From Black Box to Symbolic Insight:
Interpretable Machine Learning With T-Spline Networks

Marco Antonio Marquez-Vera
http://orcid.org/0000-0002-2969-9084
Polytechnic University of Pachuca, Mexico

Alfian Ma'arif
Universitas Ahmad Dahlan, Indonesia

Blanca Diana Balderrama-Hernández
Secretariat of Public Education, Mexico

ABSTRACT

This article explores the use of Kolmogorov-Arnold Networks as interpretable machine learning models instead of the universal approach theorem used in machine and deep learning. Emphasis is placed on architectures based on B-splines, T-splines, and FastKAN using RBFs, which allow for transparent function approximation. The article discusses how symbolic representations emerge from trained models, the role of node pruning in simplifying structure, and the potential of these techniques to uncover latent physical models or aid in scientific modeling where interpretability is essential. Also, by pruning the neural model, it is possible to simplify the interpretable model.

INTRODUCTION

Nowadays, the use of machines and deep learning makes possible the existence of artificial intelligence (AI) capable of interacting with humans. However, these kinds of systems are built with complex algorithms that cannot be interpretable, and no idea can be inferred to determine how to propose a better structure or parameters in a deterministic way (black box), so it is necessary to use data for tuning the

DOI: 10.4018/407567

Copyright ©2027, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishingis prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

algorithm. Recently, it is looked to have interpretability in the algorithms of AI to understand how the system works and to have an idea of what will happen if certain input is given to the algorithm (gray-box).

In this paper, the Kolmogorov-Arnold networks will be shown, and some concepts and applications will be also mentioned. To show definitions and concepts for this encyclopedia, they will be redacted in italics to understand the flow of information avoiding confusion on reading this chapter.

Returning to the topic, when a system or process is described using differential or difference equations, a simulation can represent the real behavior of the system under certain circumstances, these kind of models are called white-box models, and they involucrate to known all the parameters. However, not always is possible to know all the parameters values or they can be time-varying. Also, the parameters values are often difficult to obtain due physical constraints or lack of sensors (Obadina et al., 2022). When the model structure is known or it is necessary to work with time-varying parameters, it is possible to make parameter estimation and even to apply adaptive control if the system must to be controlled (Gibson and McAuley, 2025). On the other hand, black-box models are not interpretable, and they are tuned by using data to train the model in order to emulate a physical system, these models do not reveal the structure of the system, and it is very difficult, and even impossible, to know what will be happen in the black-box model behavior if a parameter is modified in the model. An example of black-box model is machine learning, for simple applications, where a perceptron can be computed for pattern recognition or to mimic certain system, the artificial neural network can be defined mathematically, but to implement a multilayer perceptron (machine learning) is very difficult even to propose the number of neurons in the hidden layer, most of the times, the artificial neural network (ANN) is evaluated using different activation functions and different number of neurons to determine the best configuration (Nelles and Isermann, 1996).

One alternative to ANN that gives black-box models, even when the accuracy of the model is adequate to the application developed, is the use of fuzzy logic. When it is not possible to have expert knowledge to propose a fuzzy model or controller, a Takagi-Sugeno inference system can be obtained using data, techniques like least squares (batch or recursive) and adaptive fuzzy systems can be used to obtain a model that describes or controls a system (Kahl and Kroll, 2020). Fuzzy rules and fuzzy sets obtained can be used to interpret how variables are related each other. For example Obadina et al. (2022) combined priori knowledge (white-box) of the system, and approach the unknown parameters values by using data (black-box) to obtain a gray-box model in order to control a robot manipulator.

Expert knowledge can be used to compute a Mamdani fuzzy model, this kind of inference system is proposed by determining the number, and kind of fuzzy sets to make a fuzzy partition of variables involved in the model. The consequents of fuzzy rules can be proposed also as fuzzy sets. In this way, it is possible to interpret the fuzzy rules, and to have an idea of the expert knowledge used to build the fuzzy model. Furthermore, Navarro-Almanza et al. (2022) made an interpretable neuro-fuzzy model by using a Mamdani inference system. A neuro-fuzzy system implements both techniques ANN and fuzzy logic to use the best characteristics of each one. Furthermore, it is possible to regard a neuro-fuzzy system like an ANN whose activation functions are described by fuzzy sets, or like a fuzzy system tuned like an ANN. The idea of Navarro-Almanza et al. (2022) was to extract a high-quality rule in terms of comprehensibility, accuracy and fidelity. Anyway, a neuro-fuzzy system can be obtained using radial basis functions (RBFs) in an adaptive fuzzy system, because Gaussian functions are derivable and interpreted as activation functions in an ANN (Kassem and Çamur, 2017).

From decades, AI was focused in using black-box models, principally based in ANN and deep learning. In 2023, a project based in using the Kolmogorov-Arnold representation theorem was presented in the

28 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/from-black-box-to-symbolic-insight/407567

Related Content

VAQoS: Architecture for End-to-End QoS Management of Value Added Web Services
M. A. Serhani, R. Dssouli, H. Sahraoui, A. Benharrefand M. E. Badidi (2006). *International Journal of Intelligent Information Technologies (pp. 37-56).*
www.irma-international.org/article/vaqos-architecture-end-end-qos/2409

A New Hybrid Model of Deep Learning ResNeXt-SVM for Weed Detection: Case Study
Brahim Jabirand Noureddine Falih (2022). *International Journal of Intelligent Information Technologies (pp. 1-18).*
www.irma-international.org/article/a-new-hybrid-model-of-deep-learning-resnext-svm-for-weed-detection/296269

Generative AI for Text to Image: A Comprehensive Survey
Shrishti Shah, Shubhasri Tadepalli, Lalitha Tanmai Vaddiparthi, Nishat Afshan Ansariand Ankit A. Bhurane (2024). *Making Art With Generative AI Tools (pp. 17-44).*
www.irma-international.org/chapter/generative-ai-for-text-to-image/343417

Trade Unions in AI-Era Workplace Harassment: A Criminological and Victimological Approach
Laura Gómez García (2025). *The New Role of Labor Unions in the AI Era (pp. 239-270).*
www.irma-international.org/chapter/trade-unions-in-ai-era-workplace-harassment/381913

A Study of Replicators and Hypercycles by Hofstadter's Typogenetics
V. Kvasnikaand J. Pospíchal (2014). *International Journal of Signs and Semiotic Systems (pp. 10-26).*
www.irma-international.org/article/a-study-of-replicators-and-hypercycles-by-hofstadters-typogenetics/104640