Science Arts & Métiers (SAM)
is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/18463

To cite this version :

Francisco Javier MONTÁNS, Rafael GÓMEZ-BOMBARELLI, Jose Nathan KUTZ, Francisco CHINESTA SORIA - Complex Algorithms for Data-Driven Model Learning in Science and Engineering - Complexity p.1-3 - 2019

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

Complex Algorithms for Data-Driven Model Learning in Science and Engineering

Francisco J. Montáns,[1] **Francisco Chinesta**,[2] **Rafael Gómez-Bombarelli**,[3] **and J. Nathan Kutz**[4]

[1] *Escuela Técnica Superior de Ingenieros Aeronáuticos, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros 3, E-28045 Madrid, Spain*
[2] *ESI Chair, École Nationale Supérieure d'Arts et Métiers ParisTech, 151 bvd. De L'Hôpital, F-75013 Paris, France*
[3] *Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*
[4] *Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA*

Correspondence should be addressed to Francisco J. Montáns; fco.montans@upm.es

In their first centuries, scientific and engineering developments were dominated by empirical understanding which encapsulated the first paradigm of scientific discovery. After the Renaissance, the scientific revolution and the development of calculus led to a new scientific viewpoint whereby physical principles, laws of nature, and engineering models were established by proposing new theoretical constructs that could be verified through specific experiments. This was the second paradigm of scientific discovery. More recently, the computational era, or the third paradigm of discovery, has allowed us to solve complex and nonlinear scientific and engineering problems that were beyond our analytically tractable methodologies. Today, there is a new fourth paradigm of discovery, which is a data-driven science and engineering framework whereby complex models and physical laws are directly inferred from data.

Therefore, there is increasing change in the objective of computational algorithms used in simulations. Until now, the purpose was to accurately discretize systems of linear and nonlinear continuum equations derived from physical laws, models, and principles frequently established prior to the computational era; these equations were inferred from observation on limited experimental data and significantly simplified to make them analytically tractable. Today, the available experimental data and the complexity of the equations are no longer a major limitation to the point that we may compute physical processes without resorting to analytical laws, principles, or models; we just need to predict the correct output from the system for a given input even when there is not a well-defined model. However, for this endeavor, we need new computational algorithms capable of learning the complex behavior of the system and of establishing those governing equations of the system directly from experimental data, with the flexibility of not having to rely on analytical equations. An example is the determination of the nonlinear behavior of solids and fluids under general conditions directly from measured data, without specifying the form of the constitutive relations. Many fields already have started to capitalize on such methods, developing algorithms for fuzzy relations, leading to data-driven decision-making by constructing purely computational predictive analytics in such complex fields as economics, consumer behavior and dynamics, security, and even web utilities. The engineering sciences are now poised to also take advantage of data-driven methods in obtaining physical principles and models which yield reliable laws and accurate predictions, using fewer hypotheses and fewer analytical relations and balancing the parametrization of physical models with the amount of available measurements.

The purpose of the special issue on data-driven model learning in science and engineering is to bring representative novel state-of-the-art contributions in this line. Many papers

were submitted from leading authors in data-driven procedures. After a rigorous review process, among them, only ten outstanding contributions were selected for the special issue. These are representative of different algorithmic approaches and application fields.

Clustering is an important aspect of unsupervised data-driven procedures. In these techniques, data groups are identified and tagged. T. Du et al., in their paper entitled "A Data-Driven Parameter Adaptive Clustering Algorithm Based on Density Peak," develop such an algorithm named DDPA-DP to avoid the influence of artificial parameters, improving the congruence and flexibility of clustering. The authors compare their proposal to other existing algorithms, using both synthetic data and real-world data obtained from the thermal power industry. The authors conclude that their proposal is advantageous in terms of accuracy of clustering and time complexity.

Reduced order modeling will be extremely important in biomechanics, for example, in obtaining workable constitutive relations of complex phenomena for simulations of tissue and cell behavior. In this line, B. Adam and S. Mitran present in "Data-Driven Finite Element Models of Passive Filamentary Networks" a procedure to substantially reduce the cost of finite element simulations of actin filaments in the cytoskeleton. They employ at the smaller scale Brownian dynamics with tens of millions of state variables. Then, applying a singular value decomposition, they reduce the problem to obtain a representative numerical constitutive equation at a much smaller scale. They show through an example that the reduced model captures the key features present in the behavior of the full model.

Data-driven algorithms have been used for long in the field of computer vision and are now being increasingly used in mechanics and material science. F. Nguyen et al., in their paper "Computer Vision with Error Estimation for Reduced Order Modeling of Macroscopic Mechanical Tests," use clustering and convolutional neural networks trained by supervised machine learning on digital images of mechanical tests on a specimen, under a variety of loading conditions. With their procedure, they identify a reduced order model. This reduced order model is fast to be employed in manufacturing procedures to make part-specific decisions.

Data-driven models bring a new ingredient in the overall modeling of complex systems. Data-driven models may be combined with classical modeling approaches and ensemble-based modeling. Therefore, new meta-modeling methods are also needed to manage, in a systematic and automatic manner, the combined modeling and application procedures themselves. S. V. Kovalchuk et al., in their paper "A Conceptual Approach to Complex Model Management with Generalized Modelling Patterns and Evolutionary Identification," present such an algorithm for mixed model management, considering the spaces of functions, parameters, and modeling approaches relating to them. They show in their paper interesting applications in metocean simulations, health care processes, and mining in social media.

One of the main aspects in the design of tires is the friction behavior, which largely affects its performance, temperature, and longevity. The friction laws are complex functions of different parameters like sliding velocity, contact pressure, and temperature. The usual approach is to establish an analytical law with fitted material parameters (e.g., the Huemer friction law). A more flexible and accurate data-driven approach is presented by A. Serafińska et al. in their paper entitled "Artificial Neural Networks Based Friction Law for Elastomeric Materials Applied in Finite Element Sliding Contact Simulations." In their work, the authors use artificial neural networks to obtain a regularized numerical nonlinear data-driven thermomechanical friction law, function of the mentioned variables, and the temperature. They obtain an excellent fitting to experimental results. They include their law in a finite element contact formulation and perform simulations in tires for different acceleration/brake conditions.

In the aircraft industry, the timely detection of internal damage in composite structures is a difficult and complex procedure. The purpose of structural health monitoring (SHM) is to be able to perform such detections and damage classification. In this field, data-driven approaches are especially useful and very promising. The paper of Tibaduiza et al. entitled "A Damage Classification Approach for Structural Health Monitoring Using Machine Learning" presents a data-driven methodology using data collected from piezoelectric sensors under different structural states and guided waves to identify damage type and location in CFRP (carbon fibre-reinforced polymer) sandwich structures and plates. Their examples include different types of damage as delamination and cracking of the skin. Their procedure consists in a hierarchical nonlinear principal component analyses with machine learning.

S. Pan and K. Duraisamy, in their paper "Long-Time Predictive Modeling of Nonlinear Dynamical Systems Using Neural Networks," develop data-driven models for nonlinear dynamical systems using feedforward neural networks with a Jacobian regularization for the loss function. The purpose of the Jacobian regularization is, for example, to improve the robustness of the model for limited data and to improve the predictions when the model is unstable. They compared their approach to a sparse identification of nonlinear dynamical systems approach with a background function library.

Also in the field of nonlinear dynamic analysis of systems, in the paper "Analyzing Nonlinear Dynamics via Data-Driven Dynamic Mode Decomposition-Like Methods," S. Le Clainche and J. M. Vega review and analyze two different approaches useful for data-driven analyses. These approaches are the high-order dynamic mode decomposition (HODMD) method, based on the classical dynamic mode decomposition, and the spatiotemporal Koopman decomposition (STKD) by linear expansion. In a nutshell, the former considers different snapshots (steps) to account for the evolutionary, nonlinear response of the system through different updating matrices, whereas the second approach considers different spatiotemporal decomposition (possibly nonlinear time frequency and spatial wave number correlations of the components). Some applications are also explored in their paper.

Koopman representations, with more general Koopman eigenfunctions, are also used by J. N. Kutz et al. in their paper "Applied Koopman Theory for Partial Differential Equations

and Data-Driven Modeling of Spatio-Temporal Systems." The authors focus their presentation in the impact of the choice of observable variables in the quality of the approximations obtained. They perform the analysis through several carefully selected examples which highlight their conclusions; these examples are the Burgers equation, the nonlinear Schrödinger equation, the cubic-quintic Ginzburg-Landau equation, and the equations from a reaction-diffusion system. They demonstrate that a poor choice of the observable variables brings approximations worse than those obtained with the classical DMD. The authors present a step-by-step procedure for a good selection of the mentioned variables.

R. Ibáñez and his coworkers present in their paper "A Multidimensional Data-Driven Sparse Identification Technique: The Sparse Proper Generalized Decomposition" a novel PGD-based data-driven identification procedure for high dimensional problems using Kriging interpolants. This procedure uses nonstructured datasets, is robust with respect to high dimensionality, and alleviates the curse of dimensionality, this latter property obtained through the method of separation of variables employed by the authors in previous publications to obtain reduced order models. To avoid nonparsimonious predictions, due to the solution being in a highly nonlinear manifold, the authors use sliced domains, collocation points, and local PGD versions. They name the method "sparse-PGD." They compare their approach to other available solution methods using different synthetic and physical problems.

Conflicts of Interest

The editors of the special issue declare that they have no conflicts of interest regarding the publication of this special issue.

Acknowledgments

The guest editorial team would like to thank the authors of the submitted papers for their interest on the special issue and for their valuable contributions; and the anonymous reviewers for their careful work, their advice to the editors, and their suggestions to improve the manuscripts. We also thank the publishing team for their assistance and support in all the process.

Francisco J. Montáns
Francisco Chinesta
Rafael Gómez-Bombarelli
J. Nathan Kutz