From Predictive Maintenance to Machine Learning

E. Rachelson

isae SUPAERO

Schedule

1. General introduction
   Now.
2. Case study
   This morning.
3. Presentations and discussion
   Beginning of the afternoon.
4. The Data Scientist point of view
   The rest of the afternoon.

Course goals

By the end of the class, you should be able to:

- explain the workflow of data analysis for Predictive Maintenance problems;
- know the main bottlenecks and challenges of data-driven approaches to Maintenance;
- link the Predictive Maintenance problems to their formal Machine Learning counterparts;
- know the main categories of Machine Learning algorithms and which formal problem they solve;
- know the name of some key methods in Machine Learning;
- know the existence of scikit-learn and its API.

Course material

https://github.com/erachelson/PredMaintenanceClass

Case study

In small groups. Several cases of failure analysis.

- Turbo-fan engine
- Air conditioning systems (HVAC)
- Truck air compressor
- Pneumatic valve

Your task:

- Prepare a synthesis of your case study (tell the story!).
- Highlight in particular:
  - nature of data (scalar, booleans, time series, images, text. . .);
  - properties of data (volume, cleanliness, dimensionnality. . .);
  - nature of the automated task (visualisation, anomaly detection, RUL prediction. . .);
  - name of the Machine Learning (and related) methods used;
  - open challenges, bottlenecks.

Presentations and discussion

Along the presentations, let's fill the table below, to build a common understanding of:

- the nature of predictive maintenance data
- the different tasks to automate
- the difficulties

Use case	Type of data	Properties of data	Task to automate	Difficulties	Comments

The Data Scientist perspective

Identified needs

We would like to build automated tools for the following tasks:

- Visualize system state
- Identify anomalies
- Predict Remaining Useful Life (RUL) / Time To Failure (TTF)
- Predict failure occurrence or probability at a given horizon

All this, in order to base our maintenance strategy on the (inferred) system state, rather than a general statistical trend.

Traditionaly, all this is based on user expertise.
Let's take a data-driven approach.

Data analysis workflow for Predictive Maintenance

1. Collect

- Sensors deployment
- Historical data collection
- Integrated storage and retrieval issues

$\rightarrow$ Extract-Transform-Load (ETL) process

Data analysis workflow for Predictive Maintenance

1. Collect
2. Analyze

- data cleaning
- feature selection / engineering
- algorithm selection
- parameters tuning

Data analysis workflow for Predictive Maintenance

1. Collect
2. Analyze
3. Predict

- Make predictions on new test cases
- Deploy solution in your operational process
- Make things usable

Data analysis workflow for Predictive Maintenance

1. Collect
2. Analyze
3. Predict
4. React

- Improve your maintenance decisions

Data analysis workflow for Predictive Maintenance

1. Collect
2. Analyze
3. Predict
4. React

Need to automate as many steps as possible in this workflow

$\rightarrow$ data-driven approaches

$\rightarrow$ Machine Learning for step 2 (and 3)

A word on data quality

- amount of data: data is often abundant but crucial data is often scarce
- noise, errors, missing data, outdated data: reliability
- high-dimensional data
- class imbalance
- heterogeneous data (scalars, booleans, time series, images, text, . . .)

All these will influence your algorithmic design or choices.

So let's talk about algorithms to see how we can solve the PM problems listed earlier.

Machines that learn?

Let's try to give a general definition.

Machines that learn?
Let's try to give a general definition.

Machine learning is a field of computer science that gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed.

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?

Image sources: Wikimedia commons

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- What price for this stock, 6 months from now?

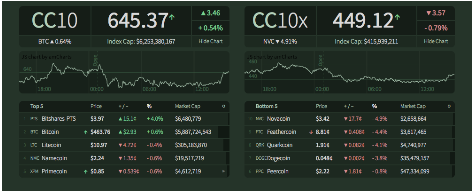


Image sources: Wikimedia commons

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- What price for this stock, 6 months from now?
- Is this handwritten number a 7?

Image sources: Seven.jpg

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- What price for this stock, 6 months from now?
- Is this handwritten number a 7?
- Is this e-mail a spam?

Enlarge your thesis!

Image sources: Iconfinder

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- What price for this stock, 6 months from now?
- Is this handwritten number a 7?
- Is this e-mail a spam?
- Can I cluster together customers? press articles? genes?

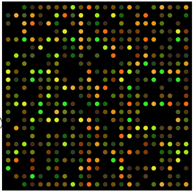



Image sources: People.jpg / Writing to Discuss: Use of a Clustering Technique / DNA microarray

ML examples

- Given 20 years of clinical data, will this patient have a second heart attack in the next 5 years?
- What price for this stock, 6 months from now?
- Is this handwritten number a 7?
- Is this e-mail a spam?
- Can I cluster together customers? press articles? genes?
- What is the best strategy when playing Counter Strike? or poker?

Image sources: CS:source / poker

ML tasks

What does ML do? 3 main tasks.

Task	Supervized Learning	Unsupervized Learning	Reinforcement Learning
Goal	Learn a function, $f(x) = y$	Find groups and correlations, $x \in C$	Optimal control, $f(x) = u \ / \ \max \sum r$
Data	$\{(x, y)\}$	$\{x\}$	$\{(x, u, r, x')\}$
Sub-task	Classification, Regression	Clustering, Density estimation, Dimensionnality reduction	Value estimation, Policy optimization
Algo ex.	Neural Networks, SVM, Random Forests	k-means, PCA, HCA	Q-learning

Evaluation criteria

Evaluating ML methods? What do we really want?

Ability to fit the training data:

- Regression: Mean Square Error
- Classification: Accuracy, TP, FP, ROC, AUC...
  cf. this Wikipedia article
- Clustering: similarity scores

Ability to generalize:

- Goal: filter out noise, avoid overfitting, generalize to unseen cases.
- ML Notions:
  - maximize margin
  - minimize difference btw class distributions (cross-entropy)

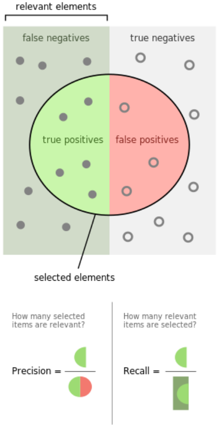


Image source: Wikimedia commons

Misconceptions and clarifications

AI: ML is only a small (currently fashionable) part of Artificial Intelligence.

BD: Big Data refers to working with datasets that have large Volume, Variety, Velocity (, Veracity, and Value).

DL: Deep Learning is Machine Learning with Deep Neural Networks.

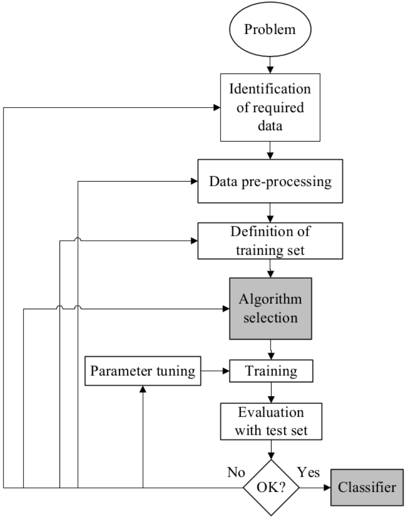
threat: ML / Data Science / Big Data are as much of a threat (to jobs, the society, the economy...) as the combustion engine was in the XIXth century.

ML software

Software:

- Many free libraries: scikit-learn, tensorflow, caffe. . . check `www.mloss.org` if you're curious.
- Free environments: Weka, RStudio. . .
- Commercial embedded solutions (more or less specialized): Matlab, IBM, emaint, Microsoft. . .

The process of (Un)Supervized Learning



From **Supervized Machine Learning: A Review of Classification Techniques**, S. B. Kotsiantis, *Informatica*, 31:249–268, 2007.

Relating PM and ML

- Visualizing system state
  $\rightarrow$ Dimensionnality reduction (Unsupervized learning)
- Detecting anomalies
  $\rightarrow$ Density estimation (Unsupervized learning)
- Predicting RUL or TTF
  $\rightarrow$ Regression (Supervized learning)
- Predicting failure in N cycles
  $\rightarrow$ Classification (Supervized learning)

Relating PM and ML

- Visualizing system state
  $\rightarrow$ Dimensionnality reduction (Unsupervized learning)
- Detecting anomalies
  $\rightarrow$ Density estimation (Unsupervized learning)
- Predicting RUL or TTF
  $\rightarrow$ Regression (Supervized learning)
- Predicting failure in N cycles
  $\rightarrow$ Classification (Supervized learning)

Thinking like a Maintenance Engineer:
How can I monitor my system to manage my maintenance operations?
Thinking like a Data Scientist:
Is this a supervized or unsupervized problem? What available data?

Now you can start discussing with data scientists to design together the most appropriate method for your data and your problem.

A word on scikit-learn

Scikit-learn = Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license
- Well documented, with lots of examples

`http://scikit-learn.org`

Let's take a look at the documentation's table of contents to grasp a few more keywords.

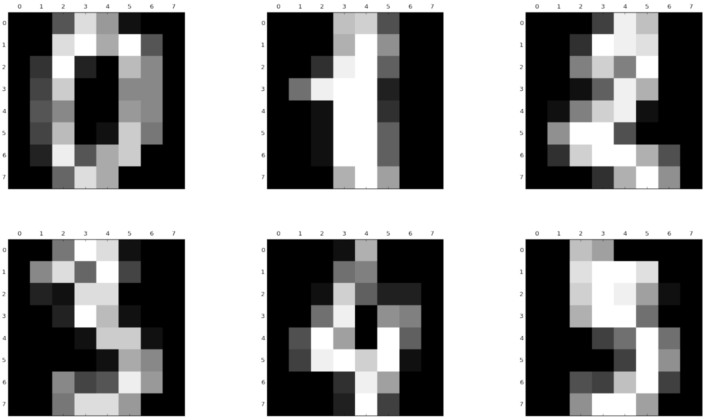
Course goals

By the end of the class, you should be able to:

- explain the workflow of data analysis for Predictive Maintenance problems;
- know the main bottlenecks and challenges of data-driven approaches to Maintenance;
- link the Predictive Maintenance problems to their formal Machine Learning counterparts;
- know the main categories of Machine Learning algorithms and which formal problem they solve;
- know the name of some key methods in Machine Learning;
- know the existence of scikit-learn and its API.

Let's finish with a focus on a practical use-case

The "digits" dataset.



Voluntarily not directly a maintenance case, because:

- the data you're likely to encounter might often surprise you;
- it's easy to visualize things with this example.