

ADVANCED DATA ENRICHMENT AND DATA ANALYSIS IN MANUFACTURING INDUSTRY %< AN EXAMPLE OF LASER DRILLING PROCESS

You Wang ^a, Hasan Tercan ^b, Thomas Thiele ^b, Tobias Meisen ^b, Sabina Jeschke ^b, Wolfgang Schulz ^{a,c}

- a. Nonlinear Dynamics of Laser Manufacturing Processes Instruction and Research Department (NLD) of RWTH Aachen University, Steinbachstraße 15, 52074, Aachen, Germany
- b. Institute of Information Management in Mechanical Engineering (IMA) of RWTH Aachen University, Dennwartstraße 27, 52068, Aachen, Germany
- c. Fraunhofer Institute for Laser Technology, Steinbachstraße, 52074, Aachen, Germany

ABSTRACT

Nowadays, the internet of things and industry 4.0 from Germany are all focused on the application of data analytics and Artificial Intelligence to build the succeeding generation of manufacturing industry. In manufacturing planning and iterative designing process, the data-driven issues exist in the context of the purpose for approaching the optimal design and generating an explicit knowledge. The multi-physical phenomena, the time consuming comprehensive numerical simulation, and a limited number of experiments lead to the so-called sparse data problems or “curse of dimensionality”. In this work, an advanced technique using reduced models to enrich sparse data is proposed and discussed. The validated reduced models, which are created by several model reduction techniques, are able to generate dense data within an acceptable time. Afterwards, machine learning and data analytics techniques are applied to extract unknown but useful knowledge from the dense data in the Virtual Production Intelligence (VPI) platform. The demonstrated example is a typical case from laser drilling process.

Keywords— sparse data problems, data analytics, machine learning, virtual production intelligence, model reduction

1. INTRODUCTION

Presently, production companies in the high-wage countries have to overwhelm the challenges of rapid responding to variant market demands, individual customer necessities and increasing labor costs. The rising complexity in the production processes motivates the generation of more reliable, efficient and flexible production planning and scheduling steps. The process parameter identification, knowledge extraction, iterative and communicative process design and multidisciplinary optimization are vital fields for production planning and decision making. Since a large amount of data are generated from machines, sensors, orders

The investigation are partly supported by the German Research Association (DFG) within the Cluster of Excellence “Integrative Production Technology for High-Wage Countries” at RWTH Aachen University.

etc. in manufacturing industry, the developed data-driven methods and methodology can be applied in manufacturing decision making processes [1]. However, the data density is not enough for extracting knowledge because the evident parameter dimensionality is not enough and “the number of experiments” is limited by time or costs.

In this work, a methodology to enrich sparse data by fast and frugal reduced models is introduced. Several typical model reduction methods such as mathematical methods, numerical methods and data-driven methods for generating reduced models are reviewed. After obtaining sufficient and dense data, machine learning methods such as clustering and classification are applied to conduct the data analysis and knowledge extraction. The example case is a laser drilling process. The detailed enrichment of the data and the data-driven decision making process are demonstrated in this example.

2. SPARSE DATA PROBLEMS IN MANUFACTURING INDUSTRY

The processes in manufacturing industry are characterized by multi-parameter models with high resolution. Moreover, the solvability of the process is considerably to be restricted by the complexity of physics. From a data analytics perspective, two main barriers are addressed in this paper, which slow down the process to extract knowledge from manufacturing processes.

The first reason is that the required number of sampling points is enormous. When the dimensionality of parameters increases, the volume of sampling space increases exponentially. That means a large amount of parameters combination is needed and the existing available data becomes sparse. Especially, in manufacturing industry, the data from different process domains becomes heterogeneous and sparse within the high dimensional parameter space.

The second reason is that the sampling process can be time consuming. The sampling process is a process of selecting or generating the dataset which can be used in knowledge extraction. The sampling points can be collected from the real experiments or the computer aided calculations. The number of real experiments is not only limited by the time restriction but also limited by the boundary of the

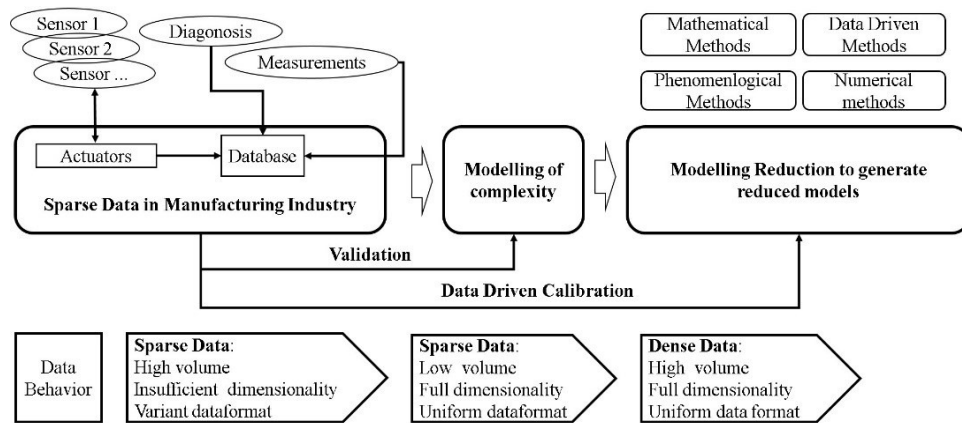


Fig. 1. Workflow to enrich sparse data to dense data

performance of machines or sensors. The computer aided calculation, also known as numerical simulation, is a powerful tool to generate sampling datasets. However, the complex numerical simulation could be time-consuming because of the high resolution and complex process physics.

3. ENRICHING SPARSE DATA BY REDUCED MODELS

Generally speaking, the reduced models can enrich the sparse data from two aspects. On the one hand, the reduced models decrease the required data volume dramatically. The “sparse data” become “dense data” because of the compendious dimensionality. On the other hand, the reduced models are also characterized by their convenient solvability, millions of datasets can be generated from reduced models in acceptable duration.

The workflow to enrich sparse data into dense data are shown in Figure 1. Firstly, the data is extracted from divergent sources in real manufacturing processes and experimental measurements. Different types of sensors, diagnosis techniques and design of experiments are intensively operated to accumulate the original sparse data. At this stage, the data are featured by its high volume, insufficient dimensionality, heterogeneous distribution and irregular data format. Thereafter, the massive mathematical and physical modelling work for the complex manufacturing process is performed and validated by the sparse data in the first stage. From the well-built complicated models, the full dimensional parameters will be involved and the data generated from multifarious models is equipped with standard data format. Since the sampling process by complex models is time consuming, model reduction techniques are applied to generate the fast and frugal reduced models and avoid the unnecessary complexity. The reduced models are derived to avoid any unnecessary complexity and to reduce the computation time of large-scale dynamical systems by inducing approximations of much lower dimensions which can produce nearly the same input-output response characteristics. Meanwhile, to ensure the accuracy and the applicability on the specific context, the reduced models are also calibrated and validated by the measured sparse data. These reduced models can simulate the complex system by preventing redundant calculation, so the sparse data can be

enriched dense enough for data-driven decision making extraction within a short phase.

The model reduction procedure adopts a top-to-down approach and starts from the original partial differential equations and derives approximated analytical solutions or a set of ordinary equations using many mathematical approaches, physical and phenomenological approaches, numerical approaches and data driven model reduction methods. Especially, there are several model reduction techniques which are convenient enough to use and worth reviewing.

3.1 Perturbation analysis

The objective of perturbation theory is to determine the behavior of the solution when one variable tends to be very small, which can lead to the split of two part of solutions for complex system. One part is the temporal solution and another part is the long term asymptotic solution. This separation of system solutions result in the reduction for the models. The typical application of this perturbation theory lies in the fields involving differential equations as well as a series of engineering problems [2] [3]. Vossen et al. used the perturbation and asymptotic analysis to describe the dynamical behaviors of the free boundaries of the melt during the laser cutting process considering the spatially distributed laser radiation. A reduced model which can generate results at real time scale was derived by perturbation analysis for the purpose of predicting the product roughness [4].

3.2 Inertia and central Manifold analysis

The inertial manifolds are connected with the long term behavior of the solutions of dissipative dynamical systems. The reduced phase space of ordinary differential equations and partial differential equations in the long time limit is named as central or inertial manifold. Schulz et al. derived a reduced model by applying inertial manifold method. This reduced model can calculate the thermal behavior in laser manufacturing processes very fast [5, 6].

3.3 Buckingham Pi theory

The Buckingham Pi theory is used to find the dimensionless groups from relevant input and output parameters. The dimensionality of the original parameters is sharply decreased and simplified by applying the dimensionless groups [7]. The major steps to perform the model reduction with assistance of Buckingham Pi theory are as following:

- Step1: Finding the dimension matrix;
- Step2: Determining the rank of the dimensions of full-space parameters;
- Step3: Finding the vectors spanning the full-space parameters;
- Step4: Finding the reference meta-models which include the dimensionless parameter groups;
- Step5: Data-driven modelling by calibration sparse data and reference meta-models;

Schulz et al. derived a reduced model by applying the Buckingham Pi theory and the steps listed above. This reduced model can calculate the heat conduction losses in laser sheet metal cutting processes rapidly [8].

3.4 Proper Orthogonal Decomposition

Proper Orthogonal Decomposition (POD) is a numeric method by searching for a low-dimensional approximate representation of the large scale dynamical systems, such as signal analysis, turbulent fluid flow and large dataset like image processing [9,10]. POD generates a set of orthonormal basis of dimensions, which minimizes the error from approximating the snapshots. It can generally give a good approximation with substantially lower dimensionality [11].

4. EXAMPLE FOR LASER DRILLING MANUFACTURING

As an example to illustrate the data enrichment by reduced models and data visualization, an advanced reduced model for sheet metal drilling has been developed by Nonlinear Dynamics of Laser Manufacturing Processes Instruction and Research Department (NLD) in RWTH Aachen University. Using this model, the final shape of the drilling holes can be calculated and described. Inside the formula (see Figure 2), the term F is the local laser fluency, z and x represent the position along z and x axis respectively, the term F_{th} is based on the heuristic concept of an ablation threshold and material dependency. The only one unknown parameter has to be calibrated and determined with experimental sparse data. Afterwards, this reduced model can be used to calculate the final shape of the drilling hole by laser sheet metal drilling. Not only the final shape of drilling hole but also the feasibility for each parameter can be indicated accurately by this reduced model

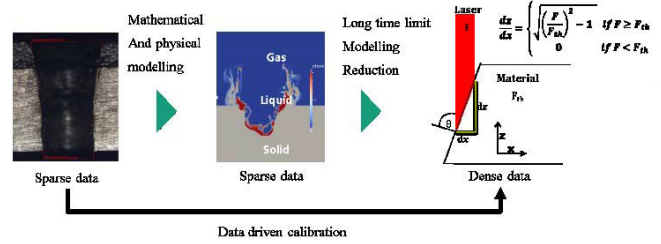


Fig. 2. The procedures to generate asymptotic drill reduced model

As a consequence, the whole asymptotic shape of the drilling hole is calculated and is illustrated. Finally, classification of sheet metal drilling can be performed by identification of the parameter region where the drilling hole achieves its asymptotic shape.

The asymptotic drill reduced model is an ordinary differential equation which can be solved within 1 second for each single run. This fast reduced model makes it possible to collect dense data in an acceptable period. By using the asymptotic drill reduced model, 10,000 sampling points can be generated within five seconds without parallel calculations. However, it takes 30 minutes to produce one sample if the complicated numerical simulation is adopted.

Table 1. Example of parameters and their range in laser drilling processes

Parameters	Ranges
Pulse Duration [tp]	0.1-1.5 [ms]
Laser Power [PL]	3-10 [kW]
Focal Position [z0]	-8-8 [mm]
Beam Radius [w0]	50-350[μm]
Rayleigh length [zR]	3-35 [mm]
Workpiece Thickness [d]	0.2-5[mm]

After the sparse data is enriched into dense data by asymptotic reduced model, the machine learning techniques are applied to conduct data analytics. Thereby the data analytics process including appropriate data visualization methods are implemented within a Virtual Production Intelligence (VPI) platform [12]. The process is described in detail in [13]. It implemented a hybrid data analytics approach with clustering and classification tree to identify parameters of the manufacturing process that result in desired outputs. The approach is shown in Figure 3.

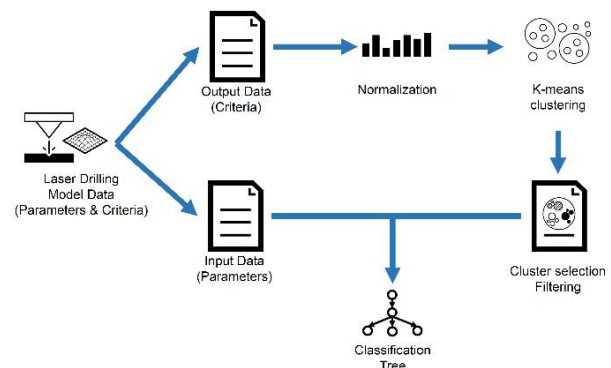


Fig. 3. Data analytics process to analyze dense data [13]

First of all, clustering is used to divide the output data of the reduced model into different regions (clusters). Thus, the user is able to select a cluster that represents desired process results. After that, the classification trees are used to identify regions of the parameters space (see also Table 1) which lead to these process result.

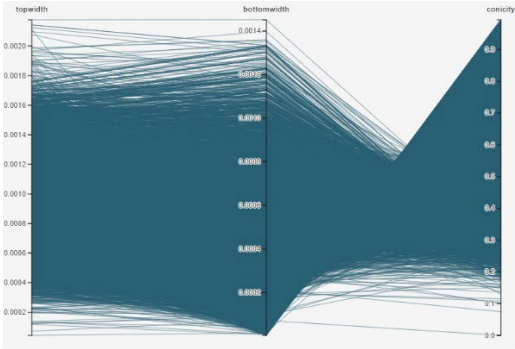


Fig. 4. Parallel coordinates visualization of 10,000 sampling points with 3 different axes (model outputs): top width, bottom width, and conicity

The output of the asymptotic drill reduced model represents the shape of the drilling hole and thus consists of three dimensions: the widths at the top and the bottom of the hole as well as the conicity. In order to depict and analyze this multi-dimensional data, a visualization technique named parallel coordinates is utilized. Figure 4 shows its implementation in the VPI platform for 10,000 laser drilling sampling points, whereas the data is generated with the fast reduce model. In the next step, the data are divided with a clustering algorithm into 4 clusters. The following figure 5 illustrates the clustering results of the K-means algorithm.

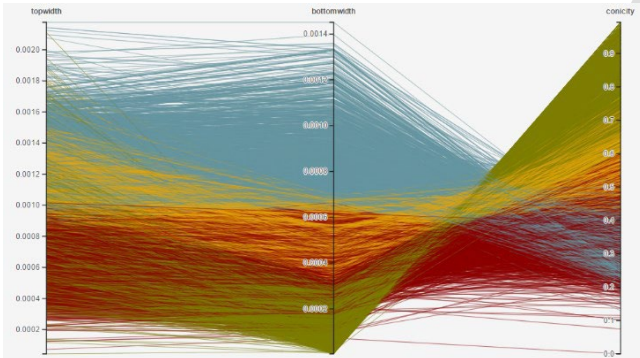


Fig. 5. Clustering results with four clusters (blue, green, red, yellow) of similar process outputs

The figure shows that 3-dimensional output space is roughly separable into different groups of sampling points, including the blue cluster with 22% of all data points as well as the yellow (26%), red (31%), and green (21%) clusters.

In the following figure, the blue and the green clusters are highlighted. It can be seen that the clusters are lying conversely and that the green one leads to high conicity values. In our application case, the high conicity is a desired process result. Thus, the green cluster is a good one, whereas the other ones are bad.

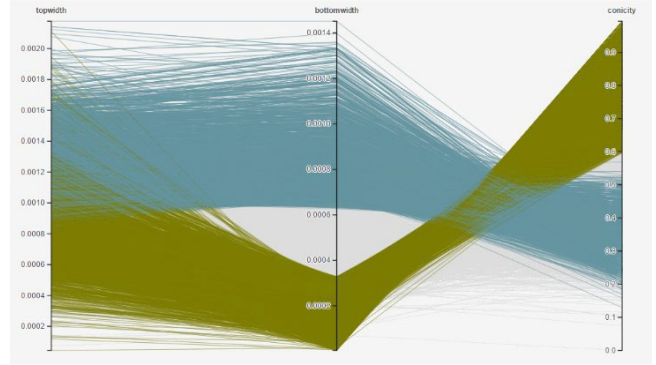


Fig. 6. Highlighting bad process results (blue cluster) and good/desired results (green) cluster in the visualization

Having identified the good (i.e. desired) output spaces as well as the bad ones, the next step is to transform the problem into a binary classification problem and to build a classification tree that is used to predict the process outcome (good/bad) on the basis of the laser drilling process parameters (see Table 1).

The following Figure 7 shows a classification tree for the desired clusters (high conicity).

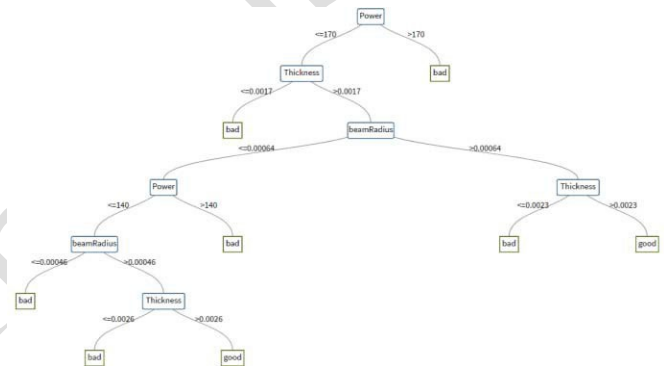


Fig. 7. Classification tree for the desired clusters (high conicity)

The tree shows that there are mainly two parameter space regions that lead to the desired results (good leaves). These two regions can be defined by the following rules (extracted from the tree):

- *Laser Power ≤ 170 & Thickness ≤ 0.0023 & Beam Radius > 0.00064*
- *Laser Power ≤ 140 & Thickness > 0.0026 & Beam Radius between 0.00046 and 0.00064*

These results show that the hybrid data analytics approach on the top of the reduced model data provides an intuitive and interpretable decision support for the laser drilling process planner. The gained knowledge, especially the identified parameter regions, can subsequently be used to further optimize the process.

6. OUTLOOK

In this paper, the methodology to enrich sparse data to dense data and analyze acquired dense data is demonstrated. In order to fully utilize the advantages of the reduced models, a

large amount of efforts should be made to generate more useful reduced models. Besides, different reduced models have different domain space, the exploration of the boundary for each reduced model is necessary for generating reliable and high quality process knowledge. These regime conditions can be determined by physical driven or data driven methods.

The comprehensive use of data from reduced models can extract the global and local knowledge of the manufacturing process. The interesting topics can be robustness analysis, global and local sensitivity analysis.

REFERENCES

- [1] Brecher C (ed) (2012) Integrative production technology for high-wage countries, *Springer*, Berlin, Heidelberg
- [2] John K. Hunter, asymptotic analysis and singular perturbation: Department of mathematics, University of California at Davis, February, 2004
- [3] J.R. King and D.S. Riley. Asymptotic solutions to the Stefan problem with a constant heat source at the moving boundary. *The Royal Society of London*, 456(1997): 1163-1174, 2000
- [4] Georg Vossen, Torsten Hermanns and Jens Schüttler, Analysis and optimal control for free melt flow boundaries in laser cutting with distributed radiation, Wiley Online library, *ZAMMZ. Angew. Math. Mech.*, 1-20(2013)/DOI 10.1002/zamm.201200213
- [5] W. Schulz and J. Michel Freie Randwertaufgaben der thermischen Materialbearbeitung: Inertiale Mannigfaltigkeiten und Dimension des Phasenraums
- [6] W. Schulz. Die Dynamik des thermischen Abtrags mit Grenzschichtcharakter. PhD Thesis, Aachen: Shaker-Verlag, 2003. Habilitationsschrift, RWTH Aachen, 1998
- [7] E. Buckingham. On physically similar systems; illustration of the use of dimensional equations. *Physical Review* 4, 4(4): 345-376, 1914
- [8] W. Schulz, D. Becker, J. Franke, R. Kemmerling and G. Herziger. *J. Phys. D: Appl. Phys.* 26 (1993) 1357-1363. Printed in the UK. Heat conduction losses in laser cutting of metals
- [9] Huanrong Li, Zhendong Luo, Jing Chen, Numerical simulation based on POD for two-dimensional solute transport problems, *Applied Mathematical Modelling* 35 (2011) 2589-2498
- [10] L. Cordier and M. Bergmann. Proper orthogonal decomposition: an overview. Post processing of Experimental and numerical Data, *Lecture Series 2003/2004*, von Karman Institut for Fluid Dynamics, pages 1-45, 2003
- [11] Anindya Chatterjee, An introduction to the proper orthogonal decomposition, Special section: *computational science, current science*, vol. 78, No. 7, 10 April 2000
- [12] R. Reinhard, C. Büscher, T. Meisen, D. Schilberg and S. Jeschke. (2012) Virtual Production Intelligence – A Contribution to the Digital Factory. In: Hutchison D, Kanade T, Kittler J et al. (eds) *Intelligent Robotics and Applications*, vol 7506. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 706–715
- [13] H. Tercan, T. Al Khawli, U. Eppelt, C. Büscher, T. Meisen and S. Jeschke (2017). Improving the laser cutting process design by machine learning techniques. *Production Engineering*, 2(11), 195-203.