models, such as predicting the desirability *per se* of job y , e.g. predicting the number of applications to job y in the next time period. By tuning the impact of these auxiliary models in the overall model, we expect to increase the diversity of recommendations (which recommendations would I get if I were a competition addict? a competition hater? if I were in urgent need of finding a job? etc).

Notably, such an enriched model is expected to shed light on the causes for the gendered differences among users behaviors, e.g. in terms of competition aversion / appetite.

THE INTERNSHIP:

The goal is: i) to define and tackle auxiliary tasks related to job recommendation, such as predicting the desirability of a job offer; ii) to investigate how the models related to the auxiliary tasks interact with the central task of predicting the desirability of a job for a particular job seeker; iii) to interpret the impact of these auxiliary models on the diverse categories of users.

The internship will be conducted in cooperation with the French public employment service.

The internship requires excellent algorithmic skills (programming environment in Python), as well as creativity.

References

- [1] Guillaume Bied, Solal Nathan, Elia Perennes, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Christophe Gaillac, Michèle Sebag: Toward Job Recommendation for All. IJCAI 2023: 5906-5914
- [2] Guillaume Bied, Christophe Gaillac, Morgane Hoffmann, Philippe Caillou, Bruno Crépon, Solal Nathan, Michèle Sebag: Fairness in job recommendations: estimating, explaining, and reducing gender gaps. AEQUITAS@ECAI 2023
- [3] Judea Pearl, Dana Mackenzie: The Book of Why: The New Science of Cause and Effect. 2018