Minería de Datos

Master en Ciencia de Datos e Ingeniería de Computadores

Universidad de Granada

Preprocesamiento y Clasificación

Autores:

Brian Sena Simons.
Miguel Garcia Lopez.
Álvaro Santana Sánchez.
Ana Fuentes Rodríguez.

Grupo:

Data Mavericks.

Índice

1. Introducción.

Se ha realizado un análisis y comparativa entre diferentes modelos para la detección de anomalías y predicción de vida útil restante (RUL por sus siglas en inglés) en compresores del sector ferroviario. Para ello, se ha utilizado el conjunto de datos (dataset) "MetroPT-3" [1]. Está publicado en "UCI Machine Learning Repository" [2] y, según la descripción, MetroPT-3 [1] es un conjunto de datos multivariantes de series temporales. Los datos provienen de sensores analógicos y digitales instalados en un compresor de tren, que miden 15 señales como presiones, corriente del motor, temperatura del aceite y señales eléctricas de las válvulas de entrada de aire. La información fue registrada a una frecuencia de 1 Hz entre febrero y agosto de 2020 (véase Tabla 1)

Variable	**Tipo**	**Mín.**	**Q1**	**Q2**	**Media**	**Q3**	**Máx.**
TP2	Numérico	-0.032	-0.014	-0.012	1.368	-0.010	10.676
TP3	Numérico	0.730	8.492	8.960	8.985	9.492	10.302
H1	Numérico	-0.036	8.254	8.784	7.568	9.374	10.288
DV_pressure	Numérico	-0.032	-0.022	-0.020	0.05596	-0.018	9.844
Reservoirs	Numérico	0.712	8.494	8.960	8.985	9.492	10.300
Oil_temperature	Numérico	15.40	57.77	62.70	62.64	67.25	89.05
Motor_current	Numérico	0.020	0.040	0.045	2.050	3.808	9.295
COMP	Numérico	0.000	1.000	1.000	0.837	1.000	1.000
DV_eletric	Numérico	0.000	0.000	0.000	0.1606	0.000	1.000
Towers	Numérico	0.000	1.000	1.000	0.9198	1.000	1.000
MPG	Numérico	0.000	1.000	1.000	0.8327	1.000	1.000
LPS	Numérico	0.000	0.000	0.000	0.00342	0.000	1.000
Pressure_switch	Numérico	0.000	1.000	1.000	0.9914	1.000	1.000
Oil_level	Numérico	0.000	1.000	1.000	0.9042	1.000	1.000
Caudal_impulses	Numérico	0.000	1.000	1.000	0.9371	1.000	1.000

Tabla 1: Información básica de los diferentes tipos de datos presentes en MetroPT-3 [1]

Este conjunto de datos tiene como objetivo principal mejorar la detección de fallos y la predicción de mantenimiento. Aunque no contiene etiquetas directas, se dispone de informes de fallos que permiten evaluar la efectividad de los algoritmos de detección de anomalías, predicción de fallos y estimación de RUL (véase la Tabla 2).

Además, se recomienda utilizar el primer mes de datos para entrenar modelos, dejando el resto para las pruebas, permitiendo también la formación incremental si fuera necesario.

Número	Inicio	Fin	Duración	Importancia
1	4/12/2020 11:50	4/12/2020 23:30	700	Alta
2	4/18/2020 00:00	4/18/2020 23:59	1440	Alta
3	4/19/2020 00:00	4/19/2020 01:30	90	Alta
4	4/29/2020 03:20	4/29/2020 04:00	40	Alta
5	4/29/2020 22:00	4/29/2020 22:20	20	Alta
6	5/13/2020 14:00	5/13/2020 23:59	599	Alta
7	5/18/2020 05:00	5/18/2020 05:30	30	Alta
8	5/19/2020 10:10	5/19/2020 11:00	50	Alta
9	5/19/2020 22:10	5/19/2020 23:59	109	Alta
10	5/20/2020 00:00	5/20/2020 20:00	1200	Alta
11	5/23/2020 09:50	5/23/2020 10:10	20	Alta
12	5/29/2020 23:30	5/29/2020 23:59	29	Alta
13	5/30/2020 00:00	5/30/2020 06:00	360	Alta
14	6/01/2020 15:00	6/01/2020 15:40	40	Alta
15	6/03/2020 10:00	6/03/2020 11:00	60	Alta
16	6/05/2020 10:00	6/05/2020 23:59	839	Alta
17	6/06/2020 00:00	6/06/2020 23:59	1439	Alta
18	6/07/2020 00:00	6/07/2020 14:30	870	Alta
19	7/08/2020 17:30	7/08/2020 19:00	90	Alta
20	7/15/2020 14:30	7/15/2020 19:00	270	Media
21	7/17/2020 04:30	7/17/2020 05:30	60	Alta

Tabla 2: Intervalos de tiempo con problemas en la compresión del aire. Nos permite evaluar la capacidad de detección anomalías de nuestros modelo.

2. Análisis Exploratorio de Datos.

3. Regresión Logística.

4. Máquinas de Vectores de Soporte.

5. Clasificador Bayesiano.

6. Árboles de clasificación.

7. Gradient Boosting.

8. Stacking.

9. AdaBoost.

10. Bagging.

Referencias

[1] Narjes Davari et al. *MetroPT-3 Dataset*. https://doi.org/10.24432/C5VW3R. UCI Machine Learning Repository. 2021.

[2] Markelle Kelly, Rachel Longjohn y Kolby Nottingham. *The UCI Machine Learning Repository*. Último acceso el 11/01/2025. 2024. URL: https://archive.ics.uci.edu.