

When to Classify Events in Open Times Series?

Youssef Achenchabe

YOUSSEF.ACHENCHABE@UNIVERSITE-PARIS-SACLAY.FR

Paris-Saclay University, Agroparistech and Orange Labs.

Alexis Bondu

ALEXIS.BONDU@ORANGE.COM

Orange Labs, Châtillon.

Antoine Cornuéjols

ANTOINE.CORNUEJOLS@AGROPARISTECH.FR

Paris-Saclay University, Agroparistech.

Vincent Lemaire

VINCENT.LEMAIRE@ORANGE.COM

Orange Labs, Lannion.

Editors: Emtiyaz Khan and Mehmet Gonen

Abstract

In numerous applications, for instance in predictive maintenance, there is a pressure to predict events ahead of time with as much accuracy as possible while not delaying the decision unduly. This translates in the optimization of a trade-off between earliness and accuracy of the decisions, that has been the subject of research for time series of finite length and with a unique label. And this has led to powerful algorithms for Early Classification of Time Series (ECTS). This paper, for the first time,

investigates such a trade-off when events of different classes occur in a streaming fashion, with no predefined end. In the *Early Classification in Open Time Series* problem (ECOTS), the task is to predict events, i.e. their class and time interval, at the moment that optimizes the accuracy vs. earliness trade-off. Interestingly, we find that ECTS algorithms can be sensibly adapted in a principled way to this new problem. We illustrate our methodology by transforming two state-of-the-art ECTS algorithms for the ECOTS scenario. Among the wide variety of applications that this new approach opens up, we develop here a predictive maintenance use case that optimizes alarm triggering times, thus demonstrating the power of this new approach.

Keywords: early decision-making; predictive maintenance; monitoring

1. Introduction

In intensive care units (Shekhar et al., 2021), in control rooms of electrical power grids (Dachraoui et al., 2013), in government councils assessing emergency situations, in many kinds of contexts therefore, it is essential to make timely decisions in absence of complete knowledge of the true outcome. The issue facing the decision-makers is that, usually, the longer the decision is delayed, the clearer is the likely outcome (e.g. whether the patient is critical or not) but, also, the higher the cost that will be incurred if only because earlier decisions allow one to be better prepared. Formally, this problem translates into optimizing online the trade-off between the earliness and the accuracy of the decision. Early Classification of Time Series (ECTS) deals with time series of finite length, and a single decision per time series.

In this paper, we study this trade-off in a new context, one where events of different classes occur in a streaming fashion, i.e. in an open time series with no predefined end. We want to predict, at the optimal time, each of them with their associated time intervals. For instance, in predictive maintenance, events may be associated with types of mechanical parts malfunctions that can be expected at given times.

Interestingly, we find that the powerful algorithms that have resulted from research on ECTS can sensibly be adapted to the new problem. In particular, this paper proposes a principled way to adapt ECTS approaches to deal with the early classification in open time series (i.e. with no time bounds). We show how the classification problem in ECTS can be transformed into a new classification problem for Early Classification in Open Times Series (ECOTS) and how the decision triggering condition should be modified.

In the classical setting, the *Early Classification of Time Series* (ECTS) problem assumes that measurements, possibly multivariate, become available over time in a time series which, at time t , is $\mathbf{x}_t = \langle x_1, \dots, x_t \rangle \in \mathcal{X}^t$ where x_t is the current measurement and each $x_j (1 \leq j \leq t)$ belongs to some input domain \mathcal{X} (e.g. the temperature and the blood pressure of a patient at time step j). It is further supposed that each time series can be ascribed to a class $y \in \mathcal{Y}$ (e.g. patient who needs a surgical operation or not). The task is to predict the class of an incoming time series by optimizing the trade-off between the expected accuracy of the prediction and the increasing cost of delaying the decision.

Formally, the ECTS problem can be stated as trying to find the best (possibly future) time τ^* to make a prediction, given that only \mathbf{x}_t has been observed:

$$\tau^* = \underset{\tau \geq t}{\text{ArgMin}} \left\{ C_m(h_\tau(\mathbf{x}_\tau)|y) + C_d(\tau) \right\} \quad (1)$$

where $C_m(h_\tau(\mathbf{x}_\tau)|y)$ is the cost of misclassification incurred at τ , and y is the true class, and $C_d(\tau)$ is the delay cost. What makes the problem possible to tackle, in the ECTS scenario, is the availability of a training set $\mathcal{S} = \{(\mathbf{x}_T^i, y_i)\}_{1 \leq i \leq m}$ of complete time series, with their associated labels. For each time step t , $t \in \{1, \dots, T\}$, using \mathcal{S} , a classifier h_t can be learned $h_t : \mathcal{X}^t \rightarrow \mathcal{Y}$, with \mathcal{X}^t being the space of truncated time series at time step t . These classifiers are learned beforehand and are considered as an input to the ECTS approaches, which essentially optimize the times at which predictions are triggered.

Overall, the ECTS problem can be seen as involving two components: (i) the set of *classifiers* h_t ($1 \leq t \leq T$) which is supposed pre-learned as seen above, and (ii) a *trigger system* which decides whether to make the prediction at the current time t or to wait for at least one more measurement x_{t+1} (see Figure 1).

While the ECTS framework is well suited to many real world problems, it is limited in several ways, as has been underscored in (Bondu et al., 2022) which defines several challenges that remain to be solved. In particular, one challenge relates to the fact that, in ECTS, (i) the time series all have a finite length and (ii) it is assumed that there is a single class associated with each of them. However, applications abound where the measurements come in an open time series with no time bounds and where different events may arise, possibly of different lengths, each with its own class (see Figure 2). The setting is thus different from ECTS. Should we then abandon the good properties and performances of the best ECTS algorithms for the new ECOTS problem?

If not, adapting ECTS approaches to the ECOTS problem raises three issues:

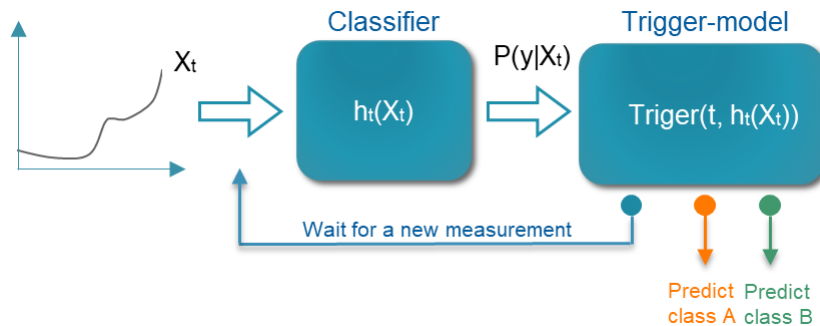


Figure 1: General schema of ECTS approaches

1. What should the *classifiers* in ECOTS do? In the ECTS framework, they take incomplete time series \mathbf{x}_t as input and make prediction about the class of the associated complete, but still unknown, time series \mathbf{x}_T . But, in ECOTS, the notion of complete time series does not make sense anymore.
2. How to solve the *earliness-accuracy* trade-off in the ECOTS setting? In the ECTS framework, the start and length of time series are known which allows one to measure the earliness. But how to define it in ECOTS?
3. How to build the *training sets* in the framework of ECOTS where there are no more individual time series with their associated class, but open time series with events of different classes and durations?

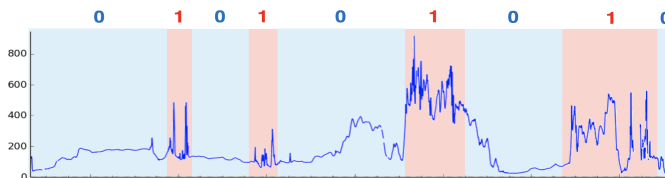


Figure 2: Example of a part of an open time series where events of possibly different lengths are here labeled with ‘0s’ and ‘1s’.

The goal of this paper is, first, to define properly the ECOTS problem and, second, to present a methodology to adapt any ECTS approach to it by answering the three questions raised above. As a result, we show how the role of the classifiers must be thought anew and transformed and how the earliness-accuracy trade-off translates to the new scenario and what the decision triggering system becomes. We illustrate our methodology by transforming two state-of-the-art ECTS algorithms for the ECOTS scenario. Among the wide variety of applications that this new approach opens up, we develop here a predictive maintenance use case that optimizes alarm triggering times, thus demonstrating the power of this new approach.

In order to avoid confusion, it is important to note that the data stream literature (Silva et al., 2013) focuses on classification of incoming data points at fixed horizon under

memory constraints and evolving properties of data. Whereas, in this paper, we focus on identifying the optimal triggering moment of the classification, the one that optimizes the *earliness-accuracy* trade-off in a stationary environment.

The paper is organized as follows. Section 2 formally draws a parallel between the ECTS problem and the ECOTS one, leading to a generic approach capable of transposing any ECTS algorithm into an ECOTS one. We then review, in Section 3, the main approaches to ECTS, and outline two competitive methods: one described in (Mori et al., 2017b) and the other, the ECONOMY method presented in (Achenchabe et al., 2021a). We then show, in Section 4, how to adapt these methods to the ECOTS problem, before comparing their performances on experiments in Section 5. The conclusion, in Section 6, underlines what has been performed in this work, and provides directions for future research.

2. ECOTS in the perspective of ECTS

This section defines the ECTS and ECOTS problems in turn, and presents our proposed methodology to adapt any ECTS approach to solve the ECOTS problem.

2.1. The ECTS problem

In the ECTS setting, each classifier h_t ($1 \leq t \leq T$) is learned from the truncated training time series up to time t : $\{(\mathbf{x}_t^j, y_j)\}_{(1 \leq j \leq m)}$ with m is the number of time series (see Figure 3). It is expected that the accuracy of the classifiers grows as t increases from the first time step $t = 1$ to the last one $t = T$.

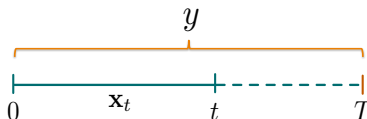


Figure 3: In the ECTS setting, the classifier h_t sees the incoming time series \mathbf{x}_t and predicts a label \hat{y} of the complete time series \mathbf{x}_T of true class y .

The problem, given an incoming time series, is to choose a time t for which the expected cost of misclassification $C_m(\hat{y}|y)$, where $\hat{y} = h_t(\mathbf{x}_t)$ and y is the true class, plus the delay cost $C_d(t)$, is minimal. Formally, the combined expected cost is given by:

$$\begin{aligned}
 f(\mathbf{x}_t) &= \mathbb{E}_{(\hat{y}, y) \in \mathcal{Y}^2}^t [C_m(\hat{y}|y)|\mathbf{x}_t] + C_d(t) \\
 &= \sum_{y \in \mathcal{Y}} P_t(y|\mathbf{x}_t) \sum_{\hat{y} \in \mathcal{Y}} P_t(\hat{y}|y, \mathbf{x}_t) C_m(\hat{y}|y) + C_d(t)
 \end{aligned} \tag{2}$$

where $\mathbb{E}_{(\hat{y}, y) \in \mathcal{Y}^2}^t [C_m(\hat{y}|y)|\mathbf{x}_t]$ is the expectancy at time t , over the variables y and \hat{y} . $P_t(y|\mathbf{x}_t)$ is the probability of the class y given an incomplete time series \mathbf{x}_t , and $P_t(\hat{y}|y, \mathbf{x}_t)$ is the probability that the classifier h_t makes the prediction \hat{y} given \mathbf{x}_t as input while y would be its true label.

The objective of the trigger function is to identify the best time t^* for triggering the decision while receiving online the measurements of the time series, the one that minimizes Equation 2. Section 3 provides a state of the art of the existing approaches to solve this problem, from very heuristic ones to more formally grounded.

2.2. The ECOTS problem

In this paper, contiguous instants with the same label are called chunks or events (see Figure 2). In the ECOTS scenario, we suppose that we have a training data set of m' labeled chunks $\{(\mathbf{x}^j, y_j)\}_{(1 \leq j \leq m')}$ coming from an open time series, where each \mathbf{x}^j has length $l_{\mathbf{x}^j}$, and where y_j is the corresponding label. The ECOTS problem consists in predicting as soon as possible these events, i.e. their class label and time limits.

2.3. Proposed transposition of ECTS approaches into ECOTS ones

We propose to simplify the ECOTS problem by considering point-wise predictions instead of predicting whole chunks (i.e. class and time limits), that is to make independent predictions of the labels associated with each single timestamp t_p in the open time series. From such predictions, chunks could be reconstituted, for instance by gluing time stamps with the same predicted class.

We posit that the unfolding time series is observed over a finite time interval which depends on the present time t . In the following, we will use the term time window and note it $\mathbf{x}_{(t-w,t)}$ at time t when its size is w (w past measurements are available at time t).

Then, the target time stamp t_p can be in the future ($t_p > t$), for example if there are warning signs that a machine will break down, or in the past ($t_p < t$) if it was necessary to wait until t_p was in the observation time window to be able to identify its class, as it can happen during a computer attack for example. Therefore, at any time t , for a target timestamp t_p , the transposed ECOTS problem is to decide whether t is the best time for the prediction \hat{y}_{t_p} , the class associated with t_p , or whether to postpone this decision to the next time step $t + 1$, which will bring a new measurement.

It is expected that, as the window of observation $\mathbf{x}_{(t-w,t)}$ comes closer to t_p , with increasing t , it is easier to make a reliable prediction about its class, but, at the same time, the cost of delaying prediction increases. (See Figure 4). Note that we assume that there is a maximal value η_M for the horizon, above which no precursor signal can be detected, and a minimal one η_m after which it is no longer useful to detect the event.

We now turn to the three questions raised in the introduction.

1- In the classical ECTS problem (see Section 2.1 and Figure 3), each classifier h_t observes a time series \mathbf{x}_t that is increasingly large as t approaches the time limit T . In the ECOTS setting, as we propose to see it, by contrast, each classifier h_η observes a sliding window $\mathbf{x}_{(t-w,t)}$ of the same number w of observations. And each one makes a prediction about the class of the time t_p positioned at a given horizon η (positive or negative) from t (see Figure 4). Given a maximal horizon $\eta_M > 0$, the first classifier that can make a prediction about a time t_p is h_{η_M} viewing the window $\mathbf{x}_{(t_p-\eta_M-w,t_p-\eta_M)}$. And the last classifier is h_{η_m} viewing the window $\mathbf{x}_{(t_p-\eta_m-w,t_p-\eta_m)}$. Thus, instead of having a set of classifiers h_t ($1 \leq t \leq T$) as in the ECTS setting, we now have a set of classifiers h_η ($\eta_m \leq \eta \leq \eta_M$) with various horizons. (See Figure 4).

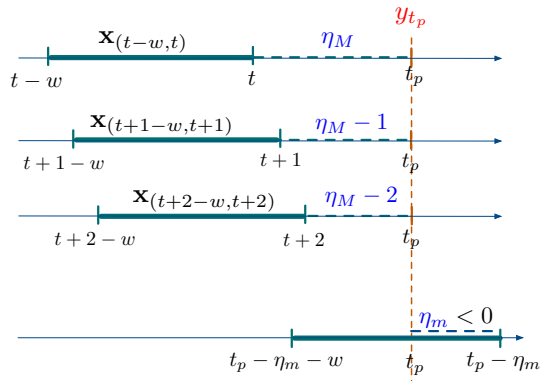


Figure 4: The fixed point in time where to make a prediction is t_p , of true label y_{t_p} . As the measurements become available from $t_p - \eta_M$ to $t_p - \eta_m$, different classifiers h_η come into play with an advancing sliding window and a diminishing horizon. The triggering system selects the best time to make a prediction \hat{y}_{t_p} .

2- The second question to solve is to adapt the earliness-accuracy trade-off to the ECOTS problem. In the proposed transposition, the gain of information over time is due to a time window $\mathbf{x}_{(t-w,t)}$ that gets closer to the target timestamp t_p , and no longer to additional measurement gathered over time as in ECTS. In addition, the single prediction triggered for each time series in ECTS is replaced by a prediction for each possible target timestamp in the open time series. These two essential but easy to carry out modifications allow to transpose most of the ECTS approaches, since the optimization criterion they use for their triggering strategy can remain essentially the same. As a proof of concept, in Section 4, we show how two state of the art methods for ECTS can be adapted to the ECOTS problem.

3- In the proposed approach, the values of the window size w , and of the bounds η_m and η_M of the horizon have to be chosen from a training set. But which training set?

In the ECOTS problem, it is assumed that the relationship between the symptoms of an event and the event itself are *stationary* over time (e.g. a given malfunction of a machine keeps the same telltale signs and the same characteristics whenever its appearance in time. Or the symptoms associated with a patient who should undergo an urgent heart operation stay the same, fortunately for the doctors, and for their training). From this property, in the same way as doctors can be trained using independent episodes in an hospital history about heart attacks, it is possible to use subsequences of the open time series, as long as they are independent, to build training datasets in order to learn the classifiers h_η (see Section 5.1).

3. Related work on trigger systems

While most of the proposed ECTS approaches rely on learning classifiers for different time steps in the time series, they differ in the way they trigger the decision to make a prediction.

Two different approaches can then be used. In the *myopic* one, at each successive time step t , the accuracy (or confidence) of the current classifier h_t on its prediction can be

assessed and it is possible to decide whether the time seems right to make a prediction. In a *non-myopic* approach, given the current incoming time series \mathbf{x}_t , the expected accuracy of all future classifiers h_τ ($t \leq \tau \leq T$) given the current input can be estimated and the best decision time be computed taking into account the delay cost. Only if t coincides with this best instant, the prediction is performed

The earliest works on ECTS were heuristics in nature as well as myopic. They did not try to explicitly optimize the cost defined in equation (2) but instead relied on the concept of an estimated confidence in the current prediction. If this estimate was above a threshold, then the decision was triggered. For instance, (Mori et al., 2017b) describes a method where the accuracy of a set of probabilistic classifiers is monitored over time, which allows the identification of time steps from whence it seems safe to make predictions. In (Parrish et al., 2013; Hatami and Chira, 2013; Ghalwash et al., 2012), various stopping rules are defined, some on them relying on a confidence level threshold. And in (Xing et al., 2009), the best time step to trigger the decision is estimated by determining the earliest time step for which the predicted label does not change, based on a 1-NN classifier.

More recently, approaches have been proposed that attempt to explicitly optimize the trade-off between the earliness and the accuracy. A notable example is (Mori et al., 2017a) where a single objective optimization criterion is defined. In (Mori et al., 2019), the authors put forward the idea of a multi-objective criterion with an associated Pareto front with multiple dominating trade-offs. However, whereas these methods are *myopic* in nature, the ECONOMY approach described in (Achenchabe et al., 2021a; Dachraoui et al., 2015) goes one step further by (i) directly optimizing the combined cost defined in equation 2 and (ii) giving a way to estimate the combined cost for future time steps, thus leading to a *non-myopic* approach that outperforms the best methods known to date.

For a recent survey, the interested reader can refer to (Gupta et al., 2020).

4. Adapting two state-of-the art ECTS approaches

In Section 2.3, we have shown how to translate an ECTS problem into an ECOTS one by modifying the definition and purpose of *the classifiers*. In the following, we demonstrate how the *triggering strategy* used in ECTS can be adapted to deal with ECOTS. For this, we consider one of the best performing myopic strategies known to date, described in (Mori et al., 2017a) and the best non-myopic approach in the literature: the ECONOMY- γ strategy described in (Achenchabe et al., 2021a). The first one relies ultimately on confidence criteria, while the second one explicitly optimizes the accuracy versus delay cost trade-off.

4.1. The SR approach

The SR approach (Mori et al., 2017a) uses a trigger-model which involves 3 parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ in order to decide if the current prediction is reliable (output 1) or if it is preferable to wait for more data (output 0):

$$\text{Trigger}^\gamma(h_\eta(\mathbf{x}_t)) = \begin{cases} 0 & \text{if } \gamma_1 p_1 + \gamma_2 p_2 + \gamma_3 \frac{t}{T} \leq 0 \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

where p_1 is the largest posterior probability estimated by the classifier h_η , p_2 is the difference between the two largest posterior probabilities, and the last term $\frac{t}{T}$ represents the proportion of the incoming time series that is visible at time t . The parameters $\gamma_1, \gamma_2, \gamma_3$ are real values in $[-1, 1]$ to be optimized using training data.

Algorithm 1 Adapted SR approach in the ECOTS scenario

Input: t : current moment.

t_p : target that belongs to $[t+\eta_m, t+\eta_M]$.

w, η_m, η_M : window size, minimum and maximum horizon.

- 1: **for all** $\eta = t_p - t, \dots, \eta_m$ (step=-1) **do**
 - 2: sliding_window = $\mathbf{x}_{(t_p-\eta-w, t_p-\eta)}$ # sliding window: adds new measurement and deletes the first one.
 - 3: compute $A = \gamma_1 p_1 + \gamma_2 p_2 + \gamma_3 \frac{\eta_M - \eta}{\eta_M - \eta_m}$ using the updated sliding window.
 - 4: **if** $A > 0$ or $\eta == \eta_m$ **then**
 - 5: return η
 - 6: **end if**
 - 7: **end for**
-

In the ECOTS problem, the trigger function for a target at horizon $\eta = t_p - t$ becomes:

$$\text{Trigger}^\gamma (h_\eta(\mathbf{x}_{(t-w, t)})) = \begin{cases} 0 & \text{if } \gamma_1 p_1 + \gamma_2 p_2 + \gamma_3 \frac{\eta_M - \eta}{\eta_M - \eta_m} \leq 0 \\ 1 & \text{otherwise} \end{cases}$$

The last term of Equation (3) is replaced by $\frac{\eta_M - \eta}{\eta_M - \eta_m}$, which represents the relative position of the current horizon η in the considered range of horizons $[\eta_m, \eta_M]$. Algorithm 1 highlights the adapted procedure of the SR approach to choose the optimal horizon for a given target t_p .

4.2. The ECONOMY- γ approach

ECONOMY- γ is a non-myopic cost-based approach (Achenchabe et al., 2021b), which is capable of estimating the expected cost of making a prediction for any time $t + \tau$ ($1 \leq \tau \leq T - t$) in the future, defined as:

$$\begin{aligned} f_\tau(\mathbf{x}_t) &= \mathbb{E}_{(\hat{y}, y) \in \mathcal{Y}^2}^{t+\tau} [C_m(\hat{y}|y)|\mathbf{x}_t] + C_d(t + \tau) \\ &= \sum_{y \in \mathcal{Y}} P_{t+\tau}(y|\mathbf{x}_t) \sum_{\hat{y} \in \mathcal{Y}} P_{t+\tau}(\hat{y}|y, \mathbf{x}_t) C_m(\hat{y}|y) + C_d(t + \tau) \end{aligned} \quad (4)$$

In practice, the terms $P_{t+\tau}(\hat{y}|y, \mathbf{x}_t)$ and $P_{t+\tau}(y|\mathbf{x}_t)$ are not tractable. A partitioning of the training data into K groups $\mathbf{g}_k \in \mathbf{G}$ (see (Achenchabe et al., 2021b)) is required to make them computable, yielding the following approximation:

$$f_\tau(\mathbf{x}_t) \approx \sum_{\mathbf{g}_k \in \mathcal{G}} P_{t+\tau}(\mathbf{g}_k|\mathbf{x}_t) \sum_{y \in \mathcal{Y}} P_{t+\tau}(y|\mathbf{g}_k) \sum_{\hat{y} \in \mathcal{Y}} P_{t+\tau}(\hat{y}|y, \mathbf{g}_k) C_m(\hat{y}|y) + C_d(t + \tau)$$

The optimal decision time, at time t , is thus estimated to be:

$$\tau^* = \underset{\tau \in \{0, \dots, T-t\}}{\text{ArgMin}} f_\tau(\mathbf{x}_t) \quad (5)$$

The idea is to estimate the cost of a decision for all future time steps, up until $t = T$, based on the current knowledge about the incoming time series \mathbf{x}_t . The decision is postponed unless $\tau^* = 0$, that is when it is expected that there will be no better trade-off in the future. If so, the prediction $h_t(\mathbf{x}_t)$ is returned and the classification process is terminated. Otherwise, the decision is postponed to the next time step, and Eq. (5) is computed again, this time with \mathbf{x}_{t+1} as input. The process goes on until a decision is made or $t = T$ at which point a prediction is forced.

Algorithm 2 Adapted ECONOMY approach in the ECOTS scenario

Input: t : current moment.

t_p : target that belongs to $[t+\eta_m, t+\eta_M]$.

w, η_m, η_M : window size, minimum and maximum horizon.

- 1: **for all** $\eta = t_p - t, \dots, \eta_m$ **do**
 - 2: compute $f_\eta(\mathbf{x}_{(t-w+1,t)})$
 - 3: **end for**
 - 4: $\eta^* = \text{ArgMin}_{\eta_m \leq \eta \leq t_p - t} f_\eta(\mathbf{x}_{(t-w+1,t)})$
 - 5: **if** $\eta^* == t_p - t$ or $t_p - t == \eta_m$ **then**
 - 6: return η^*
 - 7: **end if**
 - 8: return None
-

In ECOTS, as time t increases, the task is to label each time step t_p as it appears in the span of the horizon: $[t + \eta_m, t + \eta_M]$ (see Figure 4). While in Equation (4), the term $\mathbb{E}^{t+\tau}_{(\hat{y}, y) \in \mathcal{Y}^2} [C_m(\hat{y}|y)]$ involves the calculation of the confusion matrices for future time steps $t + \tau$ knowing the current incoming time series \mathbf{x}_t , the adaptation to ECOTS requires that the confusion matrices are now computed for the various horizons from η_m to η_M and then used to estimate the cost of decision for each horizon η :

$$f_\eta(\mathbf{x}_{(t-w,t)}) = \mathbb{E}_{(\hat{y}, y) \in \mathcal{Y}^2}^\eta [C_m(\hat{y}|y)|\mathbf{x}_{(t-w,t)}] + C_d(\eta)$$

and the best horizon:

$$\eta^* = \underset{\eta_m \leq \eta \leq t_p - t}{\text{ArgMin}} f_\eta(\mathbf{x}_{(t-w,t)})$$

The decision to classify the data point t_p is triggered at time t either because $t + \eta^* = t_p$ (i.e. corresponding to the optimal cost) or when $t_p = t + \eta_m$ (i.e. it is not possible to wait any longer), see Algorithm 2 for more details.

The cost of delay $C_d(t)$, which is an *increasing* function of t in ECTS is a *decreasing* function $C_d(\eta)$ of the horizon η in ECOTS. Indeed, as the target that we want to label approaches (η decreasing), the cost of the decision increases.

Note that the time and space complexities of the ECONOMY- γ approach adapted to ECOTS are the same than in the original approach. In the ECTS setting, at testing phase, computing the cost at each timestep is in $O(T^2)$ at worst case, and in ECOTS setting is in $O((\eta_M - \eta_m)^2)$.

5. Experiments

We have proposed a principled method to adapt any ECTS approach into an ECOTS one (see Sections 2 and 4). The aim of the experiments is to validate that the adaptation of the SR and ECONOMY- γ approaches is efficient in the ECOTS setting. In addition, we illustrate the applicability of the proposed approaches for predictive maintenance using real data from the industrial domain.

This section aims at answering the following questions:

1. How efficient is the proposed framework for adapting any ECTS approach to the ECOTS problem compared to baseline algorithms designed for the ECOTS problem?
2. How do these approaches behave when the delay cost increases, and when the misclassification cost becomes very imbalanced?
3. How these approaches adapt their decision time to the observed data?

Our *source code* is shared¹ for full reproducibility of the experiments. This also allows interested researchers to extend the experiments to other open time series datasets.

5.1. Experimental protocol

5.1.1. DATA DESCRIPTION:

We use an open real dataset (data, 2020) from one of the Schwan’s factories. It contains 100 multivariate time series corresponding to 100 machines monitored over time for a period of 1 year (January 2015 to January 2016) with measurements collected every hour. Each time series is a multi-dimensional data table whose rows indicate the temporal domain and columns include telemetry features (pressure, rotation, voltage and vibration), 5 Boolean columns encoding different types of device errors which are premises correlated with a future failure and a last column which indicates the presence or absence of a failure (the variable to be predicted). This makes 8761 rows and 10 columns for each machine. The whole dataset contains 3919 errors and 761 failures for a total number 876,100 timestamps. This dataset is extremely imbalanced with 0.08% of timestamps associated with the abnormal class (i.e. failure). There is on average 7 failures per time series, with a minimum of 0 and a maximum of 19 failures per machine during the observed year.

5.1.2. PROBLEM STATEMENT:

Traditionally, the problem of predictive maintenance is solved by fixing a horizon for predictions (e.g. if a technician needs at least 12 hours to take preventive actions before the machine fails actually, then a fixed horizon would be chosen as $\eta = 12$ hours). Our goal is to use the ECOTS approaches to make this horizon *adaptive* to the observable part of the time series at hand.

1. <https://github.com/YoussefAch/ECOTS>

5.1.3. EVALUATION CRITERION:

Ultimately, the value of using an early classification method is defined by the *average cost* incurred using it, as in (Achenchabe et al., 2021a). Given an open time series \mathcal{S} (e.g. a machine monitored over a year), observed on a finite time interval of sufficient length N . This time period is composed of time stamps $t \in [1, N]$, labeled by the class y_t . As time increases from 1 to N , the ECOTS system makes predictions for each time step t : \hat{y}_t while the true class is y_t . In addition, for each $t \in [1, N]$, the prediction is made using a classifier $h_{\eta_{t^*}}$ corresponding to the triggering horizon η_{t^*} , thus incurring a delay cost $C_d(\eta_{t^*})$. Hence, the formula:

$$AvgCost(\mathcal{S}) = \frac{1}{N} \sum_{t=1}^N (C_m(\hat{y}_t|y_t) + C_d(\eta_{t^*})) \quad (6)$$

5.1.4. COMPUTING THE COSTS IN THE EXPERIMENTS:

In real applications, the decision costs would be provided by domain experts. In order to study the behavior of the different ECOTS algorithms, a large range of values has been considered for the misclassification and the delay costs.

The cost of misclassification: Since we deal with a predictive maintenance problem, we make the assumption that the cost of missing a failure is much higher than the cost of sending the technical team. We thus consider four different misclassification costs $C_m = \begin{bmatrix} TN & FN \\ FP & TP \end{bmatrix}$, by varying the importance of false negatives:

$$C_m^{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, C_m^{(2)} = \begin{bmatrix} 0 & 10 \\ 1 & 0 \end{bmatrix}, C_m^{(3)} = \begin{bmatrix} 0 & 100 \\ 1 & 0 \end{bmatrix}, C_m^{(4)} = \begin{bmatrix} 0 & 1000 \\ 1 & 0 \end{bmatrix}.$$

The cost of delaying decision: The delay cost $C_d(\eta)$ is provided by the domain experts for an actual use case, and could be of any form. In our experiments, we set it as a linear function of horizon, with coefficient, or slope, α : $C_d(\eta) = \alpha \times \frac{\eta_M - \eta}{\eta_M - \eta_m}$. The larger α , the higher the cost of postponing the decision and the greater the incentive to make prediction for large horizons η . When α is very high, the gain in misclassification cost by waiting to be closer to the target cannot compensate for the increase of the delay cost, and it is better to make a decision early on, that is for large horizons, close to η_M . If, on the contrary, α is very low compared to the misclassification cost, it does not hurt to wait until the target t_p is close to the sliding window $\mathbf{x}_{(t_p - \eta - w, t_p - \eta)}$. Our experiments were run over a large range of values of $\alpha \in [10e-04, 10e-03, 10e-02, 1, 10, 100, 1000]$.

5.1.5. TRAINING THE COLLECTION OF CLASSIFIERS AND ECOTS ALGORITHMS

Data split and extraction: We splitted the set of time series into four parts: 50% for training the classifiers, 20% for testing the ECOTS algorithms, 15% for validating the ECOTS algorithms and 15% for estimating the confusion matrices. This split is inspired from the original paper of ECONOMY (Achenchabe et al., 2021a). Subsequences of size w were extracted from the training open time series by doing the following steps: (i) time stamps t_p , aka targets, were set within the time series, spaced with $w + \eta_M$ time units in

order to avoid overlaps between samples; (ii) $\eta_M - \eta_m$ subsequences of size w were extracted around each target, each one dedicated to the training of the classifier h_η (see Figure 4).

Choice of the parameters w , η_m , η_M : These parameters depend on the problem that is being solved and the data associated with it. One of the key ingredients of early classification methods is the information gain measured by the AUC. Generally, the expected cost of misclassification decreases as the target being classified gets closer to the sliding window. A window size of $w = 10$ has been chosen to study the behavior of the ECOTS problems, since it shows a significant information gain over various horizons using AUC. We refer the reader to the supplementary material for AUC curves as a function of horizon with different sliding window sizes. They exhibit equivalent information gain curves, which means that this dataset is not very sensitive to the choice of w . The parameter η_M can be chosen according to the AUC, for our experiments we have chosen $\eta_M = 50$ as the AUC reaches 0.5 which corresponds to the random model, while, for η_m , we chose the end of the sliding window: $\eta_m = -w$.

Training the collection of classifiers: As mentioned in Section 2.2, a set of classifiers h_η for different horizons η such that ($\eta_m \leq \eta \leq \eta_M$) has to be trained. Extracted features² from sliding windows include simple statistics: *min*, *max*, *mean*, *median* and the *count* of each type of errors. For our experiments, we trained XGboost models by fine tuning parameters within the following grid of values³: *min child weight* $\in [1, 5, 10]$, *gamma* $\in [0.5, 1, 1.5, 2, 5]$, *subsample* $\in [0.6, 0.8, 1]$, *colsample by tree* $\in [0.6, 0.8, 1]$, *max depth* $\in [3, 4, 5, 10]$. The parameter *scalePosWeight* is set to the ratio of positive examples over negative ones in order to take into account the fact that the dataset is imbalanced. The combination of parameters that minimize the total cost⁴ is chosen on a validation set (20% of the set used for training the classifiers), then the model is learned on the whole training set. For a fair comparison, the same collection of classifiers is used for all ECOTS algorithms.

ECOTS algorithms: The competing approaches considered in this paper are described below as well as their optimized hyper-parameters.

- **Late baseline:** the last classifier in the collection \tilde{h}_{η_m} is used. This is the last time that a prediction can be made.
- **Early baseline:** the first classifier in the collection \tilde{h}_{η_M} is used. This corresponds to the earliest possible prediction with the largest horizon in the future.
- **Confidence-based Classifiers (CC):** for a fixed target t_p , this method takes a decision as soon as the confidence of the classifier regarding the class of interest exceeds a given threshold, optimized as a meta-parameter for each value of α using validation set. If this never happens, then $t_p = t + \eta_m$ and the prediction is forced using \tilde{h}_{η_m} .
- **Economy- γ** (see Section 4.2): for each value of α , the number of groups K used in the method is optimized in the range $[1, 5]$ using a validation set.

2. Reproducible using our source code available in the supplementary material.

3. The interested reader can refer to the official documentation for more details: <https://xgboost.readthedocs.io/en/stable/>

4. Given the cost of false positives and false negatives, the total cost is computed on a validation set as the sum over wrongly predicted samples weighted by the corresponding cost.

- **SR** (see Section 4.1): for each value of α , the parameters γ_1, γ_2 and γ_3 were optimized in the range $[-1, -0.5, 0, 0.5, 1]^3$ using a validation set.

Note that the “late” and the “early” baselines are not adaptive, while the “Confidence-based” method adapts its decisions to the current input. One goal of the experiments is to compare these methods with ones that have been *translated* from the ECTS framework: **Economy- γ** and **SR**.

5.2. Results and analysis

In this section, detailed answers to the questions raised in the introduction of Section 5 are given, supported by numerical results.

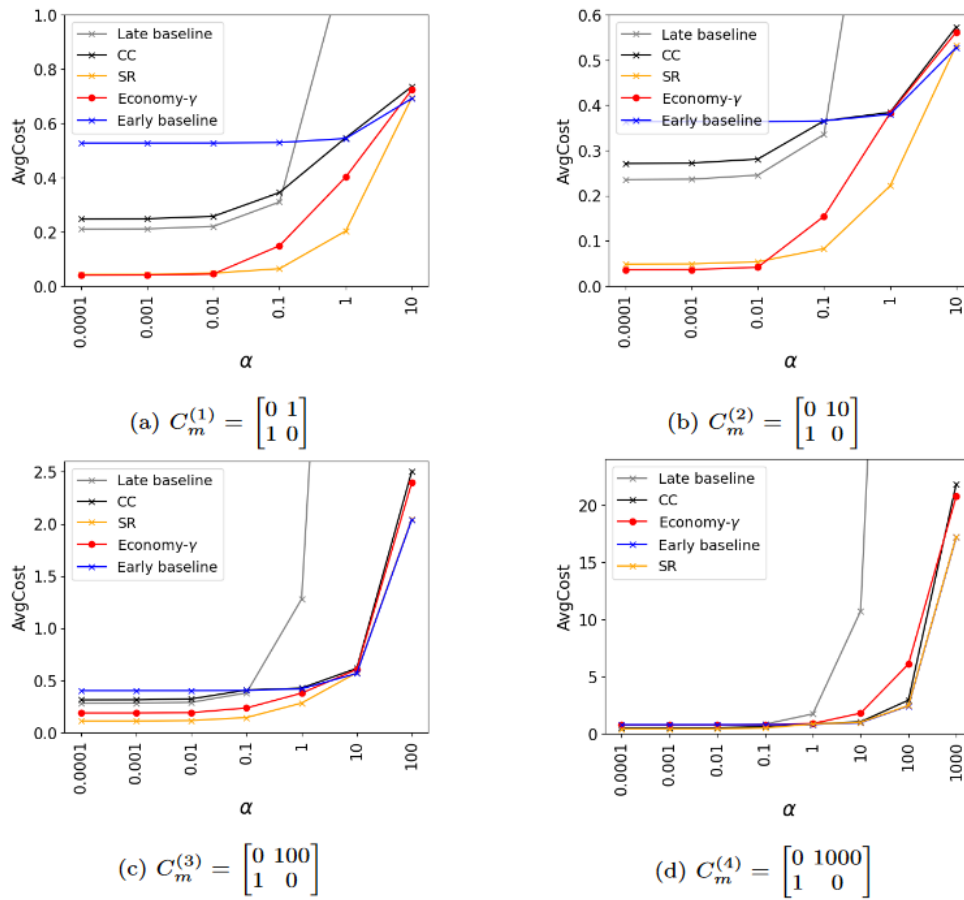


Figure 5: *AvgCost* of ECOTS algorithms computed on the test set for different values of the α parameter of the delay cost (x-axis), and for different values of misclassification cost. The maximum value of *alpha* on the x-axis is chosen in each subfigure to reach at least the maximum value in the misclassification cost matrix C_m . Figures with all values of alpha are provided in the supplementary material

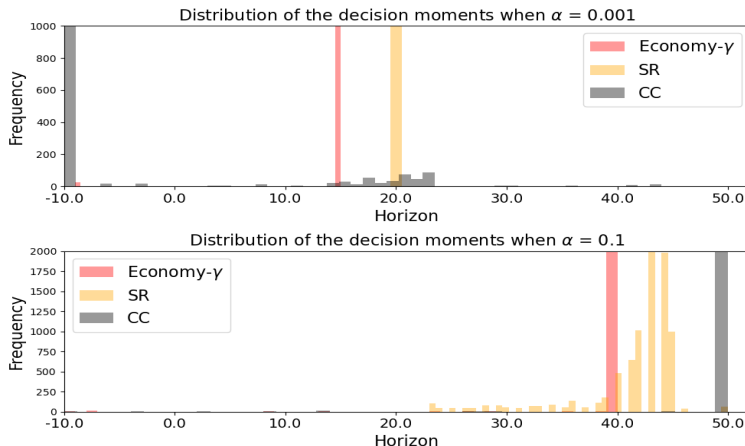


Figure 6: Distribution of the decision moments for **ECONOMY- γ** , **SR** and **CC** algorithms, for $\alpha = 0.001$ and $\alpha = 0.1$ both for $C_m^{(2)} = \begin{bmatrix} 0 & 10 \\ 1 & 0 \end{bmatrix}$.

Efficiency of the proposed framework: In Figure 5, one can note interesting patterns for the four matrices of misclassification costs and the large range of values of α and therefore of delay cost functions. When the cost of delaying decision is high ($\alpha \geq 10$), the optimal strategy is to make predictions immediately (i.e. the “early” baseline), for the largest value of the horizon η_M . When the delay cost is low, ($\alpha \leq 0.01$), taking late decisions is a good strategy even though it is not optimal (i.e. the “late” baseline). It is apparent that the **CC** method essentially switches from one baseline strategy to the other one as α increases, and therefore seems to realize the best of the two strategies adaptively.

At the same time, both methods “imported” from ECTS: **Economy- γ** and **SR**, noticeably overcome **CC**. They are able to better control the horizon of decision when α is low, thus achieving significantly better performance (see more on this in the discussion below on the ability to adapt the triggering times), and they perform as well as the competitors for high values of α . In particular, this shows that the often preferred, almost by default, confidence-based methods (e.g. **CC**) are being overtaken by more formally based methods translated from ECTS.

The experimentation on this dataset taken from a predictive maintenance problem, leads to the conclusion that the adapted **Economy- γ** and **SR** methods seem to be specially interesting under a wide range of conditions.

Effect of the delay and misclassification costs: In all the situations corresponding to the subfigures of Figure 5, the average cost sharply increases when the delay cost become very high (note the logarithmic scale on the x -axis). Indeed, decisions have to be made early so as to avoid high delay costs, but this is at the price of false positives and negatives which may incur high cost, specially for the $C_m^{(3)}$ and $C_m^{(4)}$ cost matrices.

Ability to adapt the triggering moment according to observed data: In order to better understand the properties of the adaptive approaches, we show in Figure 6 the distribution

of the decision moments of the three methods: ECONOMY- γ , CC and SR. We have chosen the scenario $C_m = C_m^{(2)}$ (The method behaves similarly for other values of $C_m = C_m^{(\cdot)}$ and additional figures are given in the supplementary material). One immediate finding is that both ECONOMY- γ and **SR** are more ready to consider intermediate horizons of prediction than **CC**. For $\alpha = 0.1$, **SR** is more prone to spread its decision horizons than ECONOMY- γ , which shows how this efficient approach adapts the horizon. It may explain its superior performance in this instance.

6. Conclusion

The dilemma of having to take a decision under time pressure is present in a wide range of domains. In the case of finite time series each involving a single decision, it has been shown that it can usefully be expressed as the problem of optimizing a trade-off between the earliness and the accuracy of the decision. In this paper, we offered ways to go beyond finite time series to address the problem of early classification of multiple events in open time series (ECOTS). We have formally defined this problem for the first time in the literature and provided a recipe that allows the transformation of *early classification of time series* (ECTS) approaches to the ECOTS problem. Adapting in this way two state-of-the-art ECTS algorithms, we applied them to a real world dataset related to predictive maintenance. The experiments attest that the new algorithms effectively optimize the earliness vs. accuracy tradeoff, exceeding the performance of heuristic-based algorithms.

It is our hope that these results can impact a wealth of applications that include health-care, predictive maintenance, autonomous driving, decision aid in agriculture, prediction of failure in cold chains, to name but a few.

In summary, this work focuses on variable-horizon adaptive prediction in a stationary environment, whereas most of the existing works on data streams consider fixed-horizon prediction, but potentially in a non-stationary environment (Bifet et al., 2018). The combination of the two problems remains to be studied.

References

- Youssef Achenchabe, Alexis Bondu, Antoine Cornuéjols, and Asma Dachraoui. Early classification of time series. *Machine Learning*, pages 1–24, 2021a.
- Youssef Achenchabe, Alexis Bondu, Antoine Cornuéjols, and Vincent Lemaire. Early and revocable time series classification. *arXiv preprint arXiv:2109.10285*, 2021b.
- Albert Bifet, Ricard Gavaldà, Geoff Holmes, and Bernhard Pfahringer. *Machine learning for data streams: with practical examples in MOA*. MIT press, 2018.
- Alexis Bondu, Youssef Achenchabe, Albert Bifet, Fabrice Clérot, Antoine Cornuéjols, Joao Gama, Georges Hébrail, Vincent Lemaire, and Pierre-François Marteau. Open challenges for machine learning based early decision-making research. *arXiv preprint arXiv:2204.13111*, 2022.
- Asma Dachraoui, Alexis Bondu, and Antoine Cornuéjols. Early classification of individual electricity consumptions. *RealStream2013 (ECML)*, pages 18–21, 2013.

- Asma Dachraoui, Alexis Bondu, and Antoine Cornuéjols. Early classification of time series as a non myopic sequential decision making problem. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 433–447. Springer, 2015.
- data. Schwan’s factory data. <https://github.com/DeeptiChevvuri/Predictive-Maintenance-Modelling-Datasets>, 2020.
- Mohamed F Ghalwash, Dušan Ramljak, and Zoran Obradović. Early classification of multivariate time series using a hybrid hmm/svm model. In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, pages 1–6. IEEE, 2012.
- Ashish Gupta, Hari Prabhat Gupta, Bhaskar Biswas, and Tanima Dutta. Approaches and applications of early classification of time series: A review. *IEEE Transactions on Artificial Intelligence*, 2020.
- Nima Hatami and Camelia Chira. Classifiers with a reject option for early time-series classification. In *Computational Intelligence and Ensemble Learning (CIEL), 2013 IEEE Symposium on*, pages 9–16. IEEE, 2013.
- Usue Mori, Alexander Mendiburu, Sanjoy Dasgupta, and Jose A Lozano. Early classification of time series by simultaneously optimizing the accuracy and earliness. *IEEE transactions on neural networks and learning systems*, 29(10):4569–4578, 2017a.
- Usue Mori, Alexander Mendiburu, Eamonn Keogh, and Jose A Lozano. Reliable early classification of time series based on discriminating the classes over time. *Data mining and knowledge discovery*, 31(1):233–263, 2017b.
- Usue Mori, Alexander Mendiburu, Isabel Marta Miranda, and José Antonio Lozano. Early classification of time series using multi-objective optimization techniques. *Information Sciences*, 492:204–218, 2019.
- Nathan Parrish, Hyrum S Anderson, Maya R Gupta, and Dun Yu Hsiao. Classifying with confidence from incomplete information. *J. of Mach. Learning Research*, 14(1):3561–3589, 2013.
- Shubhranshu Shekhar, Dhivya Eswaran, Bryan Hooi, Jonathan Elmer, Christos Faloutsos, and Leman Akoglu. Benefit-aware early prediction of health outcomes on multivariate eeg time series. *arXiv preprint arXiv:2111.06032*, 2021.
- Jonathan A Silva, Elaine R Faria, Rodrigo C Barros, Eduardo R Hruschka, André CPLF de Carvalho, and João Gama. Data stream clustering: A survey. *ACM Computing Surveys (CSUR)*, 46(1):1–31, 2013.
- Zhengzheng Xing, Jian Pei, and S Yu Philip. Early prediction on time series: A nearest neighbor approach. In *IJCAI*, pages 1297–1302. Citeseer, 2009.