Symbolic regression in materials science and engineering

Evgeniya Kabliman[1,2,3]
[1]University of Bremen, Faculty 04: Production Engineering – Mechanical Engineering and Process Engineering, Badgasteinerstraße 1, 28359 Bremen, Germany
[2]Leibniz Institute for Materials Engineering – IWT, Badgasteinerstraße 3, 28359 Bremen, Germany
[3]Technical University of Munich, TUM School of Engineering and Design, Materials Department, Chair of Materials Engineering for Additive Manufacturing, Freisinger Landstrasse 52, 85748 Garching, Germany

kabliman@iwt-bremen.de

In material science and engineering, it is essential to understand how material properties can be optimized through modification of material structure and variation of processing parameters. Often, this information is hidden, and one needs to derive the models to describe the process-structure-property relationships. The increasing use of artificial intelligence, particularly machine learning, can help uncover these hidden correlations. Herein, symbolic regression (SR) has demonstrated itself as a powerful technique to derive interpretable models. In contrast to commonly used “black-box” models such as ANN [1], SR generates open-to-inspect models in the form of mathematical expressions that allow the straightforward analysis of the correlation between the input features (such as process parameters and structural information) and the target variables (material properties).

The symbolic regression can be solved in terms of genetic programming (GP), a type of evolutionary algorithm mimicking natural evolution. GP operates with a population of programs that are iteratively improved by mimicking processes observed in natural evolution, such as survival of the fittest, recombination, and mutation [2]. Thus, the programs that fit better to an objective function (also called fitness function in GP) are selected with a higher probability for recombination, while the rest have a low probability of being selected.

The present work will review several examples of the application of symbolic regression for optimizing the material properties, in particular during the forming operations of metallic alloys [3-6]. To analyze the material response to applied deformation, the force-displacement curves (or stress-strain curves) are commonly used. The corresponding data can be obtained using various testing methods such as tensile, compression, or torsion tests. Using the testing data, the models can be derived and used for extrapolation beyond the measured values which is of particular use in numerical simulations. Having the models in the form of mathematical expressions allows direct implementation of them into numerical solvers.

Figure 1 demonstrates the comparison of measured force-displacement curves and the modeling results using several approaches [5]:

- Data-driven modeling using the symbolic regression
- Physics-based modeling based on internal state variables such as mean dislocation density
- Hybrid data-driven and physics-based modeling where the physics-based model is enhanced by using the symbolic regression

The modeling is performed for two different sets of processing parameters. While one set is used for the direct calculation of the force-displacement curves, the extrapolation was applied to the other one. The results demonstrate a deviation between the applied modeling approaches pointing to the strengths and shortcomings when using symbolic regression in material science and engineering.

References

[1] Bhadeshia H.K.D.H., ISIJ Int., 39 (10) (1999) 966-979
[2] Koza J.R, Keane M.R., Rice J.P., IEEE International Conference on Neural Networks, 1 (1993) 191-198
[3] Versino D., Tonda A., Bronkhorst C.A., Comput. Methods Appl. Mech. Engrg., 318 (2017) 981-1004.
[4] Kabliman, E., Kolody, A.H., Kommenda, M., Kronberger, G., AIP Conf. Proc. 2113 (2019) 180009
[5] Kabliman, E., Kolody, A.H., Kronsteiner, J., Kommenda, M., Kronberger, G., Appl. Eng. Sci. 6 (2021) 100052
[6] Kronberger, G., Kabliman, E., Kronsteiner, J., Kommenda, M., Appl. Eng. Sci. 9 (2022) 100080

Figures

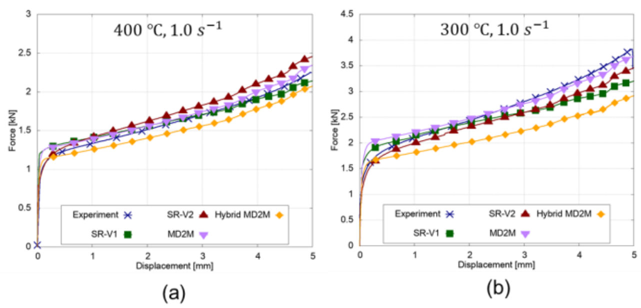

Figure 1. Comparison of the measured force-displacement curves and calculated data using the symbolic regression (SR-V1 and SR-V2), the physics-based model (MD^2M), and the hybrid data-driven physics-based model: (a) at 400 °C and strain rate 1.0 s^{-1}; (b) extrapolation to 300 °C (Reproduced from [5])