A case study : Influence of Dimension Reduction on regression trees-based Algorithms - Predicting Aeronautics Loads of a Derivative Aircraft

E. FOURNIER
Paul Sabatier University - Airbus

Supervisor(s): T. KLEIN (ENAC, IMT), F. GAMBOA (IMT), S. GRIHON (Airbus)

Ph.D. expected duration: Nov. 2016 - Nov. 2019

Adress: Airbus France 316, Route de Bayonne, Toulouse France

Email: edouard.fournier@airbus.com

Abstract: In aircraft industry, market needs evolve quickly in a high competitiveness context. This requires adapting a given aircraft model in minimum time considering for example an increase of range or of the number of passengers (e.g A330 family [1]). Besides, the design of an airframe structure is a complex task and involves nowadays many simulation activities generating massive amounts of data. Such is the case of the process of loads and stress computations [9, 8] for an aircraft (that is to say the calculations of forces and mechanical strains sustained by the structure). The overall process is run to identify load cases (i.e aircraft mission and configurations: maneuvers, speed, loading, stiffness...) that are critical in terms of stress endured by the structure and the parameters which make them critical. This is in order to make decisions on some load cases to reduce the impact in terms of weight on the structure.

Depending on the configuration, the computation of loads and stress to resize the airframe is on the critical path of the aircraft definition: this is a time consuming (approximately a year for a new aircraft variant) and costly process, one of the reason being the high dimensionality and the large amount of data - typically for an overall aircraft structure, millions of load cases can be generated and for each of these load cases millions of structural responses (i.e how structural elements react under such conditions) have to be computed. As a consequence, engineers take a lot of time in managing and analyzing data.

In an effort to continuously improve methods, tools and ways-of-working, Airbus has invested a lot in digital transformation and the development of infrastructures allowing to treat data (newly or already produced). From this standpoint, the objective is to exploit reduced order models and optimization tools in the right places of the computational process and evolve them to adapt them to the process and make them more performing. These techniques as defined in [7] even if essentially used linked to internet and business intelligence can benefit to the manufacturing industry (here aeronautics) and covers a large number of fields as shown in [10].

This study has been realized during a proof of value sprint project within Airbus to demonstrate the usefulness of statistics and machine learning approaches in the Engineering field to reduce lead time in the computation of loads and preliminary sizing of an airframe. There is a real interest especially in the today strategy of Airbus focused on derivatives with tight non-recurring cost constraints and short time-to-market objectives. The scientific challenge is to adapt the today methods to large scale use-cases required for airframe design and to create adequate methods to be more efficient and avoid such massive data as well as approaches to spare computational times.

In our case study, variants concern the maximum take-off weight of a given aircraft model and we focus on data related to the wing. Data (around 28 000 observations per weight variant) we have at our disposal are the 24 aircaft parameters (features) used in the computing chain for calculating loads along the wing (outputs of dimension 29) and are coming from the weight variants 238tons, 242t, 247t and 251t: is it possible to predict loads in a extrapolation context, i.e from 238t to 242t, 247t and 251t? In a previous internal project, it has been shown that the family of regression trees works well to predict loads for different aircraft missions in an interpolation context. Thus, the following algorithms based on decision trees will be investigated: the Classification and Regression Trees [4], Bagging [2], Random Forest [3], the Gradient Boosting [6] and AdaBoost [5]. Besides, because of the number of features and because the outputs are multidimensional, dimensional reduction techniques (PCA, polynomial fitting, combined) need to be investigated. For each algorithm, we evaluate the results of extrapolation for eight different configurations of dimensional reduction techniques on the inputs as well as on the outputs. It shows that AdaBoost with Random Forest offers promising results in average in terms of accuracy and computational time to estimate loads on which a PCA is applied only on the outputs.

Airbus pursues the increasing knowledge capitalization and the development of new methods and tools for Research and Engineering through Big Data initiatives and the promising results of the sprint project, in which this case study has been achieved, are part of the root of upcoming bigger projects about Machine Learning in the load and stress process.

References

[1] Commercial Aircraft Airbus. A330 family. 2017. Available from : http://www.aircraft.airbus.com/aircraftfamilies/passengeraircraft/a330family/.

[2] Leo Breiman. Bagging predictors. *Machine Learning*, 24:123140, 1996.

[3] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[4] Leo Breiman, Jerome Friedman, R.A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.

[5] H. Drucker. Improving regressors using boosting techniques. *Proceedings of the Fourteenth International Conference on Machine Learning*, pages 107–115, 1997.

[6] J. H. Friedman. Greedy function approximation: A gradient boosting machine. 1999.

[7] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *Internation Journal of Information Management*, 35:137–144, 2015.

[8] Keith D. Hjelmstad. *Fundamentals of Structural Mechanics*. Springer US, 2005.

[9] Frederic M. Hoblit. *Gust Loads on Aircraft: Concepts and Applications*. AIAA Education Series, AIAA, 1988.

[10] J. Manyika and al. Big data: the next frontier for innovation, competition and productivity. 2011. Mc Kinsley Global Institute.

Short biography – Graduated in 2016 as a Statistical Engineer of INSA Toulouse, I have worked mainly in the transportation industry: traffic modeling and prediction on the Ile de Re Bridge and at the Michelin Technology Center. I am pursuing a PhD in the field of Big Data for the loads and stress computation process sustained by the aircraft structure. This thesis in funded by Airbus (CIFRE convention) and supported by the Doctoral School MITT (Toulouse Mathematics, IT and Telecommunication).