2020 ASME-CIE Hackathon: Identifying, Extracting, Analyzing of Value from Large Unstructured Data Sets in Mechanical Engineering

Hilton, St. Louis, MO, USA
August 15-16, 2020

In conjunction with
IDETC/CIE 2020

Sponsored by

ASME Computers & Information in Engineering Division (CIE) &
ASME Technical Events and Content (TEC) Sector Council

ASME Manufacturing Engineering Division (MED) Centennial Celebration Endorsed Event

Click to Register for the Hackathon

($25 for Hackathon event only)

Access to the Competition Sample Datasets HERE

Meeting Location: Zoom Links TBA

Important Dates:

- **August 14th, 2020:** Deadline for hackathon sign-up
- **August 15th, 3 pm 2020:** Hackathon Kick-off
- **August 16th, 4 pm 2020:** Due for Hackathon deliverables
- **August 16th, 8 pm 2020:** Awarding ceremony

Awards:

- **First Place:** $2,000
- **Second Place:** $1,000
- **Third Place:** $500

<u>Hackathon Problem 1</u>: Generating a Data-Driven Surrogate Model for Machine Damage Accumulation

<u>Subject Matter Expert</u>

- *Christopher McComb, Assistant Professor, School of Engineering Design, Technology and Professional Programs, PennState*
- *Zhenghui Sha, Assistant Professor, Department of Mechanical Engineering, University of Arkansas*
- *Faez Ahmed, Assistant Professor, Department of Mechanical Engineering, MIT*

<u>Problem Statement</u>

The Bernard M. Gordon Learning Factory is a hands-on facility for engineering students, which provides modern design, prototyping, and manufacturing facilities. Many of the machines in the Learning Factory are instrumented using a sensor suite that provides monitoring capabilities. The readings from a heterogeneous set of sensors are used to report metrics continuously for many different machines. These sensors record a variety of values every 10 minutes, such as temperature, velocity, and acceleration. The sensors also provide a computed damage accumulation measure, which is helpful for predictive maintenance.

The objective of this problem is to create a data-driven surrogate model that makes it possible to rapidly compute the damage accumulation values based on the other measurements. This would be useful to the Learning Factory as a digital twin, enabling them to assess future usage scenarios for the machines and calculate the damage accumulation associated with those scenarios.

<u>Implicit Challenges</u>

This data and use case presents several challenges. These include:
- Is it feasible to construct a data-driven digital twin for forecasting? (Kunath et al., 2018)
- What is the best approach to identifying appropriate signals in this data with which to make predictions? (Long et al., 2019)

<u>Datasets</u>

Data is provided for several different machines, including three Bridgeport mills, one drill press, and one lathe. Each machine has several sensors, and each of these sensors collects data such as peak velocity, RMS velocity, peak acceleration, and temperature. A damage accumulation value is also computed from these data. Data is logged approximately every 10 minutes. Specifically, these files are provided for each machine:

1. **[machine name]week1-train.csv**
   This file contains training data for one week, including both independent variables (velocity, acceleration, etc.) and dependent variables (damage accumulation)
2. **[machine name]week2-train.csv**
   Same format as above, but for a second week.
3. **[machine name]week3-test.csv**
   This file contains data that you will use to make predictions for submission and scoring. Specifically, it contains independent variables, but not dependent variables.
4. **[machine name]week3-submit.csv**
   You will use this file to submit your predictions. These files are explained in more detail under the Submission section below.

A partial example of the training data is provided below, and shows one of the independent variables (peak velocity) and one of the dependent variables (damage accumulation). Every machine will have 11 independent variables and 2 dependent variables (damage accumulation in two different modes).

Machines > Lathe 1 > MIB > Y-Vertical > Peak Velocity		Machines > Lathe 1 > MIB > Y-Vertical > Damage Accumulation	
Time (UTC)	Avg(in/sec)	Time (UTC)	Avg(Damage)
2/15/20 5:08	0.0032806	2/15/20 5:08	0.9889563
2/15/20 5:18	0.0031057	2/15/20 5:18	1.0187886
2/15/20 5:28	0.0035309	2/15/20 5:28	1.012189
2/15/20 5:38	0.0023349	2/15/20 5:38	1.0183065
2/15/20 5:49	0.0040037	2/15/20 5:49	0.9838173
2/15/20 5:59	0.0019277	2/15/20 5:59	0.9948952
2/15/20 6:09	0.0037558	2/15/20 6:09	0.9767354
2/15/20 6:18	0.0027029	2/15/20 6:18	1.0100354
...	...	...	...

Submission

You will submit one CSV file for each machine (a total of 5 training files). Templates are provided to you and are named with the format **[machine name]week3-submit.csv**. Do not edit the time values in column 1 or headers in row 1. You should fill the remainder of columns 2 and 3 with your predictions, based on inputs from **[machine name]week3-test.csv** and the time values provided in the submission file.

A partial example of a filled submission file is provided below.

Time (UTC)	Machines > Lathe 1 > MIB > X-Axial > Damage Accumulation	Machines > Lathe 1 > MIB > Y-Vertical > Damage Accumulation
2/29/20 5:06	0.86683465	0.91679842
2/29/20 5:16	0.33197709	0.3755712
2/29/20 5:26	0.07547059	0.12803265

2/29/20 5:36	0.06430619	0.04984671
2/29/20 5:46	0.41965689	0.36266437
2/29/20 5:56	0.57093488	0.18604864
2/29/20 6:07	0.38318072	0.95258113
2/29/20 6:17	0.17897555	0.48990568
2/29/20 6:26	0.18985331	0.25177106
...	...	...

Judgement Rubric

It is important to note that only 30% of your score will rely on the results from your algorithm, while the rest will be based on your approach, creativity, and presentation.

Category	Criteria	Scoring
Technical Approach (40%): Methods and algorithms of the proposed data analytics and visualization	• Requirement analysis and problem formulation • Literature review and exploration of ideas • The development and design of the idea • Readiness of the idea and the approach	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Creativity and innovation (20%): New direction in field to approach to the problem	• The technology breaks new ground • The project makes a profound break from established design • The project adds a major departure from established design • The code adds a new twist on established design • The chosen technology and design is already deeply established	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Results (30% Output performance and V&V	• Objective is successfully achieved, which is measured by the Mean Squired Error and the R-squared metric over testing set.	Team with the best performance (10 pts) Team with the second-best performance (7 pts) Team with the third-best performance (5 pts) Teams within top-five performance: 3 points Rest (1 pts).
Overall Presentation (10%): Organization, structure and message conveying	• Title, headings, labels: Appropriate size, location, spelling, and content • The demonstration of teamwork • Structure and Clarity • Boarder impact of the idea to ME subfields	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

Subject Matter Expert: Mentors

Christopher McComb, Assistant Professor, School of Engineering Design, Technology and Professional Programs, PennState

Zhenghui Sha, Assistant Professor, Department of Mechanical Engineering, University of Arkansas

Faez Ahmed, Postdoctoral Fellow, Department of Mechanical Engineering, Northwestern University

References:

1. Kunath, M., & Winkler, H. (2018). Integrating the Digital Twin of the manufacturing system into a decision support system for improving the order management process. *Procedia CIRP*, 72, 225-231.
2. Long, Wen, Zhichen Lu, and Lingxiao Cui. "Deep learning-based feature engineering for stock price movement prediction." *Knowledge-Based Systems* 164 (2019): 163-173.

Hackathon Problem 2: Smart Manufacturing – In-Process Data Mining for Powder-Bed Fusion Additive Manufacturing

Subject Matter Expert

- *Yan Lu, Senior Research Scientist, Professor, System Integration Division, NIST*
- *Dehao Liu, School of Mechanical Engineering, Graduate Research Assistant, George Institute of Technology*
- *Anh Tran, Research Staff, Sandia National Laboratories*

Problem Statement

Additive manufacturing (AM) processes build parts layer-by-layer directly from 3D models. AM enables the fabrication of complex heterogeneous parts, which makes it an attractive alternative for high-value, low-volume production. However, the turnkey deployment of the technology hits consistent barriers including low part repeatability and lack of effective qualification tools. Fundamental issues exist with the understanding and control of the dynamic and stochastic nature of AM processes. In-situ monitoring for additive manufacturing is considered as the main enablers to understand AM processes, set optimal material and machine specific process parameters, and close the control loops in real-time to limit the stochastic variability introduced by the dynamic nature of the processes. For manufacturers to build quality AM parts, in-situ data has the potential to be used for quality assurance and certification, which will dramatically reduce the need for lengthy and high-cost post inspections.

The goal of this hackathon subtopic is to promote the use of data science in powder-bed fusion additive manufacturing to accelerate the understanding of the powder-bed fusion AM process, to improve PBF process monitoring and control as well as to explore in-process data-based product qualification. This will be achieved by developing sets of data analytics tools, predictive models, and process control and optimization algorithms for PBF processes. This tool will be an early step in allowing the industry to move away from 100% testing and towards born-qualified parts.

Challenges

- AM in-process data registration – how to align the multi-modal in-process monitoring data in time and space to allow for fusion and correlation [1].
- What kind of features to be extracted from the multi-level, multi-scale AM in-process monitoring data, e.g., build command, chamber monitoring trended data, co-axial images, and layerwise images, etc.
- What kind of relationship between build commands and the in-process measurements such as coaxial melt pool characteristics/layerwise surface images. (Ref: Zhuo, 2019a, 2019b)
- How to fusion the data from the multi-modal in-process measurements (Ref: A Review of Data Fusion Techniques, The Scientific World Journal Volume 2013, Article ID 704504)
- How to develop real-time control or layerwise control strategy from the data for the PBF AM processes? (Ref: Mahesh Mani 2015, Measurement Science Needs for Real-time Control of Additive Manufacturing PowderBed Fusion Processes)

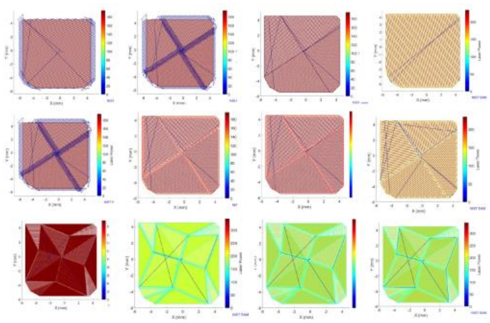
Datasets

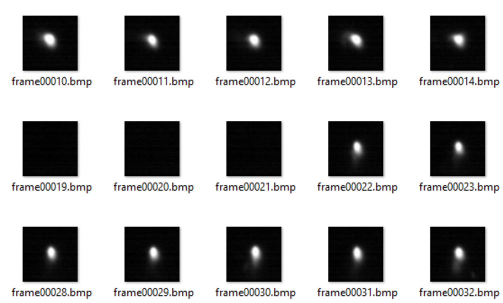
An experimental L-PBF build was conducted on the Additive Manufacturing Metrology Testbed (AMMT) at the National Institute of Standards and Technology (NIST). The AMMT is a fully customized metrology instrument that enables flexible control and measurement of the L-PBF process. Two cameras were installed for process monitoring, including a high-resolution camera that captures the layerwise images of the entire part, and a high-speed camera used to capture melt pool images. The Galvo mirror system and the beam splitter allow the high-speed camera to focus on the current laser melting spot. Emitted light from the melt pool, through an 850 nm bandpass filter (40 nm bandwidth), is imaged on the camera sensor. On AMMT both Galvo and laser command are updated on field-programmable gate array (FPGA) at 100 kHz. The digital commands are developed to specify the motion of the Galvo scanner of the L-PBF system. It is transformed into a time series of scanner positions and laser power as control commands.

Inconel 625 powder and build plate were used. The substrate has a dimension of 102 mm x 102 mm x 13 mm. Twelve rectangular parts (with chamfered corners) of dimensions 10 mm x 10 mm x 5 mm were laid on the substrate, with a minimum spacing of 10 mm between parts. Each part was built with a different scan strategy.

Data sets and data formats used for the study include:

1) Part Design model (STL file)
2) Part layout (drawing in pdf/part location in XML)
3) Process settings; camera settings; and camera calibration models (PNG, jpg, XML)
4) Data sets for build command at 100KHz for every part every layer (Xls)
5) Melt-pool images for every part, every layer at 2KHz (BMP/JPG/AVI/PNG)
6) Layerwise images, 2 per layer, before and after exposure settings (BMP/PNG)





Judgment Rubric

Category	Criteria	Scoring
Relevance to the AM Engineering Problems (20%) Which problem the developed data analytics to address?	• Identify the specific challenges the proposed methods and algorithms address • Provide a discussion of the impact the proposed data analytics methods • Discuss how the proposed methods can be transferred to AM production environment	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Technical Approach (40%) Methods and algorithms of the proposed data analytics and visualization	• Requirement analysis and problem formulation • Literature review and exploration of ideas • The development and design of the idea • Scientific soundness of the approach • The creativity of the approach • The readiness of the idea and the approach • Automated workflow: data/metadata acquisition through an open interface • Data visualization technique and quality: see Data Visualization Rubric	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Results (30%) Output performance and V&V	• The objective is successfully achieved. • Definitive conclusion with a well thought out reason or evidence backing it. • Quality: prediction accuracy, an improvement on the benchmark, and computational cost • Uncertainty quantification • Explainability • Verification and validation • Implementation discussions and Improvement directions	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Overall Presentation (10%): Organization, structure, and message conveying	• Title, headings, labels: Appropriate size, location, spelling, and content • The demonstration of teamwork • Structure and Clarity • Boarder impact of the idea to ME subfields	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

Subject Matter Expert: Mentors

Yan Lu, *Senior Research Scientist, Professor, System Integration Division, NIST*

Dehao Liu, School of Mechanical Engineering, Graduate Research Assistant, George Institute of Technology

Anh Tran, Postdoctoral Appointee, Sandia National Laboratories

Reference

[1] Vasileios Argyriou Jesús Martínez Del Rincón Barbara Villarini Alexis Roche, “Image, Video & 3d Data Registration”, John Wiley & Sons, 2015.
[2] Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A. (Eds.), Feature Extraction-Foundations and Applications, Springer, 2006.
[3] Yang, Z., Lu, Y., Yeung, H., and Krishnamurty, S., “From Scan Strategy to Melt Pool Prediction: A Neighboring-Effect Modeling Method”, CIE 2019.
[4] Yang, Z., Lu, Y., Yeung, H., and Krishnamurty, S., “Investigation of Deep Learning for Real-Time Melt Pool Classification in Additive Manufacturing.
[5] Federico Castanedo, “A Review of Data Fusion Techniques”, The Scientific World Journal Volume 2013.
[6] Mani, M., Lane, B. M., Donmez, A. M., Feng, S. C., Moylan, S. P., Fesperman, R. R., “Measurement Science Needs for Real-time Control of Additive Manufacturing PowderBed Fusion Processes”, NISTIR 8036, 2015.