Synergetic transformations in model based and data based diagnosis

Louise Travé-Massuyès

Nathalie Barbosa Roa

Data scientist and Big Data engineer
Continental Automotive France SAS
https://www.linkedin.com/in/nathaliebarbosaroa/
Fault detection, data-based, time series, manufacturing, machine learning

Xavier Pucel

Research engineer
ONERA / DTIS, University of Toulouse
https://sites.google.com/view/xavier-pucel/accueil
Autonomous robots, diagnosis, decision, verification and validation, planning and scheduling

Louise Travé-Massuyès

CNRS Research Director
LAAS-CNRS, University of Toulouse
http://homepages.laas.fr/louise

Diagnosis theories, Model-Based Diagnosis, Data-Based Diagnosis, Machine Learning, Monitoring and Health Management, Diagnosability, Sensor placement, Diagnosis architectures, Qualitative models and qualitative reasoning formalisms

Elodie Chanthery

Assistant Professor
LAAS-CNRS, INSA, University of Toulouse
https://homepages.laas.fr/echanthe/
Diagnosis, prognosis, hybrid systems, autonomous systems, distributed systems

+ DISCO team (LAAS-CNRS)

Predictive maintenance
& Monitoring and surveillance
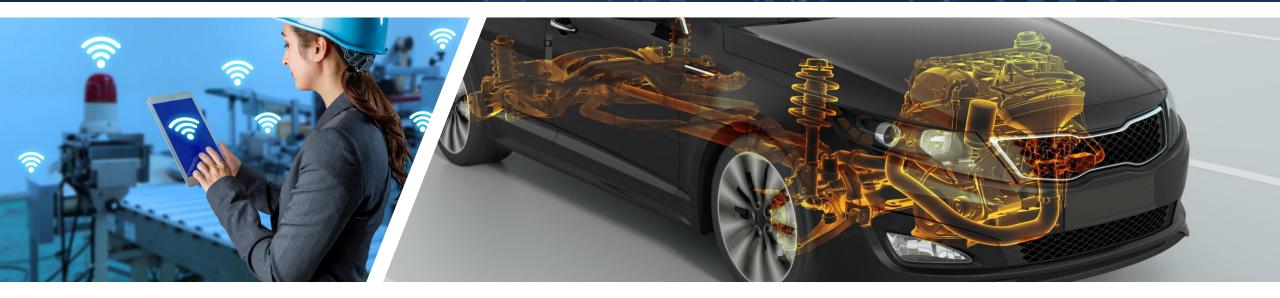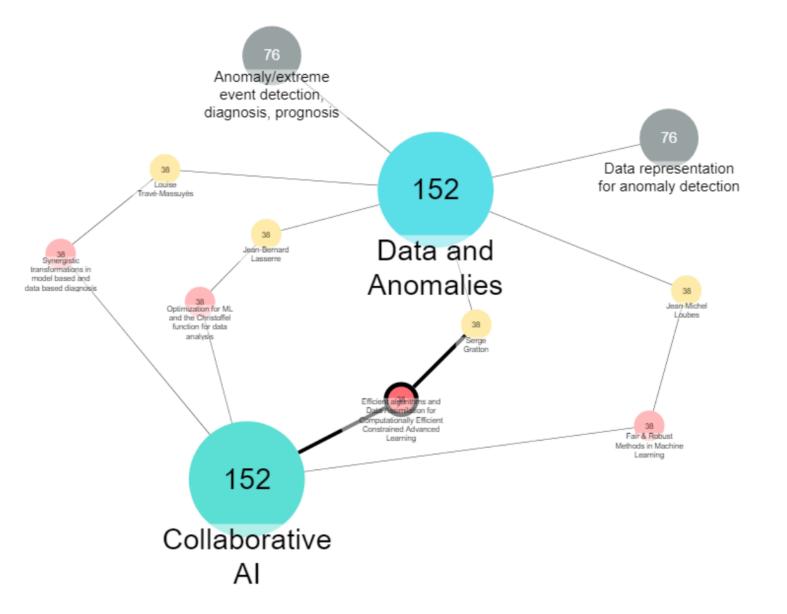
Extract knowledge from data

Integrate knowledge and data

Predictive maintenance
& Monitoring and surveillance

Core to the theme **"Data & Anomalies"** and some contributions to the **"Explainability"** and "**Automated reasoning and decision making**" themes.

What have we done ?

Escape game JEPIA and other popularization events

- Toulouse is AI "Comprendre l'intelligence Artificielle, Apprendre et Entreprendre" (L. Travé-Massuyès & N. Barbosa Roa) − Sept 2019
- Nuit Européenne des chercheurs, Table Ronde "Intelligence Artificielle: comment (bien) préparer notre avenir" − Sept 2019
- SIANE Salon des partenaires de l'Industrie, Plateau télé "ANITI et Interactions avec l'Industrie" − Oct 2019
- Petit Illustré d'IA, "Le diagnostic, c'est quoi ?", authors: Louise Travé-Massuyès and Yannick Pencolé, co-edited by CNRS Occitanie Ouest and La Dépêche du midi − December 2020
- Conference debate of the "Club Audiovisuel Numérique de Toulouse Métropole": How does AI work? (L. Travé-Massuyès) − March 24, 2020
- Posd-cast Investiga'Sciences: Industry and crafts: maintenance, same methods? (L. Travé-Massuyès) − December 7, 2020
- YouTube video in the series « Qui cherche cherche » - Louise Travé-Massuyès − March 2021

➢Afterwork "Data and Anomalies" – April 8, 2021.

➢Future Intelligence 2021, Workshop « Diagnosis and maintenance: where and when? » (L. Travé-Massuyès & E. Chanthery) – May 3, 2021

➢32rd International Workshop on Principles of Diagnosis (DX-22), Toulouse, France, September 14-16, 2022 (Chairs: E. Chanthery & P. Ribot)

➢AI lab sessions at INSA by Adrien Dorise

➢Involvement of Elodie Chanthery in the creation of two new lectures:
- Introduction to Machine Learning (M1)
- Machine Learning (M2)

➢One lecture "Diagnosis and Supervision" by Louise Travé-Massuyès at ENAC (M1)

➢ Implicit knowledge extraction from process data flows

➢ Novel methods for anomaly detection

- DTW based AD in irregularly sampled and distorted time series
- Christoffel Function based AD in data streams
- Dynamic clustering based for frugal embedded AD

➢ Machine learning architectures for fault and performance loss diagnosis

➢ Tree based diagnosis enhanced with meta knowledge

PhD theses defended

Adrien Dorise. *Embedded anomaly detection-Machine learning based radiation hardening of space electronics.* PhD thesis, Toulouse, INSA, defended on December 2, 2022.

Sepulveda Edgar. *Detection et diagnostic de défauts et des pertes de performance dans les centrales photovoltaïques de forte puissance.* PhD thesis, Toulouse, INSA, defended on February 16, 2023.

Le Toan Duong. *Leveraging process mining for knowledge extraction and process optimization.* PhD thesis, Toulouse, INSA, defended on June 26, 2023.

Ducharlet Kevin. *Détection d'anomalies dans les flux de données pour une application dans les réseaux de capteurs.* PhD thesis, Toulouse, INSA, defended on September 28, 2023.

Vignolles Amaury. . *Apprentissage de modèles pour la gestion de santé de systèmes hybrides sous incertitudes.* PhD thesis, Toulouse, INSA, defended on Novembre 15, 2022.

Gaffet Alexandre. *Approches de machine learning pour le diagnostic de processus de production automobile.* PhD thesis, Toulouse, INSA, defended on on April 4, 2023.

On going PhD theses

Predictive Maintenance and diagnosis of robots and cobots

Integration of Machine Learning and Business Knowledge for diagnosis and Prdictive Maintenance

PhD theses defended

Adrien Dorise. *Embedded anomaly detection-Machine learning based radiation hardening of space electronics.* PhD thesis, Toulouse, INSA, defended on December 2, 2022.

Sepulveda Edgar. *Dete... ...uts et des pertes de performance dans les centrales photovoltaïques ...* ...hesis, Toulouse, INSA, defended on February 16, ...

Le Toan Duong. *Leveraging process m... ...extraction ... ...timization.* PhD thesis, Toulouse, INSA, defende...

Ducharlet Kevin. *Détection d'un... ...nées pour une application dans les réseaux de capt...* ... ...ed on September 28, 2023.

Vignolles Amaury. *. Apprentis... ...systèmes hybrides sous incertitudes.* PhD thesis, ...2022.

Gaffet Alexandre. A... ...ing pour le diagnostic de processus de production automobile. Ph... ...se, I... defended on on April 4, 2023.

On going PhD theses

Predictive Maintenance and diagnosis of robots and cobots

Integration of Machine Learning and Business Knowledge for diagnosis and Prdictive Maintenance

CNES/Occitanie Region PhD
1st circle

CIFRE PhD
2nd circle

CIFRE PhD
1st circle

CIFRE PhD
2nd circle

Ministery PhD
1st circle

CIFRE PhD
2nd circle

CIFRE PhD
2nd circle

CIFRE PhD
1st circle

CNES/Occitanie Region PhD
1st circle

CIFRE PhD
2nd circle

CIFRE PhD
1st circle

CIFRE PhD
2nd circle

Ministery PhD
1st circle

CIFRE PhD
2nd circle

CIFRE PhD
2nd circle

CIFRE PhD
1st circle

A bit of what is going on

SYSTEM

inputs

u

X : state vars
(unknown)
f : faults

outputs

y

obs

Diagnosis tests:

~~**X** : state vars (unknown)~~

Model $\Longrightarrow$ Relations s.t. **R**(***obs***) = $r(t)$ such that $\lim_{t \to \infty} r(t) = 0$ for all ***obs*** consistent with normal behavior.

ANITI U Université de Toulouse

SYSTEM

inputs
u

X : state vars
(unknown)
f : faults

outputs
y

obs

Diagnosis tests:

X : state vars
(unknown)

Model $\Longrightarrow$ Relations s.t. $\mathbf{R}(\textbf{\textit{obs}}) = r(t)$ such that $\lim_{t \to \infty} r(t) = 0$ for all ***obs*** consistent with normal behavior.

A → M1 — T
A1 → F
B
C — M2 — U
D
A2 → G
E → M3 — W

inputs

outputs

T, U, W $\Longrightarrow$

Simple static polybox

	ok	Faulty M1	Faulty M2	Faulty M3	Faulty A1	Faulty A1
R1 : A*C + B*D − F = 0	0	X	X	0	X	0
R2 : E*C + B.D − G = 0	0	0	X	X	0	X
R3 : A*C − E*C − F + G = 0	0	X	0	X	X	X

Data based diagnosis

Data based diagnosis

SYSTEM

inputs
u

x : state vars
f : faults

outputs
y

$$\boldsymbol{obs}$$

Diagnosis tests:

$$\mathbf{R}(\boldsymbol{obs}) = r(t) \text{ such that } \lim_{t \to \infty} r(t) = 0$$

for all *obs* consistent with normal behavior.

Fault2 data — Nominal data — Fault1 data — **Dataset** — Fault4 data — Fault5 data — Fault3 data

Symbolic classification

Nominal data — Fault3 data — Fault1 data

Fault1 data — Fault4 data — Fault2 data — Fault5 data

Nominal data

Fault3 data — Fault1 data

Fault2 data — Fault1 data

Fault5 data — Fault4 data

Fault4 data — Fault5 data

Symbolic classification

Algorithm to estimate a function f such as:

$$x = (x_1, x_2, \ldots, xn) \in \mathbb{R}^n \xrightarrow{\quad f \quad} a \in \mathbb{R} \xrightarrow{\quad t \quad} label \in \{0, 1\}$$

known
unknown

Dataset $\mathcal{D}$ of p pairs (x, $label$)

Set of operators
(ex: +, *, -, /, sin, arctan, ln, ²)

Threshold ε

Genetic algorithm

Symbolic Regressor → Candidates c

$c(x)$

Transformer → Class 0/1

f

Symbolic classification

Evaluation of the candidate c with the log-loss fitness function

$$Fitness(c) = -\frac{1}{p} \sum_{(x,label) \in \mathcal{D}} [label * \ln\left(t(c(x))\right) + (1 - label) * \ln\left(1 - t(c(x))\right)]$$

Stopping condition

IF $Fitness(c) < \varepsilon$ **THEN** algorithm stops and f is found   **ELSE** $Mutate(c_{best})$ **AND** $Repeat$

Symbolic classification

Train a symbolic classifier to find f

Fault i data

Nominal Data + Fault j

f

t

$a \in \mathbb{R}$

Label = 1

Label = 0

Diagnosis

Meta-knowledge

Symbolic classification

Train a symbolic classifier to find f

Fault i data

Label = 1

f t

$a \in \mathbb{R}$

1

Nominal Data + Fault j

Label = 0

$t:$

-δ   0   +δ

δ is a hyperparameter of DT4X

ANITI   **U** Université de Toulouse

	ok	Faulty M1	Faulty M2	Faulty M3	Faulty A1	Faulty A1
R1 : $A*C + B*D - F = 0$	0	X	X	0	X	0
R2 : $E*C + B.D - G = 0$	0	0	X	X	0	X
R3 : $A*C - E*C - F + G = 0$	0	X	0	X	X	X

$$f_0 = B*D + E*C - G$$

B*D + E*C - G
Sample repartition during training:
fault 0 (Nominal) occurrences: 723
fault 1 (M1) occurrences: 125
fault 2 (M2) occurrences: 139
fault 3 (M3) occurrences: 155
fault 4 (A1) occurrences: 144
fault 5 (A2) occurrences: 148

$$f_1 = -A*C - B*D + F$$

$$f_2 = A*C + B*D - F$$

=0

!=0

-A*C - B*D + F
Sample repartition during training:
fault 0 (Nominal) occurrences: 723
fault 1 (M1) occurrences: 125
fault 4 (A1) occurrences: 144

A*C + B*D - F
Sample repartition during training:
fault 2 (M2) occurrences: 139
fault 3 (M3) occurrences: 155
fault 5 (A2) occurrences: 148

=0

!=0

=0

!=0

Diagnosis in this leaf: Nominal
Sample repartition during training:
fault 0 (Nominal) occurrences: 723

Diagnosis in this leaf: A1
Sample repartition during training:
fault 1 (M1) occurrences: 125
fault 4 (A1) occurrences: 144

Diagnosis in this leaf: M3
Sample repartition during training:
fault 3 (M3) occurrences: 155
fault 5 (A2) occurrences: 148

Diagnosis in this leaf: M2
Sample repartition during training:
fault 2 (M2) occurrences: 139

Nominal

Fault M1 & Fault A1

Fault M3 & Fault A2

Fault M2

ANITI — Université de Toulouse

	ok	Faulty M1	Faulty M2	Faulty M3	Faulty A1	Faulty A1
R1 : $A*C + B*D - F = 0$	0	X	X	0	X	0
R2 : $E*C + B.D - G = 0$	0	0	X	X	0	X
R3 : $A*C - E*C - F + G = 0$	0	X	0	X	X	X

$- B*D - E*C + G$

$-A*C - B*D + F$

$A*C + B*D - F$

$A*C - E*C - F + G$

-B*D - E*C + G
Sample repartition during training:
fault 0 (Nominal) occurrences: 5456
fault 1 (M1) occurrences: 368
fault 2 (M2) occurrences: 358
fault 3 (M3) occurrences: 379
fault 4 (A1) occurrences: 356
fault 5 (A2) occurrences: 364
fault 6 ([M1, M2]) occurrences: 360
fault 7 ([M1, M3]) occurrences: 363
fault 8 ([M1, A1]) occurrences: 353
fault 9 ([M1, A2]) occurrences: 363
fault 10 ([M2, M3]) occurrences: 351
fault 11 ([M2, A1]) occurrences: 365
fault 12 ([M2, A2]) occurrences: 349
fault 13 ([M3, A1]) occurrences: 361
fault 14 ([M3, A2]) occurrences: 377
fault 15 ([A1, A2]) occurrences: 364

=0

!=0

A*C + B*D - F
Sample repartition during training:
fault 0 (Nominal) occurrences: 5456
fault 1 (M1) occurrences: 368
fault 4 (A1) occurrences: 356
fault 8 ([M1, A1]) occurrences: 353

-A*C - B*D + F
Sample repartition during training:
fault 2 (M2) occurrences: 358
fault 3 (M3) occurrences: 379
fault 5 (A2) occurrences: 364
fault 6 ([M1, M2]) occurrences: 360
fault 7 ([M1, M3]) occurrences: 363
fault 9 ([M1, A2]) occurrences: 363
fault 10 ([M2, M3]) occurrences: 351
fault 11 ([M2, A1]) occurrences: 365
fault 12 ([M2, A2]) occurrences: 349
fault 13 ([M3, A1]) occurrences: 361
fault 14 ([M3, A2]) occurrences: 377
fault 15 ([A1, A2]) occurrences: 364

=0

!=0

=0

!=0

Diagnosis in this leaf: Nominal
Sample repartition during training:
fault 0 (Nominal) occurrences: 5456

Diagnosis in this leaf: M1
Sample repartition during training:
fault 1 (M1) occurrences: 368
fault 4 (A1) occurrences: 356
fault 8 ([M1, A1]) occurrences: 353

Diagnosis in this leaf: M3
Sample repartition during training:
fault 3 (M3) occurrences: 379
fault 5 (A2) occurrences: 364
fault 14 ([M3, A2]) occurrences: 377

A*C - E*C - F + G
Sample repartition during training:
fault 2 (M2) occurrences: 358
fault 6 ([M1, M2]) occurrences: 360
fault 7 ([M1, M3]) occurrences: 363
fault 9 ([M1, A2]) occurrences: 363
fault 10 ([M2, M3]) occurrences: 351
fault 11 ([M2, A1]) occurrences: 365
fault 12 ([M2, A2]) occurrences: 349
fault 13 ([M3, A1]) occurrences: 361
fault 15 ([A1, A2]) occurrences: 364

=0

!=0

Diagnosis in this leaf: M2
Sample repartition during training:
fault 2 (M2) occurrences: 358
fault 7 ([M1, M3]) occurrences: 8
fault 9 ([M1, A2]) occurrences: 10
fault 13 ([M3, A1]) occurrences: 11
fault 15 ([A1, A2]) occurrences: 10

Diagnosis in this leaf: [M2, A1]
Sample repartition during training:
fault 6 ([M1, M2]) occurrences: 360
fault 7 ([M1, M3]) occurrences: 355
fault 9 ([M1, A2]) occurrences: 353
fault 10 ([M2, M3]) occurrences: 351
fault 11 ([M2, A1]) occurrences: 365
fault 12 ([M2, A2]) occurrences: 349
fault 13 ([M3, A1]) occurrences: 350
fault 15 ([A1, A2]) occurrences: 354

PolyBox Results

Id	SklDT accuracy (%)	SklDT time (s)	DT4X accuracy (%)	DT4X time (s)
1	47.83	0.08	79.93	0.64
2	47.27	0.09	80.30	0.63
3	47.10	0.08	80.00	0.64
4	47.46	0.08	80.14	0.85
5	47.84	0.08	79.52	0.79
6	45.96	0.09	80.08	0.66
7	47.17	0.08	80.59	0.94
8	42.28	0.09	80.16	0.94
9	47.81	0.09	80.66	0.64
10	46.63	0.09	79.72	1.32

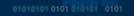Randomly generated datasets with 10 084 testing samples and 22 937 training samples

Better results + formal expressions of residuals ... But still a lot to do ...

More adventures in ANITI 2 ?