Description of the document

The submitted manuscript has been rejected once before a submission to TCML. We would like to thank our current and former reviewers for their valuable help in improving the manuscript and share how we authors have carefully addressed the past presented concerns to promote open, transparent, and ethical review process.

The four reviewers' feedback is split into six categories based on their content, presented below: Benchmarks and tests, Proofs and applicability, Algorithm's incompleteness, Language and format, SHGO, and Negative Societal impact. At the end of each category, authors detail their response and changes made to the manuscript based on the feedback.

Benchmarks and tests

Reviewers

R1: The evaluation is weak for an algorithmic/empirical work. What experimental setup or data is used to generate Fig. 4? It is important to define the interactive demo env in the main paper to complement with Fig. 4. 2.

R2: The authors describe a general scheme, but don't provide any precise RL algorithm for their proposed approach in the main paper. Some empirical evaluation is presented, but it is not clear what the experimental set-up is. Have you performed any thorough evaluation of your proposed method in an RL task?

R3: While the derived algorithm appears promising, an experimental evaluation is entirely missing. Why are there no experiments? Did you not find any suitable benchmarks?

Authors

Currently at the time of the previous submission, no suitable benchmarks or datasets were available for the testing. As of 2023, there is now a publicly available benchmark, DeepSeaTreasure v1, made by Cassimon et al. for multi-objective RL cases. This benchmark has been now utilized to test the proposed algorithm, and the tests are carefully detailed now in their own section.

Based on the feedback regarding the interactive visualisation and its implementation, the section in the manuscript regarding this matter as well as the accompanied has now been revised to a more carefully detailed format.

Proofs and applicability

Reviewers

R1: The scientific arguments are vague at times, e.g., "An additional note can be made from the trivial case where the reward transfer won't work: the reward vector of zeros where no reward can be transferred. However, as that case is practically impossible to attain, it shall be ignored here."

R2: The central terms "reward transfer" and "priority order" are not clearly defined. The paper remains vague in many aspects and is therefore largely incomprehensible. Incorrect claims are made: It is claimed that the paper proves that reward transfer has a global minimum given priority order. Apart from the fact that the terms "reward transfer" and "priority order" are not formally defined, I can see no such proof in the paper. No evidence is brought that the proposed method works. No evidence is provided that the proposed method adds value over existing methods.

R3: An in-depth theoretical analysis is also not provided (i.e. proof of convergence to the correct utility function or anything alike).

R4: You claim, “We prove that the reward transfer has a global minimum given the priority order, which can be obtained using an existing optimisation algorithm”. Where is this poof given? What evidence is provided that the proposed method is useful?

Authors

The paper was reformatted to contain regular mathematical definitions and proofs based on the interpretation of R2’s concerns. After reformatting, a more in-depth theoretical analysis from a mathematical point of view was given to the said definitions and proofs.

The proof of the minimum regarding the transfer problem was explained in a more detailed manner, to show why it is a proof, what it proves and why is that proof fundamental for the algorithm.

A new section discussing the convergence to the utility function was provided. While this convergence cannot explicitly be guaranteed, a new test designed to show approximate solutions was provided in the empirical testing. This test should provide evidence that the proposed method is useful.

Algorithm 1’s incompleteness

Reviewers

R1: Algorithm 1 seems to be incomplete. Does the Alg. 1 just depict one iteration of the process depicted in Fig. 2?

Authors

Algorithm 1 depicted one iteration of the process. Algorithm 1 has now been revised to depict the whole outline, and a longer explanation has been given in the caption.

Language and format

Reviewers

R1: The paper can be formatted better. The formatting of phrases in quotes in Sec. 4.1 needs to be corrected.

R2: The paper is not completely anonymized (acknowledgement of funding not hidden). The writing is often imprecise and contains several incorrect statements, e.g., In Related Work: "Mannion et al." -> "Mannor et al." and "Vamplev et al." -> "Vamplew et al." I don't think the sentence "Mannion et al. do not consider multi-objective problems" is correct, since the title of their paper is "A ≡ geometric approach to multi-criterion reinforcement learning". For Perny et al., I believe that they also consider the multi-objective case, not only the bi-objective case. In Section 3: The sentence "The reward is a team reward…" is a bit perplexing since it seems the authors consider a single-agent RL problem.

R4: Examples of a largely incomprehensible text:

- "Additionally, allowing differing priorities between objectives’ themselves allows for a broader spectrum of problem formulation as well, as there are indeed problems where the optimisation of one objective downgrades the optimisation of the other objective."
- “the optimal policy π is not the reward vector with the highest values, but rather the reward vector that has the user-preferred distribution.” How is a policy supposed to be a reward vector?
- “a priori info” -> “a priori information”
- “the driver must maintain security” -> “the driver must maintain safety”

- “Monte Carlo Tree search” -> “Monte Carlo tree search”
- “a RL” -> “an RL”
- “only a couple examples can be made”

Authors

The paper’s format was revised to a more mathematical one. The one-letter or one-word typos mentioned by R2 and R4 have been corrected. The discerns over multi-objective RL studies presented by Mannion et al. and Perny et al. have been more carefully detailed. While Mannion et al. consider multiple criteria, but these criteria are formulated as a linearly weighted sum, compressing the multiple objectives’ case into a single-objective case. While the theoretical work by Perny et al. has more than two objectives, their implementation only tests the bi-objective case. The phrase regarding the team reward has been revised, as one part of the proposed algorithm is no longer implemented as an RL agent. The formatting of phrases in quotes in Sec. 4.1 was corrected.

SHGO

Reviewers

R3: What is SHGO?

Authors

The development of SHGO is not intended to be our contribution, nor is the algorithm’s extensive analysis in the scope of the submitted manuscript. As such, the presentation of SHGO and a description of its implementation in an extensive manner is left to Reed et al. in their respective paper. The analysis and presentation of SHGO in the submitted manuscript, was, however, extended to better address this concern.

Negative societal impact

Reviewers

R3: Negative societal impacts are not discussed.

Authors

A report and discussion about the algorithm’s carbon emissions and power consumption has been added. At large, other negative societal impacts are not applicable to this theoretical work, and as such they are not discussed.