Data Driven Model Learners through machine learning techniques

Clara Argerich Martin*, Rubén Ibanez Pinillo† and Francisco Chinesta††

*ENSAM ParisTech
151 Boulevard de l'Hôpital, F-75013 Paris, France.
Clara.argerich_martin@ensam.eu
†IcI Ecole Centrale de Nantes
1, rue de la Noe, 44321 Nantes Cedex 3, France.
†† ESI GROUP Chair @ ENSAM ParisTech
151 Boulevard de l'Hôpital, F-75013 Paris, France.

ABSTRACT

The accuracy of most simulations relies on the trustworthiness of the so-called constitutive equations. These relations relate constraints with kinematic quantities (i.e. elastic constitutive tensor relating stresses with strains or the thermal conductivity relating thermal fluxes with temperature gradients) and are of crucial importance in order to close the set of modelling equations. Nevertheless, if the calibration of such relations is not good enough, the simulations will automatically inherit a modelling error.

On the other hand, there are plenty of techniques based on Big Data and machine learning, whose main task is to unveil hidden relationships or correlations inside a given amount of data. Therefore, from a Big Data perspective one of the main questions arising is: do we collect enough data to build models without relying on constitutive equations? If so, how do we build a model? How many data is required? The most known Big Data techniques used to classify and obtain data behaviour are neural networks, regression trees, t-sne [1,2,3] . These algorithms are often based in linear-regressions and present a black-box nature, such for example the neural network. Also, these techniques rely literally on *Big Data,* the greater the amount of collected data, the more accurate the prediction will be. In this work we present a data-driven technique based on the separability of variables and on previous work developed in the group [5].

The aim of this work is to build an algorithm able to learn a model without relying on an ad-hoc constituve equation, to be able to do so with a reduced amount of data, the so-called, *Smart-Data* and to be able to build models on data presenting non-linear behaviours. To do so the algorithm will be presented and tested for data with different nature.

REFERENCES

[1] S. Roweis and L. Saul . "Nonlinear dimensionality reduction by locally linear embedding". *Science* v.290 no.5500, pp.2323--2326, 2000..

[2] L. Vans Der Maaten and G. Hinton . "Visualizing Data using t-SNE". *Journal of Machine arning Research* 9, 2579-2605, 2008.

[3] Y. Bengio, R. Ducharme, P. Vincent and C. Jauvin. "A Neural Probabilistic Language Model" *urnal of Machine Learning Research 3*, 1137-1155, 2003.

[4] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean. "Distributed Representation of Words and Phrases an their Compositionality". *ProceedingNIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems*, Volume 2 2013

[5] R. Ibáñez, E.Abisset-Chavanne, A. Ammar, D. González, El.Cueto, A. Huerta, JL. Duval and FChinesta "A Multidimensional Data-Driven Sparse Identification Technique: The Sparse Proper Generalized Decomposition" *Complexity, Volume 2018, Article ID 5608286, 11 pages*