INTERNATIONAL DESIGN CONFERENCE - DESIGN 2016
Dubrovnik - Croatia, May 16 - 19, 2016.

A KNOWLEDGE-BASED SYSTEM FOR NUMERICAL DESIGN OF EXPERIMENTS

G. Blondet, J. Le Duigou and N. Boudaoud

Keywords: design of experiments, knowledge-based system, simulation-based design, bayesian network, artificial neural network

1. Introduction

In a product development process, Numerical Design of Experiments (NDoE) are used to improve the quality of products, by optimizing the product [Hu et al. 2008], or enhance its robustness [Patelli et al. 2012]. A NDoE is defined by an ordered sequence of simulations from a parameterized numerical model. Thus, as a numerical simulation may be very time-consuming and expensive, a NDoE may drastically accentuate the cost of study. This first issue can be solved by using specific methods. Sensitivity analysis methods [Patelli et al. 2012] are used to identify and select parameters (i.e. factors) and interactions between these parameters which have the most important influence on the studied output. Surrogate modelling methods [Castric et al. 2012] are used to replace the initial numerical model by a simpler model, able to be executed faster. Adaptive NDoE may be used to minimize the number of experiments, by defining iteratively optimal experiments [Forrester and Keane 2009].

Nevertheless, these methods increase the complexity of the NDoE process. Multiple methods, with numerous internal parameters and options, must be defined before executing the NDoE. This configuration step may require expert knowledge to define the best combination of methods which lead to an optimal NDoE process. A previous study, covering simulation data management applications for numerical simulation process enlightened a lack of support for NDoE process data. It concluded that capitalizing and sharing best practices may shorten the configuration step of this process [Blondet et al. 2015]. It aims to minimize the time spent by designers with no added-value.

The knowledge, concerning the NDoE process configuration, can be capitalized and reused by Knowledge-Based Systems (KBS). KBS are a specific branch of Artificial Intelligence [Kiritsis 1995], which aims to solve problems faster than humans, by exploiting data and knowledge more efficiently. A KBS can analyze knowledge by different reasoning strategies, as symbolic, statistical and connexionist (networks) approaches, and more recently distributed intelligence (e.g. multi-agent systems, swarm intelligence, etc.). Specific methods and systems must be used to gather, classify, trace and deliver a large amount of data and knowledge for different stakeholders in extended enterprises during long periods. While some KBS had been developed to configure a DoE process [Lorenzen et al. 1992], they did not cover numerical aspects nor all types of methods and objectives used by NDoE.

This paper presents a KBS for NDoE process, to reuse capitalized data of every previously executed process. Such a system aims to shorten the configuration step of the NDoE process by proposing improved configurations to the designer for a specific study. Section 2 presents the global architecture of the proposed KBS. This section focuses on the inference engine selection. Main inference methods are compared regarding to our need of a decision-aid system to help the designer to configure the NDoE process. Section 3 details the inference strategy, regarding to the inference engine selected in Section 2. Section 4 shows a use case to illustrate the behavior of this KBS.

2. Knowledge-based systems

A KBS can support designer's decision by analyzing capitalized data, representing the knowledge in a company, and propose new solutions or advice. This section presents the KBS structure and defines the inference engine used to propose new NDoE process configurations.

2.1 Specifications of the KBS for NDoE process

The goal is to find the best configuration of the NDoE process, regarding to its objective (surrogate modeling, optimization, sensitivity analysis, etc.), constraints (e.g. execution time, accuracy, etc.) and capitalized knowledge. As an example, a designer may not know how to choose a specific type of surrogate model. According to capitalized knowledge and data already known by the designer, the KBS system would deduce a good candidate of type of surrogate model.

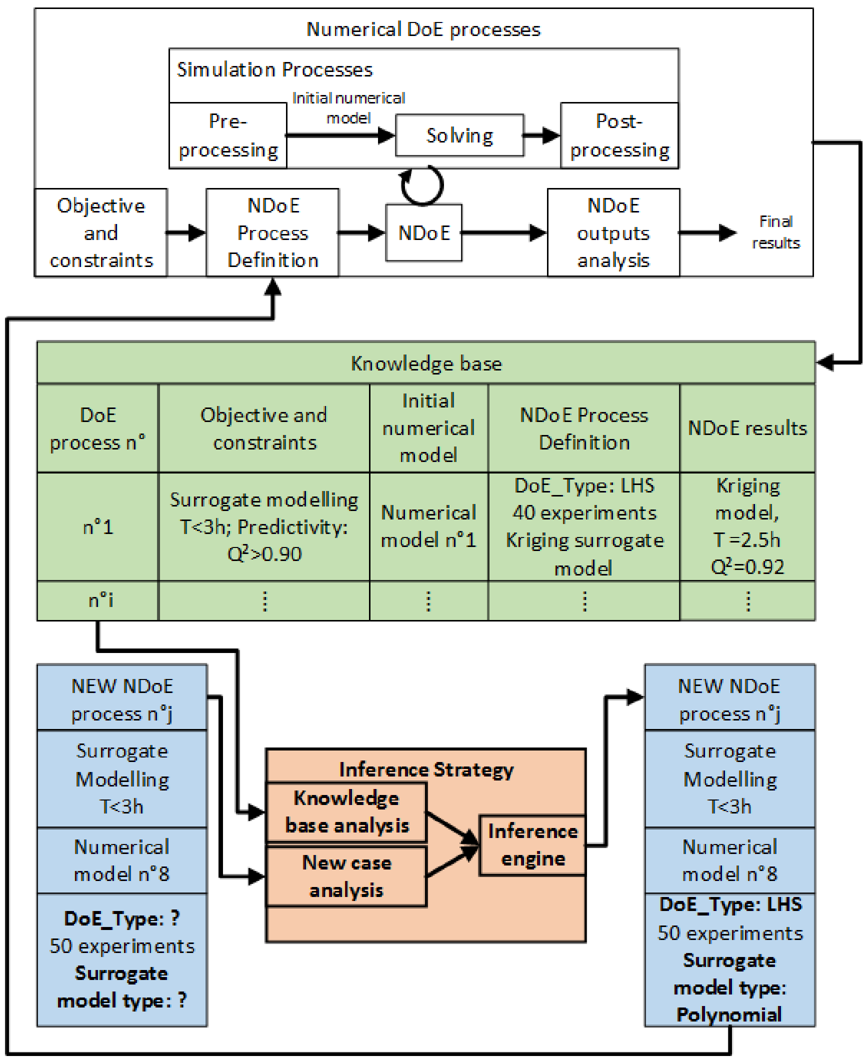


Figure 1. The proposed KBS structure for numerical DoE process

Figure 1 shows the general architecture of the proposed KBS, which is composed of:

- A NDoE process, based on a validated parameterized numerical model and a simulation process able to execute each experiment. The process is entirely defined (e.g. objective, constraint, type of metamodel, factors and outputs, etc.) and executed. NDoE outputs obtained are analyzed to fulfill the objective of the NDoE process. For instance, a NDoE can be used for a sensitivity analysis to determine the influence of factors and their interactions on the studied output [Da

Veiga 2014]. Constraints could be, in this case, a time limit and a minimum accuracy on sensitivity indices estimations.

- A knowledge base, to gather every previous configurations of NDoE processes. All data needed to define completely the process are stored, as well as results;
- An inference engine, to deduce new solutions from the knowledge base, regarding to the new incompletely defined NDoE process;
- A user interface, not represented here, is used to ensure a bi-directional communication with the user. The user is able to define a known part of the NDoE process, to ask the system for a complete configuration with explanations, and to validate or modify the proposed configuration.

In order to help the designer to configure the NDoE process, the inference engine must be selected regarding to several criteria:

- Learning: the system has to learn from the capitalized knowledge each time the knowledge base is updated.
- Speed: the learning step and the reasoning step have to be as fast as possible. The time spent by the system must be lower than the time spent by the designer.
- A priori knowledge support: the system should be able to be enriched by designer's knowledge to simplify the inference process.
- Inference on continuous and discrete variables: the inference engine aims to determine numerical parameters and to select methods
- Explanation: the system must be able to explain clearly its reasoning to the user.

2.2 Comparison of inference methods

To analyze the knowledge base, a large amount of inference engines exists. This subsection focuses on several methods used in the context of mechanical design and simulation. These methods will be discussed according to previously described criteria.

2.2.1 Rule-based system

Some expert systems were developed and deal with DoE to bring support for researcher in their experimental approach [Weiner 1992]. A specific expert system was designed to configure a DoE process [Lorenzen et al. 1992]. All of these developments, based on predefined rules, do not support NDoE. The scope of covered methods is very limited. Another KBS was developed for finite element meshing [Dolšak 2002] to propose optimal solution for several mesh parameters. A large amount of rules, obtained by a machine learning method (inductive logic programming), composed the knowledge base. This KBS was developed to shorten the time-consuming task of meshing. Another applications of rule-based systems exist, for instance for design automation [Naranje and Kumar 2014]. The main limitation of this approach is the requirement of an exhaustive and accurate definition of rules. A second limitation is its boolean property, which can dismiss several solutions which can be potentially good.
Fuzzy logic alleviates these limitations by taking into account the inaccuracy of the knowledge. It has been used, for instance, for predicting finite-element-analysis results [Subba Rao and Pratihar 2007].

2.2.2 Machine learning

Machine learning techniques are able to learn dynamically from updated knowledge-base to deduce rules. As each case in the knowledge base is already classified and each data are labelled, only known methods based on supervised learning, and applied for mechanical product development processes, are presented here.

1. Artificial neural networks

Artificial neural networks were used, for instance, to predict gasoline engine performance [Tasdemir et al. 2011]. This application shows a close efficiency between artificial neural networks and fuzzy logic methods. Another application for manufacturing uses this method for feature recognition and process planning [Ding and Matthews 2009]. To select each operation, tools, etc., binarized outputs vectors are defined. Artificial neural networks have been also used to determine the best surrogate modelling method according to a specific problem [Cui et al. 2016]. The main drawback of this method is its "black box" property. It is impossible to explain the reasoning of a KBS using artificial neural networks.

2. Decision trees

They can be used for feature selection, for instance, to identify important sources of fault [Sugumaran and Ramachandran 2011]. These methods are non-parametric (no hypothesis on data distribution), non-linear, robust and easy to understand. Over-fitting effects can be limited by pruning methods for instance [Sahin et al. 2012]. Decisions trees are limited to discrete variables and it cannot be used for regression tasks. It is also impossible to pre-define the structure of the tree, as this step is computed automatically. Thus, a decision tree cannot be enriched by user's knowledge [Bayat et al. 2009].

3. Bayesian networks

This probabilistic approach may be used for classification and detection. As an example, it may be used to predict the quality of a machining process [Correa et al. 2009]. Bayesian networks can represent uncertain knowledge with clarity. The structure of the network can be defined by the user according to its knowledge. Bayesian networks are able to determine as discrete as continuous variables.

All of these machine learning models fulfill the learning property needed, but they may presents more difficulties to be determined because of over-fitting phenomenon and the number of their internal parameters. Some of them are unable to explain the reason of the proposed decision, as neural networks. Another drawback is the difficulty to compute numerical internal parameter of methods used in the NDoE process, except for bayesian networks.

2.2.3 Case-based reasoning

This reasoning method reuses or adapts former solutions for a new problem. It was applied, for instance, for finite element analysis to support adaptive mesh processes [Khan et al. 2014]. This method does not need any model and thus, there is no learning step like machine learning methods. The first step can be based on similarity measures, ranking, k-nearest neighbor algorithms, etc. Cases can be reused by adapting the most similar case (transformational reuse) or by reusing the method used to obtain this similar case (derivational reuse). For transformational reuse, a domain-specific transformational operator and rules are used [Lopez de Mantaras et al. 2005]. Case-based reasoning provides clear explanations of its reasoning process. The main drawback of this method is the obligation to define specific rules for case adaptation.

2.2.4 Hybrid systems

Hybrid KBS combine several approaches in their inference engine. A large amount of different hybridization has been developed in the last decade [Sahin et al. 2012]. For instance, the ANFIS method (Artificial Neuro-Fuzzy Inference System) combines fuzzy logic and neural networks [Jang 1993], [Roy and Pratihar 2013]. However, this approach keeps the unexplainability of a neural network.

2.2.5 Comparison

All of these methods reviewed above need to be compared to define the inference strategy of the KBS. Results are summarized in Table 1. The speed criterion is unable to discriminate any method, as it depends on the complexity of each model, the size of the dataset, the size of the output and computational resources. The only method which respects all others criteria is bayesian network: able to learn from a dataset, able to determine the value of discrete and continuous variable, able to be enriched by personal user's knowledge and able to explain reasoning with clarity. However, the distribution of each variable must be set or estimated to compute conditional probabilities. While it could be easily set for discrete variables, continuous variable are more difficult to be modelled. To alleviate this difficulty, a mixed approach is proposed, dealing with discrete variables by a bayesian network and with continuous variables by an artificial neural network. This approach is detailed in the next section.

Table 1. Comparison of inference methods

Class of reasoning	Rule-based reasoning		Machine learning			Case-based reasoning	Hybrid systems
Methods	Formal logic	Fuzzy logic	Neural networks	Decision trees	Bayesian networks		ANFIS
Learning	N	N	Y	Y	Y	Y	Y
Continuous/discrete variables	N/Y	N/Y	Y/Y	N/Y	Y/Y	Y/Y	Y/Y
A priori knowledge	Mandatory		Impossible		Possible	Mandatory	Mandatory
Explainability	+	++	--	++	++	++	--

3. KBS for numerical DoE

This section aims to define the inference strategy able to support the NDoE process configuration. Data describing the NDoE process are firstly described to design the knowledge base. Based on inference methods comparison, the inference strategy is specified according to the knowledge base specifications.

3.1 Knowledge base

The knowledge base contains a structured description of each NDoE process. Data describing the process can be organized in several categories (Figure 2):

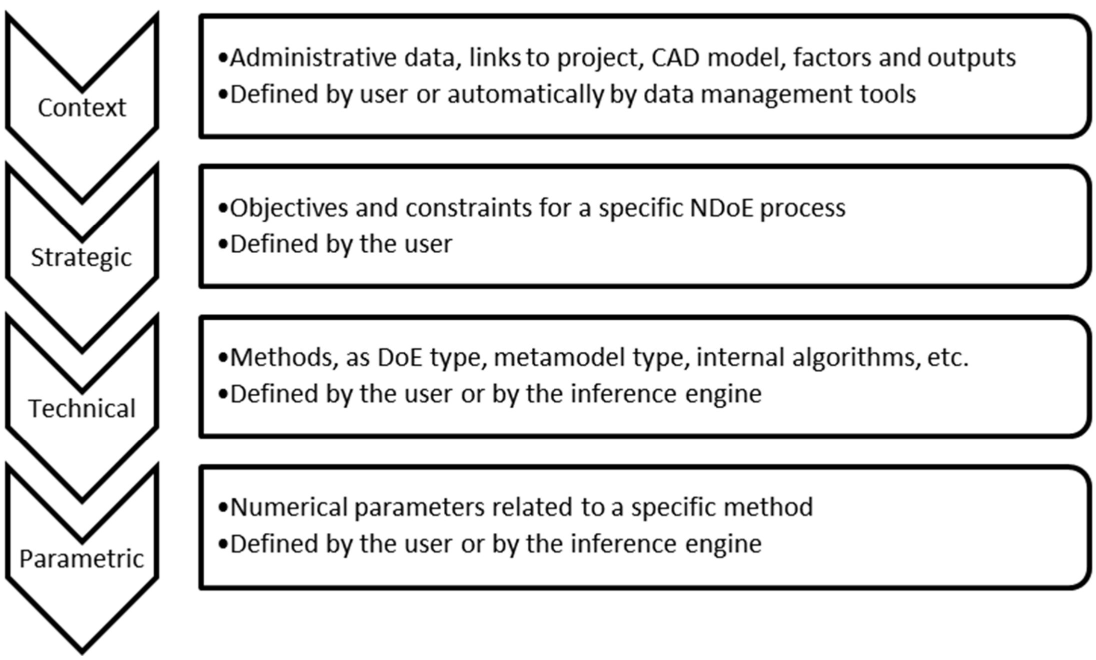


Figure 2. Data classification

- Context data: it covers identification data, authors, dates, and project, product, numerical model linked with the NDoE process. It concerns also the number and types of factors involved in the DoE, studied outputs, the computational cost of one simulation (assessed by the system if necessary), the hardware used. Context data are mandatory to execute the KBS. These data must be defined manually or retrieved in a data management system;
- Strategic data: it is specific to the objective of the NDoE process (sensitivity analysis, surrogate modelling, robustness analysis, optimization, exploration, etc.) and constraints.

- Technical data: it gathers type of methods used according to strategic data. For instance, for the surrogate modelling objective, the type of DoE and surrogate model has to be selected. It concerns also methods used by these methods. For instance, a method to compute polynomial coefficients for the polynomial chaos surrogate model must be selected, and several exists;
- Parametric data: each numerical datum needed by each method, as the degree of a polynomial model or the number of iteration of an optimization algorithm.

In addition to these types of data, NDoE process performance indicators are also taken into account. They are specific for each objective and they are useful to compare processes to each other. For instance, the Q^2 coefficient is used to measure the surrogate model predictivity (i.e. its ability to predict good results from unknown inputs).

3.2 Inference methodology

The proposed KBS can operate in two ways. The KBS aims to complete the definition of a DoE process according to already defined context, strategic, technical and parametric data. The KBS can also diagnose the efficiency of a user-defined configuration by predicting process performance indicators, like accuracy or time.

The inference strategy is user-centered. He/she is able to define input data manually, change several aspect of the configuration proposed by the KBS and even to run the NDoE process even it seems to be inefficient. Moreover, performance indicators are interpreted by the user, not by the KBS, in order to avoid an inherent bias in the inference engine. For instance, a NDoE process declared as inefficient would penalized wrongly several methods, while an inefficient configuration could be composed of potentially very efficient methods, and it also depends on the context, as the initial numerical model complexity (non-linear properties, multi-physics computations, etc.). This ability is let to the user, regarding to its own requirements.

The proposed inference methodology is based on 2 main steps, illustrated in Figure 3 and in Figure 4. All missing data needed to complete the configuration are predicted by a bayesian network except continuous variables, which is predicted by an artificial neural network.

3.2.1 DoE process definition

A preliminary step consists in defining context and strategic data demanded to start inferences. These data may be gathered by a data management system deployed in the company or defined manually by the user. These data are essential to ensure traceability and to define clearly the context of an optimal NDoE process configuration.

3.2.2 Methods selection

A bayesian network is used to deduce the best combination of methods. This bayesian network must take into account the influence of each technical and parametric data.

Some classes of methods will be used depending on the objective and on the user's will. For instance, while a NDoE must be chosen in every cases, a sensitivity analysis method cannot be selected if the objective is surrogate modelling. Moreover, metaheuristics (for adaptive DoE) can be used for several objectives but they are optional. From this observation, some variables can be ignored or are mandatory according to the objective of the NDoE process. It is more convenient to define one bayesian network for each objective, to avoid a useless complex network. The structure of the network (i.e. stochastic dependencies) is modelled according to expert knowledge.

Only available data, corresponding to the right objective, are used to compute the probability distribution of each variable. There is no hypothesis on probability distributions of variables.

While it is generally difficult to model the probability distribution of continuous parameters, all parametric data must be included into the bayesian network as their influences can be significant. At this step, continuous variable are included as discrete variable composed of several continuous intervals. This discretization is used to prepare the third step, by estimating the most probable interval for needed parametric data.

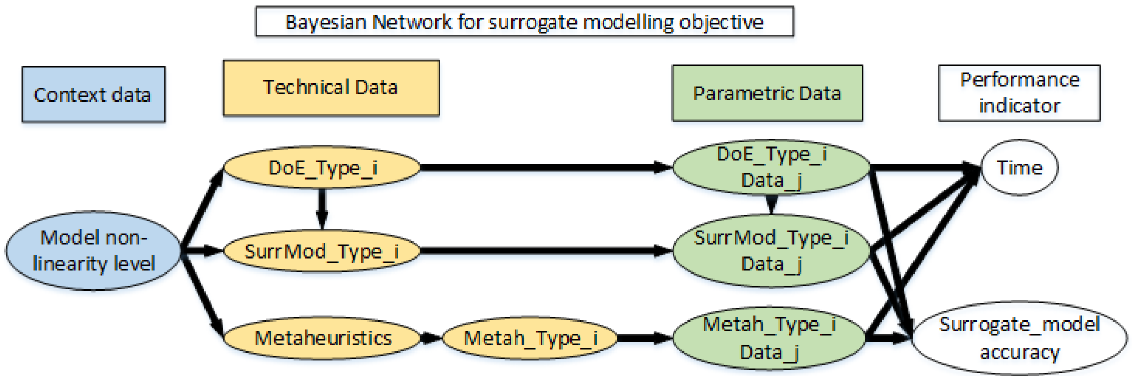


Figure 3. Bayesian network for NDoE process

3.2.3 Parametric data computation

To complete the configuration of the NDoE process, an artificial neural network is used to compute a value accurately for continuous parametric data. From most probable intervals obtained from the bayesian network, a subset of the set of case determined in the first step is extracted. As the bayesian network, the neural network is pre-built according to the same input type taken into account in the bayesian network, for each type objective. A binarization is used to represent selected and non-selected methods. The new set of case is used to train the neural network.

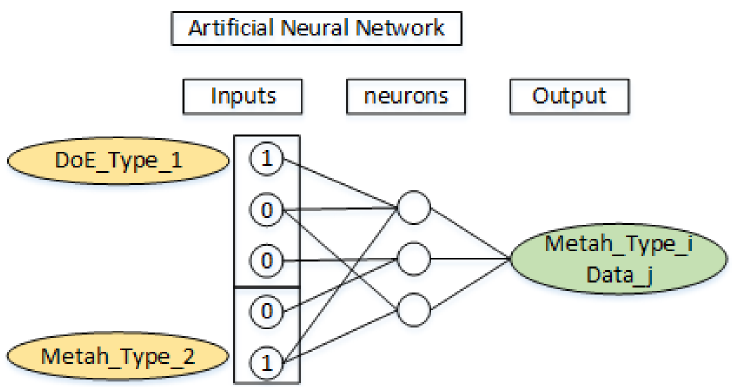


Figure 4. Artificial neural networks for NDoE process

3.2.4 Implementation

The KBS is supported by different tools:

- Virtuoso server (http://virtuoso.openlinksw.com/), containing each case of DoE process;
- Netica (http://www.norsys.com), to compute a Bayesian network;
- Uranie (http://sourceforge.net/projects/uranie/), to compute parametric data by artificial neural networks, and to execute the DoE process;
- Code_Aster (http://web-code-aster.org), a finite-element analysis software, to run simulations.

This system will be supported by an user interface.

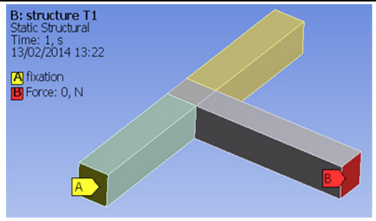
4. Use-case

This use-case presents an application of the proposed KBS to a very specific case. It aims to illustrate the running of the bayesian network based on a simple test-case and limited to specific choices.

4.1 Problem definition

The user wants to obtain an accurate and predictive surrogate model of the finite-element model of a T-shape beam. This model has a very low computational cost, in order to test our KBS fast.

Table 2. Description of the model used for the use-case

Factors	Model	Output
E: Young's modulus (MPa) Fx and Fy: Force (N) on B		S : maximal value of the first invariant of the strain tensor

The bayesian network used for this use-case is illustrated by Figure 5. A limited number of variables and states were considered for the bayesian network. Only one numerical model and objective (here, surrogate modelling) are considered, so that these variables were not included into the network. The database is composed of 500 NDoE processes, described by the DoE type, the number of experiments, the type of surrogate model and one parameter for each type of surrogate model (degree for polynomial chaos and the number of neurons for neural networks). Each DoE process produces two performance indicators, representing accuracy (RMSE) and predictivity (Q^2) of each computed surrogate model.

4.2 Results

Results are integrated into the bayesian network in Figure 5.

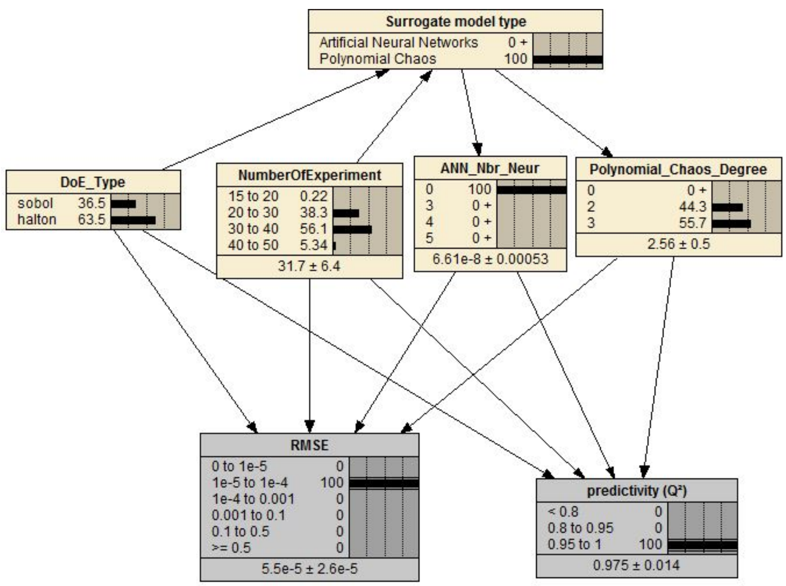


Figure 5. Bayesian network results to choose the best configuration

Conditional probabilities allow the user to request the most efficient combination of method to get maximal accuracy and predictivity. This figure represents the network after the learning step and after the user set his/her requirements. As represented in grey nodes, he/she wants a surrogate model with RMSE between 1E-05 and 1E-04, and a Q^2 between 0.95 and 1 (i.e. a very good surrogate model). The network translate this by a probability of 100% and computes, thanks to conditional probabilities, the probably most efficient configuration: a Halton type DoE with 30 to 40 experiments, combined with a polynomial chaos with degree 3. This configuration is the best according to known NDoE processes and the context data.

5. Discussion

The proposed KBS aims to shorten the NDoE process by reusing capitalized knowledge and avoiding designers to loose time by searching for an optimal configuration. It is important to compare the time spent to set the KBS, the time needed for the learning step and the time really saved for designers. This development is more adapted to companies which need and have already applied a data management strategy to capitalize data.

Also, it may be useless if there is only one user of NDoE process in the whole company. A lonely user will be the only expert in the company and he/she will be the only source of data for the KBS. This KBS offers a framework to share, retrieve and analyze knowledge. It would benefit extended companies, with several teams in different locations, using NDoE processes for different goals. A theoretical threshold may also exist when the knowledge base becomes to be too huge. In this case, the learning step would be too long, although in this case, the knowledge base would be highly filtered and sorted before the learning step to avoid a too huge learning data set.

The proposed system can be considered by most of the users as a black-box, but an expert is necessary to enrich the KBS with, for instance, a new kind of variable.

6. Conclusion and future works

Numerical DoE process could be very complex to be configured. A KBS is proposed to help the designer for this task. This approach aims to reuse capitalized knowledge to shorten the preparation step. This KBS is based on a knowledge base, to support data capitalization, classification and requests, and an inference engine for data analysis. The proposed inference methodology is based on a mixed approach, combining bayesian network and artificial neural networks. This approach will be compared to a bayesian network in which the continuous variables are directly modelled or estimated from data.

Each aspect of the KBS will be extended, as the bayesian network structure and the knowledge base, to validate this proposed KBS and to cover a large scope of methods and configurations. Learning algorithms will be compared and a validation strategy will be set to avoid over-fitting. Furthermore, the structure can be also determined by a learning step, while it is currently set according to expert knowledge.

These developments extend previously developed expert systems capabilities, by combining uncertain expert knowledge and empirical data analysis, and apply an inference method used in other fields to the numerical simulation process. Such a KBS could make the use of NDoE easier for companies, leading to better products in a shorter and shorter product development process.

Acknowledgement

This work is done in the French FUI project SDM4DOE. We also thank all consortium partners for their contribution during the development of ideas and concepts proposed in this paper.

References

Bayat, S., Cuggia, M., Rossille, D., Kessler, M., Frimat, L., "Comparison of Bayesian network and decision tree methods for predicting access to the renal transplant waiting list", Studies in Health Technology and Informatics, No.150, 2009, pp. 600–604.

Blondet, G., Boudaoud, N., Le Duigou, J., "Simulation Data Management for Adaptive Design Of Experiment", Proceedings of Qualita'2015. Nancy, France, 2015.

Blondet, G., Le Duigou, J., Boudaoud, N., Eynard, B., "Simulation data management for adaptive design of experiments: A litterature review", Mechanics & Industry, No.16, Vol.6, 2015, p. 611.

Castric, S., Cherfi, Z., Blanchard, G. J., Boudaoud, N., "Modeling Pollutant Emissions of Diesel Engine based on Kriging Models : a Comparison between Geostatistic and Gaussian Process Approach", Proceedings of 14th IFAC Symposium on Information Control Problems in Manufacturing, INCOM'12, Vol.2006, No.14, 2012, pp. 1708–1715.

Correa, M., Bielza, C., Pamies-Teixeira, J., "Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process", Expert Systems with Applications, Vol.3, No.36, 2009, pp. 7270–7279.

Cui, C., Hu, M., Weir, J. D., Wu, T., "A recommendation system for meta-modeling: A meta-learning based approach", Expert Systems with Applications, No.46, 2016, pp. 33–44.

Da Veiga, S., "Global sensitivity analysis with dependence measures", Journal of Statistical Computation and Simulation, Vol.7, No.85, 2014, pp. 1283–1305.

Ding, L., Matthews, J., "A contemporary study into the application of neural network techniques employed to automate CAD/CAM integration for die manufacture", Computers and Industrial Engineering, Vol.4, No.57, 2009, pp. 1457–1471.
Dolšak, B., "Finite element mesh design expert system", Knowledge-Based Systems, Vol.5-6, No.15, 2002, pp. 315–322.
Forrester, A. I. J., Keane, A. J., "Recent advances in surrogate-based optimization", Progress in Aerospace Sciences, Vol.1-3, No.45, 2009, pp. 50–79.
Hu, W., Yao, L. G., Hua, Z. Z., "Optimization of sheet metal forming processes by adaptive response surface based on intelligent sampling method", Journal of Materials Processing Technology, Vol.1-3, No.197, 2008, pp. 77–88.
Jang, J.-S. R., "ANFIS: adaptive-network-based fuzzy inference system", IEEE Transactions on Systems, Man, and Cybernetics, Vol.3, No.23, 1993, pp. 665–685.
Khan, A. A., Chaudhry, I. A., Sarosh, A., "Case Based Reasoning Support for Adaptive Finite Element Analysis: Mesh Selection for an Integrated System", Applied Physics Research, Vol.3, No.6, 2014, pp. 21–39.
Kiritsis, D., "A review of knowledge-based expert systems for process planning. Methods and problems", The International Journal of Advanced Manufacturing Technology, Vol.4, No.10, 1995, pp. 240–262.
Lopez de Mantaras, R., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Watson, I., "Retrieval, reuse, revision and retention in case-based reasoning", The Knowledge Engineering Review, Vol.03, No.20, 2005, p. 215.
Lorenzen, T. J., Truss, L. T., Spangler, W. S., Corpus, W. T., Parker, A. B., "DEXPERT: an expert system for the design of experiments", Statistics and Computing, Vol.2, No.2, 1992, pp. 47–54.
Naranje, V., Kumar, S., "A knowledge based system for automated design of deep drawing die for axisymmetric parts", Expert Systems with Applications, Vol.4, No.41, 2014, pp. 1419–1431.
Patelli, E., Murat Panayirci, H., Broggi, M., Goller, B., Beaurepaire, P., Pradlwarter, H. J., Schuëller, G. I., "General purpose software for efficient uncertainty management of large finite element models", Finite Elements in Analysis and Design, No.51, 2012, pp. 31–48.
Roy, S. S., Pratihar, D. K., "Adaptive neuro-fuzzy expert systems for predicting specific energy consumption and energy stability margin in crab walking of six-legged robots", Journal of Intelligent and Fuzzy Systems, Vol.3, No.24, 2013, pp. 467–482.
Sahin, S., Tolun, M. R., Hassanpour, R., "Hybrid expert systems: A survey of current approaches and applications", Expert Systems with Applications, Vol.4, No.39, 2012, pp. 4609–4617.
Subba Rao, A. V., Pratihar, D. K., "Fuzzy logic-based expert system to predict the results of finite element analysis", Knowledge-Based Systems, Vol.1, No.20, 2007, pp. 37–50.
Sugumaran, V., Ramachandran, K. I., "Effect of number of features on classification of roller bearing faults using SVM and PSVM", Expert Systems with Applications, Vol.4, No.38, 2011, pp. 4088–4096.
Tasdemir, S., Saritas, I., Ciniviz, M., Allahverdi, N., "Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine", Expert Systems with Applications, Vol.11, No.38, 2011, pp. 13912–13923.
Weiner, D. J., "Expert systems to aid in the formulation of hypotheses and design of experiments in biomedical research", Mathematical and Computer Modelling, Vol.6-7, No.16, 1992, pp. 185–198.

Gaëtan Blondet, Phd
Sorbonne Universités, Université de Technologie de Compiègne, Department of Mechanical Systems Engineering, UMR7337 Roberval
Rue du docteur Schweitzer CS 60319, 60203 Compiègne, France
Email: gblondet@utc.fr