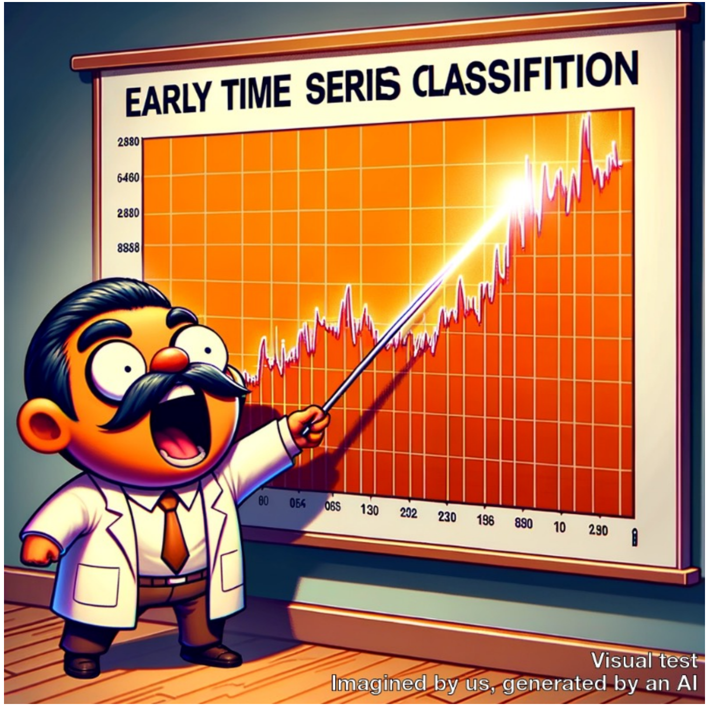
Early Decision Making

Tutorial PFIA 2024

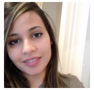


Our research history and team


Antoine Cornuéjols


Alexis Bondu


Vincent Lemaire


Asma Dachraoui

2012 : start working on Early Classification of Time Series (@EDF) with *Asma Dachraoui, Phd student*

2015 : development of the first non-myopic approach


Youssef Achenchabe

2019 : second Phd student, *Youssef Achenchabe* (@Orange)

2020 : improvement of non-myopic approaches

2022 : - extension to revocable decisions
- extension to open time series (i.e. data stream)

2023 : Extension of ECTS to a more general problem, called ML-EDM

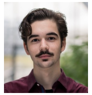

Albert Bifet


Joao Gama


Georges Hébraill


Pierre-François Marteau

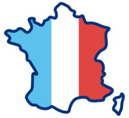

Fabrice Clérot

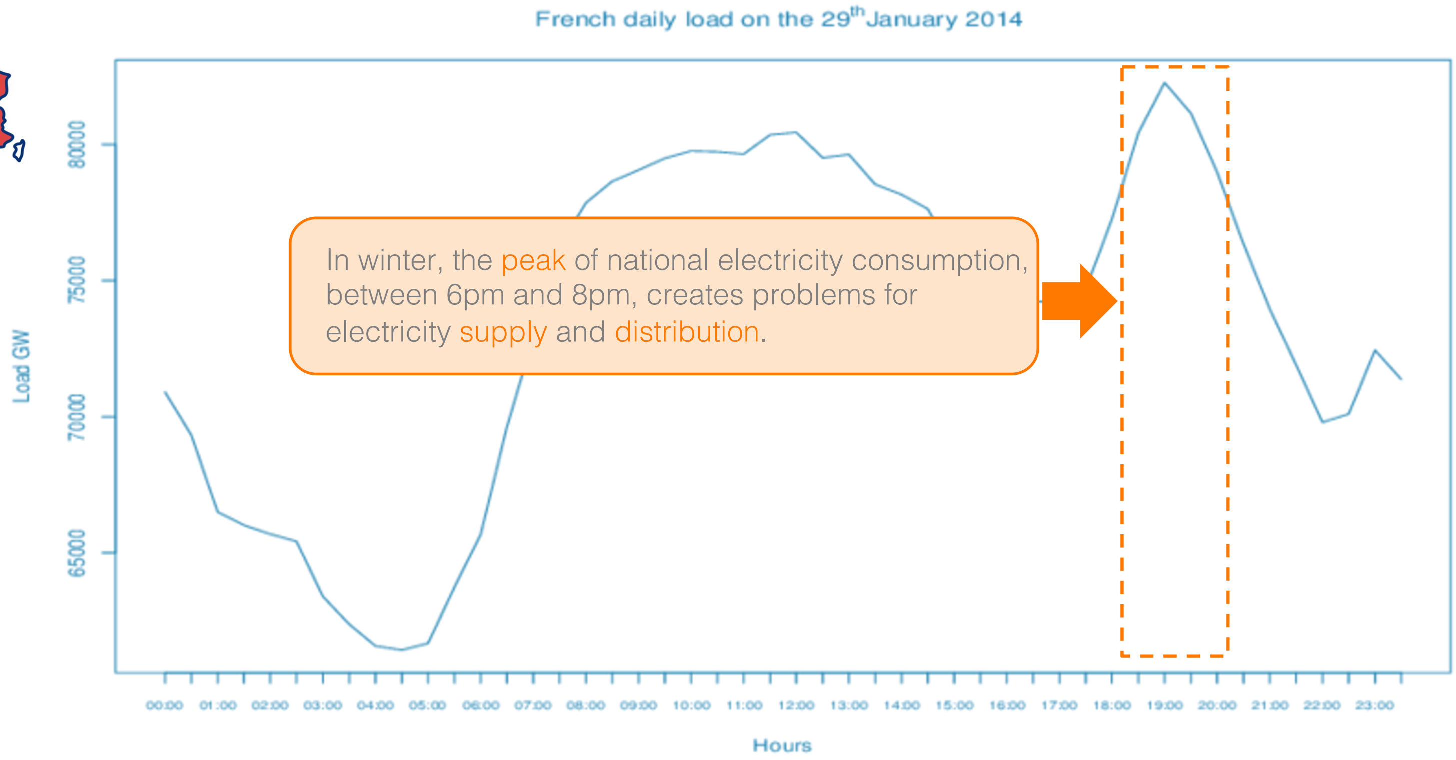

Aurélien Renault

2024 : third Phd student, *Aurélien Renault* (@Orange) working on non-stationary environment

Motivations

EDF's original target application

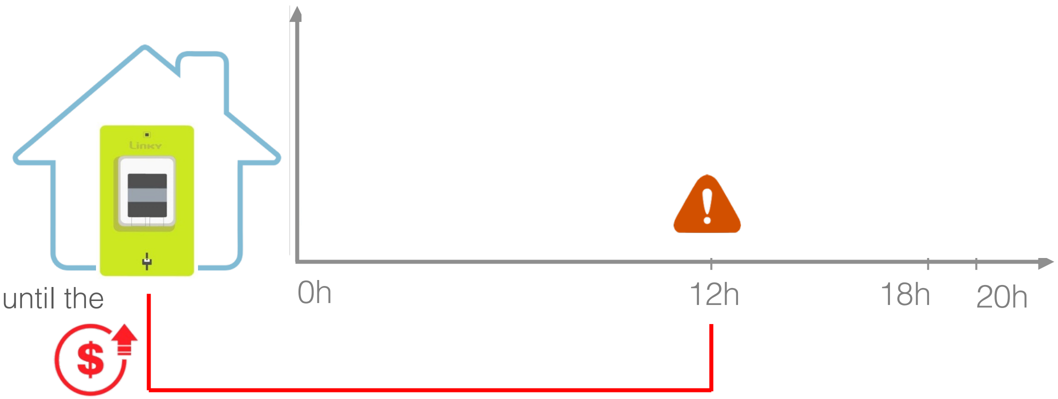




In winter, the peak of national electricity consumption, between 6pm and 8pm, creates problems for electricity supply and distribution.

EDF's original target application

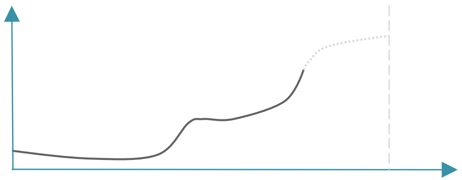
Early targeting of households involved in the evening peak, between 18pm and 20pm.



Price increase signal until the end of the day.

Early detection of high consumption between 6pm and 8pm.

The question is: When to make a decision ?



Compromise between Earliness *vs.* Accuracy

Too early: few data is available, the decision is not reliable.

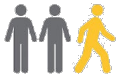
Good trade-off: a correct compromise between earliness and quality of the decision.

Too late: a cost is incurred when the decision is delayed.

Early but inaccurate decisions …

… Accurate but too late decisions

Many applications with time pressure to decide



Churn: when to contact potential churners?

Waiting for further information … before they leave Orange.

Fraud: when to block a bank card?

Waiting for further information … before the bank account is empty.

Network monitoring: when to detect a malfunction?

Waiting for further information … before the breakdown is established.

Outline

#1 Introduction to ECTS

#2 Proposed taxonomy

#3 State of the art

#4 Experiments

#5 Tutorial

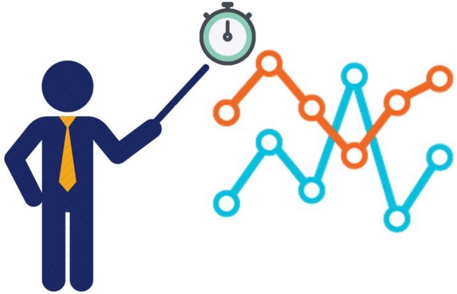
#6 Generalization to ML-EDM

We provide :

Python library:
- 10 competing approaches,
- 3 baselines.

Collection of 35 datasets:
- non-z-normalized
- stratified train / test split

#1 Introduction to ECTS



Bibliographic origins

- First publication to mention the ECTS problem in 2008 [1]
- But, ECTS is rooted in the optimal stopping problem [2] (1969)
  - Here, the online decision to make is simply to stop receiving new pieces of information.
  - A value is observed at each time step and a reward is received.
  - The objective is to maximize the reward by stopping receiving new values at proper time.

e.g. 1 – The Shepp's urn problem

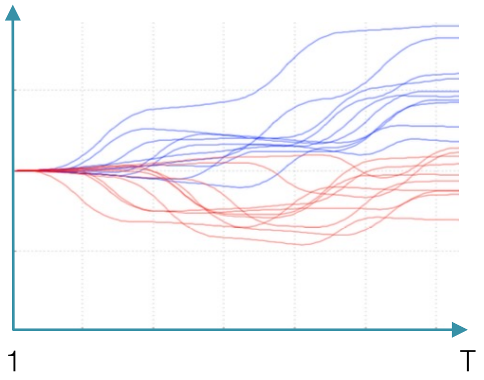
e.g. 2 – The secretary problem

[1] *Xing, Z., Pei, J., Dong, G., & Yu, P. S. (2008, April). Mining sequence classifiers for early prediction. In Proceedings of the 2008 SIAM*
[2] *Shepp, L. A. (1969). Explicit solutions to some problems of optimal stopping. Ann. Math. Statist, 40(993-1010), 227.*

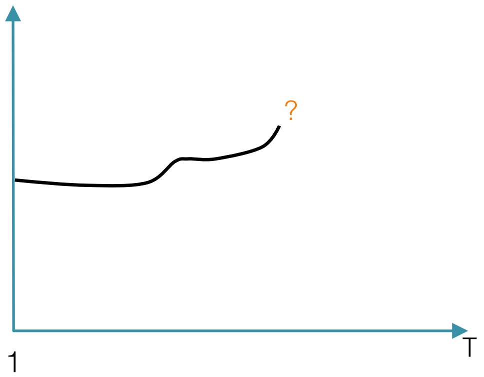
ECTS is an Optimal Stopping problem coupled with a classification task

- The training set is made up of time series of the same length
- The full length, T, corresponds to the maximum time horizon to trigger the decision
- Each label is assigned to a full length time series
- At runtime, the input time series is progressively observed (with non i.i.d measurements)
- The objective is to trigger a reliable decision as soon as possible, before the time T
- A single irrevocable decision can be triggered

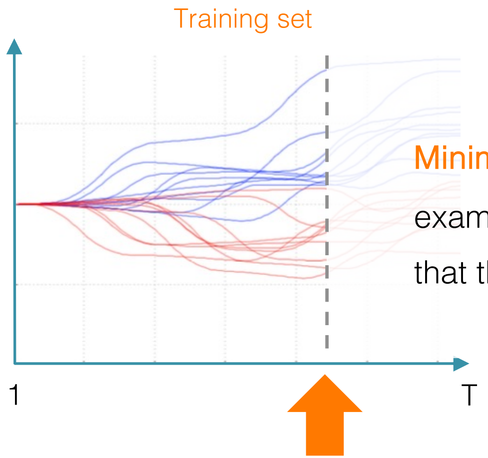
The training set



At runtime



Back to the first “ECTS” approach



Minimal Prediction Length, defined of all training examples, such that the predictions are the same that the ones using the full-length time series.

Shortcomings

- There is no adaptation of the trigger moment to the observed time series
- No trade-off between Accuracy and Earliness

[1] *Xing, Z., Pei, J., Dong, G., & Yu, P. S. (2008, April). Mining sequence classifiers for early prediction. In Proceedings of the 2008 SIAM*

Ground truth

Two decision costs are considered :

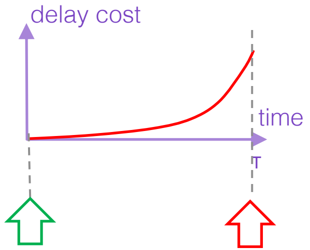
1 - miss-classification cost;

$$C_m(\hat{y}|y) : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$$

$\hat{y}/y$	1	0
1	10	5
0	100	0

2 - and the cost of delaying the decision.

$$C_d(t) : \mathbb{R} \rightarrow \mathbb{R}$$



The example of fraud detection:

The worst case scenario would be to miss an anomaly and trigger the prediction at the last moment.

In the best case, the potential for an anomaly is ruled out as soon as possible.

Ground truth

Two decision costs are considered :

1 - miss-classification cost;

$$C_m(\hat{y}|y) : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$$

$\hat{y}/y$	1	0
1	10	5
0	100	0

2 - and the cost of delaying the decision.

$$C_d(t) : \mathbb{R} \to \mathbb{R}$$

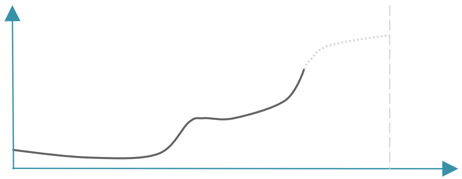
- Considering the following 'time dependent" Loss Function:

$$\mathcal{L}(\hat{y}, y, \hat{t}) = \mathrm{C}_m(\hat{y}|y) + \mathrm{C}_d(\hat{t})$$

- The Empirical Risk is defined as:

$$AvgCost = \frac{1}{N}\sum_{i=0}^{N} C_m(\hat{y}_i|y_i) + C_d(\hat{t}_i)$$

General form of ECTS models



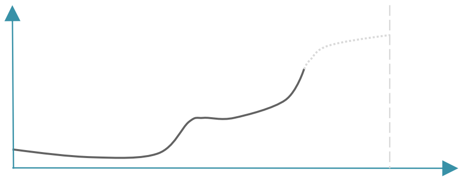
Incomplete time series: $\mathbf{x}_t = \langle x_1, \ldots, x_t \rangle$

ECTS model: $s(\mathbf{x}_t)$

Trigger time: $\hat{t} \in [1, T]$

Predicted class: $\hat{y} \in \mathcal{Y}$

General form of ECTS models



An ECTS function involves a predictor : $\hat{y}(\mathbf{x}_t)$

The optimal triggering time depends on this predictor : $t^\star = \underset{t \in [1,T]}{\arg\min} \, \mathcal{L}(\hat{y}(\mathbf{x}_t), t, y)$

Optimal time requires a complete knowledge of the time series!

The optimal triggering function is is defined as follows :

$$s^\star(\mathbf{x}_t) = \begin{cases} \emptyset & \text{if extra measures are queried;} \\ y^\star = \hat{y}(x_{t^\star}) & \text{when prediction is triggered at } t = t^\star \end{cases}$$

The objective is to minimize the true risk by finding the best triggering function, such that :

$$\underset{s \in \mathcal{S}}{\arg\min} \, \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_{\mathcal{X}}} \left[\mathcal{L}(\hat{y}(\mathbf{x}_{\hat{t}}), y, \hat{t}) \right]$$

Triggering predictions is an online problem!

Proxy metrics exist in the literature

- The Harmonic Mean:

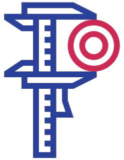
$$HM = \frac{2 \times Accuracy \times (1 - Earliness)}{Accuracy + (1 - Earliness)}$$

with $Accuracy = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}(\hat{y}_i = y_i)$ and $Earliness = \frac{1}{M \times T}\sum_{i=1}^{M} \hat{t}_i$

- The Cost Function:

$$CF = \alpha \times (1 - Accuracy) + (1 - \alpha) \times Earliness$$

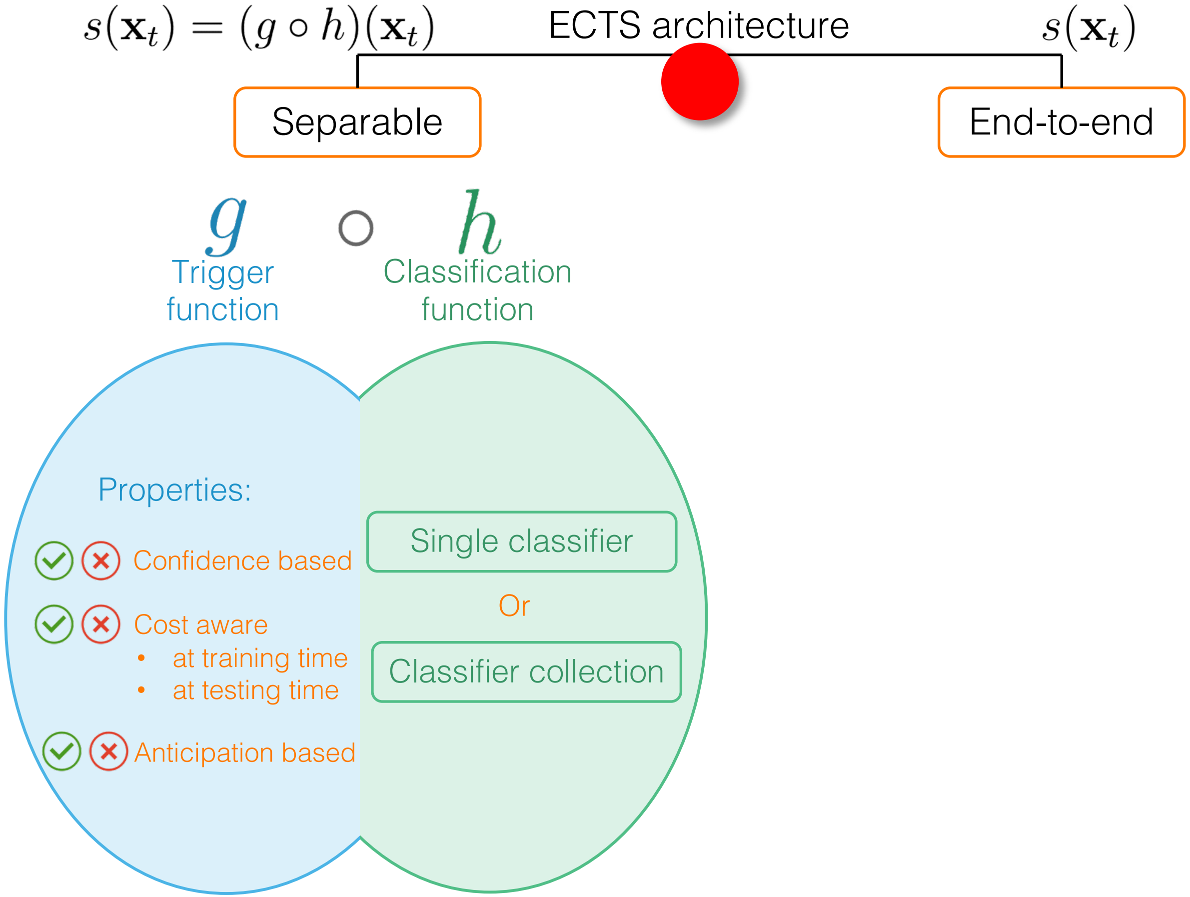
Witch equals to AgvCost when $\mathrm{C}_m(\hat{y}, y) = \mathbb{1}(\hat{y} = y)$, $\mathrm{C}_d(t) = \hat{t}/T$ and $\alpha = 1$



Why do certain approaches to literature fail to use empirical risk ? 😬

#2 Proposed taxonomy

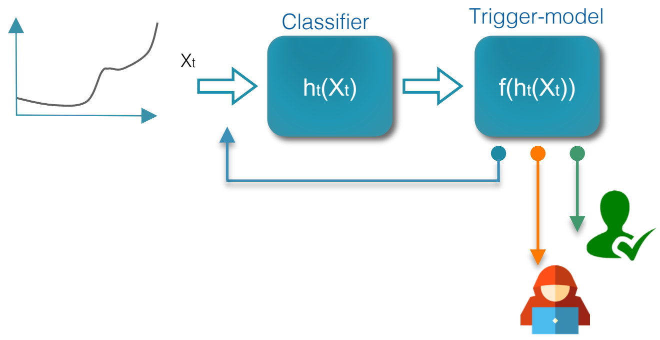
A new ECTS taxonomy



Types of architecture :

Separable

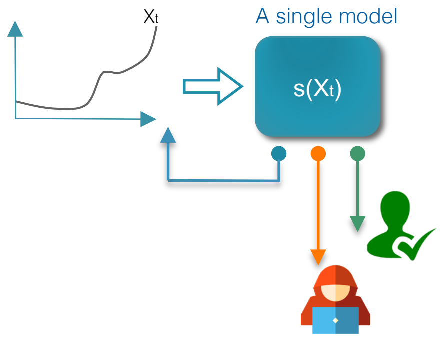
$$s(\mathbf{x}_t) = (g \circ h)(\mathbf{x}_t)$$



- **A baseline example**: Proba_Threshold consists in triggering the prediction as soon as the estimated class probability exceeds a certain threshold.

End-to-end

$$s(\mathbf{x}_t)$$



- **Example**: Early Distinctive Shapelet Classification (EDSC) finds a set of sub-series (i.e. shapelets) that allow for early discrimination between different classes..

Types of architecture :

Separable

$$s(\mathbf{x}_t) = (g \circ h)(\mathbf{x}_t)$$

- The major part of the literature.
- Involves two independent training.
- The trigger function is fed from the classifier output.

Separable RL approaches

Set of actions:

$$\{\text{'}decide\ \ now\text{'},\ \text{'}postpone\ \ decision\text{'}\}$$

We focus now on **separable** approaches

End-to-end

$$s(\mathbf{x}_t)$$

- RL and Deep Learning *(sometime mixed)*

End-to-end RL approaches

Set of actions:

$$\{\text{'}postpone\ \ decision\text{'}, c_1, \ldots, c_N\}$$

Two distinct functions (or modules),
but one common training stage.

Outline

$s(\mathbf{x}_t) = (g \circ h)(\mathbf{x}_t)$ ECTS architecture $s(\mathbf{x}_t)$

Separable

End-to-end

$g \circ h$

Trigger function

Classification function

Properties:

- Confidence based
- Cost aware
  - at training time
  - at testing time
- Anticipation based

Single classifier

Or

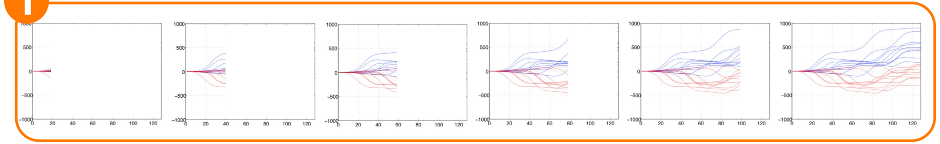
Classifier collection

Classification function

Training a collection of time-indexed classifiers

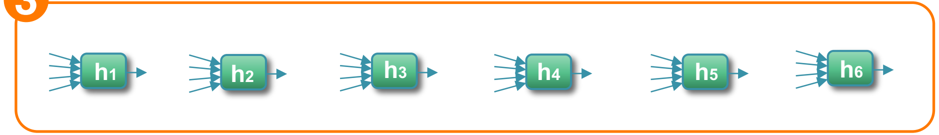
- The most used in the literature.

1 Building a collection of truncated datasets

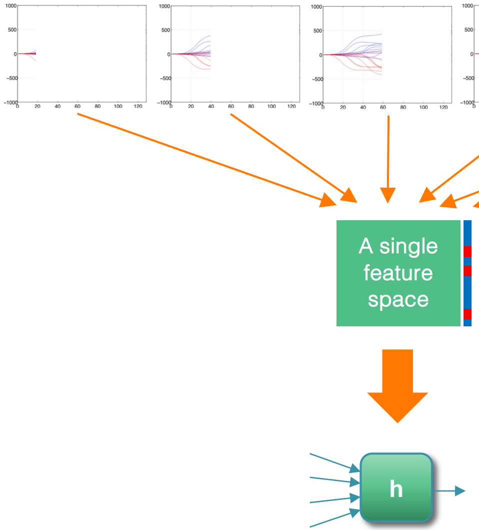


2 Independent features extraction *(or representations)*

3 Independent classifier training



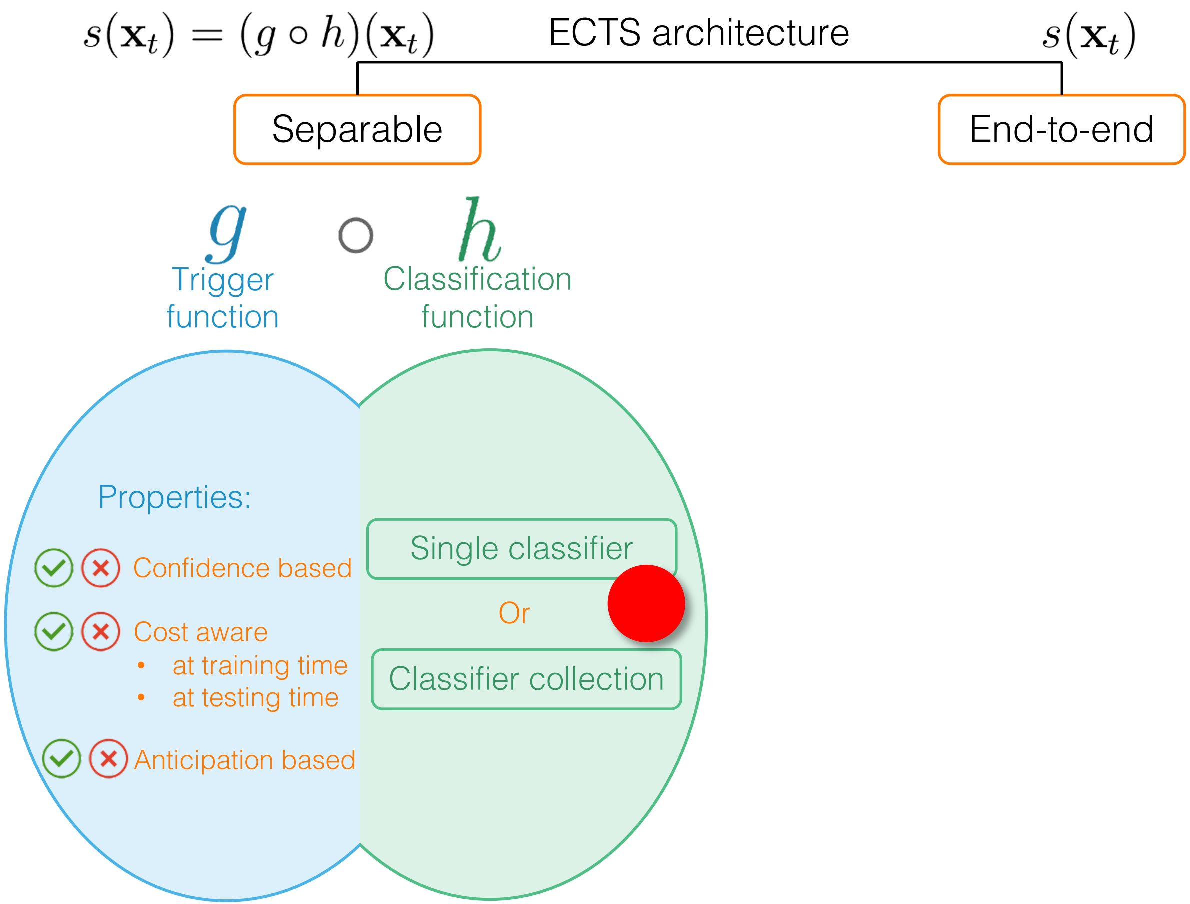
Training a single classifier



- Specific feature engineering
- Or learning a representation (deep)

- Training a single classifier

Outline

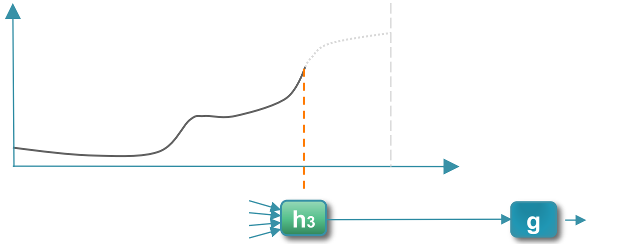


$s(\mathbf{x}_t) = (g \circ h)(\mathbf{x}_t)$
ECTS architecture
$s(\mathbf{x}_t)$
Separable
End-to-end
g
Trigger function
$\circ$
h
Classification function
Properties:
Confidence based
Cost aware
• at training time
• at testing time
Anticipation based
Single classifier
Or
Classifier collection

Confidence based approaches

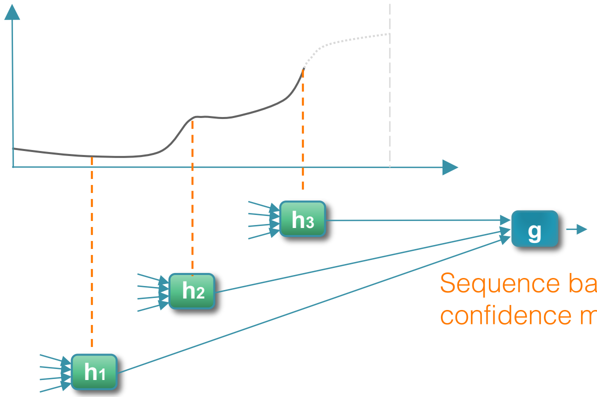
Instant-based trigger function

- The most used in the literature.
- Monitor a confidence metric over time;
- By considering only the last value classifier output.
- Trigger prediction when a threshold is reached.
- **A baseline example**: Proba_Threshold.



sequence-based trigger function

- Considering the sequence of classifier outputs.
- Specifically designed confidence metric for ECTS.



Cost awareness

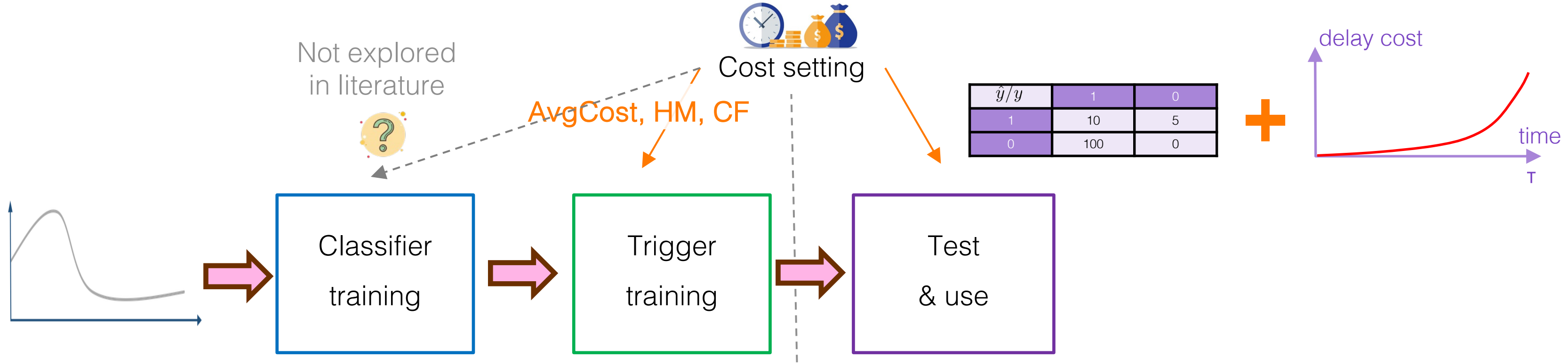
Approaches in the literature differ in the requirement *(or not)* to be informed of costs, at two different stages:

At training time

- A scalar metric is available which combines the two objectives of accuracy and earliness.
- This metric is often used to optimize the trigger parameters.

At testing time

- The cost functions themselves are available (i.e. Cm and Cd)



$\hat{y}/y$	1	0
1	10	5
0	100	0

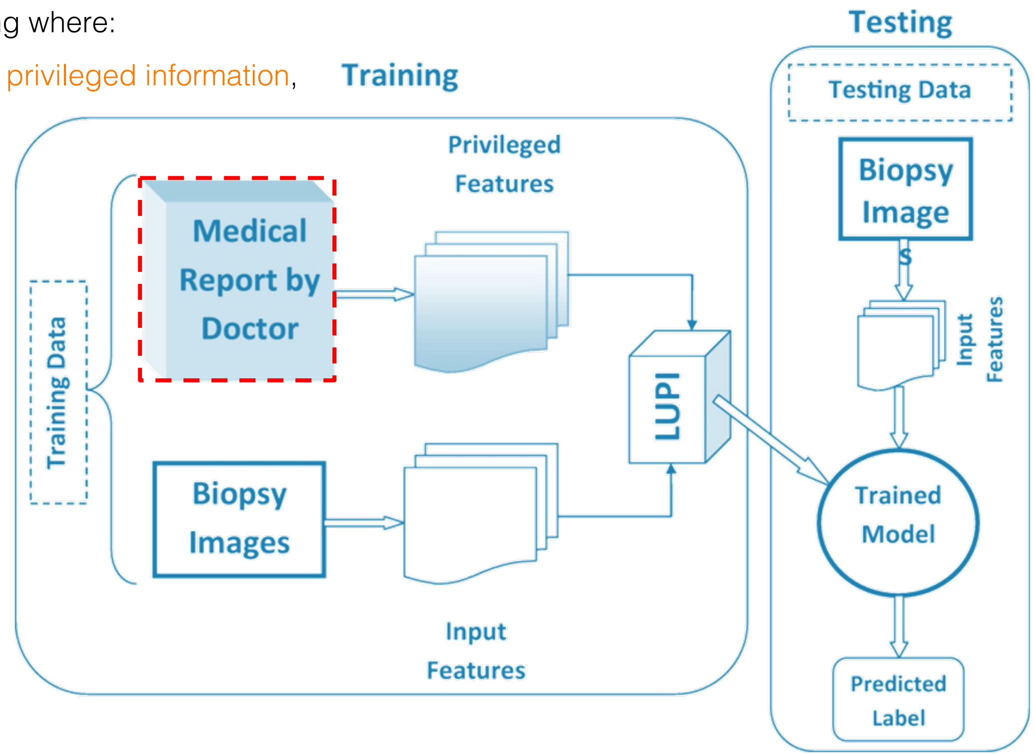
Anticipation-based

Is the current time step is better than all the future ones ?

ECTS can be cast as a LUPI problem [1]

Learning Using Privileged Information is a setting where:

- The learner can benefit at the training time of privileged information,
- That will not be available at test time.



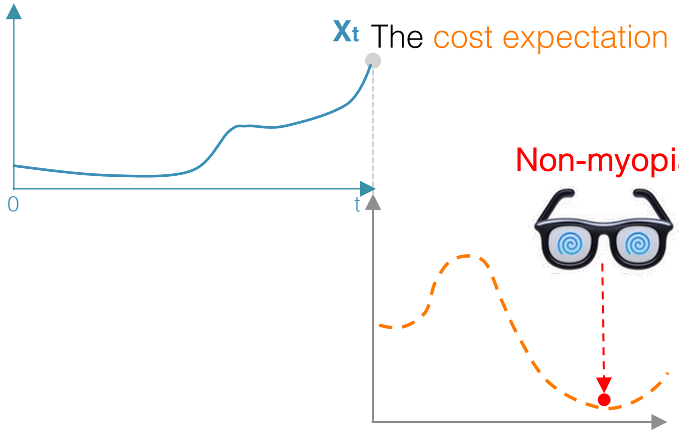
[1] Vapnik, V., & Vashist, A. (2009). A new learning paradigm: Learning using privileged information. Neural networks, 22(5-6), 544-557l

Anticipation-based

Is the current time step is better than all the future ones ?

In the ECTS problem

- LUPI makes possible to learn what are the likely future of an incoming time series;
- For example, to estimate the cost expectation for all future time steps.



The cost expectation is estimated for future time steps.

Non-myopia anticipates the best future triggering time.

[1] Vapnik, V., & Vashist, A. (2009). A new learning paradigm: Learning using privileged information. Neural networks, 22(5-6), 544-557l

Proposed ECTS taxonomy

$s(\mathbf{x}_t) = (g \circ h)(\mathbf{x}_t)$ ECTS architecture $s(\mathbf{x}_t)$

Separable

End-to-end

g $\circ$ h

Trigger function

Classification function

Properties:

- Confidence based
- Cost aware
  - at training time
  - at testing time
- Anticipation based

Single classifier

Or

Classifier collection

We will now use the proposed taxonomy to characterize the bulk of the literature.

#3 State of the art

Literature approaches

References	Classifier(s) (collection ✓)	End2end	Confidence	Anticipation	Cost awareness
EDSC (Xing et al., 2011)	Shapelet	✓	✓		✗
ECTS' (Xing, Pei, & Yu, 2012)	1NN		✓	✓	✗
Reject (Hatami & Chira, 2013)	SVM		✓		✗
RelClass (Parrish et al., 2013)	QDA, Linear SVM		✓	✓	✗
iHMM (Antonucci et al., 2015)	HMM		✓		✗
2step/NoCluster (Tavenard & Malinowski, 2016)	Linear SVM (✓)			✓	train & test
ECDIRE (Mori et al., 2017b)	Gausian Process (✓)		✓		✗
Stopping Rule (Mori et al., 2017a)	Gaussian Process (✓)		✓		train
EARLIEST (Hartvigsen et al., 2019)	LSTM	✓			train
ECEC (Lv et al., 2019)	WEASEL (✓)		✓		train
DDQN (Martinez et al., 2020)	MLP?	✓		✓	train
TEASER (Schäfer & Leser, 2020)	WEASEL (✓)		✓		train
ECONOMY-γ-max (Zafar et al., 2021)	XGBoost + tsfel (✓)			✓	train & test
DETSCNet (Chen et al., 2022)	TCN	✓			train
CALIMERA (Bilski & Jastrzebska, 2023)	MiniROCKET (✓)			✓	train
ELECTS (Rußwurm et al., 2023)	LSTM	✓			train
SOCN (Lv et al., 2023)	FCN		✓		train

Some interesting and well performing approaches …

The ASAP and ALAP baselines

- ASAP triggers all predictions at the first time step.
- ALAP triggers all predictions at the last time step.

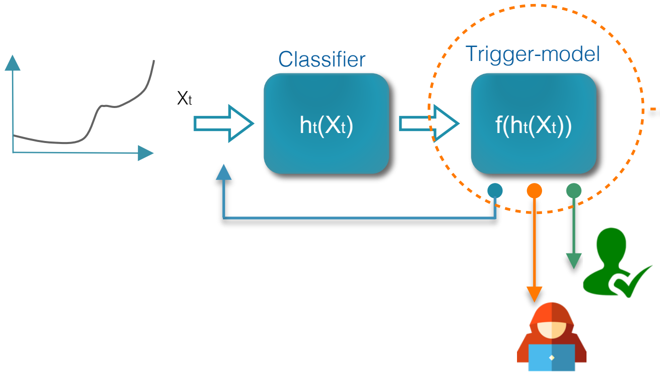
Do these baseline approaches disqualify some literature approaches?

- Not well studied in the literature !
- However, the performance of these baselines is an essential point of comparison.

Proba Theshold baseline

ID

- Separable ✓
- Collection of classifiers ✓
- Confidence based ✓
- Anticipation based ✗
- Cost aware ✓ @train @test



$$f\left(h_t(\mathbf{x}_t)\right) = \begin{cases} 0 & \text{if } \boxed{\max_{y \in Y}(\hat{p}(y|X_t)) \leq \gamma} \\ 1 & \text{otherwise} \end{cases}$$

Trigger time: we make a prediction as soon as the probability of the predicted class reaches a certain threshold.

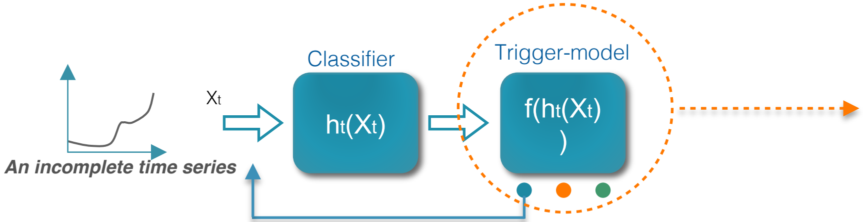
Can we perform better than this simple confidence based approach ?

- Not well studied in the literature !
- Seems to be a natural idea for datascientists

The Stopping Rule approach [1]

ID

- Separable ✓
- Collection of classifiers ✓
- Confidence based ✓
- Anticipation based ✗
- Cost aware ✓ @train @test



Hyper-parameters in [-1,1]

$$f\left(h_t(\mathbf{x}_t)\right) = \begin{cases} 0 & \text{if } \gamma_1 p_1 + \gamma_2 p_2 + \gamma_3 \frac{t}{T} \le 0 \\ 1 & \text{otherwise} \end{cases}$$

Reliability

- the largest posterior probability estimated by the classifier
- the difference between the two largest posterior probabilities

Earliness

- the proportion of the incoming time series that is available at time « *t* »

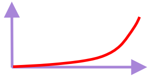
[1] Mori, Usue, et al. "Early classification of time series by simultaneously optimizing the accuracy and earliness." IEEE transactions on neural networks and learning systems (2017)

The ECONOMY approach 1/3

Non-myopia: The objective is to estimate the cost expectation at a fixed future potential triggering time $t+\tau$

	1	0
1	10	5
0	100	0

misclassification cost



$$f_\tau(\mathbf{x}_t) = \mathbb{E}^{t+\tau}\boxed{[\mathrm{C}_m|\mathbf{x}_t]} + \boxed{\mathrm{C}_d(t+\tau)}$$

delay cost

$$\boxed{\sum_{y\in\mathcal{Y}}} P_{t+\tau}(y|\mathbf{x}_t) \boxed{\sum_{\hat{y}\in\mathcal{Y}}} P_{t+\tau}(\hat{y}|y,\mathbf{x}_{t+\tau})\, \mathrm{C}_m(\hat{y}|y)$$

REJECTED

We don't have the true class

the predicted class

The best future triggering moment:

$$\tau^* = \underset{\tau\in\{0,\dots,T-t\}}{\mathrm{ArgMin}}\ f_\tau(\mathbf{x}_t)$$

Triggering condition:

$$\tau^* = 0$$

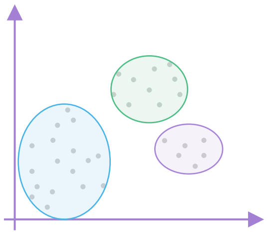
[2] Achenchabe, Youssef, et al. "Early Classification of Time Series. Cost-based Optimization Criterion and Algorithms." 2020, Machine Learning journal

The ECONOMY approach 2/3

Tractable formalization

$$f_\tau(\mathbf{x}_t) = \sum_{\mathfrak{g}_k \in \mathcal{G}} \boxed{P_t(\mathfrak{g}_k|\mathbf{x}_t)} \sum_{y \in \mathcal{Y}} \boxed{P_t(y|\mathfrak{g}_k)} \sum_{\hat{y} \in \mathcal{Y}} \boxed{P_{t+\tau}(\hat{y}|y, \mathfrak{g}_k)} \mathrm{C}_m(\hat{y}|y) + \mathrm{C}_d(t+\tau)$$

We need to build a partition of data



Prior of the class y within the group g^k

Tractable using the confusion matrices of the classifiers, within each group, and for each time horizon

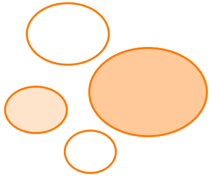
[2] Achenchabe, Youssef, et al. "Early Classification of Time Series. Cost-based Optimization Criterion and Algorithms." 2020, Machine Learning journal

The ECONOMY approach 3/3

ID

- Separable ✓
- Collection of classifiers ✓
- Confidence based ✗
- Anticipation based ✓
- Cost aware ✓ @train @test

$$f_\tau(\mathbf{x}_t) = \sum_{\mathfrak{g}_k \in \mathcal{G}} \boxed{P_t(\mathfrak{g}_k|\mathbf{x}_t)} \sum_{y \in \mathcal{Y}} P_t(y|\mathfrak{g}_k) \sum_{\hat{y} \in \mathcal{Y}} P_{t+\tau}(\hat{y}|y,\mathfrak{g}_k)\, \mathrm{C}_m(\hat{y}|y) + \mathrm{C}_d(t+\tau)$$

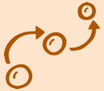


1. How to build the data partition ?
2. How to estimate the probability of belonging to each group ?

Best design choices

Among all the tested algorithms [2], the best one is ECONOMY-γ

confidence level discretization



MC model to anticipate groups

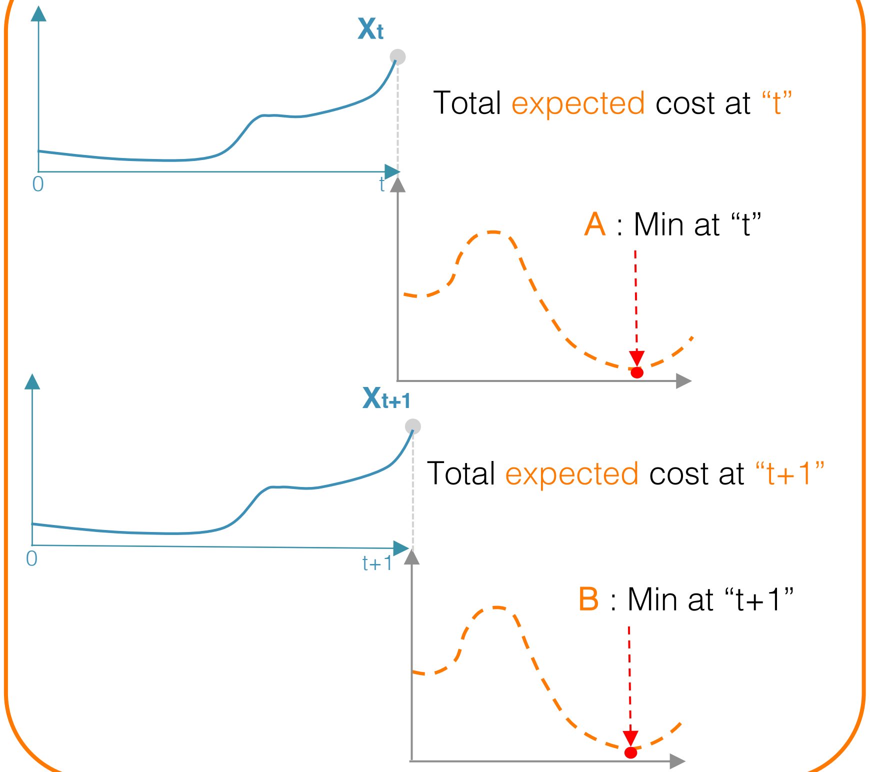
1. It build a data partition by discretizing the output confidence level of each classifier
2. It estimates the probability of belonging to each group at a future time step $t+\tau$ by using a Markov Chain model.

The best ECTS approach !

[2] Achenchabe, Youssef, et al. "Early Classification of Time Series. Cost-based Optimization Criterion and Algorithms." 2020, Machine Learning journal

The CALIMERA approach

Train set : building a regression target = B - A



- Separable
- Collection of classifiers
- Confidence based
- Anticipation based
- Cost aware @train @test

ID

Key ideas :

1. This is a non-myopic approach
2. That uses a well defined regression problem
3. A collection of time indexed regression models is trained
4. In order to detect when the min value is being exceeded.

Shortcoming : when estimating the expected cost

- The misclassification cost matrix is not fully browsed
- Only the predicted class is considered

The ECEC approach

ID

- Separable ✓
- Collection of classifiers ✓
- Confidence based ✓
- Anticipation based ✗
- Cost aware ✓ @train @test

- The objective is to limit variation in the sequence of predicted class values.
- Precision is used as a confidence measure (probability of the true class, given the predicted class)

$$r_H(y \,|\, \hat{y}) = \frac{\left\| \{X_i \,|\, H(X_i) = \hat{y} \,\&\, Y_i = y\} \right\|}{\left\| \{X_i \,|\, H(X_i) = \hat{y}\} \right\|}$$

Number of examples with predicted class y hat and true class y / Number of examples with predicted class y hat

- Confidence in the sequence of predictions is estimated using the independence assumption:

$$C_t(H_t(X)) = 1 - \prod_{k=1}^{t} \boxed{\left(1 - r_{H_k}\left(y = H_t(X) \,|\, H_k(X)\right)\right)}$$

Probability that the ith prediction will differ from the last one.

independence assumption within the sequence of predictions.

- A decision threshold is applied to this sequential confidence measure

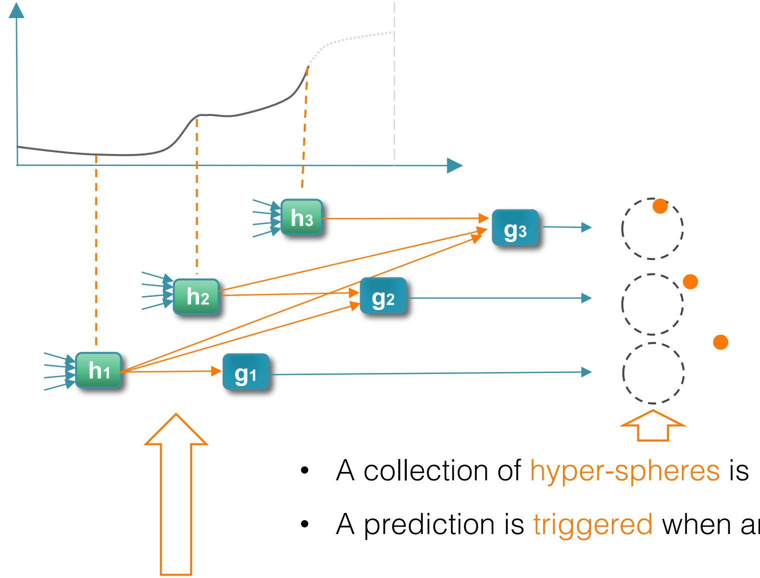
Lv, J., Hu, X., Li, L., & Li, P. (2019). An effective confidence-based early classification of time series. IEEE Access, 7, 96113-96124.

The TEASER approach

ID

- Separable
- Collection of classifiers
- Confidence based
- Anticipation based
- Cost aware @train @test

- Triggering (or not) the predictions becomes a classification problem,
- Aiming to discriminate between good and bad predictions.
- But, the balance of the problem drastically changes depending on the triggering time.
- Thus, the authors proposed to use a collection of one-class SVM.



- A collection of hyper-spheres is learned, representing good predictions.
- A prediction is triggered when an observation falls within it.

The input space dimension of one-class SVMs increases with time.

Schäfer, P., & Leser, U. (2020). TEASER: early and accurate time series classification. Data mining and knowledge discovery, 34(5), 1336-1362.

#4 Experiments

All implemented approaches in the provided library

References	Classifier(s) (collection ✓)	End2end	Confidence	Anticipation	Cost awareness
EDSC (Xing et al., 2011)	Shapelet	✓	✓		✗
ECTS' (Xing, Pei, & Yu, 2012)	1NN		✓	✓	✗
Reject (Hatami & Chira, 2013)	SVM		✓		✗
RelClass (Parrish et al., 2013)	QDA, Linear SVM		✓	✓	✗
iHMM (Antonucci et al., 2015)	HMM		✓		✗
2step/NoCluster (Tavenard & Malinowski, 2016)	Linear SVM (✓)			✓	train & test
ECDIRE (Mori et al., 2017b)	Gausian Process (✓)		✓		✗
Stopping Rule (Mori et al., 2017a)	Gaussian Process (✓)		✓		train
EARLIEST (Hartvigsen et al., 2019)	LSTM	✓			train
ECEC (Lv et al., 2019)	WEASEL (✓)		✓		train
DDQN (Martinez et al., 2020)	MLP?	✓		✓	train
TEASER (Schäfer & Leser, 2020)	WEASEL (✓)		✓		train
ECONOMY-γ-max (Zafar et al., 2021)	XGBoost + tsfel (✓)			✓	train & test
DETSCNet (Chen et al., 2022)	TCN	✓			train
CALIMERA (Bilski & Jastrzebska, 2023)	MiniROCKET (✓)			✓	train
ELECTS (Rußwurm et al., 2023)	LSTM	✓			train
SOCN (Lv et al., 2023)	FCN		✓		train

Four additional approaches, including two that have not been used in experiment due to scalability or performance issues.

Motivation: common pitfalls in the literature

- Variety of setups: with different levels of cost awareness.

- Thus, evaluation issues:

  - Globally hard to define what is « good » ECTS approach.
  - No agreement on evaluation metrics,
  - No baselines comparison
  - Copy-paste of the published results
  - Very simple cost setting
  - Unfair comparisons (e.g. comparing trigger functions, with different classifiers)

- And numerous methodology shortcoming :

  - Z-normalization
  - Biased train/test split
  - No calibration

➔ big accumulated debt
➔ no clear vision of the performance of existing approaches

Reproducibility, Protocol and Evaluation

Trigger function : we provide a modular python library implementing 10 competing approaches and 3 baselines.

Classifier : this choice is fixed, and a collection of calibrated Minirocket estimators is used.

Datasets : UCR Archive for time series classification AND a new non z-normalized collection.

Baselines : blind baselines ASAP and ALAP and proba_threshold.

Evaluation is performers using the *AvgCost* metric which represents the ground truth.

$$AvgCost = \frac{1}{N}\sum_{i=0}^{N} \alpha \times C_m(\hat{y}_i^{\hat{t}}|y_i) + (1-\alpha) \times C_d(\hat{t}_i)$$

We use an additional parameter alpha to control the relative importance of both costs

We can find in the literature …

A rather simple costs definition :

$$C_m(\hat{y}|y) = \mathbb{1}(\hat{y} \neq y)$$
$$C_d(t) = t/T$$

That simplifies the metric: $AvgCost = \alpha \times (1 - Accuracy) + (1-\alpha) \times Earliness$

Extensive experiments

1. Standard cost setting

C_d linear - C_m balanced

2. Anomaly detection cost setting

C_d exponential - C_m imbalanced

3. Conplementary results

Ablation/substitution studies

Derived from 1.

1.2 Removing calibration

1.3 Impact of base classifier

1.5 Impact of z-normalization

Derived from 2.

2.1 Exponential delay cost only

2.2 Imbalanced cost only

2.3 Standard cost setting

Used datasets:

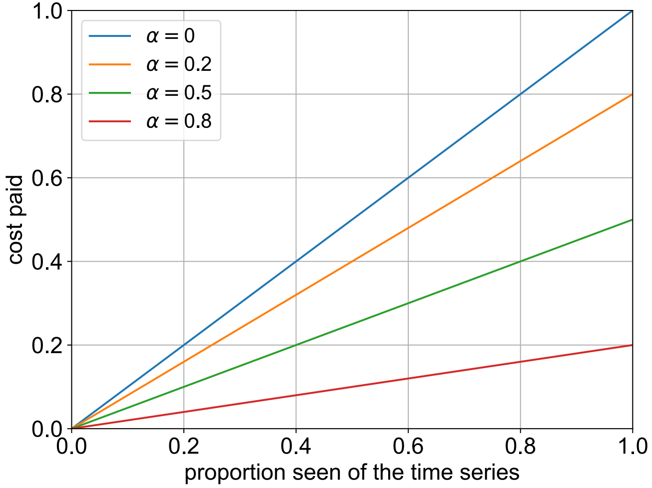
Classically used datasets collection

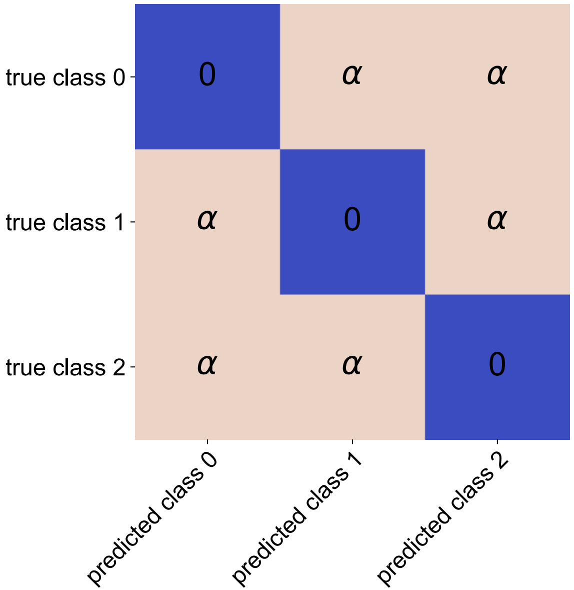
Proposed, non z normalized, datasets collection

Z normalized version of the proposed datasets collection

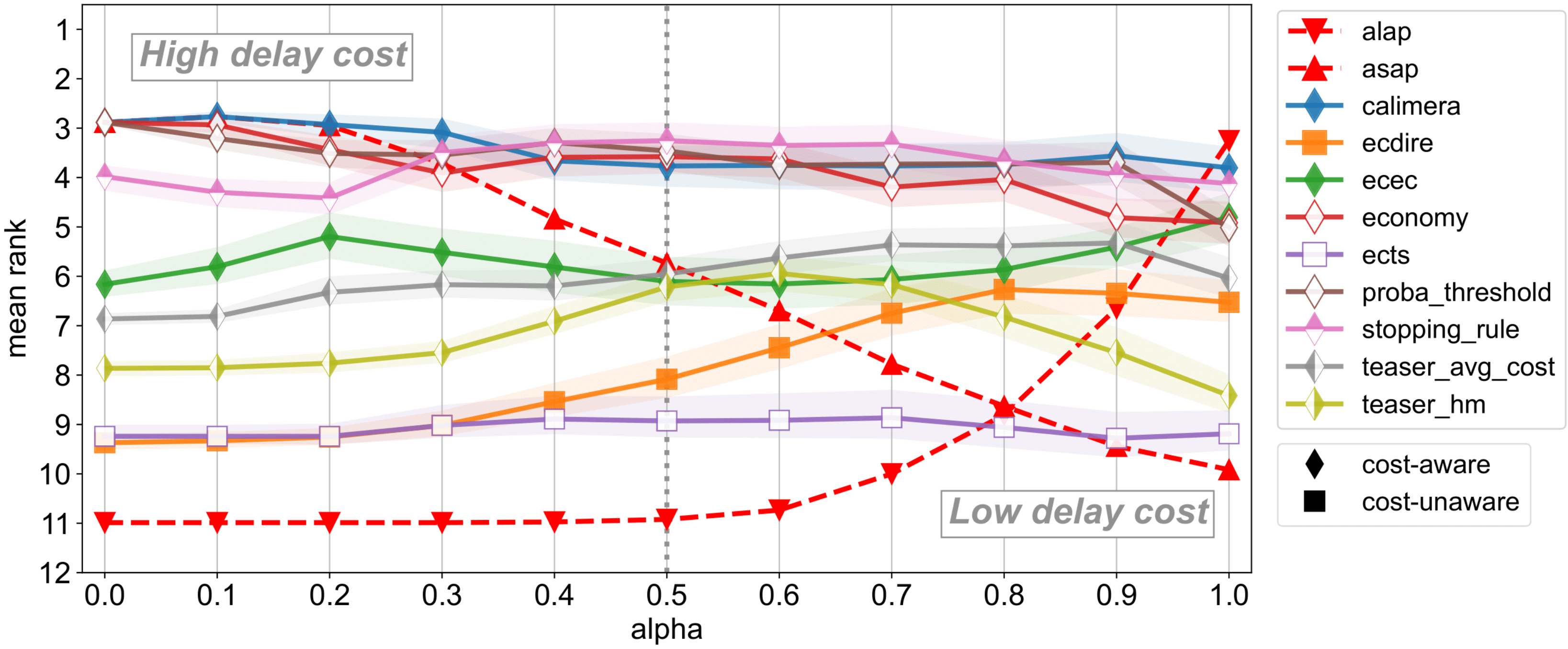
Imbalanced version of datasets collection

Experiments :
Balanced misclassification and linear delay costs



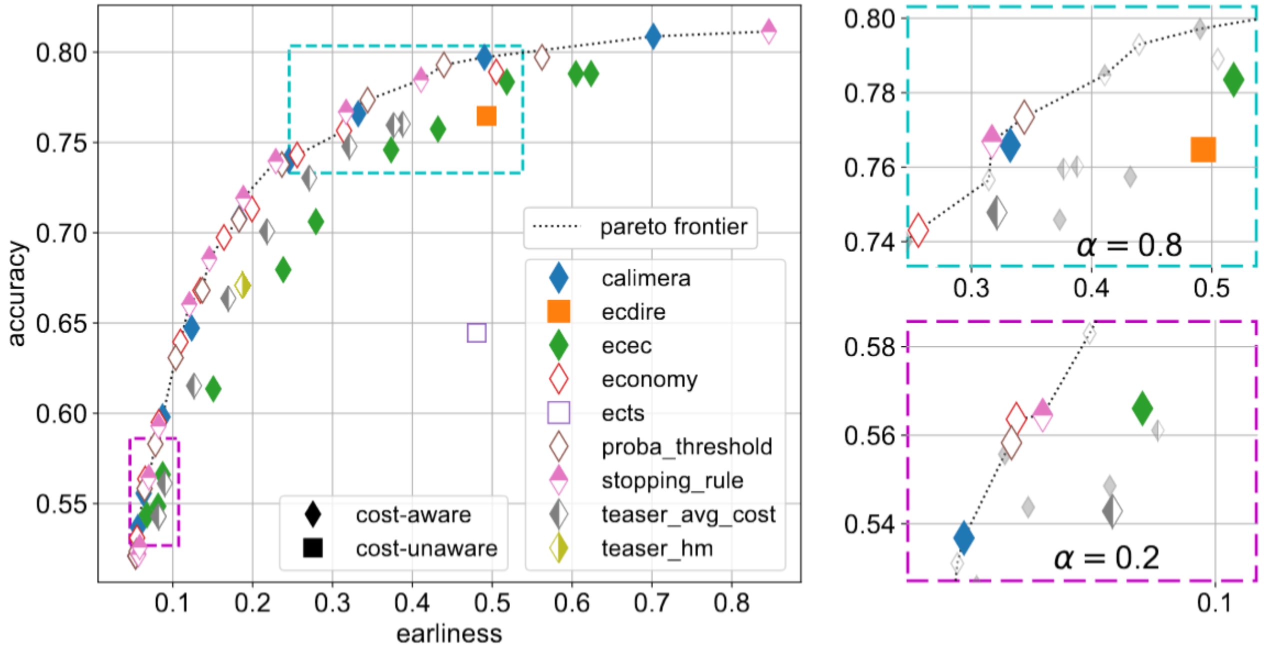


Experiments :
Balanced misclassification and linear delay costs



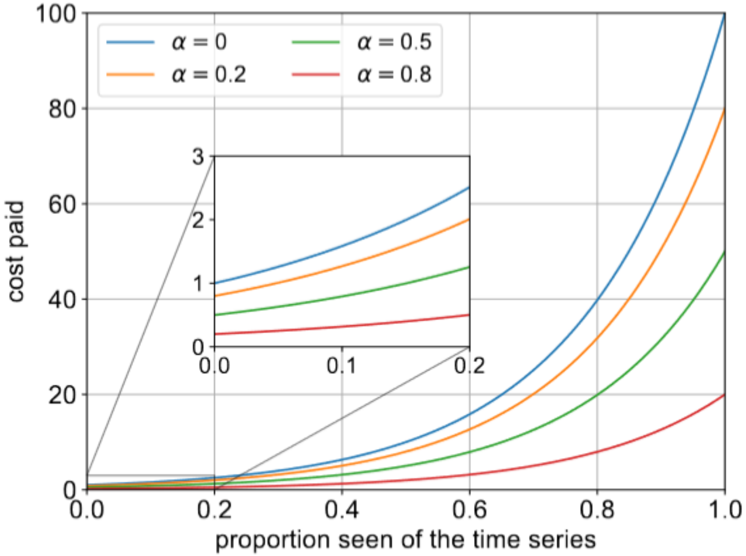
- The ASAP and ALAP baselines disqualify numerous approaches for a large range of alpha.
- Proba_threshold belongs to the top group in this balanced cost setting.

Experiments :
Balanced misclassification and linear delay costs

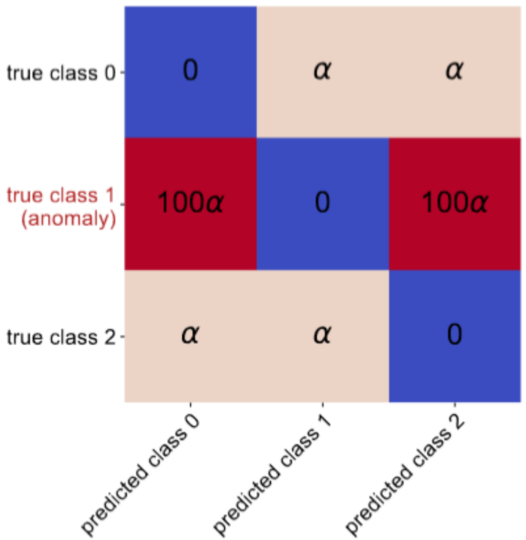


- Different compromise between earliness and accuracy are reached by competing approaches.
- Most of approaches are close to the Pareto front.

Experiments :
Unbalanced misclassification and exponential delay costs

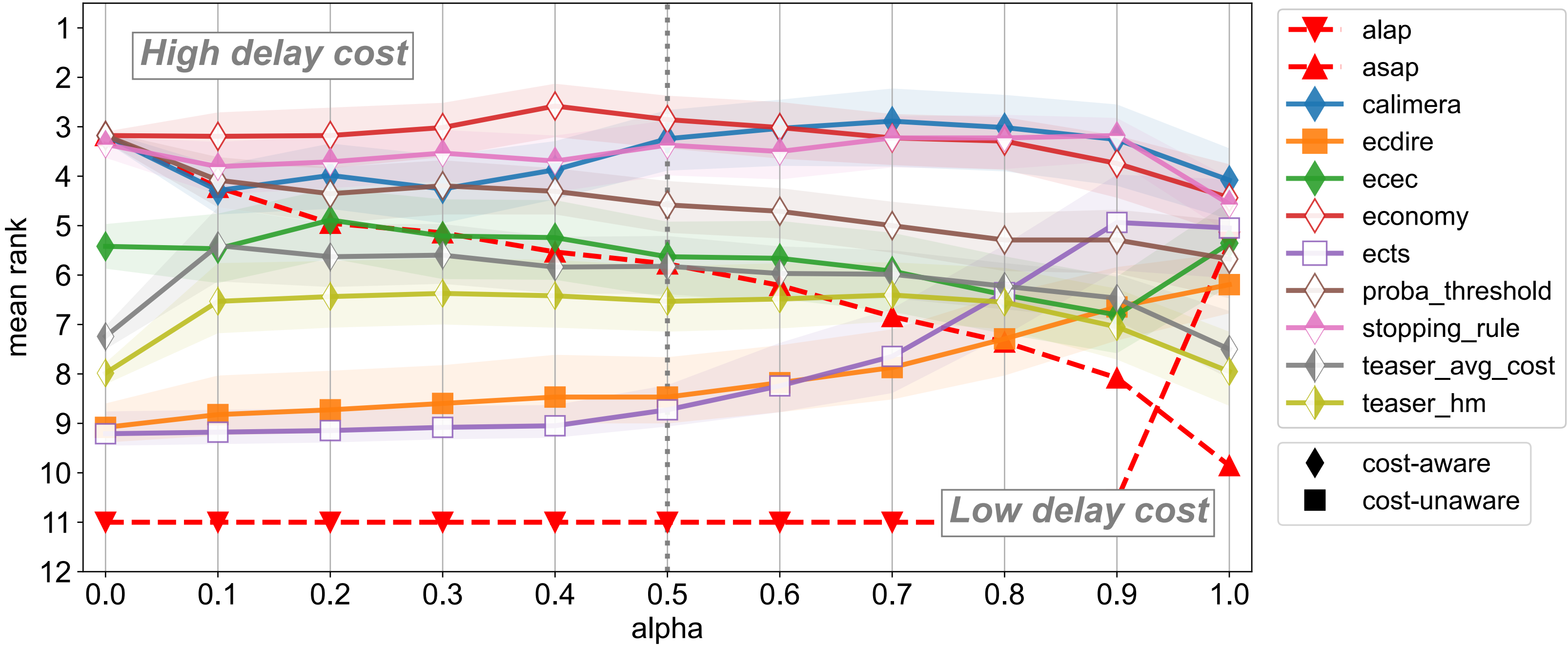


(a) Exponential delay cost



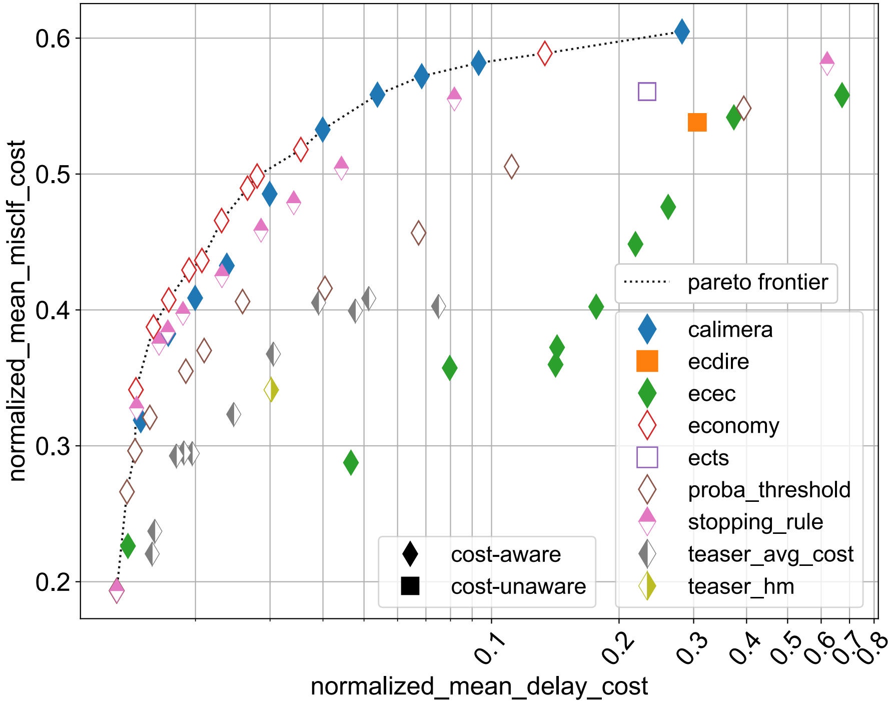
(b) Misclassification cost matrix

Experiments :
Unbalanced misclassification and exponential delay costs



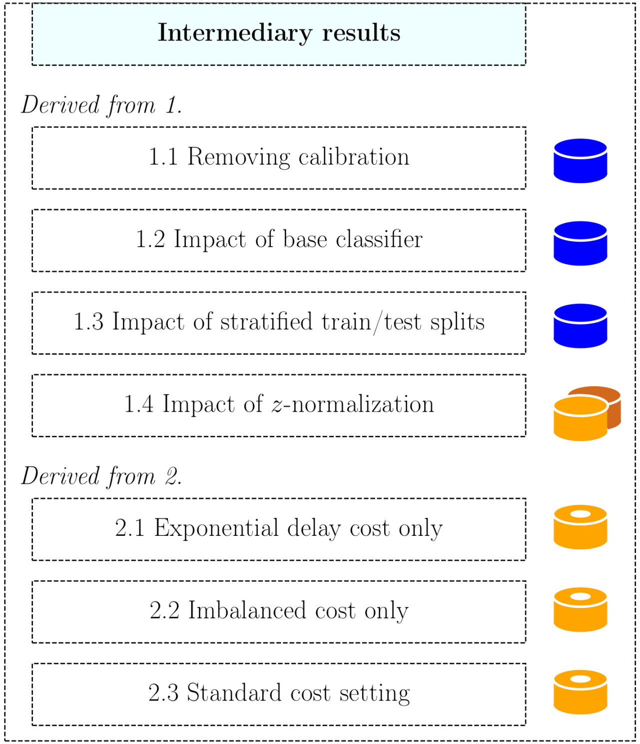
- Proba_threshold becomes easier to beat in this unbalanced cost setting.
- The non-myopia seems to be an advantageous property (see the top group)

Experiments :
Unbalanced misclassification and exponential delay costs



- The Pareto front confirms the advantage of non-myopia

Other experiments :



→ The baseline proba_threshold is penalized

→ Low impact of classifier choice

→ Less significance in statistical tests

→ No significant temporal leakage revealed

Future works :

- Evaluate competing approaches for additional cost setting
    - Cost drift
    - Noisy costs
- Cost sensitive classification stage
- Improve existing ECTS approaches (ex: Calimera).
- Generalization of the ECTS problem … (see next slides)

#5 Python Library

Open source library

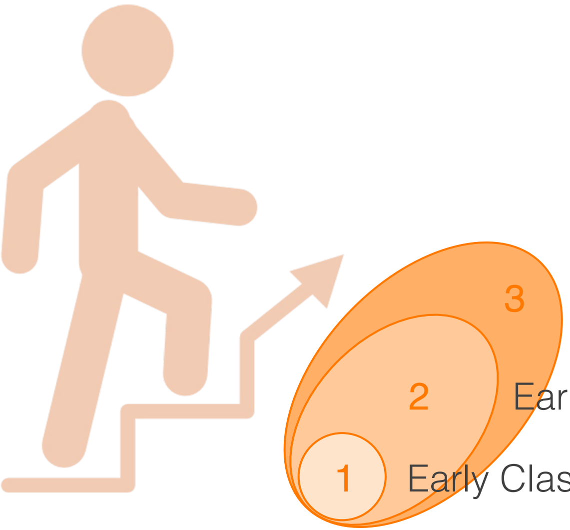
https://github.com/ML-EDM/ml_edm/

#6 ML based Early Decision Making

video playlist

Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Let's present a progression of three problems, toward ML-EDM



3 Machine Learning based Early Decision-Making (ML-EDM)

2 Early Decision to be located in time

1 Early Classification of Time Series

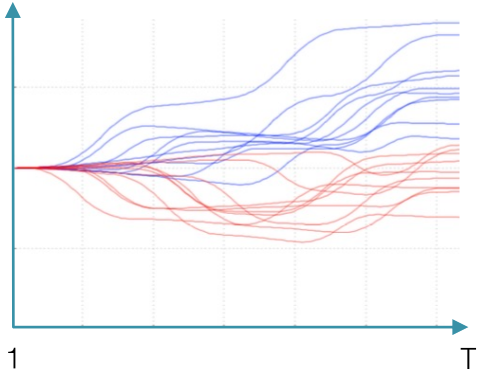
Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Problem #1 : ECTS is our starting point

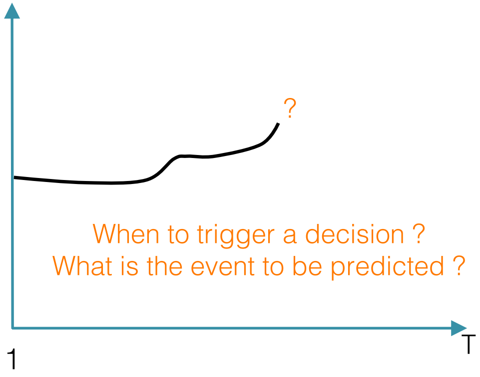
ID card of ECTS

- Learning task: a classification problem
- Data type: time series
- Decision-making period: a fixed time interval [1,T]
- Decision type: (i) unique, (ii) related to the entire period [1,T], (iii) and final.

The training set



At runtime



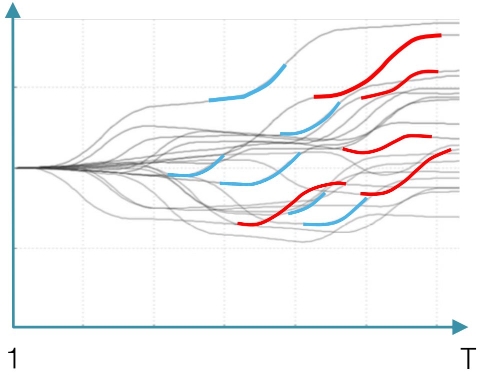
Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Problem #2 : Let 's localize our decisions, while keeping them early …

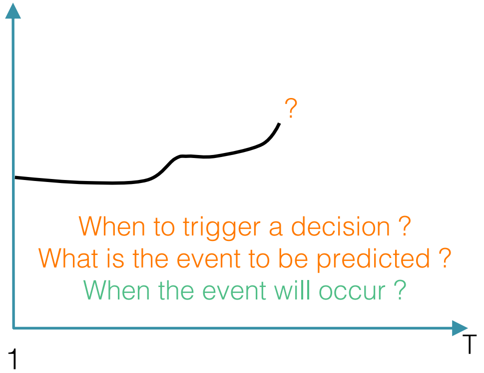
ID card of Early Decision to be Located in Time

- **Learning task:** a classification problem
- **Data type:** time series
- **Decision-making period**: a fixed time interval [1,T]
- **Decision type:** (i) unique, (ii) related to a sub period [s,e] in [1,T], (iii) and final.

The training set



At runtime



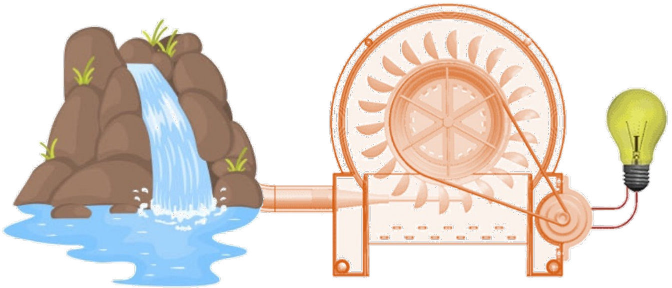
Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Problem #2 : Let 's localize our decisions, while keeping them early …

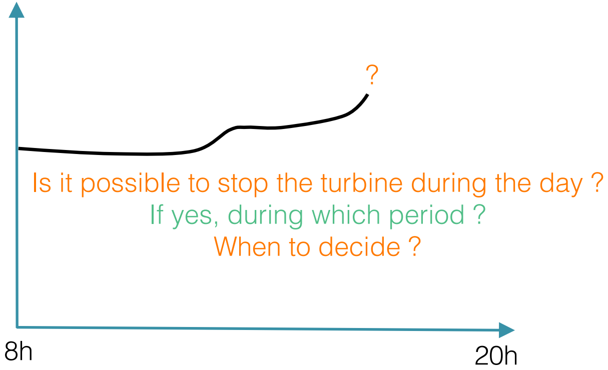
ID card of Early Decision to be Located in Time

- Learning task: a classification problem
- Data type: time series
- Decision-making period: a fixed time interval [1,T]
- Decision type: (i) unique, (ii) related to a sub period [s,e] in [1,T], (iii) and final.

Example: In a hydroelectric dam, turbines can be stopped for maintenance operations, only if the electricity demand is low enough.



At runtime



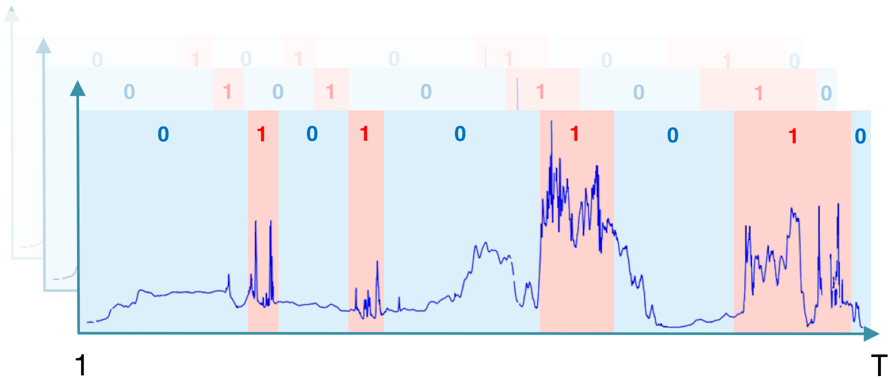
Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Problem #3 : Let 's consider multiple early decisions with ML-EDM

ID card of ML-EDM: Early Multiple Decision to be Located in Time

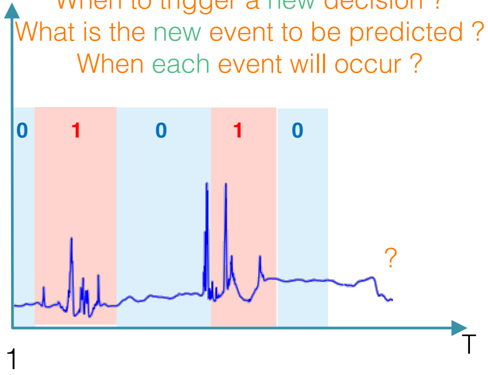
- **Learning task:** a classification problem
- **Data type:** time series
- **Decision-making period**: a fixed time interval [1,T]
- **Decision type:** (i) multiple, (ii) related to a sub period [s,e] in [1,T], (iii) and final.

The training set made of such examples:



At runtime

When to trigger a new decision ?
What is the new event to be predicted ?
When each event will occur ?

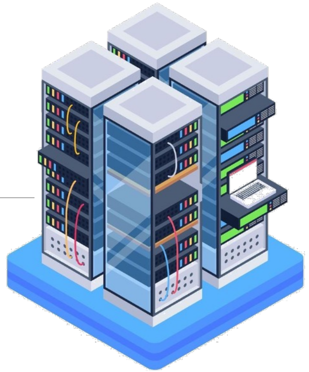


Open challenges for machine learning based early decision-making research. *ACM SIGKDD Explorations Newsletter*, *24*(2), 12-31.
Bondu, A., Achenchabe, Y., Bifet, A., Clérot, F., Cornuéjols, A., Gama, J., ... & Marteau, P. F. (2022).

Problem #3 : Let ‘s consider multiple early decisions with ML-EDM

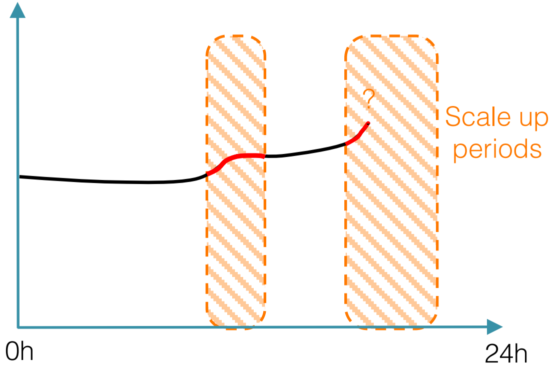
ID card of **ML-EDM**: Early Multiple Decision to be Located in Time

- **Learning task:** a classification problem
- **Data type:** time series
- **Decision-making period**: a fixed time interval [1,T]
- **Decision type:** (i) multiple, (ii) related to a sub period [s,e] in [1,T], (iii) and final.



Example: Using a cloud platform, identify as soon as possible the periods when it is necessary to scale up or down the number of nodes.

At runtime



will there be scale up (or down) periods?
If yes, during which periods?
When to decide?

Overview of the proposed ML-EDM Challenges

We have defined what the ML-EDM problem is ….

ID card of **ML-EDM**

- **Learning task:** a classification problem
- **Data type:** time series
- **Decision-making period**: a fixed time interval [1,T]
- **Decision type:** (i) multiple, (ii) related to a sub periods [s,e] in [1,T], (iii) and final.

Challenges have been proposed

Conclusion : overview of the proposed challenges

#1- Extending non-myopia to **unsupervised** approaches

#2- Addressing **other supervised** learning tasks

#3- Early **Weakly** Supervised Learning

#4 – Data type agnostic ML-EDM

#5 – Online prediction to be located in time

#6 – Online accuracy vs. earliness trade-off

#7 – Management of non-stationarity in ML-EDM

#8 – Reactivity vs. stability dilemma for revocable decisions

#9 – Non-myopia to revocation risk

#10 – Time-dependent decision costs

- Related to the learning task
- Related to the online setting
- Related to revocable decisions

One more thing ?

#5 – Online prediction to be located in time

#6 – Online accuracy vs. earliness trade-off

#8 – Reactivity vs. stability dilemma for revocable decisions

#9 – Non-myopia to revocation risk

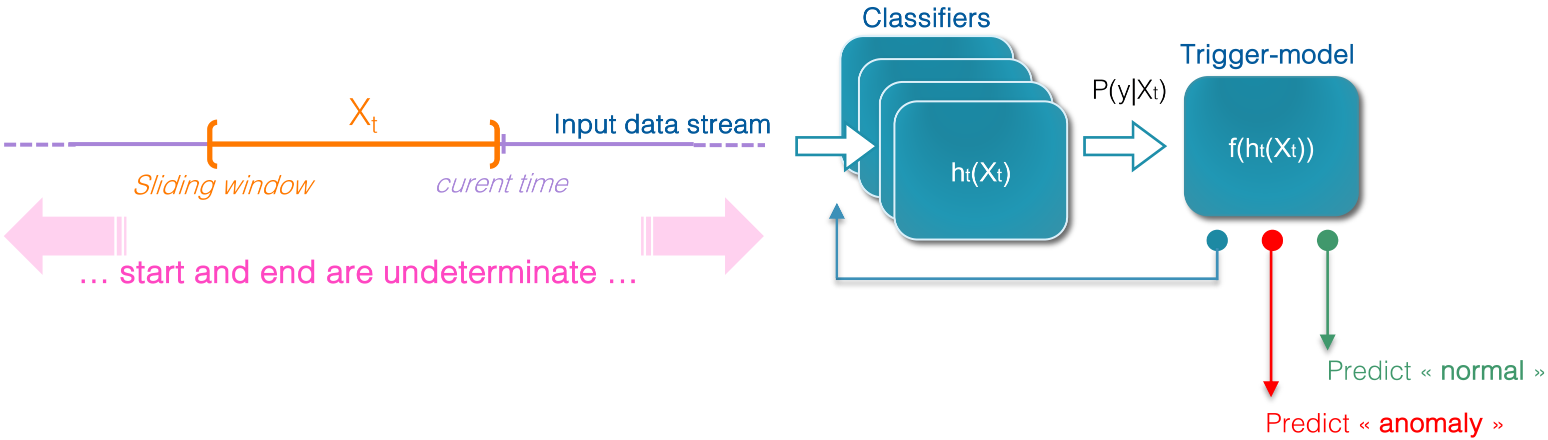
#2 ECTS with revocable decisions

#1 Early decision on open time series

Appendix #1

Early decision on open time series

Data stream can be considered

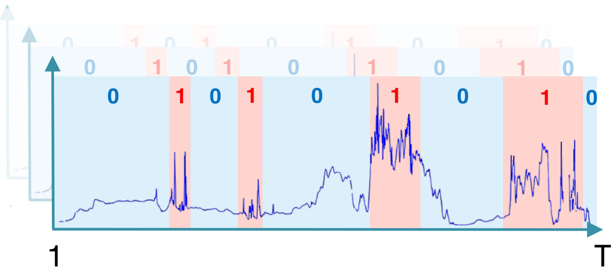


Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, December). When to Classify Events in Open Times Series? In *Asian Conference on Machine Learning (ACML)*.

Data stream should be considered, to enable important applications

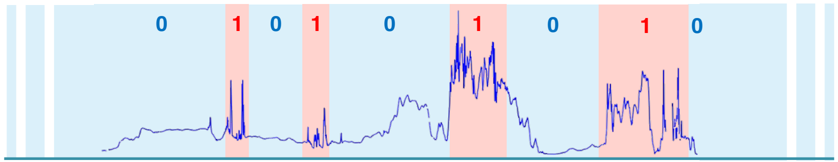
Example: continuous monitoring.

Train set (ML-EDM)



Historical data is split into periods of size T to form a collection of training examples

Train set (online ML-EDM)



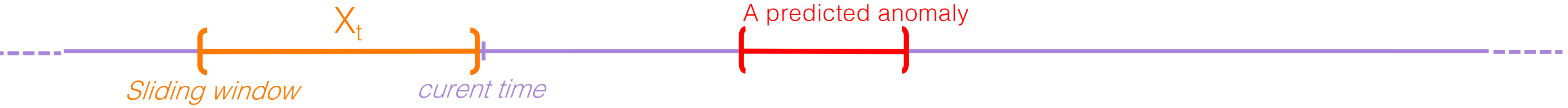
Historical data is considered in its fullness, in a continuous way (i.e. without any a priori splitting)

Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, December). When to Classify Events in Open Times Series? In *Asian Conference on Machine Learning (ACML)*.

A complex kind of prediction

Challenge #5 : Online prediction to be located in time

1. We need to predict a class label (e.g. anomaly)
2. We need to predict the associated time interval

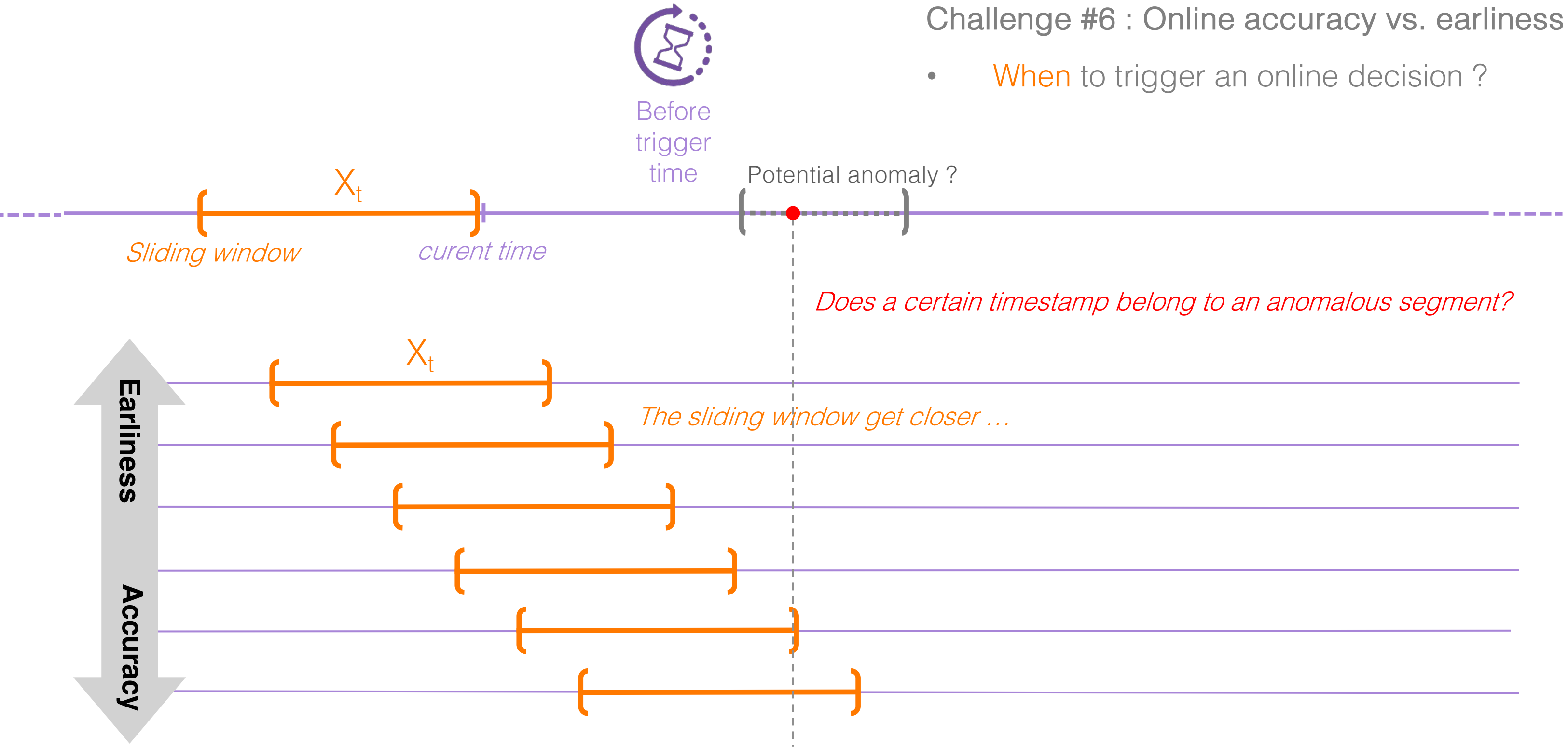


Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, December). When to Classify Events in Open Times Series? In *Asian Conference on Machine Learning (ACML)*.

What is precocity in this case?

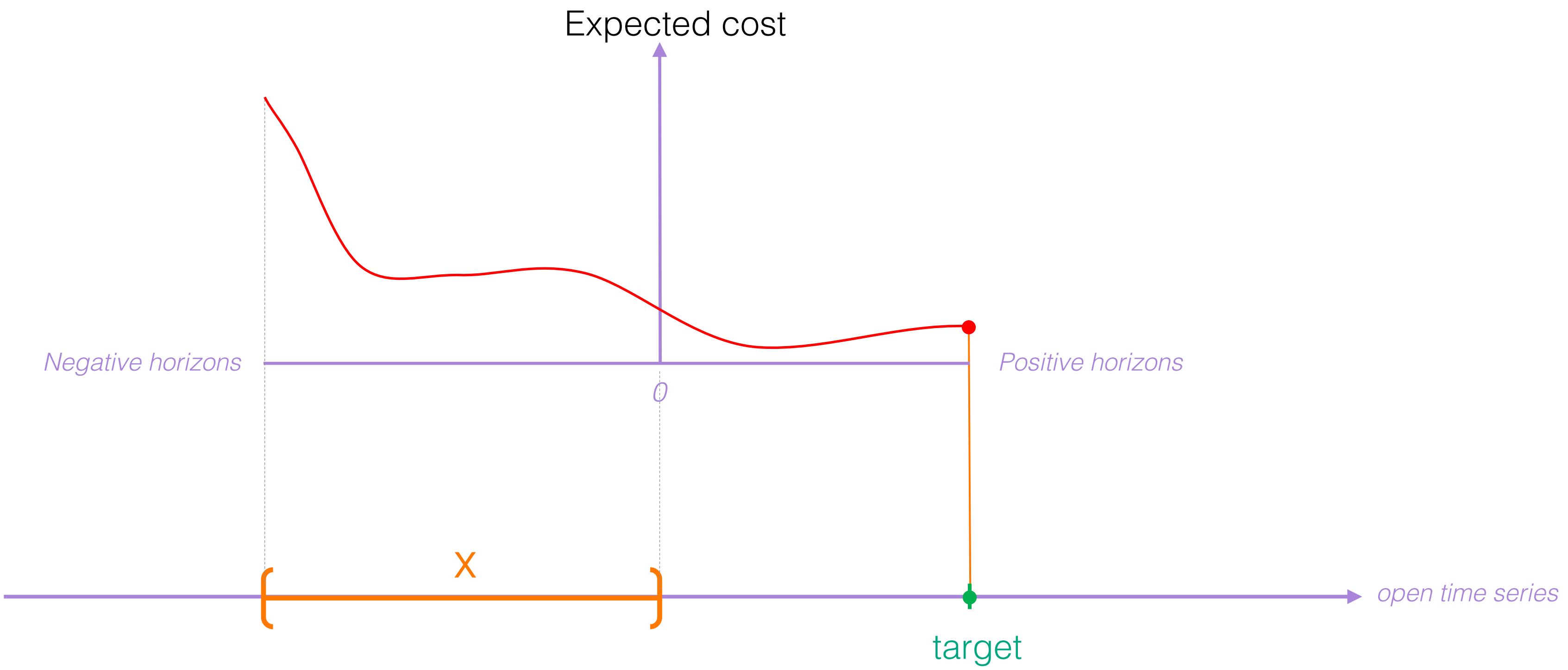
Challenge #6 : Online accuracy vs. earliness trade-off

- When to trigger an online decision ?

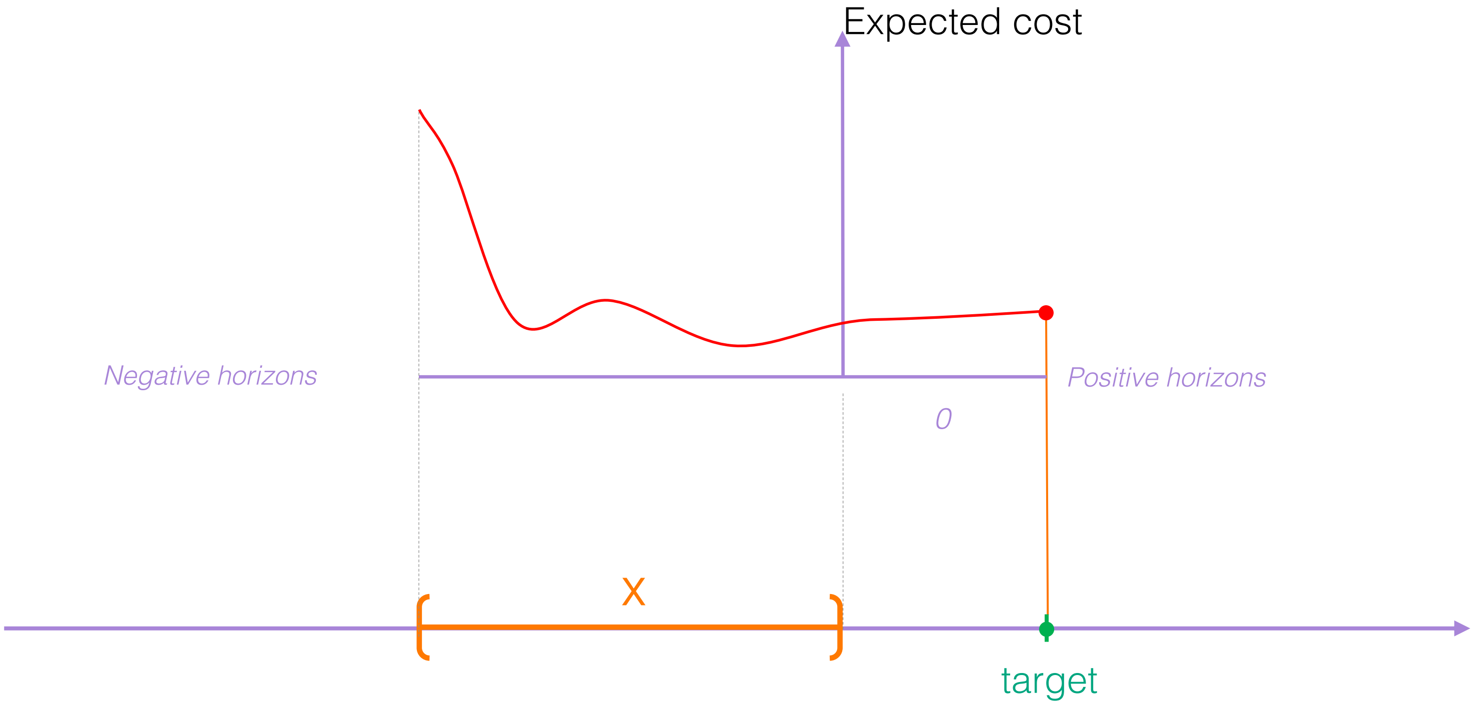


Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, December). When to Classify Events in Open Times Series? In *Asian Conference on Machine Learning (ACML)*.

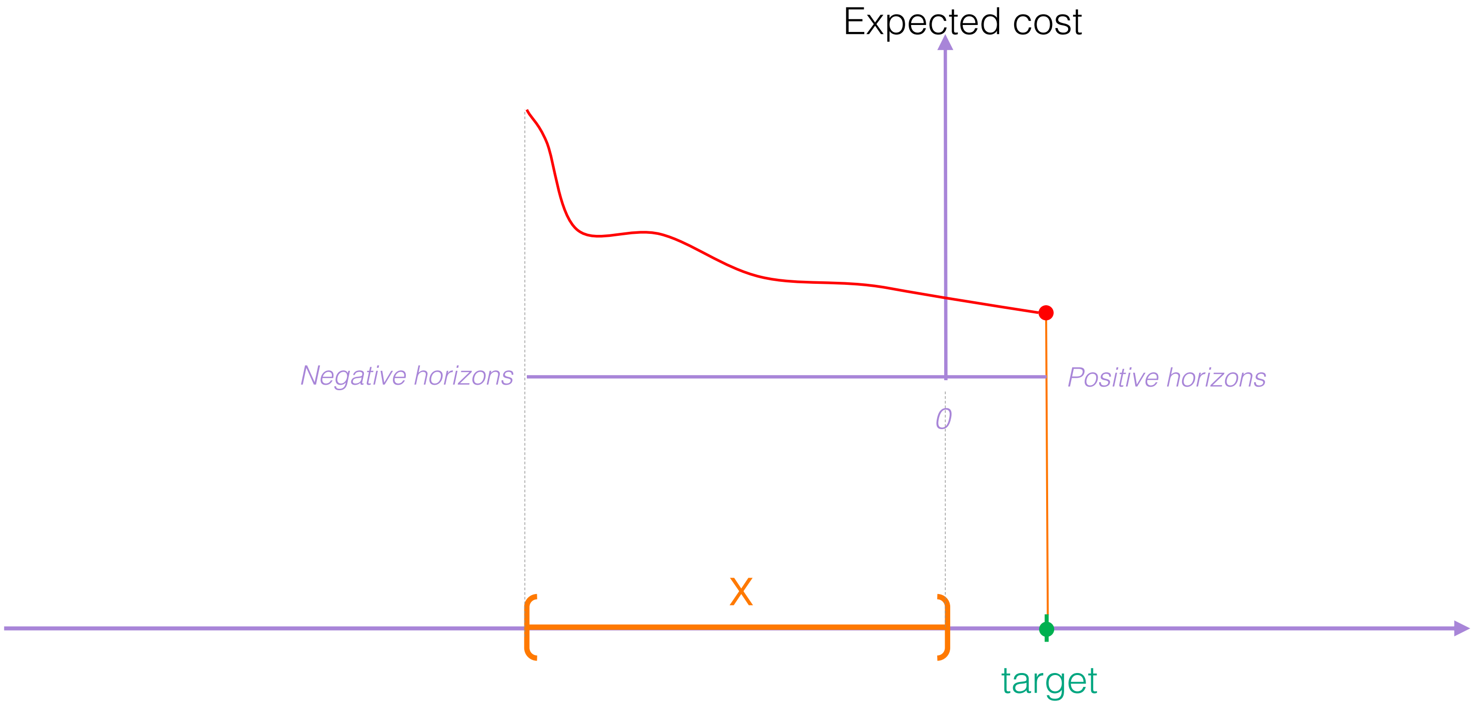
What is non-myopia in this case?



What is non-myopia in this case?



What is non-myopia in this case?



Experiments : Dataset

- Real dataset from one of the schwan's Factories (Frozen food industry corporation).
- 100 multivariate open time series associated with 100 machines monitored for a period of 1 year, labeled periods of failure.
- Features include pressure, rotation, voltage, vibration, presence of a device error.

Source: Google images

Experiments : Competing approaches

- Baselines:
  - Late baseline
  - Early baseline
- Adapted methods from ECTS:
  - Economy-γ [Achenchabe et al.]
  - SR [Mori et al.]
- CC (Consecutive classifiers): basic threshold based approach

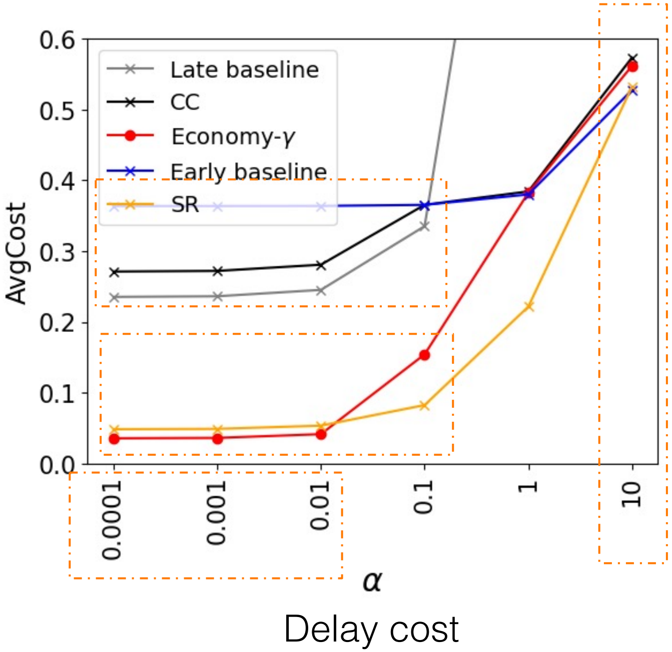
Experimental results

- When the cost of delaying decision is high, the optimal strategy is the ``early'' baseline.

- When the delay cost is low, taking late decisions is a good strategy even though it is not optimal.

- The CC method essentially switches from one baseline strategy to the other one as the delay cost increases.

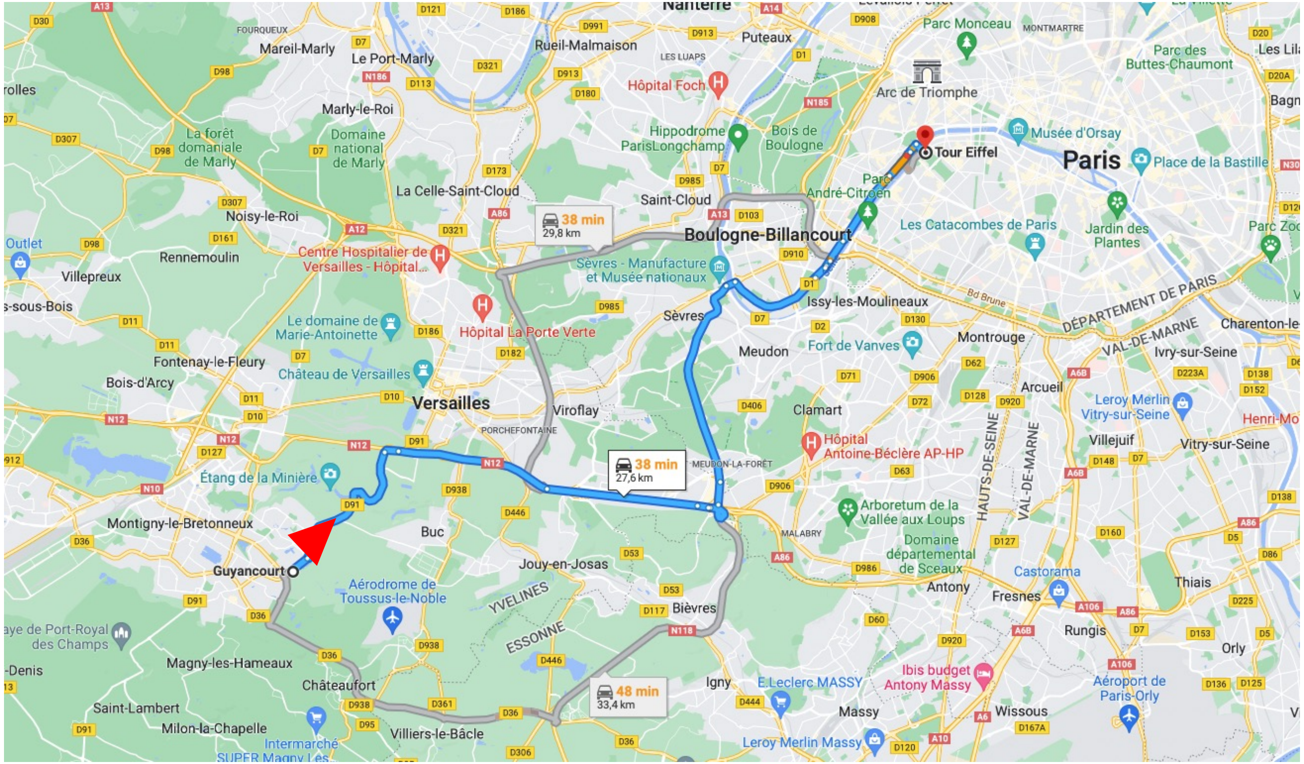
- Both methods adapted from ECTS: Economy-**γ** and SR, noticeably overcome CC.



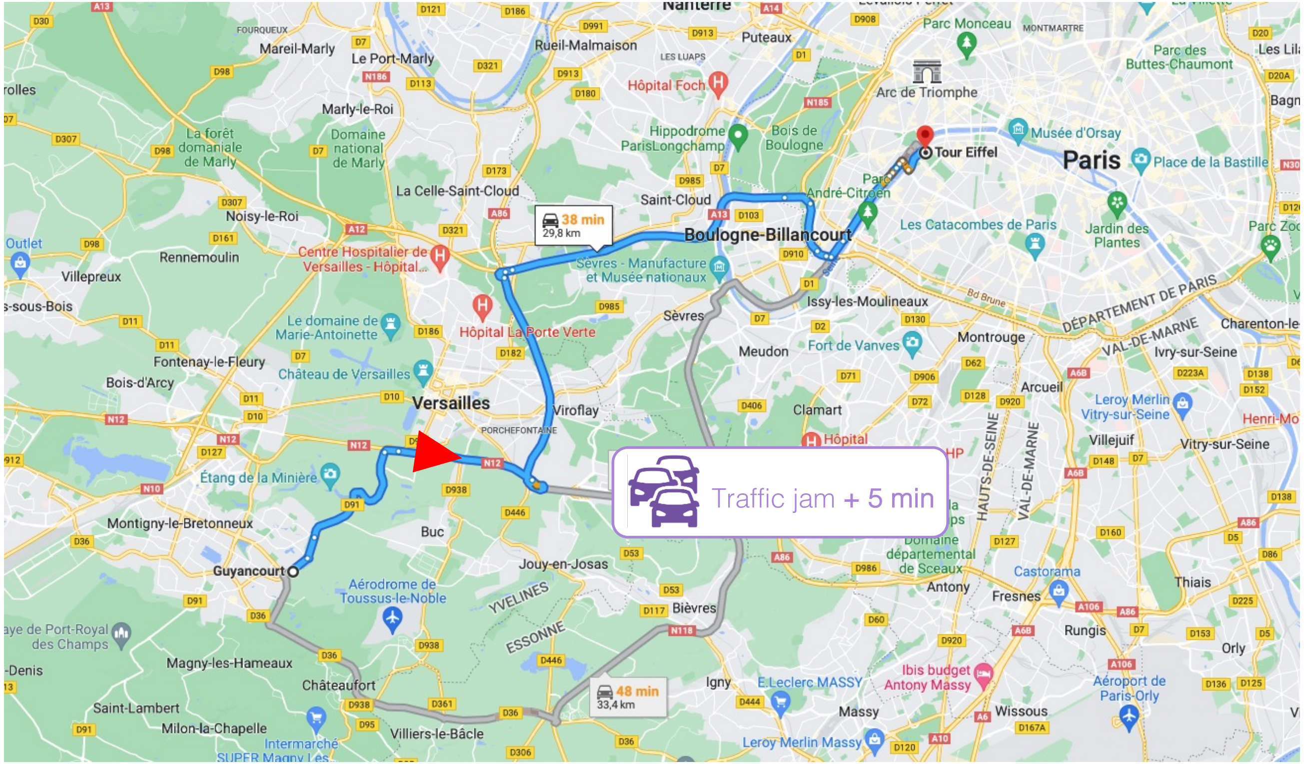
Appendix #2

ECTS with revocable decisions

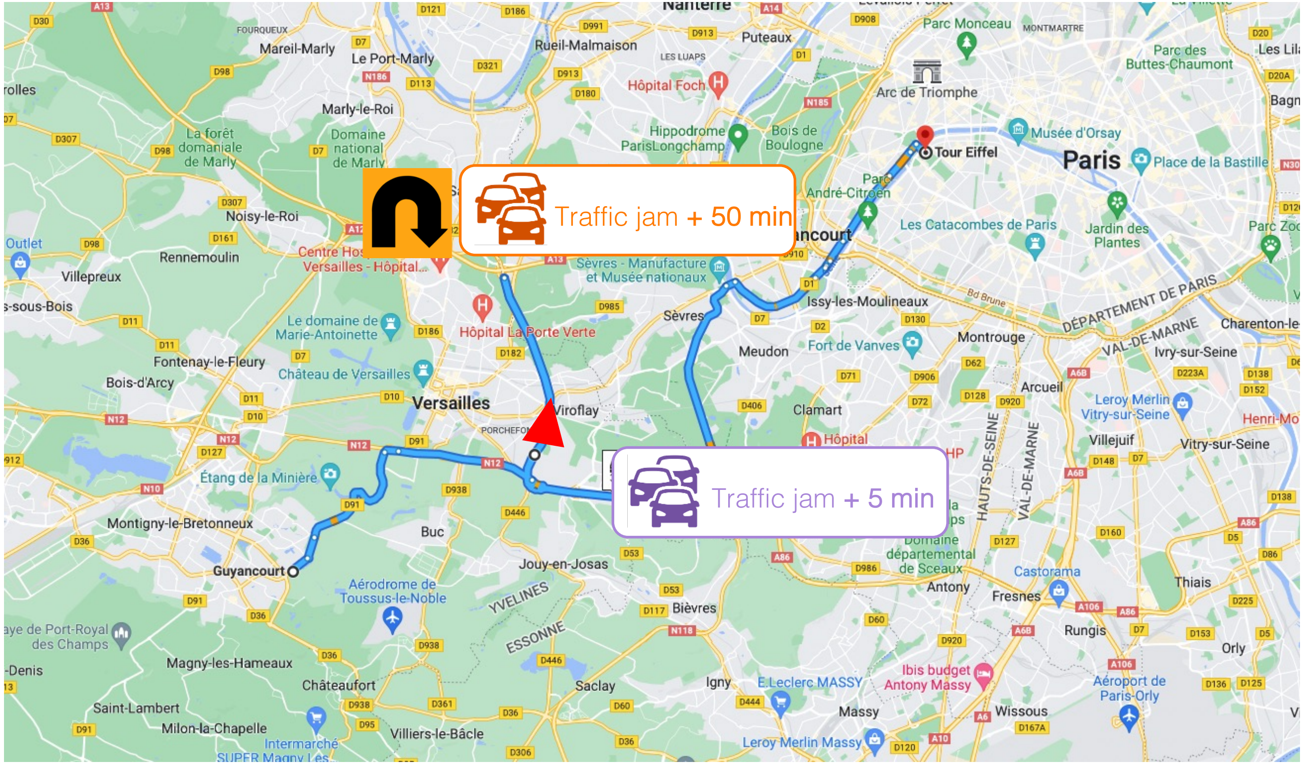
What is revoking a decision?



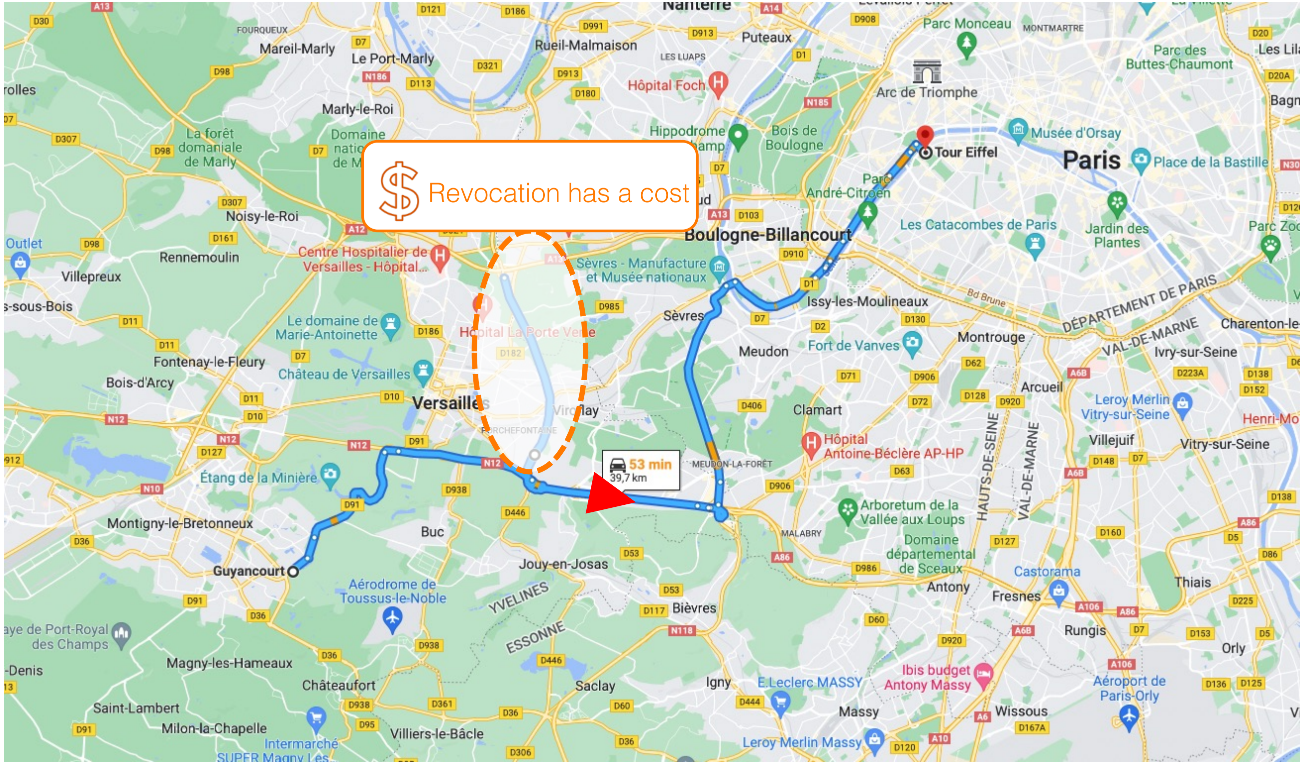
What is revoking a decision?



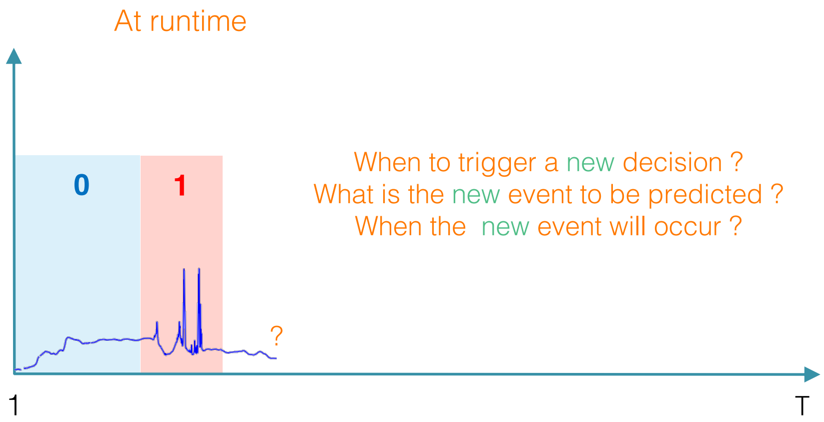
What is revoking a decision?



What is revoking a decision?

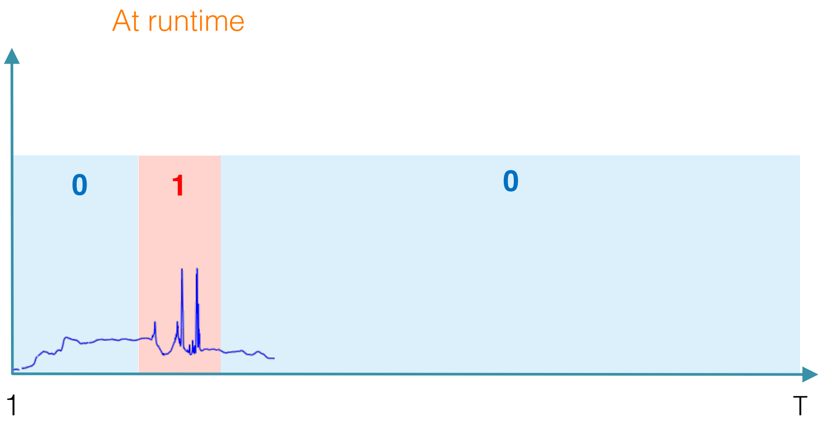


Back to ML-EDM, with revocable decisions

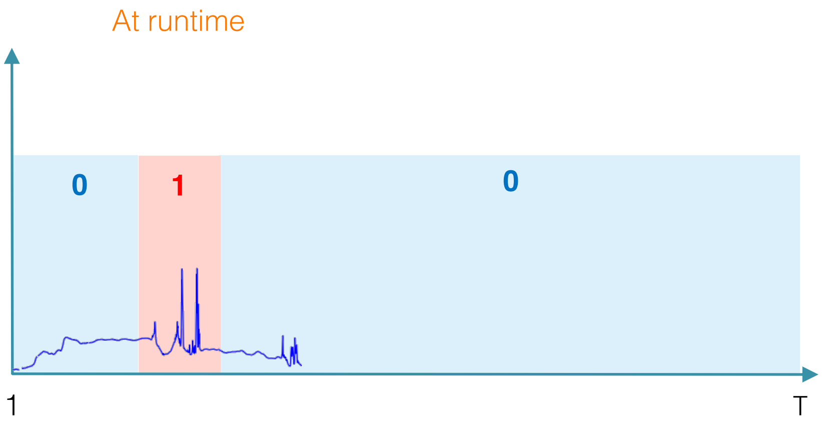


When to trigger a new decision ?
What is the new event to be predicted ?
When the new event will occur ?

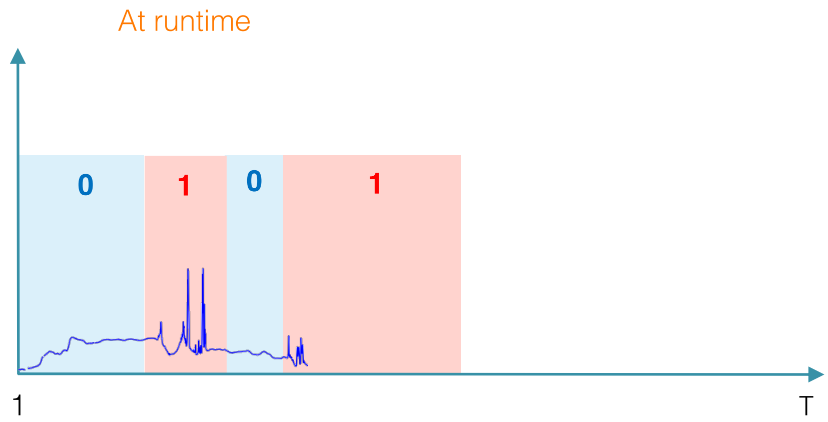
Back to ML-EDM, with revocable decisions



Back to ML-EDM, with revocable decisions

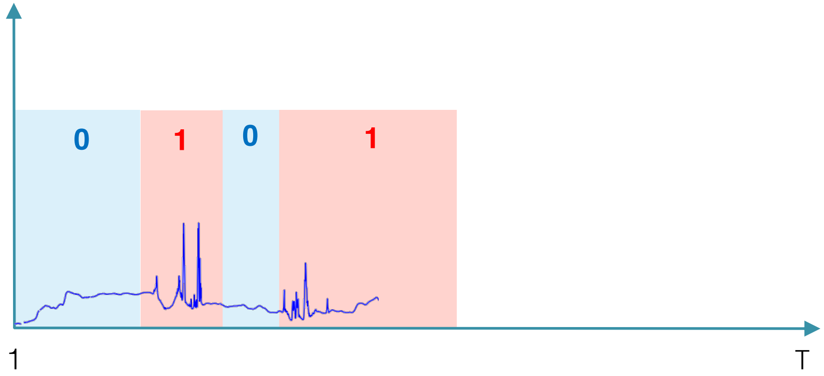


Back to ML-EDM, with revocable decisions

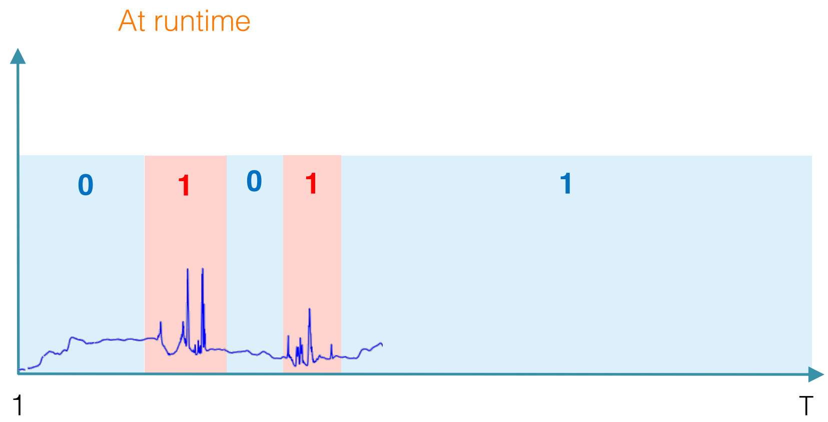


Back to ML-EDM, with revocable decisions

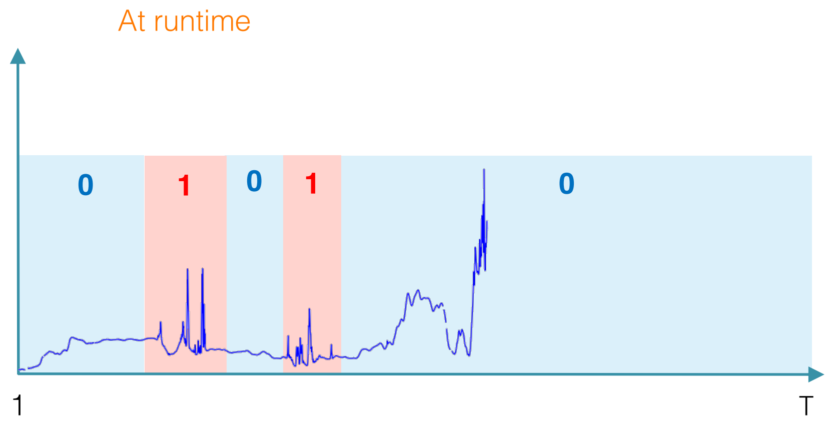
At runtime



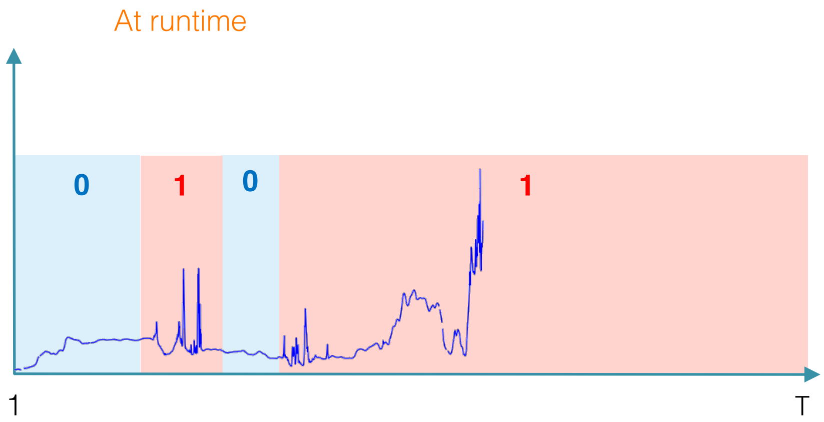
Back to ML-EDM, with revocable decisions



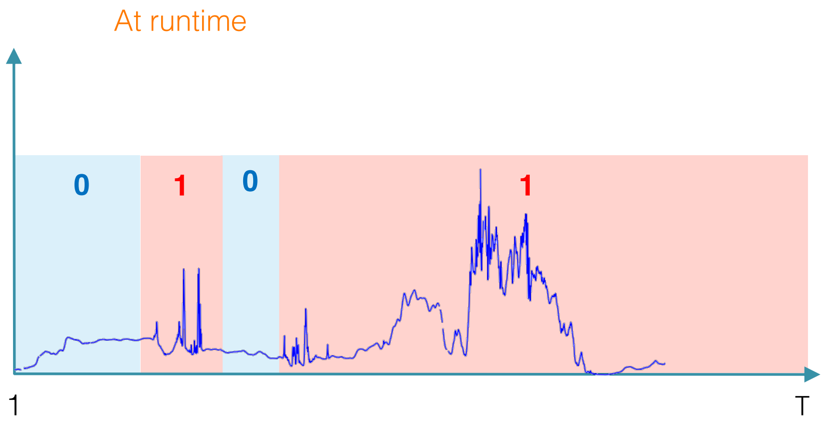
Back to ML-EDM, with revocable decisions



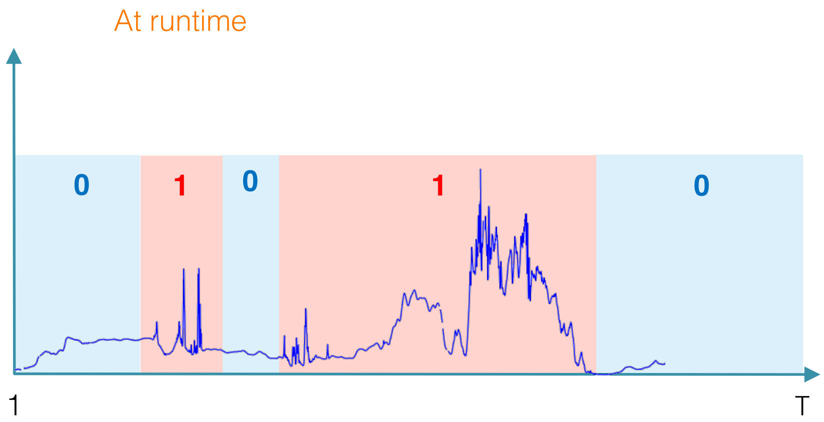
Back to ML-EDM, with revocable decisions



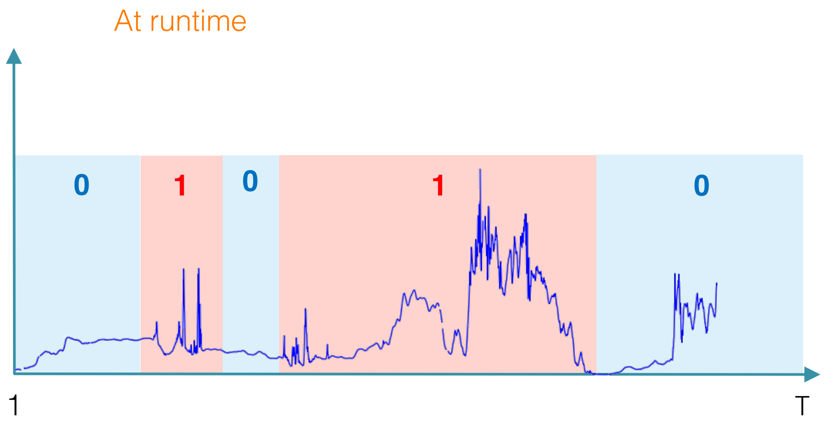
Back to ML-EDM, with revocable decisions



Back to ML-EDM, with revocable decisions

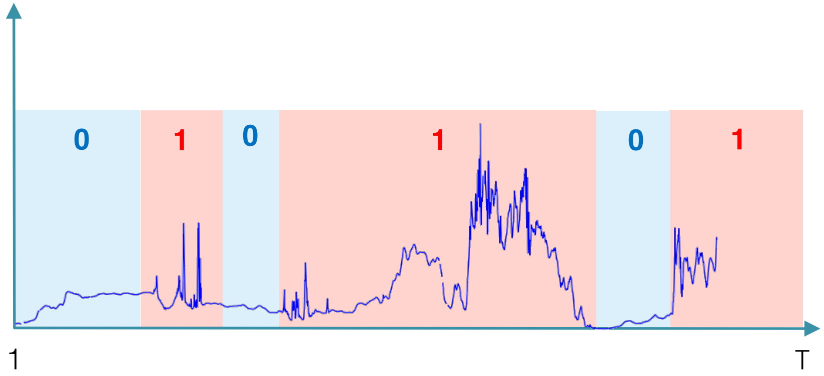


Back to ML-EDM, with revocable decisions

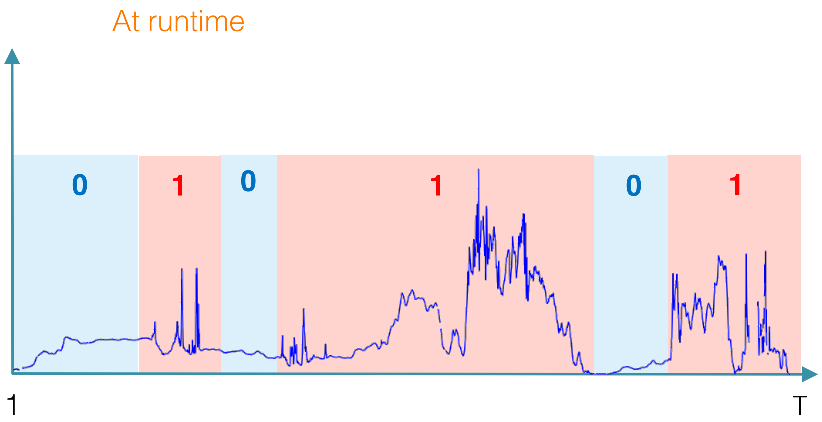


Back to ML-EDM, with revocable decisions

At runtime



Back to ML-EDM, with revocable decisions



Back to the simple case of ECTS,
with the ECONOMY approach extended to revocable decisions

An additional input :

Miss classification costs

$\hat{y}/y$	1	0
1	10	5
0	100	0

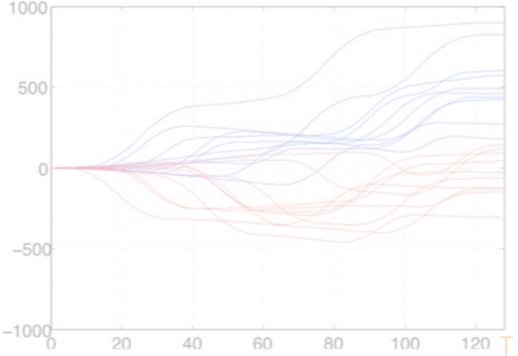
Cost of changing a decision

$\hat{y}_1/\hat{y}_2$	1	0
1	0	5
0	1	0

$$C_{cd}(\hat{y}_1|\hat{y}_2) : \mathcal{Y} \times \mathcal{Y} \to R$$

Delay cost

The training set



Challenge #8: Reactivity vs. stability dilemma for revocable decisions

- How to combat too frequent changes and unstable decisions?

[4] Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, July). Early and revocable time series classification. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Back to the simple case of ECTS,
with the ECONOMY approach extended to revocable decisions

What's new in the decision criterion ?

$$f_\tau^{\text{rev}}(\mathcal{D}_k, \tilde{t} \mid \mathbf{x}_t) = \mathbb{E}_{(\hat{y},y)\in\mathcal{Y}^2}^{t+\tau}\left[\mathrm{C}_m(\hat{y}|y) \mid \mathbf{x}_t\right] + \sum_{\substack{i=1 \\ \hat{y}_{t_i},\hat{y}_{t_{i+1}}\in\mathcal{D}_k}}^{k-1} \mathrm{C}_{cd}(\hat{y}_{t_{i+1}}|\hat{y}_{t_i}) + \mathbb{E}_{\hat{y}\in\mathcal{Y}}^{t+\tau}\left[\mathrm{C}_{cd}(\hat{y}|\hat{y}_{t_k}) \mid \mathbf{x}_t\right] + \mathrm{C}_d(\tilde{t})$$

Expected misclassification cost at ***t+tau***

Delay cost

We consider the entire sequence of decisions

The cost of **past** decision changes

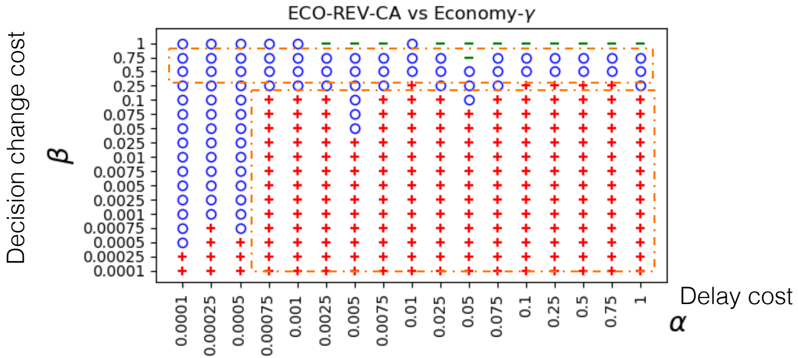
Expected cost of changing a decision at the **future** time steps « t+tau »

Challenge #9: Non-myopia to revocation risk

- How to delay a decision that risks being revoked later?

[4] Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, July). Early and revocable time series classification. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Back to the simple case of ECTS,
with the ECONOMY approach extended to revocable decisions

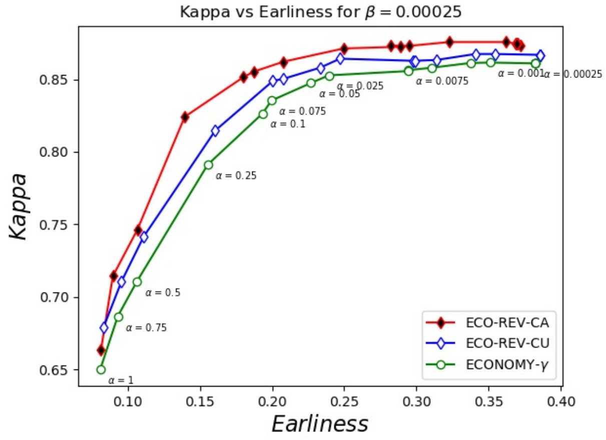


+ : Statistically significant difference in favor ECO-REV-CA

– : Statistically significant difference in favor of ECO-gamma

o : No significant difference

Results of statistical testing significance for multiple combinations of delay and misclassification costs



Predictive performance against earliness for multiple values of the delay cost

[4] Achenchabe, Y., Bondu, A., Cornuéjols, A., & Lemaire, V. (2022, July). Early and revocable time series classification. In *2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.