

Industry Talk

Memo by Marina Sanchez Del Villar

New Frontiers in Machine-Data Privacy and Governance

Speaker: Lou Zhang (MachineMetrics)¹

This event has been organised by the Technological Change and Society Interdisciplinary Research Cluster

Lou Zhang is the Director of Data Science at MachineMetrics, a venture-capital funded start-up based in Massachusetts, USA. MachineMetrics collects and aggregates information coming from thousands of manufacturing machines. Their customers are big and small corporations, mostly based in the United States, but some are also abroad.

1. How does a company like MachineMetrics obtain the data?

The problem

In a typical machine shop, each machine has an operator assigned. The current setup lacks central visibility: only by constantly looking at a machine's control panel can one determine when a machine is not working. This is not an efficient procedure, as production shortages can go unnoticed until final production is tabulated.

The way to get the data

Every industrial machine has a closed-loop feedback system. The system consists of sensors embedded in the machine that help assess the resistance against a tool, which in turn determines how much power is needed to perform a task (like cutting, for example).

MachineMetrics taps into this feedback loop to get the data. For this, it places two devices in the electrical cabinet of the machine: first, an edge device, which is a light-weight computer that collects data directly from the control of the machine (like for example, the motor speed); and a second device that collects the electrical signals coming off the machine, normally used to monitor power consumption.

Additionally, MachineMetrics places tablets outside of each machine that allow the operators to enrich the data coming from the machine. For example, operators are usually prompted to

¹ The recording of this talk is available at: <https://www.youtube.com/watch?v=DvLrrYCGaHE>

signal the reasons for down-times. The involvement of operators helps MachineMetrics in labelling the information coming from the internal controls, an important preliminary step before feeding the data into any artificial intelligence algorithm.

MachineMetrics has five years of data warehoused, which allows them to see aggregate trends (like the dip and recovery following covid, or to quantify the lower utilization during bank holidays). The collection of data from different sources (clients) has complementarities.

MachineMetrics offers their client standard subscription contracts of around USD 100 per machine. The customers pay the subscription and provide the data, and in exchange they receive data analytics on their performance.

2. What information is available to the clients?

The raw data are available via Amazon Web Services (AWS) for the customer to download. Additionally, MachineMetrics provides basic monitoring information and advanced analytics.

The basic monitoring statistics are usually displayed in a dashboard. The dashboard contains the aggregated results of the real-time data collected from the machines. It helps to quickly identify down-times and monitor equipment status.

Advanced analytics allows to discern a machine's real-time conditions directly from the data coming from the control and feedback systems. An example of an advanced analytics application is their tool monitor.

In a nutshell, the power-signals emitted by a faulty motor-part are different from those emitted by a working motor-part. Their tool monitor is based on a model that distils each tool's conditions into a single number per part. The tool monitor allows the customer to evaluate the quality of new motors (like for example, power usage), and helps detect broken tools, predict the wear of a tool, and optimize on tool replacement.

MachineMetrics also provides a benchmarking service that shows manufacturers how their production ranks vis a vis other MachineMetrics clients. The data are available at a very disaggregated level, for each type of machine across all different customers.

3. Different data protection regimes in EU vs US and their implications

In the United States, MachineMetrics has the right to use a customer's data to improve the algorithms for all customers, but customers retain the right to request their data for download and for them to be deleted.

The introduction of GDPR in Europe has caused the industry to make changes on data storage. Before GDPR, the data for all MachineMetrics clients were stored together, while now they are stored in individual *buckets*. Additionally, the European Union customers can request that their data not be used to improve the algorithm for other customers.

4. System infrastructure and governance structure of BigTech

Before cloud services, companies had their own servers onsite, which was a considerable expense. AWS and their server farms give start-ups the ability to quickly scale up. Start-ups rent units of computing depending on their needs, so they use AWS's physical infrastructure to store data and run computations.

This practice implies that start-ups are outsourcing their data management to a third-party organization, exposing them to the outsider's data management practices. If such practices are unsecure, or the redundancies insufficient, start-ups risk losing data or customers.

Changes in BigTechs' data governance also affect start-ups. For example, Facebook recently changed its privacy laws, which implied that the advertising algorithm changed. The advertisers, many of which were start-ups, struggled to reach their usual customer base.

5. Interactions with other companies

The data MachineMetrics collects could be used for other applications: predicting more accurately supply conditions, building manufacturing indices, or building better machines by machine manufacturers. MachineMetrics does not consider the sale of these data externally for two reasons: first, it would not benefit their annual-recurring revenue, which is their KPI vis-à-vis their venture capitalists; and second, they have acquired an obligation with their customers to only use the data for the betterment of their algorithms.

Big machine manufacturers are not yet conducting this data collection in-house. They have good domain knowledge and good computer engineers, but often lack good data scientist, and this is where a company like MachineMetrics can be useful. Machines manufacturers are interested in partnering with MachineMetrics for their AI algorithms. MachineMetrics is in conversations with manufacturers to add their software capabilities to the machines, so that data collection can go directly via the machine software, as opposed to the current add-on.