Macroscopes of Human behavior: the case of biomedical sensing

Nicolas Vayatis

Joint work with

- PhD students: Rémi Barrois-Muller, Thomas Moreau*, Alice Nicolaï, Charles Truong*, Aliénor Vienne
- Researchers: Julien Audiffren, Ioannis Bargiotas, Juan Mantilla, Laurent Oudre*
- MD-PhDs: Catherine de Waele, Damien Ricard, Pierre-Paul Vidal, Alain Yelnik

Introduction

ML in the Real world

- Modeling
  - Definition of the objective (and constraints)
  - Value of automatic decisions for human experts
- Information
  - Access to relevant data
  - Data preparation
- Scaling-up
  - Learning-to-learn
  - Monitoring the pipeline

ML in healthcare

- Modeling objective:
  - automatic diagnosis?
  - therapy recommendation?
- Data sources:
  - clinical trials, social security, hospital
  - mainly aggregated or reduced data
- Scaling-up?
  - Not yet there
  - Some industrial failures
  - Why is it hard?

What is the relevant level of study of Human behavior?

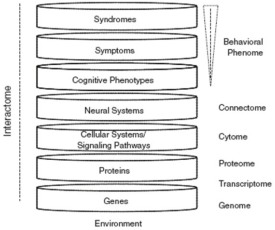
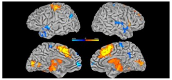
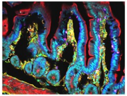
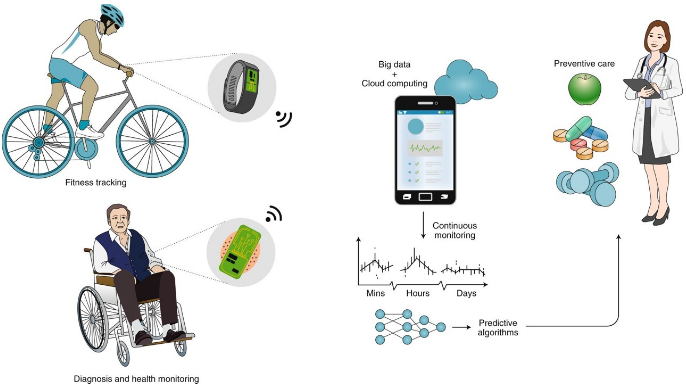


Figure 1– From Jagaroo, (Vinoth & Santangelo, 2016). The road from genome to phenome.





There are other options...



Rationale behind the project

Statement of work

- Central research question:
  - Empirical, but quantitative, study of Human behavior - regular, pathological, altered - through its sensory-motor transformations
- Main assumption:
  - All Humans are different from each other but have constant behavior over time
- Conditions of study:
  - Individual follow-up in 'natural' conditions, with 'light' sensors to allow access to large cohorts
- Challenge:
  - Find and define standards for protocols, data, evaluation

The context of this project

- Objective:
    - Assessing gait and posture for neurology, ENT and rehabilitation in the field (consultation, hospital)
- State-of-the-art:
    - Dozens of quantitative studies using classical statistical methods focusing on 2 to 5 features with around 100 subjects per study
- Few public data available
    - mostly aggregated, essentially healthy subjects.

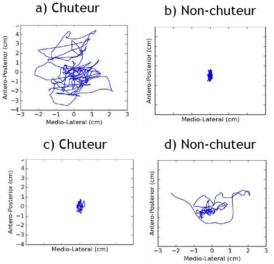
Where it all started
In 2012, a students' project

Question asked: How to have objective values for neurological (Romberg) tests in routine consultation?

The case of posturography

→ Précision : $\approx$ 0.8 mm.
→ Fréquence 60 Hz

→ Précision : 0.1 mm.
→ Fréquence 500 Hz

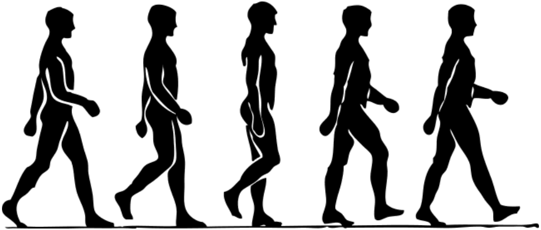


- Romberg test: eyes open/close for 20s (result of a consensus)
- SoA: use AMTI force plate (10kE), around 100 patients per cohort, about 5 average features per recording,
- Our project: use WBB (80E) in 10 consultations, 3k recordings in less than 6 months, more than 1000 features (including local ones) used to predict the risk of fall.

What we achieved
It took five years...

- A full - and valid - pipeline of data acquisition and processing from sensors to the clinician dashboard that operates in routine consultation and discovery of new markers of balance disorders (phenomic codes)
- Databases (to be published), publications (Sensors, IEEE Trans. SP, ICML, PlosOne, Frontiers, ...), opensource software and licensed patents
- Inclusion in a social security program for preventing frail states in the elderly population
- Last but not least: an interesting blend of cultures between mathematicians, computer scientists, engineers, ergonomists, neuroscientists, clinicians, psychologists

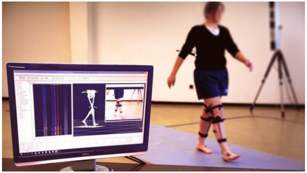
Why gait and posture?



- Most frequent and dynamic human activity
- Marker of several troubles: neurologic, orthopedic, rheumatologic...
- Strongly affects everyday life: risk of fall, frailty, mobility, loss of autonomy...
- Important cause of morbidity, high cost for public health

Quantification of walking ability

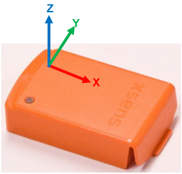
The study of walking



- Traditionally: clinical assessment made by the clinician, functional tests, questionnaires
  - (+) Easy to execute, requires clinical expertise
  - (-) Lack of precision, difficult to compare sessions
- Platforms for measuring locomotion: sensing floor, video and optical systems
  - (+) High precision, extraction of many parameters, objective quantification
  - (-) High cost, hard to use

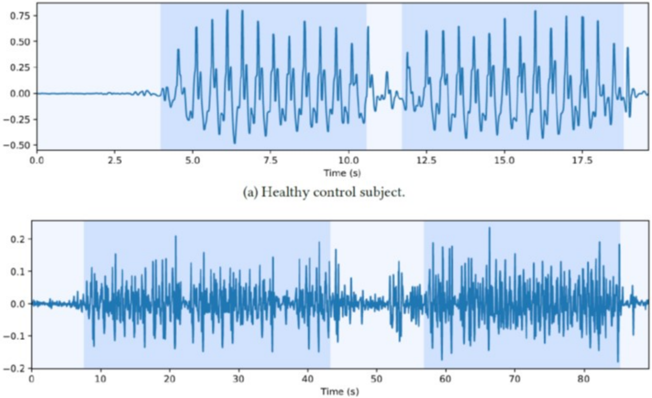
Sensing protocol

- Protocol at confort speed: stop (6 sec), directed walk (10 m), U-turn, directed walk (back), stop
- Four wireless inertial units (IMU): left foot, right foot, lower back, head
- Nine signals per sensor : linear accelerations (3D), angular speed (3D), magnetic fields (3D)



Play/Pause

Accelerometric signals on a walk exercise



(a) Healthy control subject.

(b) Osteoarthritis patient.

Figure 1.3: Vertical acceleration (m/s^2) of the lower back sensor for two different subjects. Alternating colours mark the consecutive phases: “Stand”, “Walk”, “Turnaround”, “Walk” and “Stop”.

Signal characteristics

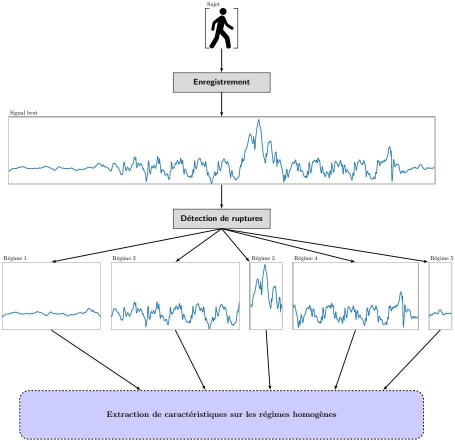
- Nonstationary signals
  $\rightarrow$ How to detect and categorize different regimes (stop, walk, U-turn...) ?
- Repeated but irregular patterns
  $\rightarrow$ Location and shape?
- A particular pattern of interest : the step
  $\rightarrow$ Locate precisely beginning and end of each step?

A sample of research topics

1. Segmentation
2. Dictionary learning

Topic #1 - Segmentation

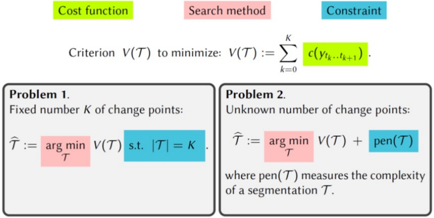
Goal of the segmentation method



- Find automatically the changepoints (start to walk, walk to U-turn, U-turn to walk, walk to stop) under weak or no supervision

Review on changepoint detection

- Modular view on the complete literature for offline changepoint detection



- More than 150 references with methodological contributions, thousands of application papers...
- Selective review to appear in Signal Processing + Python package *ruptures*, by Truong, Oudre, V.

Optimal method

Cost function | **Search method** | Constraint

How to minimize over the space of segmentations? Exhaustive enumeration is impractical: $\mathcal{O}(T^K)$ segmentations with K changes.

Optimal resolution

For Problem 1 (fixed K) and Problem 2 (unknown K), based on **dynamic programming**, using

$$\min_{|\mathcal{T}|=K} V(\mathcal{T}) = \min_{1\leq t<T} \left[c(y_{0..t}) + V(\widehat{\mathcal{T}}^{K-1}_{t..T})\right]$$

where $\widehat{\mathcal{T}}^{K-1}_{t..T}$ is the optimal segmentation of $y_{t..T}$ with $K-1$.

Complexity.

- Problem 1: $\mathcal{O}(T^2)$ (Bellman, 1955).
- Problem 2: $\mathcal{O}(T)$ (Killick et al., 2012).

Approximation method

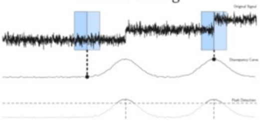
Cost function | Search method | Constraint

Approximate resolution

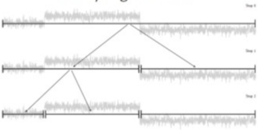
Based on *single* change point detection on sub-signals $y_{a..b}$:

$$\min_{a<t\leq b} c(y_{a..t}) + c(y_{t..b}).$$

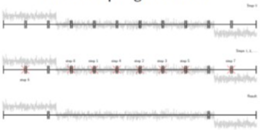
Window sliding



Binary segmentation



Bottom-up segmentation



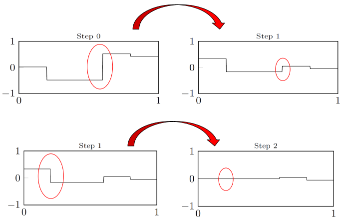
Requirements for the gait data

- Computational cost - almost real time and should operate on a clinicians laptop or surface
- Versatility - Ability to adapt to a wide range of protocols, sensors and patients.
- Automatic calibration - No hyperparameter can be tuned in routine.

A two-step greedy strategy

Follows the principle of OMP (Mallat, Zhang, 1993).

- Step 1: Detection of a single changepoint in the signal
- Step 2: Removal of the detected changepoint by projection



- Stop when K changepoints have been detected
- Use a kernel in the cost function
- Linear complexity for each detection/projection
- Consistency results available

Results

Experimental evaluation. On the *Gait* signals (Spectrogram).

	Approximate		Exact	
Metric	gkCPD	gCPD	Opt (c_{L_2})	Opt (c_{rbf})
Hausdorff	**1.13 (±0.67)**	1.34 (±0.58)	1.80 (±2.35)	1.29 (±1.06)
RandIndex	**0.91 (±0.02)**	0.90 (±0.03)	0.92 (±0.02)	0.91 (±0.02)
F1 score	**0.85 (±0.15)**	0.73 (±0.20)	0.85 (±0.13)	0.83 (±0.15)

Hausdorff is in second. The margin of F1 score is $M = 1$ sec.

Execution time comparison. Time to process 100 signals from *Gait*.

- gkCPD: 1 min 30 sec
- Opt (c_{rbf}): $>$ 2 days.

Comments.

- gkCPD is more accurate than gCPD for all metrics.
- gkCPD is close to Opt for F1 score.
- gkCPD is faster than Opt.

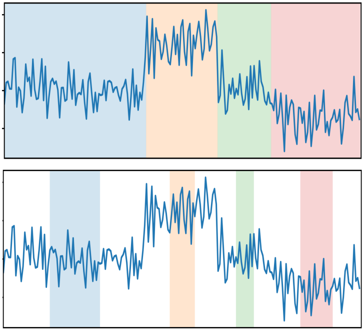
Further contributions

Unknown number of changepoints (TOV-EUSIPCO'17)

- Supervised procedure to determine the smoothing parameter
- Need fully annotated signals (timestamp of changepoint)

Kernel/metric learning (TOV-ICASSP'19)

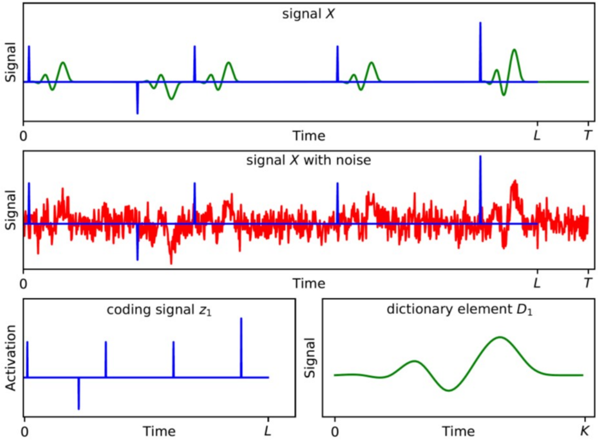
- Semi-supervised procedure to learn the kernel
- Need partially annotated signals (not changepoints)



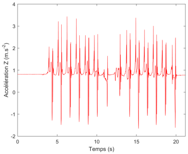
Full vs. partial annotations

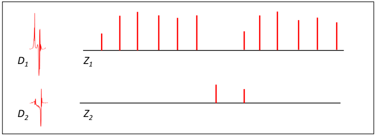
Topic #2 - Dictionary learning

Locating patterns



Model for sparse convolutional coding





Consider d-dimensional signals X of length T

- Patterns $\boldsymbol{D}_k$ with length W
- Activations Z_k of length $L = T - W + 1$

$$X[t] = \sum_{k=1}^{K} (Z_k * \boldsymbol{D}_k)[t] + \mathcal{E}[t], \qquad \forall t \in [\![0, T-1]\!]$$

where $\mathcal{E}$ independent and centered noise signal

Resolution by alternating optimization

- **Dictionary** learning

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}\in\Omega}{\operatorname{argmin}} \frac{1}{N}\sum_{n=1}^{N} \frac{1}{2}\left\| X^{[n]} - \sum_{k=1}^{K} Z_k^{[n]} * \boldsymbol{D}_k \right\|_2^2$$

where Ω : set of normalization constraints.

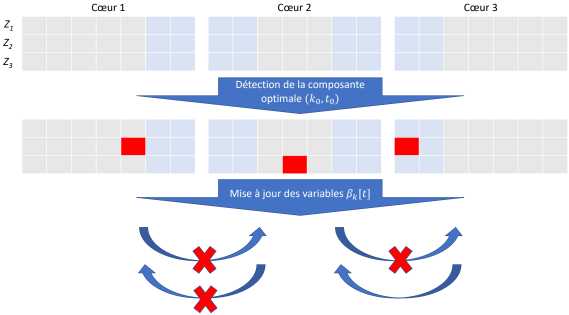
- Convolutional sparse coding

$$Z^* = \underset{Z=(Z_1,\ldots Z_K)}{\operatorname{argmin}} \frac{1}{2}\left\| X - \sum_{k=1}^{K} Z_k * \boldsymbol{D}_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|Z_k\|_1$$

Looks straightforward but...

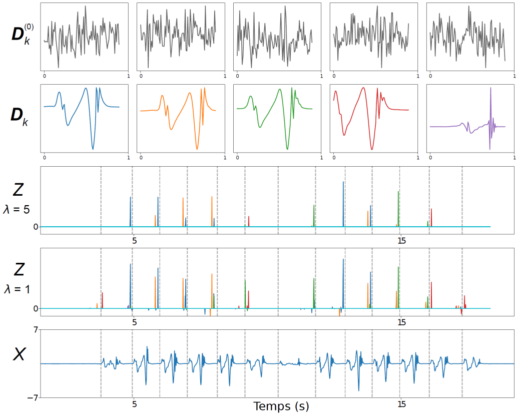
- Signals can be very long
  $\rightarrow$ How to speedup sparse convolutional coding?
- Parallelization strategy:
  - Each worker processes one subsegment
  - Use message passing in case of interferences

Basic idea of message passing



- Chose subsegment length much larger than pattern size to minimize the amount of interferences
- Significant acceleration and convergence proof under weak assumptions

Results (1/2)



- Unsupervised learning of repeated and relevant patterns
- Can be used for very long signals (ambulatory, ECG...)

Results (2/2)

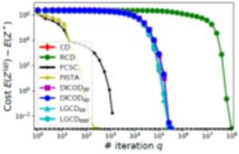


Figure 2. Evolution of the loss function for DICOD, LGCD, CD, FCSC and FISTA while solving sparse coding for a signal generated with default parameters relatively to the number of iterations.

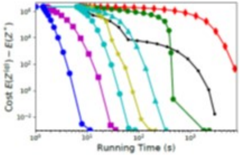


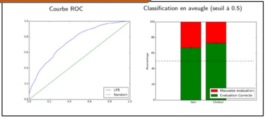
Figure 3. Evolution of the loss function for DICOD, LGCD, CD, FCSC and FISTA while solving sparse coding for a signal generated with default parameters, relatively to time. This highlights the speed of the algorithm on the given problem.

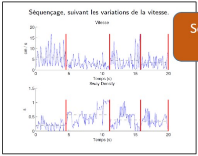
- Superlinear acceleration with respect to the sequential implementation

What we learned

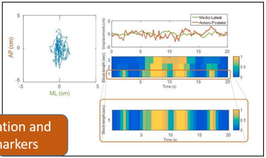
The meat is not in predicting

Prediction models and statistical performance

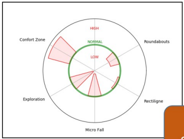




Segmentation of locally stationary signals



Robust representation and structured biomarkers



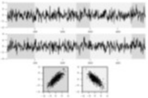
Multifactorial and explainable assessment

Old is not dead

ruptures : change point detection in Python

Schematic view of the ruptures package.

Online documentation



Simulated 2D signal

```
import ruptures as rpt

# signal generation
signal, bkps = rpt.pw_normal(n_samples=500, n_bkps=4)

# change point detection
algo = rpt.Dynp(model="rbf").fit(signal)
result = algo.predict(n_bkps=4)
```

Python code.

ML in healthcare

- Modeling objective:
  - Do not aim at diagnosis: the future is about prevention
  - Developing proper metrology of Human body is already challenging and useful
- Data sources:
  - Nowcasting requires individual and fresh data
  - Wearable and ambient sensors have to be considered within full pipelines including sophisticated preprocessing and machine learning layers designed with field experts (clinicians, ergonomists, neuroscientists)
- Scaling-up?
  - A political issue...
  - Too big for startups?
  - Also a software project, the easier to fail...

No data, no party

- Importance of preprocessing for using advanced ML
- Access to raw and synchronized data in healthcare monitoring systems is THE issue

Connecting people

- Matching agendas between clinicians and ML people
- Who has the power?
- Evaluation of careers out of disciplinary silos
- New forms of cooperation between academia, industry and... social security

Thank you!