Proposal

Xiaoyu Liu (xliu969); Tianhang Li (tli425); Hongyi Liu (hliu557); Xiaofeng Wang (xwang2443); Rui Huang (rhuang95)

Goal

In this project, we plan to analyze the data set of cutting tools. According to the variables describing signals, vibration and location, this project aims at predicting the life span of a cutting tool. Moreover, we will predict the remaining life span of a cutting tool based on the value of variables at a certain time.

Data

The data set is about the remaining life length of cutting tools. (URL: https://iaiinstitute.com/competitions/17/eventData) Several kinds of signals are collected while the cutting tools are using. In particular, the data are separated into two parts, training data and testing data. The training data is from 3 run-to-failure cutting tools, and their collection starts from the normal machining process. The testing data are from 5 cutting tools. In the training data, the signals from the controller (PLC) and add-on vibration and current sensors are collected. The PLC signal sampling frequency is 33 Hz, and the vibration sensor sampling frequency is 25600 Hz. In the testing data, time duration of experiments are the same but start points on each cutting tools are different. The unit of the timestamp is minutes. Due to a large amount of sensor data, only 1-minute data out of every 5 minutes are provided. Data are provided in a sequence: 1.csv, 2.csv.... n.csv. The last file is the data collected from the last 5 minutes prior to the end of the experiment.

Var Name	Description	Data Type
time	time	hh:mm:ss:ms
spindle_load	spindle load of cutters	float
x	machine coordinate in X	float
y	machine coordinate in Y	float
z	machine coordinate in Z	float
csv_no	corresponding sensor file	string
vibration_1	vibration signal in X	float
vibration_2	vibration signal in Z	float
vibration_3	vibration signal in Y	float
current	current signal of the first phase	float

```
plc=read.csv("plc.csv",nrows=3)
print(plc)
```

```
##          time spindle_load        x       y        z csv_no
## 1 13:45:00:13     0.064089 -636.5643 46.7288 -380.029      1
## 2 13:45:00:29     0.064089 -636.5643 46.7288 -380.029      1
## 3 13:45:00:44     0.064089 -636.5643 46.7288 -380.029      1
```

```
sensor=read.csv("1.csv",nrows=3)
print(sensor)
```

```
##   vibration_1 vibration_2 vibration_3  current
## 1  0.07457337 -0.06353944  0.11931301 49.77427
```

```
## 2  0.05529791  0.05562192  0.03636705 50.66998
## 3  0.03687642 -0.03255993 -0.04572380 50.62143
```

Statistical Methods

Firstly, extract characters in time series data. Use AR, MA, or ARIMA model to transform the time series of each variable in each time window. Then calculate the important statistics of it such as range, maximum and so on. Then use xgboost, linear regression, decision tree models to predict.

Plan and Computational Tools

Firstly, we will separate the data set according to the time window it belongs to. Then we will use parallel computation to extract the information in each variable in time series data.

After extracting the characters in parallel jobs, we will aggregate them to a complete data set and run models on it to predict the final result.