Multi-objective reinforcement learning: an ethical perspective

Timon Deschamps, Rémy Chaput, Laëtitia Matignon
LIRIS, Université Claude Bernard Lyon 1

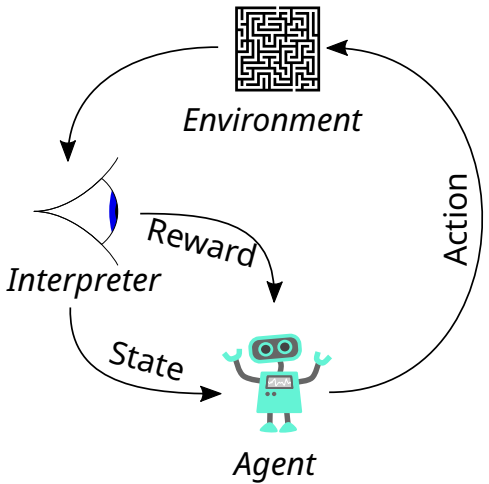
MODeM Workshop @ ECAI - 20/10/2024

Reinforcement learning (RL)

An agent interacts with an environment to learn to **maximize rewards** from its experience.

MDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$

MOMDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{R}, \gamma \rangle$
$\rightarrow$ rewards are vectors in $\mathbb{R}^m$

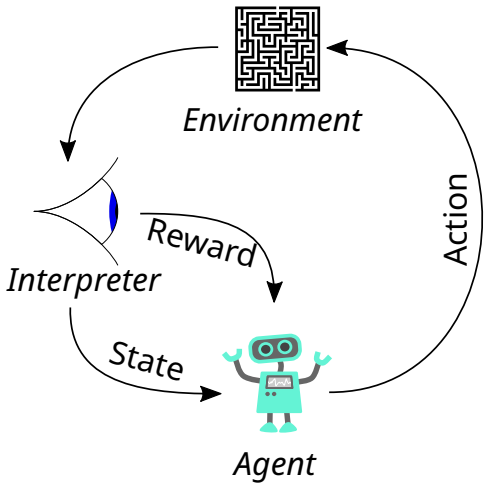


Reinforcement learning (RL)

An agent interacts with an environment to learn to **maximize rewards** from its experience.

MDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$

MOMDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{R}, \gamma \rangle$
$\rightarrow$ rewards are vectors in $\mathbb{R}^m$

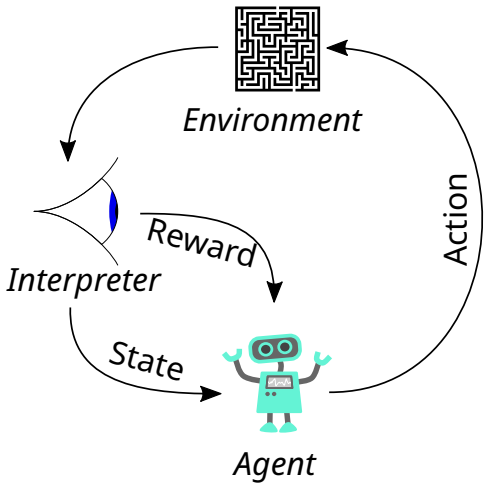


Reinforcement learning (RL)

An agent interacts with an environment to learn to **maximize rewards** from its experience.

MDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$

MOMDP : $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathbf{R}, \gamma \rangle$
$\rightarrow$ rewards are vectors in $\mathbb{R}^m$



Multi-objective RL (MORL)

Essential to model real world problems [4, 26].

Example: autonomous car balancing speed and comfort ($m = 2$)

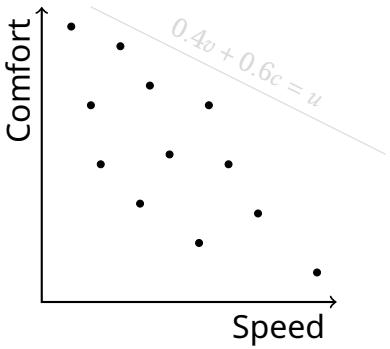
A user has preferences : $w_s = 0.4, w_c = 0.6$

Multi-objective RL (MORL)

Essential to model real world problems [4, 26].

Example: autonomous car balancing speed and comfort ($m = 2$)

A user has preferences : $w_s = 0.4, w_c = 0.6$

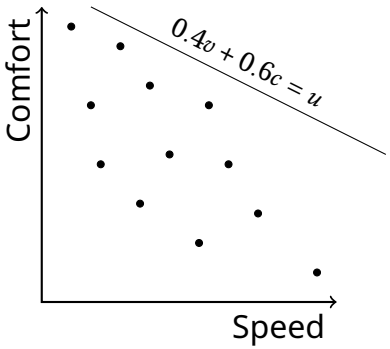


Multi-objective RL (MORL)

Essential to model real world problems [4, 26].

Example: autonomous car balancing speed and comfort ($m = 2$)

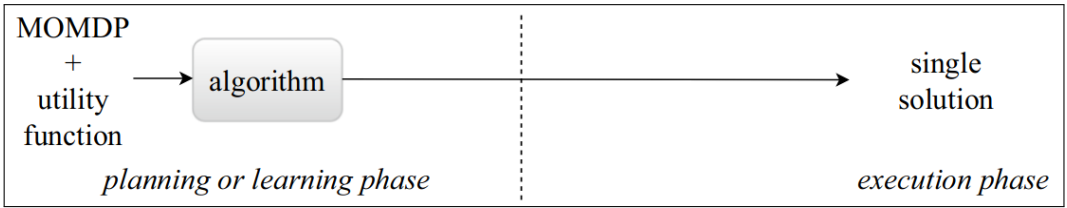
A user has preferences : $w_s = 0.4, w_c = 0.6$



Multi-objective RL: utility functions

Utility function $u : \mathbb{R}^m \rightarrow \mathbb{R}$:

- linear : $u(\mathbf{r}) = \mathbf{w}^\top \mathbf{r}$
- general : $u(\mathbf{r})$ is monotonically increasing

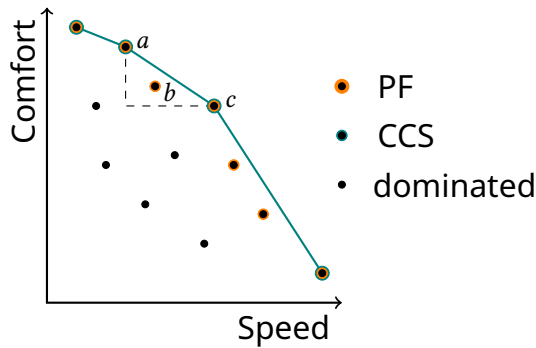


From Hayes et al. [4]

What do we do when u is unkown?

Multi-objective RL: theory

- Single/multi-policy algorithms
- Pareto-dominance $\pi \succ_P \pi'$



A taxonomy of MORL

	single policy (known u)		multiple policies (unknown u)	
	deterministic	stochastic	deterministic	stochastic
linear u	one policy in Π_{DS}: DQN [9], REINFORCE [20]		CCS of policies in Π_{DS}: Envelope [29], PG-MORL [28], PD-MORL [2], CN [1]	
general u	one policy in Π_{D}: EUPG [17], MOCAC [14], Q-steering [23]	mixture of policies in Π_{DS}: π-mix [22], S-rand [27]	PCS of policies in Π_{D}: PQL [10], PCN [13]	mixture of policies in Π_{DS}: CAPQL [6], π-mix [22], S-rand [22]

Taxonomy from Roijers et al. [19], Hayes et al. [4].

MORL and ethics

Machine ethics: artificial agents with ethically-aligned behaviors

Normative ethics:

consequentialism virtue ethics deontology

Vamplew et al. [24] argue that the multi-objective aspect is essential to develop human-aligned AI.

MORL and ethics

Machine ethics: artificial agents with ethically-aligned behaviors

Normative ethics:

consequentialism virtue ethics deontology

Vamplew et al. [24] argue that the multi-objective aspect is essential to develop human-aligned AI.

MORL and ethics

Machine ethics: artificial agents with ethically-aligned behaviors

Normative ethics:

consequentialism virtue ethics deontology

Vamplew et al. [24] argue that the multi-objective aspect is essential to develop human-aligned AI.

MORL: ethical properties

- **User-centric**: explicit consideration of the user
- **Adaptative**: adapt to the evolution of users and of society
- **Normative**: ability to follow a set of norms
- **Multi-agent**: account for, and collaborates with other agents

MORL: ethical properties

- **User-centric**: explicit consideration of the user
- **Adaptative**: adapt to the evolution of users and of society
- **Normative**: ability to follow a set of norms
- **Multi-agent**: account for, and collaborates with other agents

MORL: ethical properties

- **User-centric**: explicit consideration of the user
- **Adaptative**: adapt to the evolution of users and of society
- **Normative**: ability to follow a set of norms
- **Multi-agent**: account for, and collaborates with other agents

MORL: ethical properties

- **User-centric**: explicit consideration of the user
- **Adaptative**: adapt to the evolution of users and of society
- **Normative**: ability to follow a set of norms
- **Multi-agent**: account for, and collaborates with other agents

MORL: ethical properties - 2

MORL methods	UC	A	N	MA
CN [1], DMCRL [11], Q-steering [23]	✓	✓		
MAEE [15]			✓	✓
GUTS [18], MORAL [12], DWPI [7], QSOM-MORL [3]	✓			
EE [16], TLO [25]			✓	
MO-MIX [5], PRBS/D [8], moral rewards [21]				✓

Thank you!

References I

[1] Axel Abels, Diederik Roijers, Tom Lenaerts, Ann Nowé, and Denis Steckelmacher. Dynamic Weights in Multi-Objective Deep Reinforcement Learning. In *ICML*, 2019.

[2] Toygun Basaklar, Suat Gumussoy, and Umit Y. Ogras. PD-MORL: Preference-Driven Multi-Objective Reinforcement Learning Algorithm. In *ICLR*, 2023.

[3] Rémy Chaput, Laetitia Matignon, and Mathieu Guillermin. Learning to identify and settle dilemmas through contextual user preferences. In *ICTAI*, 2023. doi: 10.1109/ICTAI59109.2023.00075.

[4] 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. A practical guide to multi-objective reinforcement learning and planning. *AAMAS*, 2022. ISSN 1573-7454. doi: 10.1007/s10458-022-09552-y.

[5] Tianmeng Hu, Biao Luo, Chunhua Yang, and Tingwen Huang. MO-MIX: Multi-Objective Multi-Agent Cooperative Decision-Making With Deep Reinforcement Learning. *IEEE PAMI*, 2023. ISSN 0162-8828, 2160-9292, 1939-3539. doi: 10.1109/TPAMI.2023.3283537.

[6] Haoye Lu, Daniel Herman, and Yaoliang Yu. Multi-Objective Reinforcement Learning: Convexity, Stationarity and Pareto Optimality. In *ICLR*, 2022.

[7] Junlin Lu, Patrick Mannion, and Karl Mason. Inferring Preferences from Demonstrations in Multi-objective Reinforcement Learning: A Dynamic Weight-based Approach. In *ALA (AAMAS)*, 2023.

[8] Patrick Mannion, Sam Devlin, Jim Duggan, and Enda Howley. Reward shaping for knowledge-based multi-objective multi-agent reinforcement learning. *The Knowledge Engineering Review*, 2018. ISSN 0269-8889, 1469-8005. doi: 10.1017/S0269888918000292.

[9] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *NIPS*, 2013.

[10] Kristof Van Moffaert and Ann Nowé. Multi-Objective Reinforcement Learning using Sets of Pareto Dominating Policies. *JMLR*, 2014. ISSN 1533-7928.

[11] Sriraam Natarajan and Prasad Tadepalli. Dynamic preferences in multi-criteria reinforcement learning. In *ICML*, 2005.

References II

[12] Markus Peschl, Arkady Zgonnikov, Frans A Oliehoek, and Luciano C Siebert. Moral: Aligning ai with human norms through multi-objective reinforced active learning. In *AAMAS*, 2022.

[13] Mathieu Reymond, Eugenio Bargiacchi, and Ann Nowé. Pareto Conditioned Networks. In *AAMAS*, 2022.

[14] Mathieu Reymond, Conor F. Hayes, Denis Steckelmacher, Diederik M. Roijers, and Ann Nowé. Actor-critic multi-objective reinforcement learning for non-linear utility functions. In *AAMAS*, 2023. doi: 10.1007/s10458-023-09604-x.

[15] Manel Rodriguez-Soto, Maite Lopez-Sanchez, and Juan A. Rodriguez-Aguilar. Multi-objective reinforcement learning for designing ethical multi-agent environments. *Neural Computing and Applications*, 2023. ISSN 1433-3058. doi: 10.1007/s00521-023-08898-y.

[16] Manel Rodriguez-Soto, Roxana Rădulescu, Juan A Rodriguez-Aguilar, and Maite Lopez-Sanchez. Multi-objective reinforcement learning for guaranteeing alignment with multiple values. In *ALA (AAMAS)*, 2023.

[17] Diederik Roijers, Denis Steckelmacher, and Ann Nowe. Multi-objective Reinforcement Learning for the Expected Utility of the Return. In *ALA (AAMAS)*, 2018.

[18] Diederik M. Roijers, Luisa M. Zintgraf, Pieter Libin, and Ann Nowé. Interactive multi-objective reinforcement learning in multi-armed bandits for any utility function. In *ALA (AAMAS)*, 2020.

[19] Diederik Marijn Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. A Survey of Multi-Objective Sequential Decision-Making. *JAIR*, 2013. ISSN 1076-9757. doi: 10.1613/jair.3987.

[20] Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*, 1999.

[21] Elizaveta Tennant, Stephen Hailes, and Mirco Musolesi. Modeling moral choices in social dilemmas with multi-agent reinforcement learning. In *IJCAI*, 2023.

[22] Peter Vamplew, Richard Dazeley, Ewan Barker, and Andrei Kelarev. Constructing Stochastic Mixture Policies for Episodic Multiobjective Reinforcement Learning Tasks. In *Advances in Artificial Intelligence*. 2009.

[23] Peter Vamplew, Rustam Issabekov, Richard Dazeley, Cameron Foale, Adam Berry, Tim Moore, and Douglas Creighton. Steering approaches to pareto-optimal multiobjective reinforcement learning. *Neurocomputing*, 2017.

References III

[24] Peter Vamplew, Richard Dazeley, Cameron Foale, Sally Firmin, and Jane Mummery. Human-aligned artificial intelligence is a multiobjective problem. *Ethics and Information Technology*, 2018.

[25] Peter Vamplew, Cameron Foale, Richard Dazeley, and Adam Bignold. Potential-based multiobjective reinforcement learning approaches to low-impact agents for ai safety. *Engineering Applications of Artificial Intelligence*, 2021.

[26] Peter Vamplew, Benjamin J Smith, Johan Källström, Gabriel Ramos, Roxana Rădulescu, Diederik M Roijers, Conor F Hayes, Fredrik Heintz, Patrick Mannion, Pieter JK Libin, et al. Scalar reward is not enough: A response to silver, singh, precup and sutton (2021). *Autonomous Agents and Multi-Agent Systems*, 36(2):41, 2022.

[27] Kazuyoshi Wakuta. A note on the structure of value spaces in vector-valued Markov decision processes. *Mathematical Methods of Operations Research*, 1999. ISSN 1432-2994, 1432-5217. doi: 10.1007/PL00020907.

[28] Jie Xu, Yunsheng Tian, Pingchuan Ma, Daniela Rus, Shinjiro Sueda, and Wojciech Matusik. Prediction-Guided Multi-Objective Reinforcement Learning for Continuous Robot Control. *ICML*, 2020.

[29] Runzhe Yang, Xingyuan Sun, and Karthik Narasimhan. A Generalized Algorithm for Multi-Objective Reinforcement Learning and Policy Adaptation. In *NeurIPS*, 2019.