CentraleSupélec

université PARIS-SACLAY

MICS

PhD proposal:
Explaining Fairness in Preference-Based Assignments

Keywords: Computational social choice, Explainable AI, Multi-Agent Systems, Fair division, Fair matching.

1 Context

Many real-life applications deal with *preference-based assignments*. In such multi-agent problems, agents have preferences over elements (activities, resources, or even other agents), and these preferences must be aggregated into a collective decision which is an assignment of agents to these elements. Preference-based assignments include well-known problems of collective decision, such as the *allocation of indivisible resources* (assignment of students/teachers to courses, design of schedules, division of inheritance or household tasks, division of resources payed in common, etc.), or the *formation of coalitions* (formation of clubs or teams, formation of working groups, construction of strategic or military international alliances, etc.). These problems are fundamental in Computational Social Choice (COMSOC) [Brandt et al., 2016], subfield of Artificial Intelligence (AI) which studies the algorithmic aspects of collective decision.

Nowadays, with the increasing use of algorithms and AI tools in systems governing our life choices (job recruitment, insurance covering, universities assignment), important decisions for the agents can be made in preference-based assignments. When agents express preferences over other elements to be matched with, it is key to ensure the *fairness* of the assignment: no agent should feel unequally treated. Therefore, to ensure confidence and participation in the system, it is crucial to guarantee that the algorithms used for computing these assignments are fair to the agents.

Justifying that a given solution is fair is related to the ability of explaining decisions. *Explainable Artificial Intelligence (XAI)* is a hot topic in AI [Barredo Arrieta et al., 2020], which has even become a political and legal concern [Goodman and Flaxman, 2017]. In COMSOC, there is a long tradition of axiomatic characterizations of preference aggregation methods, which can be seen as rule-based explanations. Another recent line of research uses computer-aided methods [Geist and Peters, 2017] in order to derive justification of the outcomes of voting systems. While most of these works focus on voting scenarios [Boixel and Endriss, 2020, Boixel et al., 2022], only a few are specifically dedicated to preference-based assignments [Loustalot Knapp, 2022, Nizri et al., 2022].

2 Objective

The goal of this PhD is to investigate the explainability of fairness in preference-based assignments. The idea is to derive automatic justification for fairness of assignments with respect to some fairness criteria. Computer-aided tools such as SAT solving can be used.

3 Environment

The PhD is fully funded for 3 years (part of the ANR project APPLE-PIE) and will start around October 2023 (dates can be adapted). The PhD student will be welcomed in the MICS lab at CentraleSupélec

(3 rue Joliot Curie, 91190 Gif-sur-Yvette), and supervised by Anaëlle Wilczynski and Wassila Ouerdane (MICS, CentraleSupélec).

4 Candidate

We are looking for interested candidates with a Master (or engineer) degree (Bac+5 level) in Computer Science or Applied Mathematics. Solid skills in algorithmics are required, and a good knowledge in (algorithmic) game theory or computational social choice is appreciated. Most of all, the candidate should be interested by questions related to the contribution of AI to social justice.

5 How to apply

The interested candidates must send an email to Anaëlle Wilczynski (anaelle.wilczynski@centralesupelec.fr) and Wassila Ouerdane (wassila.ouerdane@centralesupelec.fr) with the following documents:

- a Curriculum Vitae,
- a motivation letter (max 2 pages),
- a transcript of the available grades for the current year and the past year,
- [optional] at most two recommendations.

References

A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion*, 58: 82–115, 2020.

A. Boixel and U. Endriss. Automated Justification of Collective Decisions via Constraint Solving. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS-20)*, pages 168–176, 2020.

A. Boixel, U. Endriss, and R. de Haan. A Calculus for Computing Structured Justifications for Election Outcomes. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22)*, 2022.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

C. Geist and D. Peters. Computer-aided methods for social choice theory. In *Trends in Computational Social Choice*, pages 249–267. AI Access, 2017.

B. Goodman and S. Flaxman. European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3):50–57, 2017.

D. Loustalot Knapp. Justification of Matching Outcomes. Master's thesis, University of Amsterdam, 2022.

M. Nizri, N. Hazon, and A. Azaria. Explainable Shapley-Based Allocation (Student Abstract). In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22)*, 2022.