Identification of explanatory variables for DMU preparation process evaluation by using machine learning techniques

Florence Danglade 1, Jean-Philippe Pernot 2, Philippe Véron 3, Lionel Fine4

(1) : LSIS - UMR CNRS 7296 Arts et Métiers ParisTech, Aix-en-Provence, France
E-mail : Florence.Danglade@ensam.eu

(2) : LSIS - UMR CNRS 7296 Arts et Métiers ParisTech, Aix-en-Provence, France
E-mail : Jean-Philippe.Pernot @ensam.eu

(3) : LSIS - UMR CNRS 7296 Arts et Métiers ParisTech, Aix-en-Provence, France
E-mail : Philippe.Veron @ensam.eu

(4) : AIRBUS Group Innovations Suresnes France
E-mail : lionel.fine@airbus.com

Short Abstract: Being able to estimate a priori the impact of DMU preparation scenarios for a dedicated activity would help identifying the best scenario from the beginning. Machine learning techniques are a means to a priori evaluate a DMU preparation process without to perform it by predicting its criteria of evaluation. For that, a representative database of examples must be developed that contains the right explanative and output variables. However, the key explanative variables are not clearly identified. This paper proposes a method for the selection of the most significant explanatory variables among all the database variables. In addition to using these variables for learning, this will allow to formalize the knowledge.

Key words: Process evaluation, CAD model preparation, knowledge formalization, machine learning, explanatory variables.

1- Introduction

The development of a product includes a multitude of activities like analysis, sizing, product optimization, process simulation or prototyping. Each activity uses a specific Digital Mock-Up (DMU) of the product which haves a more or less accurate level of details. The preparation process of an original product to a representation for a dedicated activity involves a chain of operations that are obtained with different tools that need to take into account many control parameters.
Today, even if methods and tools exist, DMU preparation processes are complex tasks that are often based on expert knowledge and are not well formalized. They require a huge amount of time when considering CAD (Computer-Aided Design) models composed of several hundreds of thousands of parts.
The performance evaluation of the prepared DMU will help to know a priori the cost and quality of the preparation. Thus, being able to estimate a priori the impact of DMU adaptation scenarios on the simulation results would help identifying the best scenario right from the beginning.
Machine Learning techniques [M1] can be used to estimate a priori quality criteria of a DMU preparation process from carefully selected examples that contain explanatory (input) and output variables to predict. An overall approach for the use of machine learning for evaluation of simplification impact is given in section 2. Output variables are the preparation process quality criteria such as the costs (i.e. execution times) and the errors induced by the simplifications on the dedicated activity (like errors on analysis results). Explanatory variables are extracted from the original and 3D prepared models, and completed with data characterizing the preparation processes whose impact has to be estimated. The section 3 describes these explanatory variables.
The main challenge to be taken up, which will be addressed in this paper, is the identification of key explanatory variables that are extracted from DMU and preparation processes data.
Indeed, given a particular objective of DMU preparation the explanatory variables are different and are often not known.
Moreover, if we want to evaluate the quality of the process without having to perform it we may find ourselves with a large number of unknown variables for a new case. These intermediate variables will be estimated by learning. It is desirable that their number is limited.
So we propose to extract maximum data from DMU, preparation processes and the dedicated activity. Then the most determinant variables to characterize the variables to predict will be selected within the framework in section 4. Some experimental results are discussed in section 5.
The knowledge of the most significant variables for objective preparation will allow us to:

- reduce the number of necessary examples ;
- reduce the number of variables that are not known for a new case (data extracted from prepared DMU);
- increase the accuracy of predictions ;
- reduce the learning time ;
- formalize the knowledge.

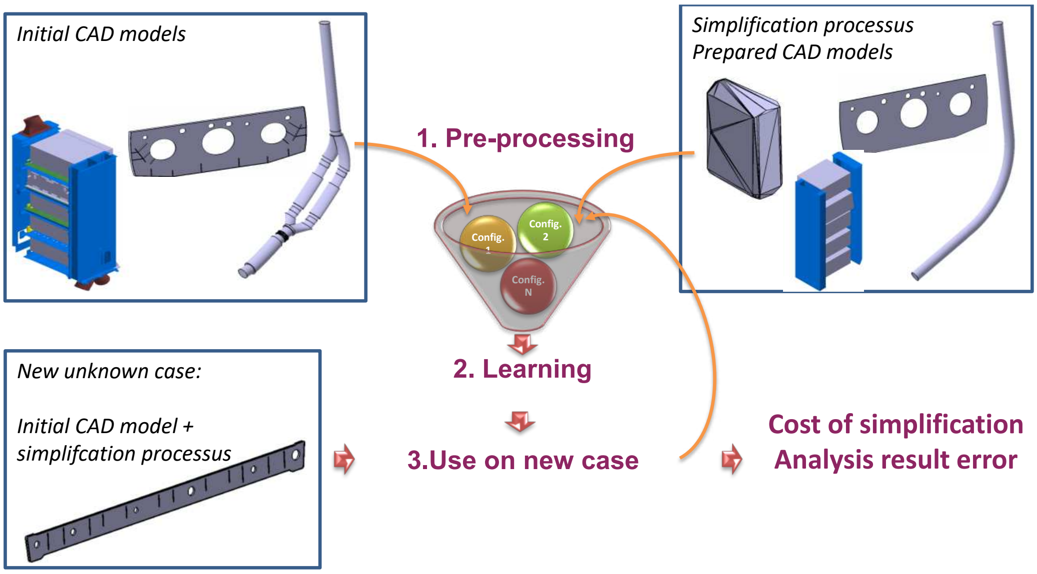

Figure 1. General approach for preparation process evaluation by using machine learning techniques.

2- Related work

2.1 – Machine learning for evaluation of simplification impact

Danglade and al. propose an approach (Figure 1) to estimate CAD model simplification impact on analysis ([DP1], [DV1]). The first step consists in building a database of examples for various component configurations. For this, key data are extracted from initial CAD model, from simplified CAD model previously computed, from analysis case and from simplification process. All these data are compiled in a matrix whose rows are component configurations and the columns are vectors of input and output variables. In the columns, vectors of variables include two main output variables (analysis result error and simplification cost) and explanative variables proposed by an expert for a specific preparation objective. Then data are adapted to machine learning technique like Neural Network [D3]. In a second step, machine learning techniques are used for carrying out classifiers for the prediction of output variables and intermediates variables. In a final step, the impact of simplification on analysis for a new case is estimated.
This paper addressed the identification of most determinant variables, in order to adapt the proposed approach to all preparation objectives and when the relevant variables are not known.

2.2 – Criteria to estimate the simplification impact on simulation result

Most determinant variables to identify are key criteria to estimate the impact on preparation result. The comparison of an original and a simplified model is a means to evaluate this impact [IJ1]. For that we can measure the similarity between models by calculating Minkowski distance, Hausdorff distance or a correlation index. Another method is to calculate differences between the original and simplified models in geometric criteria like volume, area, compactness, curvature, number of faces, number of features and so on.

3- Pre-selected explanatory variables

Explanatory variables are extracted from CAD models, preparation process description and simulation information. The explanative variables database should be as complete as possible in order to best characterize the output variables.

3.1 - Preparation process description

The preparation process of a DMU consist usually to simplify the CAD model [TB1], to adapt it to the dedicated activity and if necessary to the mesh it. The preparation process is described using vectors of parameters that specify which operators are used, theirs parameters and the used tools.

3.2 - CAD and meshed models description

The variables describing the CAD model (original and simplified), the adapted models and meshes are characterized by geometric parameters of size (area, volume, volume of the bounding box, number of parts, …) and of shape (compactness, curvatures, number of faces, number of details, number of mesh elements,…). Ways to describe these characteristics are various, they can be raw (without treatment), a mean value (calculated from values of each parts or details), a maximal value, a dimensionless on value or treated by normalization. So, CAD and meshed models are described by a great number of variables described according to different ways. For a new case, the only known variables are those that characterize the original models.

3.3 - Original and simplified models comparison

The comparator factors between original and simplified models give the level of simplification of the models. They are Hausdorff distance and benefits (Equation 1) between geometrical characteristics of original $C(M_0)$ and simplified

$C(M_i)$ CAD models (area, volume, number of faces, compactness,…).

$$Benefit = \frac{C(M_i) - C(M_0)}{C(M_0)} \quad (1)$$

3.4 - Influence factors on preparation objective

Information extracted from the dedicated activities after preparation are influence factors of the preparation on the activities. These factors quantify the geometrical changes due to simplification on the dedicated activities after preparation. They take into account the distances and positions of the simplified components relative to the boundary conditions or targets of analysis. In order to take into account of the size of different parts of a component, in addition to distance and position factors, moments have been proposed. This moment (Equation 2) is determined from the distance $BCD(C_j^n)$ between each sub-assembly C_j^n and its nearest boundary condition and the area $Area(C_j^n)$ of the sub-assembly.

$$Moment(M_i) = \sum_{1}^{N} \left(\left(BCD(C_j^n) \right)^2 . Area(C_j^n) \right) \quad (2)$$

Eventually the database contains more than 250 explanative variables. The proposed methodology will therefore have to ensure the completeness of the variables.

4- Framework to select most relevant explanatory variables

4.1 - Method for selection of explanatory variables

Not knowing the most important factors, many variables have been proposed (over 300). The initial database $\mathbf{x}_{\mathbf{base}}^{\mathbf{q}}$ contains q vectors of input variables.

The selection of variables ensures the quality of the classification and helps to formalize knowledge. The adopted method (Figure 2) is to first remove correlated variables and to select the most relevant variables from common selection algorithms.

4.2 - Correlated variables removing

After data processing (aberrant values removing, normalization, and discretization) and identifying correlated variables. The less correlated variables $x_{\overline{cor/y}}$ with the variables to predict are removed to the vectors of the base X_{base}.

4.3 - Relevant explanative variables selection

For each variable to predict y, the explanatory variables x_{base} are classified according to their influence on the variable y. Relevant explanative variables x_{exp} are selected by a stepwise backward, or forward, regression algorithm [CV1]. This consists in eliminating (if backward) or adding (if forward) one by one a relevant variable according to its rank ($Rank[x_{base}, y]$). Models are evaluated by the average quadratic error $AQE(x_{exp}^q)$ (Equation 3), where y^n is the actual variable for example n and p^n is the predicted variable. Variables are removing or adding from the initial q variables models giving a q' variables model. The operation is repeated until the q' variables model is not better than the q variables model. If the evaluation criteria have not reached an acceptable threshold and no longer changes, the completeness of the explanatory variables is called into question. It will be necessary to identify new input variables.

$$AQE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y^n - p^n)^2} \quad (3)$$

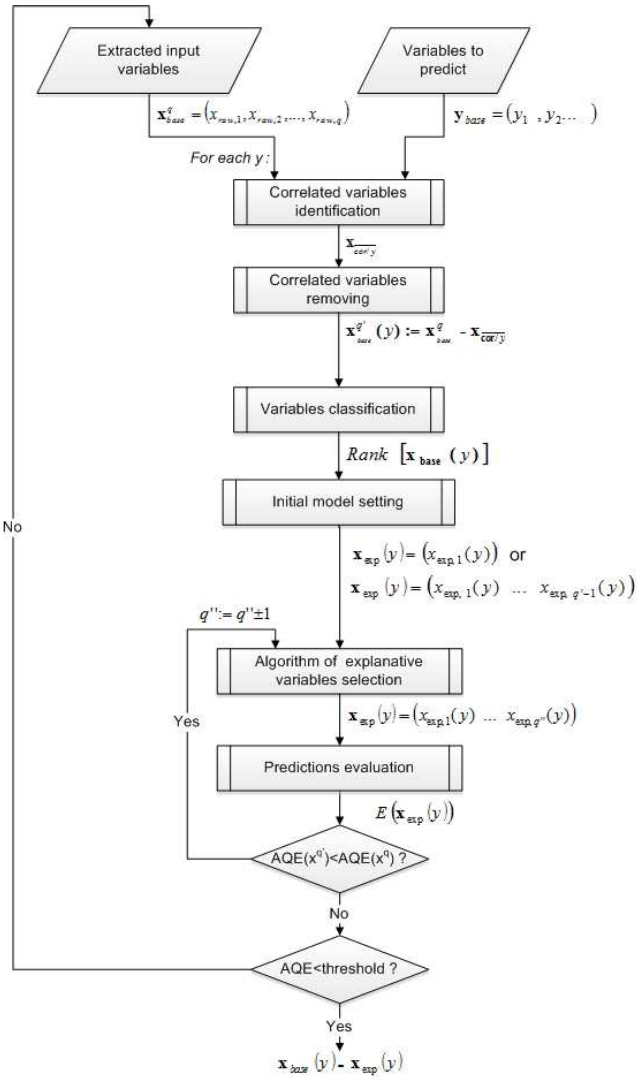

Figure 2 : method for relevant explanative variables selection.

5 – Results

The proposed approach was applied to the explanative variables identification for the estimation of the simplification impact on convective heat transfer analysis on complex CAD models. The main output variables were the analysis result error (y1), the cost of the preparation (y2) and the cost of simulation (y3).

The analysis and removing of correlated variables has reduced up to 6% AQE error on the predictions and has reduced up to 50% the learning time.
Finally, 36 known variables for a new case and 33 intermediate variables were selected for predicting the output variable y1. 34 known variables for a new case were selected for predicting the output variables y2 and y3. The rate of correct classifications on new cases are 90% for the prediction of the analysis result error (y1) and 100% for the costs of simulation and preparation (y2 and y3).

6 – Conclusion

Using machine learning techniques for a priori evaluation of the quality of a DMU preparation process requires carefully identifying the explanatory variables. The selection of variables ensures the quality of the classification and helps to formalize knowledge. A method was proposed to select variables correlated to be removed and to select variables from common selection algorithms. The completeness of the explanatory variables was validated by classification tests on new cases. It is therefore possible to identify the criteria that influence the result of a preparation of a DMU activity when they are not known. Finally, it will be possible to use machine learning techniques to evaluate a DMU preparation process.

7- References

[CV1] A. Cornuéjols, C.Vrain, and J.-D. Zucker. Panorama de l'intelligence artificielle, volume 2 Algorithmes pour l'intelligence artificielle. Cépaudès, Toulouse, 2014.

[DP1] F. Danglade, J.-P. Pernot, P. Véron. On the use of machine learning to defeature cad models for simulation. In Computer-Aided Design and Applications 11 (2014) 358–368

[DV1] F.Danglade, P. Véron , J.-P. Pernot, L. Fine. Prediction of CAD model defeaturing impact on heat transfer FEA results using machine learning techniques. Proceedings of Joint Conference on Mechanical, Design Engineering & Advanced Manufacturing, Toulouse, France, June 18th–20th, 2014

[D3] G. Dreyfus. Réseaux de neurones : méthodologie et applications. Algorithmes. Eyrolles, 2004.

[IJ1] N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani. Three-dimensional shape searching : state-of-theart review and future trends. Computer-Aided Design, 37(5) :509 – 530, 2005.

[M1] T. M. Mitchell, Machine Learning, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 1997.

[TB1] A. Thakur, A. G. Banerjee, and S. K. Gupta, "A survey of CAD model simplification techniques for physics-based simulation applications," Comput.-Aided Des., vol. 41, no. 2, pp. 65 – 80, 2009.