Decoating Performance Evaluation for Aluminium Manufacturing

Problem statement

Aluminium decoating consists in removing any kind of coatings, paints, lacquers as well as other organic contaminants and humidity from the scrap, by continuous thermal pyrolysis in order to obtain, together with the pre-treatment, clean and hot shreds, ready to be melt with the highest yield. Besides, such decoating process benefits include lower emissions, improved control of the remelting process, reduced dross generation and greater melt cleanliness.

Tilting Rotary Furnaces model, with an oxy-fuel combustion system has been widely used in Aluminium manufacturing for so many years. The system delivers high quality of metal and minimumize maintenance effort. However, there is still quality issue that can be improved with predictive maintenance. Increasing maintenance frequency is helpful, but leading to more downtime and maintenance cost.

The goal of this project is determining maintenance time based on the model predictions. Let's assume we will do maintence after each bad cycle. If we accurately predict the bad cycle, 30 minutes downtime will be good enough to get system back to normal and the cost is 16,000 USD. If the bad cycle was not captured, the maintenance will take 1 hour downtime and loss 32,560 USD. Also, the extra labor cost will be 300 USD/hr. Low quality metals sent to the downstream remelting process can result in much higher black dross level. Proper disposal for each kg of black dross costs at least $60. Poor metal quality resulted in 50kg more black dross than good metal. Therefore, timely alarms of poor quality decoating cycle are critical and potentially avoid business losses.

Data

There are three tables.

a) Sensor.csv - Sensing data from OSIPi, which is time series data including 27 columns. The columns starting with "B_" are from sensors. Cycle ID is unique for each cycle during the current period. Column "Good/Bad" is the label showing this cycle is good or bad. Column timestamp is the cycle finish time. The sampling rate is not constant because OSIpi system archived small changing data point, only "big changing" points will be record in the database.

b) Sensor_high*freq.csv - Sensing data from controller, which has higher frequency with non constant sampling rate. It is also time series data including 8 columns. The columns staring with "B*" are from controllers. Column "Percent" represent the carrier occupation percentage. The carrier will pass the recycled metal through the decoater. For different percentages, different mount of fuel, airflow, etc. could be different. You will find two columns named "Percent_min" and "Percent_max" in "Percent_reference.csv". If Percent_min < Percent < Percent_max, the corresponding combustion related values should be applied.

c) Percent_reference.csv - Combustion related data which provides recipes for different carrier percentages. This data is not time series data.

Notes

a) With these tables, you can build a classification model to predict good or bad cycle.

b) The timestamps are in the same time zone.

c) The model performance will be validated by a holdout dataset which is not provided in this folder. Therefore, we are expecting the delivered solution should not require running through the notebook from the beginning to the end. Please use the following template to organize your functions.

d) REST API is encouraged but not required.

Python Enviroment with Package Versions

Please use a Python 3.7+ environment with the following packages: Pandas==1.2.4 scikit-learn==0.24.2 numpy==1.17.4 matplotlib==3.1.1

In [ ]:

```
class DataPreprocessing(data):
    def __init__(self, columns = None):
        self.columns = columns

    def outlierRemoval(self, threshold = 0.05):
        return output

    def catergoricalEncoding(self, how='one hot'):
        return output
```