A Reinforcement Learning Framework For Studying Group And Individual Fairness

Extended Abstract

Alexandra Cimpean
Vrije Universiteit Brussel
Brussels, Belgium
ioana.alexandra.cimpean@vub.be

Catholijn Jonker
Technische Universiteit Delft
Delft, The Netherlands
c.m.jonker@tudelft.nl

Pieter Libin
Vrije Universiteit Brussel
Brussels, Belgium
pieter.libin@vub.be

Ann Nowé
Vrije Universiteit Brussel
Brussels, Belgium
ann.nowe@vub.be

ABSTRACT

Reinforcement learning is a commonly used technique for optimising objectives in decision support systems for complex problem solving. When these systems affect individuals or groups, it is essential to reflect on fairness. As absolute fairness is in practice not achievable, we propose a framework which allows to balance distinct fairness notions along with the primary objective. To this end, we formulate group and individual fairness in sequential fairness notions. First, we present an extended Markov decision process, fMDP, that is explicitly aware of individuals and groups. Next, we formalise fairness notions in terms of this fMDP which allows us to evaluate the primary objective along with the fairness notions that are important to the user, taking a multi-objective reinforcement learning approach. To evaluate our framework, we consider two scenarios that require distinct aspects of the performance-fairness trade-off: job hiring and fraud detection. The objectives in job hiring are to compose strong teams, while providing equal treatment to similar individual applicants and to groups in society. The trade-off in fraud detection is the necessity of detecting fraudulent transactions, while distributing the burden for customers of checking transactions fairly. In this framework, we further explore the influence of distance metrics on individual fairness and highlight the impact of the history size on the fairness calculations and the obtainable fairness through exploration.

KEYWORDS

reinforcement learning; automated decision support; fairness framework; trustworthy AI

ACM Reference Format:
Alexandra Cimpean, Catholijn Jonker, Pieter Libin, and Ann Nowé. 2024. A Reinforcement Learning Framework For Studying Group And Individual Fairness: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

Fair and balanced automated decision support is essential, to avoid discrimination or favouritism towards individuals and groups. This is crucial in a wide array of applications, such as finance [9], job hiring [15, 16], epidemic mitigation [3, 4, 8] and fraud detection [12]. Fair decision support systems allow stakeholders to make informed decisions, taking into account an appropriate performance-fairness trade-off. This is important, as advice that is proposed by a decision support system might severely impact individuals and groups. Therefore, it is vital to study this matter to enable a wider acceptance of algorithms that support decision makers. As fairness requirements depend on the problem context and the decision maker's concerns, a framework should be capable of dealing with multiple fairness notions, that encompass the ethical considerations of the problem domain and the stakeholders. Consequently, it is important to develop a framework that considers fairness based on sensitive features (e.g., race and gender) and their combinations.

Recent work on fairness in RL has focused on single fairness notions in application-specific solutions [2, 6, 7, 14, 17, 20] and typically relies on reward shaping [2, 10]. However, such approaches do not suffice for real-world decision support problems, as the desired performance-fairness trade-off cannot be described upfront by stakeholders. Furthermore, real-world problems typically require multiple, possibly conflicting, fairness notions [11]. To this end, a multi-objective approach is essential to manage the main objective and to consider multiple fairness notions simultaneously [5]. We propose a formal fairness framework that is capable of dealing with multiple fairness notions. We experimentally evaluate this framework in job hiring and credit card fraud detection settings.

2 FAIRNESS FRAMEWORK

We define the fairness framework and highlight its requirements and suitability regarding distinct problem settings.

2.1 fMDP and the fairness history

A sequential decision process can be formally described as a Markov Decision Process (MDP) [19], consisting of a set of states $\mathcal{S}$, a set of actions $\mathcal{A}$, a set of rewards $\mathcal{R}$ and a transition function $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$ describing the probability of a next state $\mathbf{s}_{t+1}$ and reward r_t given the current state $\mathbf{s}_t$ and action a_t. We extend this standard MDP to an fMDP to encode a feedback signal f_t, that concerns an indication whether the chosen action a_t was correct at time t.

Existing fairness notions typically concern fair treatment between individuals or groups. We introduce the following notation regarding individuals and groups. $\mathcal{I}_t$ refers to the set of individuals involved in the decision process at time t and we use $i_t \in \mathcal{I}_t$ to refer to an individual of that set. In the job hiring setting, $\mathcal{I}_t$ refers to the set of candidates who applied for the job at time t and for which a decision (i.e., hire or reject the applicant) should be made. We refer to the set of all individuals involved in the decision process from the start $t = 0$ up to time T as $\mathcal{I}^T$. We define $\mathcal{G}_{g,t} \subseteq \mathcal{I}_t$ as the individuals of $\mathcal{I}_t$ that make up group g. We refer to all individuals involved in the decision process until time T, that belong to group g, as $\mathcal{G}_g^T$. For ease of notation, we assume that groups are predefined and can be empty. In the job hiring setting, $\mathcal{G}_g^T$ refers to the group of men or women, who applied for a job until time T. Given the fMDP, we assume that a state $\mathbf{s}_t$ provided to the RL agent encodes the individuals $\mathcal{I}_t$ and groups $\mathcal{G}_t$ involved in the decision at time t. Furthermore, the agent's action a_t encodes the decision impacting the involved individuals $\mathcal{I}_t$ and groups $\mathcal{G}_t$, and the feedback $\mathcal{I}_t$ and $\mathcal{G}_t$ specifies the correctness of that decision.

Given an fMDP, we define a history $\mathcal{H}^T$ until time T of past interaction tuples and their feedback regarding the ground truth:

$$\mathcal{H}^T = \{\mathbf{s}_t, a_t, r_t, f_t\}_{t=0}^T \tag{1}$$

We define the encountered states, selected actions and feedback from history $\mathcal{H}^T$ until time T respectively as $\mathcal{H}_S^T$, $\mathcal{H}_A^T$ and $\mathcal{H}_f^T$. Consequently, $\mathcal{H}_S^T$, $\mathcal{H}_A^T$ and $\mathcal{H}_f^T$ are also defined in terms of groups $\mathcal{G}^T$ and individuals $\mathcal{I}^T$.

2.2 Fairness notions

We formally define a fairness notion $\mathscr{F}$ as a power set $\mathscr{P}$ over $\mathcal{G}^T$ groups (Equation 2) and $\mathcal{I}^T$ individuals (Equation 3), given the history of encountered states $\mathcal{H}_S^T$, chosen actions $\mathcal{H}_A^T$ and feedback $\mathcal{H}_f^T$ until time T:

$$\mathscr{F} : \mathscr{P}(\mathcal{G}^T) \times \mathscr{P}(\mathcal{H}_S^T) \times \mathscr{P}(\mathcal{H}_A^T) \times \mathscr{P}(\mathcal{H}_f^T) \hookrightarrow \mathbb{R}^- \tag{2}$$

$$\mathscr{F} : \mathscr{P}(\mathcal{I}^T) \times \mathscr{P}(\mathcal{H}_S^T) \times \mathscr{P}(\mathcal{H}_A^T) \times \mathscr{P}(\mathcal{H}_f^T) \hookrightarrow \mathbb{R}^- \tag{3}$$

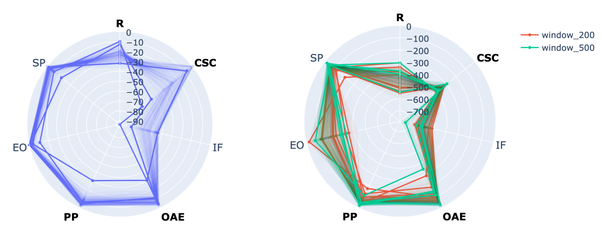
The fairness notion $\mathscr{F}$ is defined as the negative absolute difference in treatment between groups or individuals. The closer $\mathscr{F}$ is to zero, the smaller the difference in treatment is between the groups or individuals. While $\mathscr{F}$ may be intractable due to limitations of defining exact fairness [6], we propose to approximate it with $\hat{\mathscr{F}}$. For a future fairness objective, $\mathscr{F}$, and by extension its approximation $\hat{\mathscr{F}}$ provide a foundation for a reward signal that can be used with a multi-objective RL approach.

3 RESULTS

As both the job hiring and fraud detection scenario deal with a reward and multiple fairness objectives, the number of policies with suitable trade-offs can scale exponentially. To this end, we use Pareto Conditioned Networks (PCN) [13], as PCN trains a single neural network to approximate all non-dominated policies, by applying supervised learning techniques to improve the policy. We report the learned non-dominated coverage sets for all fairness notions and the main reward [5], across 10 seeds after 500 000 steps. Meaning, the reward vector consists of: the main reward (R) and the fairness notions statistical parity (SP), equal opportunity (EO), overall accuracy equality (OAE), predictive parity (PP), individual fairness (IF) and consistency score complement (CSC) [11, 21].

For the job hiring scenario, we train an agent to hire and maintain a well-performing team of 100 employees from the Belgian population [18], where episodes last for a maximum of 1000 timesteps. We ask the agent to optimise four objectives: {R, SP, EO, IF}. We implement the fairness history as a sliding window of 100 timesteps and use the Bray-Curtis distance metric for the individual fairness notions. Figure 1a shows a representative set of policies from the non-dominated coverage sets. Note how the some of the best learned policies are close to 0 for all group fairness notions, indicating the agent has learned policies which can satisfy more fairness notions than initially requested. In contrast, due to the impact of how the individual fairness notions are defined, it is possible for the agent to find larger differences in non-dominated values among them. However, obtaining a higher CSC or IF comes at the cost of the main reward and some additional group fairness notions.

For the fraud detection scenario, we assume the default parameters of the MultiMAuS simulator [22], but increase the frequency of fraudulent transactions to approximately 10%. We let the agent check transactions for a week, resulting in at most 1000 transactions per episode. We ask the agent to optimise on four objectives: {R, OAE, PP, CSC}. Figure 1b shows the learned trade-offs for two different history sizes. The policies learned by the agent across both window sizes follow similar trade-offs with regards to the reward and the fairness notions. Note that individual fairness is low for both IF and CSC. The largest contributor to this effect is the different base rates for fraudulent transactions between individuals, indicating the agent has mostly focused on improving the requested group fairness notions, at the cost of individual fairness.



(a) Job hiring **(b) Fraud detection**

Figure 1: Representative policies from the non-dominated coverage sets, with requested objectives in bold.

ACKNOWLEDGMENTS

Alexandra Cimpean receives funding from the Fonds voor Wetenschappelijk Onderzoek (FWO) via fellowship grant 1SF7823N. All experiments were performed on the VSC high performance computing infrastructure [1].

REFERENCES

[1] Vlaams Supercomputing Center. 2023. Hydra hardware. https://www.vscentrum.be https://www.vscentrum.be.

[2] Jingdi Chen, Yimeng Wang, and Tian Lan. 2021. Bringing fairness to actor-critic reinforcement learning for network utility optimization. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications* (Vancouver, BC, Canada). IEEE Press, Vancouver, BC, Canada, 1–10. https://doi.org/10.1109/INFOCOM42981.2021.9488823

[3] Alexandra Cimpean, Timothy Verstraeten, Lander Willem, Niel Hens, Ann Nowé, and Pieter Libin. 2023. Evaluating COVID-19 vaccine allocation policies using Bayesian m-top exploration. *arXiv preprint arXiv:2301.12822* (2023), 26.

[4] Ezekiel J. Emanuel, Govind Persad, Adam Kern, Allen Buchanan, Cécile Fabre, Daniel Halliday, Joseph Heath, Lisa Herzog, R. J. Leland, Ephrem T. Lemango, Florencia Luna, Matthew S. McCoy, Ole F. Norheim, Trygve Ottersen, G. Owen Schaefer, Kok-Chor Tan, Christopher Heath Wellman, Jonathan Wolff, and Henry S. Richardson. 2020. An ethical framework for global vaccine allocation. *Science* 369, 6509 (2020), 1309–1312. https://doi.org/10.1126/science.abe2803

[5] Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. 2022. A practical guide to multi-objective reinforcement learning and planning. In *AAMAS* (2022/04/13), Vol. 36. 26.

[6] Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. 2017. Fairness in Reinforcement Learning. In *ICML (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, Sydney, Australia, 1617–1626. https://proceedings.mlr.press/v70/jabbari17a.html

[7] Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. 2016. Fairness in Learning: Classic and contextual bandits. *Advances in Neural Information Processing Systems* 29 (2016), 325–333. arXiv:1605.07139

[8] Pieter J. K. Libin, Arno Moonens, Timothy Verstraeten, Fabian Perez-Sanjines, Niel Hens, Philippe Lemey, and Ann Nowé. 2021. Deep Reinforcement Learning for Large-Scale Epidemic Control. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*, Yuxiao Dong, Georgiana Ifrim, Dunja Mladenić, Craig Saunders, and Sofie Van Hoecke (Eds.). Springer International Publishing, Cham, 155–170.

[9] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. 2018. Delayed Impact of Fair Machine Learning. In *ICML*, Vol. 80. PMLR, Stockholm, Sweden, 3150–3158.

[10] Weiwen Liu, Feng Liu, Ruiming Tang, Ben Liao, Guangyong Chen, and Pheng Ann Heng. 2020. *Balancing Between Accuracy and Fairness for Interactive Recommendation with Reinforcement Learning*. Vol. 12084 LNAI. Springer International Publishing, Cham. 155–167 pages. https://doi.org/10.1007/978-3-030-47426-3_13 arXiv:2106.13386

[11] Karima Makhlouf, Sami Zhioua, and Catuscia Palamidessi. 2020. On the applicability of ML fairness notions. , 32 pages. arXiv:2006.16745

[12] Dennis Soemers, Ann Nowé, Tim Brys, Kurt Driessens, and Mark Winands. 2018. Adapting to Concept Drift in Credit Card Transaction Data Streams Using Contextual Bandits and Decision Trees. *AAAI* 32, 1 (2018), 7831–7836.

[13] Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. 2022. Pareto Conditioned Networks. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems* (Virtual Event, New Zealand) *(AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1110–1118.

[14] Manel Rodriguez-Soto, Maite Lopez-Sanchez, and Juan A Rodriguez-Aguilar. 2021. Guaranteeing the Learning of Ethical Behaviour through Multi-Objective Reinforcement Learning. *ALA* (2021), 9.

[15] Candice Schumann, Samsara N. Counts, Jeffrey S. Foster, and John P. Dickerson. 2019. The Diverse Cohort Selection Problem. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* 2 (2019), 601–609. arXiv:1709.03441

[16] Candice Schumann, Jeffrey S. Foster, Nicholas Mattei, and John P. Dickerson. 2020. We need fairness and explainability in algorithmic hiring. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* 2020-May, Aamas (2020), 1716–1720.

[17] Umer Siddique, Paul Weng, and Matthieu Zimmer. 2020. Learning fair policies in multiobjective (Deep) reinforcement learning with Average and Discounted Rewards. *ICML* 119 (13–18 Jul 2020), 8864–8874. https://proceedings.mlr.press/v119/siddique20a.html

[18] STATBEL. 2023. Employment and unemployment. https://statbel.fgov.be/en/themes/work-training/labour-market/employment-and-unemployment#figures

[19] Richard S. Sutton, Andrew G. Barto, and et al. 2018. *Reinforcement Learning : An Introduction*. MIT Press. 526 pages.

[20] Paul Weng. 2019. Fairness in reinforcement learning. *CoRR* abs/1907.10323 (2019), 5. arXiv:1907.10323

[21] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning Fair Representations. In *ICML (Proceedings of Machine Learning Research, Vol. 28)*, Sanjoy Dasgupta and David McAllester (Eds.). PMLR, Atlanta, Georgia, USA, 325–333.

[22] Luisa M Zintgraf, Edgar A Lopez-Rojas, Diederik M Roijers, and Ann Nowé. 2017. MultiMAuS: a multi-modal authentication simulator for fraud detection research. In *29th European Modeling and Simulation Symp.(EMSS 2017)*. Curran Associates, Inc., 360–370.