A Benchmark for Early Time-Series Classification

Petro-Foti Kamberi  
Institute of Informatics & Telecommunications, NCSR “Demokritos”, Athens, Greece

Evgenios Kladis  
Institute of Informatics & Telecommunications, NCSR “Demokritos”, Athens, Greece

Charilaos Akasiadis  
Institute of Informatics & Telecommunications, NCSR “Demokritos”, Athens, Greece

Abstract

The objective of Early Time-Series Classification (ETSC) is to predict the class of incoming time-series by observing the fewest time-points possible. Although many approaches have been proposed in the past, not all techniques are suitable for every problem type. In particular, the characteristics of the input data may impact performance. To aid researchers and developers with deciding which kind of method suits their needs best, we developed a framework that allows the comparison of five existing ETSC algorithms, and also introduce a new method that is based on the selective truncation of time-series principle. To promote results reproducibility and the alignment of algorithm comparisons, we also include a bundle of datasets originating from real-world time-critical applications, and for which the application of ETSC algorithms can be considered quite valuable.

2012 ACM Subject Classification  Computing methodologies → Machine learning algorithms

Keywords and phrases  Time-series analysis, Classification, Benchmark

Digital Object Identifier  10.4230/LIPIcs.TIME.2023.18

Category  Extended Abstract


Supplementary Material  Software: https://github.com/xarakas/ETSC
archived at swh:1:dir:9b337165e0682df9285e5f66c99e4722a71c2988

Funding  This work has received funding from the EU project CREXDATA (under grant agreement No. 101092749).

1 Extended Abstract

Latest technological advancements drive the generation of large volumes of time-series data. Sea vessels, for example, utilize integrated sensory and telecommunication devices to continuously report trajectory information in the form of time-series data. This abundance of data is leveraged by machine learning techniques to address various problems. In the life sciences domain, simulators are incorporated to test the effectiveness of new experimental drugs “in-silico”. Such simulations often require long time and large amounts of computational resources, which, in the case of unsuccessful drug treatment cases being simulated, are consumed in vain [1]. It would be desirable to be able to predict such outcomes early on by observing the simulations course as time-series, and to terminate not interesting trials so to speed up the whole drug discovery process. To this end, the domain of Early Time-Series Classification (ETSC) has an objective to classify time-series at the earliest point possible, before the entire series is observed [5].

Meanwhile, despite the numerous proposed methods for ETSC, there is a notable absence of a dedicated experimental evaluation and comparison framework in this field. Furthermore, ETSC methods are predominantly evaluated and compared against only a limited set of
alternative algorithms. In order to fill this gap, we have developed a framework that allows
to empirically compare five existing approaches as well as a newly introduced one, using
a curated bundle of datasets from real-world applications. This framework is utilized to
highlight the ETSC algorithms merits and shortcomings when applied to cases with different
features, e.g. dataset size, observation variability, class imbalance, etc. The framework can
be easily extended to include more datasets and algorithms, and is openly available online [2].

Evaluation metrics. In the ETSC domain, apart from the predictive performance (accuracy
and F1-score), there is also the objective to optimize the earliness of the generated predictions.
These two metrics can also be combined into a single metric as a harmonic mean [9]. Training
and testing times are also of interest to consider in real-world applications.

Algorithms. Our framework includes five existing ETSC algorithms, i.e. ECEC [7],
ECONOMY-K [3], ECTS [10], EDSC [11], and TEASER [9]. ECEC calculates confidence
thresholds above which a class label prediction is considered to be reliable. ECONOMY-K
performs clustering on the training data, and estimates the cost of having to observe more
time points to generate a prediction. ECTS utilizes nearest neighbors and reverse nearest
neighbors sets for its decisions. EDSC extracts shapelets of the training data that are then
matched with the input test data. TEASER trains classifiers on overlapping prefixes of the
training data, and then applies an one-class SVM that validates the class label prediction.

In addition, we propose a new method that can be configured to utilize different state-of-
the-art full time-series classification algorithms for ETSC. It relies on iteratively truncating
time-series into prefixes of gradually increasing length and then applying Minirocket [4],
MLSTM [6], and WEASEL [8] to the truncated examples for predicting the corresponding
class labels. Thus, for each dataset, a fixed earliness is determined throughout the training
phase. Although this might be suboptimal in the sense that for particular instances the
prediction could have been generated earlier than for others, it constitutes a comprehensive
baseline for diagnosing if applying ETSC to particular datasets and domains would be
successful, i.e. if accurate predictions can be generated earlier, before the whole time-series
is observed. Our approach can incorporate any full time-series classification algorithm.

Datasets. We have collected 10 publicly available datasets from the UEA & UCR Time-
series Classification Repository, and also introduced two new cases, one originating from the
drug discovery domain, and the other from the field of maritime intelligence. In turn, these
datasets are categorized according to particular characteristics: size, coefficient of variability,
levels of class imbalance, number of distinct class labels, and number of variables.

Comparison results. Judging by our experimental results, we can see that ECEC achieves
the most accurate and early predictions for datasets with lengthy time-series, but it requires
higher training times. When the number of examples in the dataset increases, the MLSTM
variant of our method competes with ECEC and TEASER in terms of the harmonic mean
between accuracy and earliness, but it has higher training times compared to both. In
applications with high variance in measurements and high class imbalance, ECEC and
MLSTM achieve the highest harmonic mean scores. In multi-class classification cases,
MLSTM is the best choice with the lowest earliness scores, followed by Minirocket, which has
high accuracy and reduced training times. For the rest of the datasets we tested, Minirocket
is the most suitable algorithm for ETSC in terms of harmonic mean. It has very low
earliness scores and training times, although its predictive accuracy is worse than ECEC,
ECONOMY-K, and ECTS, which however achieve higher earliness scores.
As concerns future work, we are already in the process of further enriching our framework with additional ETSC algorithms and more datasets. We believe that establishing consolidated benchmarking procedures will be of great benefit for the ETSC community.

References