Red Shirt Dev Tour

IoT Edge – AI – Lab – 2nd Series

Vincent THAVONEKHAM
Igor LEONTIEV
Taras CHIEN
Artem SHEIKO

Paris 2018

Contents

About

Learn how to use Azure IoT Edge to create a predictive maintenance model from imbalanced dataset using Machine Leaning supervised learning models such as Support Vector Machine and Gradient Boosting Classifiers, to develop modules to monitor information from many drone's sensors, create some features using C# and use obtained dataset to predict a drone breakdown in this series of Hands-On Labs.

Introduction

Predictive maintenance is very broad as a matter of fact it applies to many areas of manufacturing. That's why it is highly related to the new wave around Internet of Things. When the talk is about Internet of Things, it's not enough to just collect all of data from machines and store it safely. The real value is in what proprietor does with that data, in what degree he can use that data to learn new things about these machines.

The reason predictive maintenance is so important, especially to manufacturers, is that it will reduce the operational costs significantly. Today it costs a typical manufacturer (type SME) literally billions of dollars to maintain their machines. Manufacturer has to manage large fleet of support engineers who go out and monitor those things on a regular basis (human cost, fuel, the process of finding a replacement for broken parts, installation of updates, ...). But, typically, even in that case, it's hard to predict when machine can break. Even with a good regular maintenance cadence the machine can break anytime in between that leads to a loss of time and money. Furthermore, in some cases if manufacturer does not respond quickly, he can lose customers.

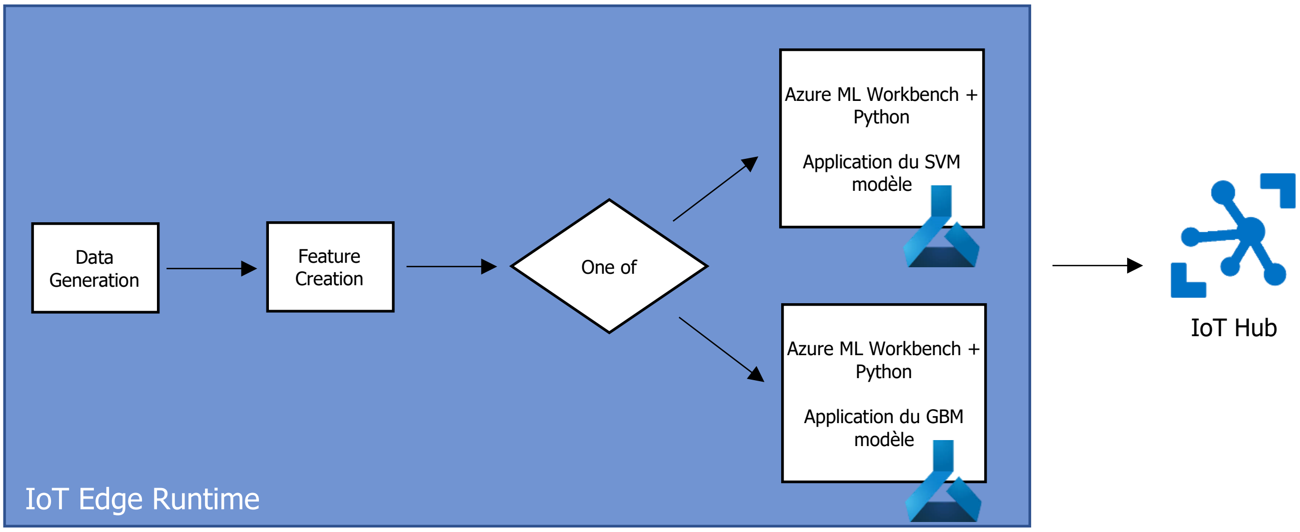
That's why predictive maintenance is a necessary and a very effective tool to stream line the maintenance processes and reduce costs. Key point of a scenario that avoids a situation with an unexpected failure is to collect data, every few minutes or every hour from machines and run it through a Machine Learning model. Model will predict the probability of a machine breakdown and therefore the manufacturer will be able to turn off the machine the next day or two, repair it, reallocate production capacities and ultimately delight the client in place of upset him. If manufacturer does it right and if he uses the right predictive models, he could prevent customer dissatisfaction significantly. So, manufacturer will deal with the problem even before it happens.

Azure IoT Edge is an Internet of Things (IoT) service that is based on top of IoT Hub. This service helps to evaluate significant parameters of the degradation of a machine and to act on the faulty element as close as possible. This also makes it possible to identify a degradation in the case of a variable lifetime of a device.

In this Hands-On Lab series, we will use drone as device to monitor. Drones are increasingly capable but also, increasingly complex. Different systems on the aircraft generate large amounts of data. Pilots are also required to be -aware of their flight environment, and, at the same time, focused on flying the aircraft. To understand the health of a drone and ensure safe flights, advanced analysis of a wide variety of data is required.

Solution

The architecture of the solution is presented below.



I. Data Understanding

Sensors indicators are presented in the table 1.1. These indicators values may indicate an issue with the drone, such as blade, motor or battery issues. Some of the following values can be affected by multiple factors including pilot actions (flying style, altitude changes etc.) and ambient conditions (humidity, temperature).

Table 1.1.

Measurable indicators

X	Drone's physical orientation in space
Y	
Z	
Heading	Heading drone degree. The initial value can be set for the ground station.
Start Distance, m	Distance from take-off point
Flight Altitude, m	Distance to the ground
Minutes per battery	How many minutes the drone can fly until the battery drains.
Battery, %	Percent charged
Battery Voltage	Difference in electric potential between the positive and negative terminals of a battery
Cell 1 voltage	Difference in electric potential between the positive and negative terminals of a cell
Cell 2 voltage	
Cell 3 voltage	
Cell 4 voltage	
Km per battery	How many kilometres expected the drone can fly until the battery drains.

Signal Strength, milliwatts	Signal strength based on the connection to the remote
Flight time, seconds	Flight duration
Response Time, seconds	How quick the drone responds to the remote commands
Ambient humidity, %	Ambient drone environment humidity
Ambient temperature, °C	Ambient drone environment temperature
Ground station latitude	Ground Station physical orientation coordinates
Ground station longitude	
Timestamp	Timestamp, when data where obtained

Note that even a perfect battery would have minor deviations and it is normal. Lower values are better. Higher values may provide an early sign that the battery is not as efficient.

Fewer turn records provides less accurate model quality, that's why to provide a dataset to create a Machine Learning model a dataset of 28800 records was collected.

II. Data Preparation and Feature Creation

Features to create are shown in the Table 1.2

Table 1.2.

Indicators to calculate

Pitch	Pitch degree
Yaw	Yaw degree
Roll	Roll degree
Latitude	Drone's coordinates
Longitude	
Is minor Deviation Cell 1, boolean	Minor deviation is when a cell differs more than 0.05 V from the other cells. The total amount of deviations per cell is then divided by the total amount of flight minutes to get the number of minor deviations per minute.
Is minor Deviation Cell 2, boolean	
Is minor Deviation Cell 3, boolean	
Is minor Deviation Cell 4, boolean	
Is major deviation Cell 1, boolean	Major deviation is when a cell differs more than 0.15 V from other cells. An abnormal battery will: 1. Have most of the major deviations in one cell 2. There will be multiple major deviations per minute 3. The deviations continue longer than some period (30 seconds, 2 minutes etc.) If a battery shows such symptoms, then this may impact the battery life.
Is major deviation Cell 2, boolean	
Is major deviation Cell 3, boolean	
Is major deviation Cell 4, boolean	

III. Modelling and Evaluating

A. SMOTE oversampling for an imbalanced dataset & Support Vector Machine Classification¶

1. Key notes

Dataset description

Initial dataset is historical data about the previous drone flights. There are 28800 records and 27 features. 1920 records are marked as malfunctions. This makes an imbalanced dataset with a 15:1 ratio of success against unsuccessful flights. The flight result has a simple unproblematic / fail status, that is encoded by -1 and 1 respectivement.

Objective

There are many different approaches to classify imbalanced data. One of the strategies is to rebalance the dataset by oversampling the minority class or / and undersampling the majority class. This is done to improve the true positive rate of the minority class. For this lab, we will look at rebalancing the dataset using Synthetic Minority Over-sampling Technique (SMOTE), which oversamples the minority class, and at SVM classification to predict the state of flight.

Methodology

The sklearn has many methods for oversamplng / undersampling. We will use the SMOTE method introduced in 2002 [1]. With SMOTE, synthetic examples are interpolated along the line segments joining some / all of the k minority class nearest neighbors. In the experiment, the oversampling rate is varied between 10-70% with 10% increments. The percentage represents the final minority class fraction after oversampling: if the majority class has 1000 data points and the minority class 50, at 10% the minority class will have 100 data points after oversampling. After SMOTE application, the rebalanced data is classified using a SVM.

2. Pre-processing

The data represents sensors measurements. Some of the records are missing, some measurements are identical and so not useful for prediction. We will remove those columns with high missing count or constant values.

The Random Forest variable importance is used to rank the features in terms of their importance. For the random forest, we will impute the remaining missing values with the median for the column. We will additionally scale the data that is applied to the SVM. We will use the sklearn preprocessing module for both imputing and scaling.

[1] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Oversampling TEchnique. Journal of Artijcial Intelligence Research, 16:321-357

3. SVM Classification

The SVM is sensitive to feature scale so the first step is to center and normalize the data. The train and test sets are scaled separately using the mean and variance computed from the training data. This is done to estimate the ability of the model to generalize.

4. Results & Discussion

5. References

B. Using the Gradient Boosting Classifier for an imbalanced data set

1. Key notes

2. Pre-processing

3. SVM Classification

4. Discussion

5. References

VI. Deployment