Genetic Programming and Evolvable Machines (2025) 26:15
https://doi.org/10.1007/s10710-025-09513-w

BOOK REVIEW

A review of "Symbolic Regression" by Gabriel Kronberger, Bogdan Burlacu, Michael Kommenda, Stephan M. Winkler, and Michael Affenzeller, ISBN 978-1-138-05481-3, 2024, CRC Press.

William G La Cava[1,2]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

From its humble beginnings as a demonstration of genetic programming (GP), symbolic regression (SR) has become a field of study in its own right. Part of its appeal is that provides a unique approach to interpretable machine learning, filling a sizeable gap between overly simple, white-box modeling approaches and overly complex, black-box ones.

Whereas the primary gateway to learning about SR was once an introduction to genetic algorithms and GP, it is often now sought by practitioners to develop interpretable models in different fields, from materials science to physics to medicine. This appeal, in turn, has brought SR into the spotlight of a broader machine learning research community who has never heard of GP, and is used to reaching to other methodologies (e.g., deep learning) for problem solving. In other words, SR has become for many a gateway to GP, rather than the reverse.

In this light, "Symbolic Regression", by authors Kronberger, Burulacu, Kommenda, Winkler, and Affenzeller, is a timely and important contribution to the field, which before now lacked an introductory work that centered the science of SR. The authors are affiliated with the Heuristic and Evolutionary Algorithms Laboratory (HEAL), University of Applied Sciences Upper Austria, which has a long and prolific history of both applied and basic SR research.

Given the background of the authors, it is natural that the textbook itself is written primarily for scientists and engineers who are looking to apply SR to their problems to gain insight into the underlying processes at play. The intended audience sets this book apart from other introductory works, which tend to focus on a computer science audience and develop the algorithms comprising GP from the ground up, treating SR as a convenient example of how it may be used.

✉ William G La Cava
william.lacava@childrens.harvard.edu

1 Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA

2 Department of Pediatrics, Harvard Medical School, Boston, MA, USA

Published online: 29 March 2025

As an engineering graduate student in the 2010s interested in modeling dynamical systems with SR, my introduction to the field was a piecemeal assortment of texts, from Koza's famous first book [1], Poli et al.'s interpretable GP field-guide [2], and Sean Luke's excellent treatment of algorithms [3], to recent conference and journal papers. Applying *the right* SR methods to real-world systems *in the right way* was largely a process of trial and error, as it often tends to be. Thankfully, this book brings together the topics most central to applying SR, with a focus on the challenges and pitfalls that often arise. It will undoubtedly be a valuable resource to SR practitioners and students who want to do robust applied work.

The first three chapters motivate SR, provide the basics of supervised learning, and then define SR in finer detail. An introductory reader may benefit from reading these in order 2, 1, 3, since chapter 2 defines some concepts mentioned in chapter 1. Chapter 4 covers evolutionary computation and GP as one approach to SR, drawing on the authors' collective years of expertise. Chapter 5 covers model inspection and validation, and is perhaps the most interesting as it demonstrates how to use statistical modeling approaches to reason about SR-generated models, covering topics like model selection criteria and error decomposition. Chapter 6 covers advanced topics, including non-GP approaches to SR and constant optimization, and Chapter 7 provides a number of valuable and open-source examples of SR's application to real-world problems. The end of each chapter points the reader to appropriate references to dig into related concepts that have been touched upon.

Overall, I would recommend "Symbolic Regression" to any data scientist interested in using SR in real-world applications. Although the authors purposefully avoid specific software examples to attempt to keep the coverage as general as possible, I do think a next edition could benefit from inclusion of open-source code examples to allow readers to reproduce and engage with the text. Regardless, this textbook is likely to make the use of SR algorithms more data-driven and robust as the field continues to expand into new domains.

References

1. J.R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection* (MIT Press, Cambridge, MA, USA, 1992)
2. R. Poli, W.B. Langdon, N.F. McPhee, A Field Guide to Genetic Programming. Published via http://lulu.com and freely available at http://www.gp-field-guide.org.uk, (2008). (With contributions by J. R. Koza)
3. S. Luke, Essentials of Metaheuristics, 2nd edn. Lulu, (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.