Evolutionary Computation and Explainable AI: a year in review

Jaume Bacardit[1][0000−0002−2692−7205]
Alexander E.I. Brownlee[2][0000−0003−2892−5059]
Stefano Cagnoni[3][0000−0003−4669−512X]
Giovanni Iacca[4][0000−0001−9723−1830]
John McCall[5][0000−0003−1738−7056]
David Walker[6][0000−0001−8686−4253]

[1] Newcastle University, Newcastle u/o Tyne, UK jaume.bacardit@newcastle.ac.uk
[2] University of Stirling, Stirling, UK alexander.brownlee@stir.ac.uk
[3] University of Parma, Parma, Italy stefano.cagnoni@unipr.it
[4] University of Trento, Trento, Italy giovanni.iacca@unitn.it
[5] Robert Gordon University, Aberdeen, UK j.mccall@rgu.ac.uk
[6] University of Plymouth, Plymouth, UK david.walker@plymouth.ac.uk

Abstract. In 2022, we organized the first Workshop on Evolutionary Computation and Explainable AI (ECXAI). With no pretence at completeness, this paper briefly comments on its outcome, what has happened since then in the field, and our expectations for the near future.

Keywords: Explainable Artificial Intelligence · Evolutionary Computation and Optimisation · Machine Learning.

1 Explainable AI: facts and motivations

The increasing adoption of black-box algorithms, including Evolutionary Computation (EC)-based methods, has led to greater attention to generating explanations and their accessibility to end-users. This has created a fertile environment for the application of techniques in the EC domain for generating both end-user- and researcher-focused explanations. Furthermore, many XAI approaches in Machine Learning (ML) rely on search algorithms – e.g., [10] – that can draw on the expertise of EC researchers.

Important questions that automated decision-making techniques (such as EC and ML) have raised include: (1) Why has the algorithm obtained solutions in the way that it has? (2) Is the system biased? (3) Has the problem been formulated correctly? (4) Is the solution trustworthy and fair?

The goal of XAI and related research is to develop methods to interrogate AI processes and answer these questions. Our position is that, despite the differences in the problem formulation (ML vs. optimisation), using or adapting XAI techniques to explain EC-based processes that tackle search problems will improve such methods' accessibility to a wider audience, increasing their uptake

and impact. As well as this, we posit that EC can play a crucial role in improving the state-of-the-art XAI techniques within the wider AI community.

Perhaps the most crucial reason why explainability is important is **trust**. The research community is already largely convinced of the value of EC approaches and is keen to increase the uptake of EC tools and methods by non-EC experts. Central to this is convincing users that they can trust the solutions that are generated by knowing *what* makes that solution better than something else, which might be synonymous with knowing *why* the solution was chosen.

Extending this theme is that of **validity**. EC methods, and optimisers in general, only optimise the target function. Explaining why a solution was chosen might clarify whether it is solving the actual problem or just exploiting an error or loophole in the problem's definition, which can lead to surprising or even amusing results [8], but can also simply yield frustratingly incorrect solutions.

EC is stochastic, which makes noise in the generated solutions virtually unavoidable. Thus, another motivation is whether we can explain which characteristics of the solution are crucial: its **malleability**. This property could be assessed by answering the question: "Which variables could be refined or amended for aesthetic or implementation purposes?".

Finally, when we define a problem, it is often hard to fully codify all the real-world goals of the system. We want something that is mathematically optimised but also something that corresponds to the problem owner's hard-to-codify intuition. By incorporating XAI into interactive EC, we could make it easier for the problem owner to interact with the optimiser [15].

Based mainly on these considerations and aiming to support these research lines, we decided to foster a tighter interaction between EC-based and other AI methods by organizing a dedicated workshop at GECCO 2022 [2], that will be iterated in 2023, inviting participants to focus their contributions on two main, complementary, issues, namely: (1) How EC can contribute to XAI?; and (2) How XAI can be used to explain EC-based solutions?

2 A year in review

Several papers were discussed at the ECXAI22 workshop. An introductory paper [2] outlined the broad research questions around EC and XAI, as noted above, and gave a brief literature review of recent work. Two papers explored routes for possible explainability in EC, one describing the mining of surrogate models for characteristics like the sensitivity of the objectives to each variable and inter-variable relationships for bitstring-encoded benchmark functions [13], and the other proposing Population Dynamics Plots to visualise the progress of an EA, to allow the lineage of solutions to be traced back to their origins and provide a route to explaining the behaviour of different algorithms [16].

Other papers explored the use of EC in providing explainability for ML systems, in particular, related to the assessment of Neural Transformers Trained on source code [11], the evolution of interpretable restriction policies for pandemics control [4], and the optimisation of explainable rule sets [12] and Learning Classifier Systems (LCS) through feature clustering [1].

2.1 Other recent relevant papers and new trends

Genetic Programming (GP) has long been claimed to produce better explainable models than other ML methods for its capacity to evolve complex symbolic expressions that intrinsically define their semantics. For the same reasons, GP has also the capacity to be a useful tool for post-hoc explanation of black-box models. These two viewpoints are extensively discussed in [9]. GP is also used in [14], aiming at deriving a context-aware approach, which essentially means developing a system that can decompose the main problem into a set of sub-problems (contexts) and finding specific solutions to each of them. According to the authors, this approach results in prediction models that are smaller and easier to interpret than those obtained by the evolutionary learning algorithms without context awareness.

Recent work has also focused on the use of evolutionary learning to induce decision trees combined with reinforcement learning (RL), both for discrete [5] and continuous action spaces [4]. A similar approach has been applied for the automated analysis of ultrasound imaging data [6], and, more recently, in the domain of multi-agent reinforcement learning (MARL) tasks [3].

One of the most appealing recent trends currently emerging relates to the use of Quality Diversity (QD) evolutionary algorithms, such as MAP-Elites, to search for a multitude of diverse policies for RL tasks [7].

3 An invitation and concluding remarks

The intersection between XAI and EC is evidently an emerging area, as demonstrated by the steady stream of recent publications and interest in the ECXAI workshop at GECCO 2022. Several, quite different approaches have been reported in the literature since then, and the topic is ripe for cross-fertilisation of ideas between the EC and XAI communities.

On the one hand, using XAI to attempt to explain the behavior and outcomes of EC techniques seems to be a viable way to attract attention on these optimisation methods and present them as a rigorous, robust alternative for solving complex optimisation problems. Using XAI in EC may also be a way to free the field from the excessive use of metaphors, that has been heavily criticized in the past few years, focusing more on the analysis of the algorithmic functionalities.

On the other hand, using EC for developing or augmenting XAI seems another important direction that deserves to be explored in the future. In this sense, the seminal works that have just been published on the use of QD algorithms for interpretable RL, or of EC for interpretable MARL, are encouraging.

We invite further participation in the ECXAI workshop at GECCO 2023[7].

References

1. Andersen, H., Lensen, A., Browne, W.N.: Improving the search of learning classifier systems through interpretable feature clustering. In: Proceedings of the Genetic

[7] https://ecxai.github.io/ecxai/workshop-2023

and Evolutionary Computation Conference Companion. p. 1752–1756. GECCO '22, ACM, New York, NY, USA (2022)
2. Bacardit, J., Brownlee, A.E.I., Cagnoni, S., Iacca, G., McCall, J., Walker, D.: The intersection of evolutionary computation and explainable ai. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1757–1762. GECCO '22, Association for Computing Machinery, New York, NY, USA (2022)
3. Crespi, M., Custode, L.L., Iacca, G.: Towards interpretable policies in multi-agent reinforcement learning tasks. In: Bioinspired Optimization Methods and Their Applications: 10th International Conference, BIOMA 2022, Maribor, Slovenia, November 17–18, 2022, Proceedings. pp. 262–276. Springer, Cham (2022)
4. Custode, L.L., Iacca, G.: Interpretable AI for policy-making in pandemics. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1763–1769. GECCO '22, ACM, New York, NY, USA (2022)
5. Custode, L.L., Iacca, G.: Evolutionary learning of interpretable decision trees. IEEE Access **11**, 6169–6184 (2023)
6. Custode, L.L., Mento, F., Tursi, F., Smargiassi, A., Inchingolo, R., Perrone, T., Demi, L., Iacca, G.: Multi-objective automatic analysis of lung ultrasound data from COVID-19 patients by means of deep learning and decision trees. Applied Soft Computing **133**, 109926 (2023)
7. Ferigo, A., Custode, L.L., Iacca, G.: Quality diversity evolutionary learning of decision trees. arXiv:2208.12758 (2022)
8. Lehman, J., Clune, J., Misevic, D.: The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities (2020), arXiv:1803.03453
9. Mei, Y., Chen, Q., Lensen, A., Xue, B., Zhang, M.: Explainable Artificial Intelligence by Genetic Programming: A survey. IEEE Transactions on Evolutionary Computation pp. 1–1 (2022), In press, Early Access
10. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?". Explaining the predictions of any classifier. In: International Conference on Knowledge Discovery and Data Mining. ACM SIGKDD, New York, NY, USA (2016)
11. Saletta, M., Ferretti, C.: Towards the evolutionary assessment of neural transformers trained on source code. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1770–1778. GECCO '22, ACM, New York, NY, USA (2022)
12. Shahrzad, H., Hodjat, B., Miikkulainen, R.: Evolving explainable rule sets. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1779–1784. GECCO '22, ACM, New York, NY, USA (2022)
13. Singh, M., Brownlee, A.E.I., Cairns, D.: Towards explainable metaheuristic: Mining surrogate fitness models for importance of variables. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1785–1793. GECCO '22, ACM, New York, NY, USA (2022)
14. Tran, B., Sudusinghe, C., Nguyen, S., Alahakoon, D.: Building interpretable predictive models with context-aware evolutionary learning. Applied Soft Computing **132**, 109854 (2023)
15. Virgolin, M., De Lorenzo, A., Randone, F., Medvet, E., Wahde, M.: Model learning with personalized interpretability estimation (ML-PIE) (2021), arXiv:2104.06060
16. Walter, M.J., Walker, D.J., Craven, M.J.: An explainable visualisation of the evolutionary search process. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1794–1802. GECCO '22, ACM, New York, NY, USA (2022)