Enhancing empirical modelling in environmental science with knowledge discovery and genetic programming

Mohammad Sadegh Khorshidi [a], **Mostafa Gandomi** [b], **Mohammad Reza Nikoo** [c], **Danial Yazdani** [a], **Fang Chen** [a] and **Amir H. Gandomi** [a]

[a] *Faculty of Engineering and IT, University of Technology Sydney, Australia*
[b] *Faculty of Civil Engineering, University of Tehran, Iran*
[c] *Department of Civil and Architectural Engineering, Sultan Qaboos University, Muscat, Oman*
Email: gandomi@uts.edu.au

Abstract: Genetic programming (GP) has shown great promise in empirical modelling for environmental science, particularly in complex systems such as climate, flood, and environmental modelling. However, the success of GP largely depends on the quality and quantity of data used for training. In this regard, knowledge discovery (KD) can significantly improve GP's ability to model complex interactions (Grin and Gandomi 2021). KD is the process of discovering new knowledge or insights from existing data, often through data mining and machine learning techniques. KD can be used in conjunction with GP to identify relevant variables, patterns, and interactions within a dataset, which can then be used to improve the accuracy and generalization of GP models. By discovering new knowledge, KD can also help GP to avoid overfitting and capture more complex relationships between variables.

One area where GP with KD can be particularly useful is climate modelling, which is a complex and challenging problem with many variables and interactions that must be taken into account. GP can be used to optimize a population of models based on inputs such as atmospheric composition, ocean currents, and land surface processes. By evolving the structure of the models over time, GP can capture complex interactions between variables and generate accurate predictions of climate phenomena. Using KD with GP can help identify new relevant data and knowledge, which can further improve the accuracy and generalization of these models (Khorshidi et al 2023). Another area where GP with KD can be applied is flood modelling, which involves the prediction of flood events based on a range of factors such as precipitation, topography, and land use. GP can be used to model the complex interactions between these factors and predict flood events more accurately. The use of KD with GP can help to identify new relevant data and knowledge, such as the impact of changing land use or the interactions between different precipitation patterns, which can further improve the accuracy and generalization of these models. Similarly, GP with KD can be used in environmental modelling to model complex interactions between various environmental factors, such as air quality, soil erosion, and water pollution. The use of KD with GP can help to identify new relevant data and knowledge, such as the impact of changing environmental policies or the interactions between different pollutants, which can further improve the accuracy and generalization of these models.

The combination of KD and GP represents a powerful approach to empirical modelling in environmental science, particularly in complex systems. By using KD to identify new knowledge and insights about the system, GP can be used to model these interactions more accurately and effectively, improving our understanding of complex phenomena and helping to make more informed decisions about environmental management and policy. The use of KD with GP has the potential to significantly improve the accuracy and generalization of models, leading to better predictions and insights about complex environmental systems.

REFERENCES

Grin, A.V. Gandomi, A.H., 2021. Advancing Genetic Programming via Information Theory. In 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 1991-1998). IEEE.

Khorshidi, M.S., Yazdani, D., Mańdziuk, J., Nikoo, M. R., Gandomi, A.H., 2023. A Filter-Based Feature Selection and Ranking Approach to Enhance Genetic Programming for High-Dimensional Data Analysis. In 2023 IEEE Congress on Evolutionary Computation (CEC). IEEE.

Keywords: *Environmental modelling, genetic programming, variable interactions, knowledge discovery*