Explainable Artificial Intelligence: A Genetic Programming Approach

IEEE CEC 2025 Tutorial

Length: 2 hours

Tutorial level: Advanced

Yi Mei, Associate Professor, Victoria University of Wellington
Qi Chen, Senior Lecturer, Victoria University of Wellington
Andrew Lensen, Senior Lecturer, Victoria University of Wellington
Bing Xue, Professor, Victoria University of Wellington
Mengjie Zhang, Professor, Victoria University of Wellington

Topic Description

Although machine learning has achieved great success in many real-world applications, it is criticised as usually behaving like a black box, and it is often difficult, if not impossible, to understand how the machine learning system makes the decision/prediction. This could lead to serious consequences, such as the accidents of the Tesla automatic driving cars, and biases of the automatic bank loan approval systems.

To address this issue, Explainable AI (XAI) is becoming a very hot research topic in the AI field due to its urgent needs in various domains such as finance, security, medical, gaming, legislation, etc. There have been an increasing number of studies on XAI in recent years, which improves the current machine learning systems from different aspects.

In evolutionary computation, Genetic Programming (GP) has been successfully used in various machine learning tasks including classification, symbolic regression, clustering, feature construction, and automatic heuristic design. As a symbolic-based evolutionary learning approach, GP has an inherent great potential to contribute to XAI, as a GP model tends to be interpretable. However, the interpretability in GP is not as straightforward as one expect it to be, and the models evolved by GP can still be huge and complex, thus less interpretable.

This tutorial will give a brief introduction on how one may achieve better model interpretability in XAI using GP. To this end, we will first briefly introduce XAI and GP. Then we will introduce the GP techniques/strategies that could lead to better model interpretability. We follow the common taxonomy of XAI, and divide the techniques into intrinsic and post-hoc methods. In addition, we will also review some visualisation methods by/for GP to improve interpretability. The tutorial is concluded with some discussions on the current trend, challenges, and potential future research directions.

Outline

This **2-hour** tutorial will be composed of the following parts:

1. Brief introduction to XAI [25 mins]

2. Brief introduction to GP [20 mins]
3. How to approach better model interpretability through GP [60 mins]
4. Challenges and Future Directions [10 mins]

Potential Audience

We expect 50 participants to this tutorial.

This tutorial will be interested to all researchers who are interested in XAI, GP, and more general evolutionary machine learning and optimisation, as well as practitioners who want to solve their real-world problems with more interpretable solutions.

For GP researchers, this tutorial provides a summary of how to improve the explainability and interpretability of the evolved GP models. For general AI and machine learning researchers, this tutorial demonstrates the potential to achieve better XAI by using a GP approach. For practitioners with their real-world prediction and decision-making tasks, this paper can provide interpretable solutions (i.e., models obtained by GP-based approaches) for solving their tasks.

Activities

We plan to have some interactive activities to show some GP-evolved models and ask the audience to judge their interpretability, and show some examples of the models evolved by GP with interpretability strategies.

Other Information

Older similar tutorials were held in IEEE CEC 2021, 2022 and 2024. This proposed tutorial will have better structure, richer content, and more recently updated literature than the previous ones.

The tutorial has been very successful when being held in the previous conferences. Although held online in 2021 and 2022, we had 50+ online attendees, which was one of the largest tutorials in the conference. XAI is a trendy topic and AI and machine learning, and we believe this tutorial can attract more EC, especially GP researchers to contribute to this interesting new area.

This tutorial is largely based on our survey paper:

Yi Mei, Qi Chen, Andrew Lensen, Bing Xue, and Mengjie Zhang. Explainable artificial intelligence by genetic programming: A survey. *IEEE Transactions on Evolutionary Computation*, 27(3):621-641, 2023.

Presenters

Yi Mei (Main presenter): Yi.mei@ecs.vuw.ac.nz
Victoria University of Wellington

Yi Mei is an Associate Professor/Reader at the School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand. His research interests include evolutionary computation and machine learning for combinatorial optimisation, hyper-heuristics, genetic programming, automatic algorithm design, and explainable AI. Yi has more than 250 fully refereed publications, including the top journals in EC and Operations Research (OR) such as IEEE TEVC, IEEE Transactions on Cybernetics, European Journal of Operational Research, ACM Transactions on Mathematical Software, and top EC conferences (GECCO). He won an IEEE Transactions on Evolutionary Computation Outstanding Paper Award 2017, GECCO Best Paper Awards in 2022, 2023 and 2024, the EuroGP Best Paper Award 2022, and a GECCO Humies Silver Award. He is the Chair of IEEE CIS Travel Grant subcommittee, Chair of IEEE Taskforce on Evolutionary Scheduling and Combinatorial Optimisation, and Chair of IEEE New Zealand Central Section. He is an Associate Editor/Editorial Board Member of 7 international journals, including the IEEE Transactions on Evolutionary Computation, IEEE Transactions on Artificial Intelligence, and Journal of Scheduling. He is a Fellow of Engineering New Zealand and an IEEE Senior Member.

Qi Chen: qi.chen@ecs.vuw.ac.nz
Victoria University of Wellington

Qi Chen is currently a Senior Lecturer in Artificial Intelligence in the Centre for Data Science and Artificial Intelligence and the School of Engineering and Computer Science at Victoria University of Wellington. Her research focuses on data mining, machine learning, evolutionary computation, symbolic regression, feature manipulation. She has over 70 papers published in fully referred international journals and conferences and most of them are on symbolic modeling. Dr Chen has been serving as a program committee member of over ten international conferences including AAAI, IEEE CEC, IEEE SSCI, Australian AI and SEAL. She is serving as a reviewer of over ten international journals including IEEE Transactions on Cybernetics and IEEE Transactions on Evolutionary Computation.

Bing Xue: bing.xue@ecs.vuw.ac.nz
Victoria University of Wellington

Bing Xue is a Fellow of IEEE and Fellow of Engineering New Zealand. She is currently Professor of Artificial Intelligence, Deputy Head of School for Engineering and Computer Science, Deputy Director of Centre for Data Science and Artificial Intelligence, at Victoria University of Wellington (VUW). Her research focuses mainly on machine learning and evolutionary computation, such as evolutionary deep learning, feature selection, and image analysis, and their real-world applications in aquaculture, marine science, biology, healthcare, forest, and others. She has over 400 fully refereed publications and leading several prestigious research grants. She has been organising many international conferences, such as General Chair of PRICAI 2025, IVCNZ 2025, and EuroGP 2025,

Conference Chair of IEEE CEC 2024 and EuroGP 2024. She has also served as an Associate Editor of several international journals, such as IEEE TEVC, IEEE TAI, IEEE CIM, and ACM TELO.

Mengjie Zhang: mengjie.zhang@ecs.vuw.ac.nz
Victoria University of Wellington

Mengjie Zhang is a Fellow of Royal Society of New Zealand, a Fellow of Engineering New Zealand, a Fellow of IEEE, an IEEE Distinguished Lecturer, currently Professor of Computer Science at Victoria University of Wellington, where he heads the interdisciplinary Evolutionary Computation and Machine Learning Research Group. He is the Director of the Centre for Data Science and Artificial Intelligence at the University. His research is mainly focused on AI, machine learning and big data, particularly in evolutionary learning and optimisation, feature selection/construction and big dimensionality reduction, computer vision and image analysis, scheduling and combinatorial optimisation, and evolutionary deep learning and transfer learning. Prof Zhang has published over 900 research papers in refereed international journals and conferences. He received the "EvoStar/SPECIES Award for Outstanding Contribution to Evolutionary Computation in Europe" in 2023. Since 2007, he has been listed as a top five (currently No. 3) world genetic programming researchers by the GP bibliography (http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/index.html). He is also a Clarivate Highly Cited Researcher in the field of Computer Science -- 2023. Prof Zhang is currently the Chair for IEEE CIS Awards Committee. He is also a past Chair of the IEEE CIS Intelligent Systems Applications Technical Committee, the Emergent Technologies Technical Committee and the Evolutionary Computation Technical Committee, a past Chair for IEEE CIS PubsCom Strategic Planning subcommittee, and the founding chair of the IEEE Computational Intelligence Chapter in New Zealand.