Introduction to Data Science

Homework 01

0811033 黃勁瑋

2023/09/22

Homework Description

- Find at least one data set that you plan to study for your future homeworks and final project.
- Explain the features in your data set.
- Discuss possible problems you plan to investigate based on the data sets you select.

Part 1. Selected Dataset Basic Information

The "Bosch CNC Machining Dataset" from UCI Machine Learning Repository is selected for this homework [1]. Figure 1 is the screenshot of the introduction of this dataset on UCI Machine Learning Repository webpage. The data is stored on Github and can be accessed with the link "github.com/boschresearch/CNC_Machining" [2].

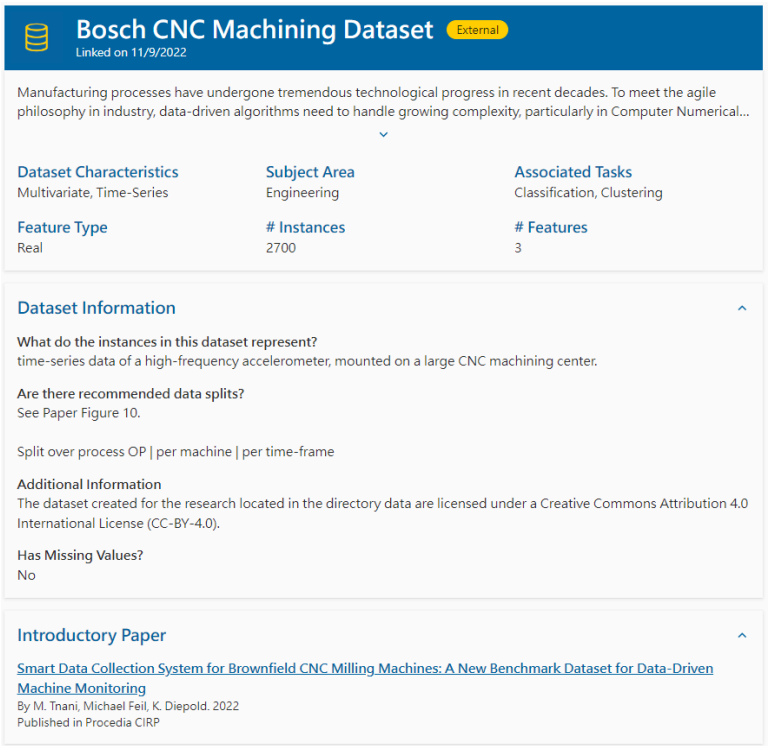

Fig. 1: Screenshot of the "Bosch CNC Machining Dataset" from UCI Machine Learning Repository website [1].

The dataset contains vibration data from a real production environment of 3 different CNC machines undergoing 14 different kinds of tool operation on aluminum within a certain time period. The vibration has direction on the x, y, and z axis. The raw data is collected by a Boch CISS sensor attached on the rear end of each CNC machine's spindle housing. Each set of vibrational data is labeled with OK or NOK depending on whether this set of data is recorded when the machine is operating finely or under abnormal situations. After the raw data was collected, the data is then processed by the procedures introduced in the paper "Smart data collection system for brownfield CNC milling machines: A new benchmark dataset for data-driven machine monitoring" [3] to generate usable and machine learning (ML) training friendly data which we see on the Github page. The data provided is in .h5 format, which requires the python packages h5py to process. There is a script on Github demoing how to load the data [2].

Part 2. Importances behind the Selected Dataset

CNC machining is a manufacturing process which is applicable for a wide range of different materials and shapes. When failures occur in CNC machining, the part manufactured during the current run cannot meet the design spec and has to be reworked or discarded. Failures may generate extra cost for a factory. The cause of failure can be affected by many parameters such as tool wear, chatter, improper tool clamping, chip jamming, geometries, materials and many other physical conditions. According to ChatGPT, "failure prediction in CNC (Computer Numerical Control) machining is a crucial step in improving manufacturing efficiency, reducing downtime, and enhancing product quality. By implementing advanced monitoring and predictive maintenance techniques, you can minimize unexpected machine failures and optimize production processes [4]."

Traditionally, failure prediction in CNC factories is done by experience and try-and-error methods. Sometimes there is even no failure prediction action for some factories since they simply use their cutting tools until they break on the CNC machine and control part quality fully by after-machining inspection. The reason behind this is because there are too many parameters that could affect failures. A better solution of failure prediction in CNC machining may be using data driven machining learning methods.

The "Bosch CNC Machining Dataset" provides comprehensive data for solving the failure prediction problem in CNC machining. The data range is wide and not limited to a small number of machines and tools, which indicates the final model trained by these data could have better scalability . According to the dataset owner, the goal of this dataset is to

"enhance the scalability of machine learning in real-world applications by presenting a dataset containing the main challenges that hinder the reliability of ML algorithms in the manufacturing environment [2]." A machine learning model used for predicting CNC machining failure beforehand can be trained based on this selected dataset.

References

[1] "Bosch CNC Machining Dataset," *UCI Machine Learning Repository*, https://archive.ics.uci.edu/dataset/752/bosch+cnc+machining+dataset (accessed Sep. 21, 2023).

[2] MATnani. "Bosch Research/CNC_MACHINING: Data Set for Process Monitoring on CNC Machines." *GitHub*, https://github.com/boschresearch/CNC_Machining. Accessed 21 Sept. 2023.

[3] TNANI, Mohamed-Ali; FEIL, Michael; DIEPOLD, Klaus. Smart data collection system for brownfield CNC milling machines: A new benchmark dataset for data-driven machine monitoring. Procedia CIRP, 2022, 107: 131-136.

[4] ChatGPT, "Introduction to Failure Prediction in CNC Machining," OpenAI. [Online]. Available: [Insert the URL of the page where you found this response]. Accessed: September 21, 2023.

[5] BLACK, J. Temple; KOHSER, Ronald A. DeGarmo's materials and processes in manufacturing. John Wiley & Sons, 2017.