

OTTO-VON-GUERICKE-UNIVERSITY MAGDEBURG
Faculty of Computer Science



MASTER THESIS

A comparison of Time Series Classification Algorithms based on their ability to learn on diminishing time series

AUTHOR:
ISMAIL, WAHBA

MATRICULATION NUMBER:
217526

EXAMINER AND SUPERVISOR:
PROF. DR. MYRA SPILIOPOULOU
KNOWLEDGE MANAGEMENT AND DISCOVERY LAB
INSTITUTE OF TECHNICAL AND BUSINESS INFORMATION SYSTEMS
OTTO-VON-GUERICKE-UNIVERSITY MAGDEBURG

2ND SUPERVISOR:
NAME
INSTITUTE
UNIVERSITY

day.month.year

Wahba, Ismail:

A comparison of Time Series Classification Algorithms based on their ability to learn
on diminishing time series

Master Thesis, Otto-von-Guericke-University Magdeburg, year.

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pellentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

Acknowledgement

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pellentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

Table of Contents

List of Figures	i
List of Tables	ii
List of Abbreviations	iii
1 Introduction	1
1.1 Goal of the Thesis	2
1.2 Structure of the Thesis	2
2 Concepts and Terminology	3
2.1 Time Series Data	3
2.1.1 Nature of Time Series Data	3
2.2 Time Series Classification	4
2.3 Early Time Series Classification	4
2.4 notes	5
3 Appendix	8
4 Declaration of Authorship	12

List of Figures

List of Tables

List of Abbreviations

eTSCA	Early Time Series Classification Algorithms
HM	Harmonic Mean
i.i.d	Independent and Identically Distributed
MPL	Minimum Prediction Length
TSC	Time Series Classification
TSCA	Time Series Classification Algorithms

1 Introduction

Time Series Classification is a field of machine learning that has grabbed the attention of many researchers in the last decade. Time series data exists, by nature, in numerous real scenarios; medical examination records of patients{reference}, signal processing{reference}, weather forecasting{reference} and astronomy{reference} are some of them.

Classification of time series data has been tackled with different objectives; the first is concerned with the accuracy of classification as well as space and time complexity. This objective is referred to simply as Time Series Classification (TSC). While the second objective adds the factor of earliness as a primary goal and is referred to as Early Time Series Classification (eTSC).

Numerous algorithms have been introduced to tackle the problem of Time Series Classification. According to the [2], these algorithms can be divided, based on their technique, into six groups.

Whole time series algorithms{reference} compare two time series, usually by employing an elastic distance measure between all data points of both time series. Phase dependent interval algorithms{reference} operate by extracting informative features from intervals of time series, they are more suitable for long and noisy data than whole time series algorithms. Phase independent interval algorithms{reference} are used when a class can be identified using a single or multiple patterns regardless of when they occur during the time series. Dictionary based algorithms{reference} consider the number of repetitions of patterns as a factor of classification and not just simple occurrence of one. Ensembles{reference} combine the power of different algorithms, either of different or same core technique, then make the final classification decision based on voting. In addition to the previous algorithms, there are also deep learning time series algorithms which build classifiers using generative as well as discriminative models.

On the other hand, Early Time Series Classification algorithms are designed to deal with less data in order to achieve earliness of prediction, but of course this comes with a price of accuracy. Many of the ideas applied in TSC have also been applied in eTSC; including 1-NN with Minimum Prediction Length (MPL){reference}, Phase independent intervals{reference}, generative classifiers{reference} and ensembles{reference}.

Both, Time Series Classification Algorithms (TSCA) and Early Time Series Classification Algorithms (eTSCA), have introduced well performing algorithms in terms of their respective performance measures. Their algorithms have been tested on publicly available archives{reference}; to benchmark their performance on a diverse set of datasets with different characteristics.

According to [2], based on the "No free lunch theorem", no specific algorithm has proven to prevail over all others. This means that different problems with different datasets would require a choice between the algorithms based on how they perform on them, specially for non-public or non-experimented datasets. In this thesis, we tackle this idea; by offering a framework that runs state-of-the-art algorithms on the provided dataset and provides analyses about the performance of each algorithm.

Also due to their different objectives, TSCA and eTSCA have been dealt with as two different families. Which leaves studying the relationship between both algorithm families an open area for research. We study the relationship between TSCA and eTSCA, by extending TSCA to deal with earliness as a main objective and compare how they perform in an early time series classification problem context.

Goals

1.1 Goal of the Thesis

This master thesis had two main goals. The first goal was to create a testbed for comparing different algorithms on a non-public dataset. While the second one was to study the relationship between the two families of algorithms; TSCAs and eTSCAs.

The first goal was motivated by {reference to the great bake-off}, one of the most comprehensive review papers in the time series field. With it's release, Bagnall et. al has set the foundation methodology for accurately benchmarking the performance of TSCAs for the ,at that time, currently existing and for algorithms that will be developed in the future. In their experiment, they have used 85 datasets publicly available from UCR and UEA, the biggest two data archives. Our goal was to offer a testbed, which can be used on private datasets. It runs state of the art algorithms, then provides analysis about their classification performance. The provided analysis can help, based on empirical evidence, choose the best fitting algorithm in accordance with the problem at hand.

As for the second goal, we extended the study of relationship between TSCAs and eTSCAs. Both families offer a wide variety of algorithms, but have different objectives and thus have different approaches in their learning processes. TSCAs focus primarily on the accuracy of the classification. In order to achieve this goal, full utilization of the whole time series data is done to achieve the highest possible accurate results. While eTSCAs objective tries to maximize both accuracy and earliness together, which is hard to attain because of the contradicting nature between both{reference}. This is why eTSCAs try to learn with as least possible data points as possible while maintaining classification accuracy. This study investigated the ability of TSCAs to perform in a simulated early classification context. TSCAs were trained on shortened training data, while keeping record of models' accuracy measure in comparison to a baseline utilizing complete training data points.

1.2 Structure of the Thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pelentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

2 Concepts and Terminology

This Chapter discusses the definitions and background of the topics mentioned in this thesis. We discuss the nature of time series data, the two problems of time series classification (TSC) and early time series classification (eTSC) then present the different techniques encompassed by them.

2.1 Time Series Data

A time series is a finite sequence of ordered observations, either based on time or another aspect [1, 2]. The existence of the time component makes time series an abundant form of data that covers various domains like; medicine, finance, engineering and biology [15]. A time series dataset is a collection of time series instances.

Definition 1 A set T of n time series instances, $T = \{T_1, T_2, \dots, T_n\}$.

Each of the time series instances T_i consists of a sequence of observations.

Definition 2 A time series T_i , of length L is represented as $T_i = [t_{i1}, t_{i2}, \dots, t_{iL}]$.

Time series data can come in different forms. It is important to comprehend what different forms the data can take and what implicit assumptions they convey; to be able to choose the suitable algorithms and tools to deal with it.

The first form is when the observations of instances capture a singular value, this is referred to as univariate time series.

Definition 3 Univariate time series T_i , of length L is represented as $T_i = [t_{i1}, t_{i2}, \dots, t_{iL}]$. With t_{ij} as a real valued number.

While the other form is when multiple measurements are captured by the observations. According to [16], it is essential to differentiate between the two ways multiple time series can be generated; panel data and multivariate time series data.

If more than one variable is being observed during a single experiment, with each variable representing a different measurement; this is called multivariate time series.

Definition 4 Multivariate time series T_i , of length L is represented as $T_i = [t_{i1}, t_{i2}, \dots, t_{iL}]$. With t_{ij} having M dimensions, each is a univariate time series.

While panel data is when the same kind of measurements is collected from independent instances; like different patients or diverse industrial processes.

For panel data, it is possible to assume that the different instances are i.i.d, but this assumption doesn't hold for observation of a single instance. The same goes for multivariate time series, individual univariate observations are assumed to be statistically dependant.

2.1.1 Nature of Time Series Data

Having discussed the dependency assumptions in time the different forms of time series data. It is this dependency that makes time series data challenging for conventional machine learning algorithms, which are used for tabular and cross-sectional data. Tabular and cross-sectional data assume observations to be independent and identically distributed (i.i.d) [16].

If we were to tabularize time series data; convert it into a tabular form by considering each observation as an individual feature. Then it would be possible to apply conventional machine learning algorithms, under the implicit modelling assumption that observations are not ordered. This means that if the order of the features was changed, still the model result will not change. This assumption can work for some problems, but it doesn't have to work for all problems.

2.2 Time Series Classification

Time series classification is a subtype of the general classification problem, which considers the unique property of dependency between adjacent features of instances [5]. The main goal of time series classification is to learn a function f , which given a training dataset $T = \{T_1, T_2, \dots, T_n\}$ of time series instances along with their corresponding class labels $Y = \{y_1, y_2, \dots, y_n\}$ where $y_i \in \{1, 2, \dots, C\}$, can predict class labels for unseen instances [6].

Time series classification has been studied with different objectives, some papers focused on attaining the highest accuracy of classification as the main goal [13, 12, 4, 15, 24, 7], while other papers focused on attaining lower time complexity [22, 2, 26, 21, 23].

In this master thesis, we are more interested in assessing the results in terms of accuracy than time complexity. We define accuracy like [25]; as the percentage of correctly classified instances for a given dataset D , either being a training or testing dataset.

Definition 5 *Accuracy* = $\frac{\text{number of correct classifications}}{|D|}$

2.3 Early Time Series Classification

On another side, early time series classification is also a classification problem which considers the temporal nature of data, but with a slightly different objective and used for different scenarios other than time series classification.

eTSC's main objective is to learn a model which can classify unseen instances as early as possible, while maintaining a competitive accuracy compared to a model that uses full length data or to a user defined threshold [29]. Which is a very challenging objective; due to the, naturally, contradicting nature of earliness and accuracy. In general, the more data is made available for the model to learn the better accuracy it can attain [18, 27, 28, 19]. This is why many eTSC researches consider it as a problem of optimizing multiple objectives.

eTSC is needed in situations in which waiting for more data to arrive can be costly or when making late decisions can cause unfavorable results [17, 20, 14]. This is why eTSC has been applied in various domains like early medical diagnosis [10, 8], avoiding issues in network traffic flow [3], human activity recognition [30, 11] and early prediction of stock crisis [9].

We follow the definition of earliness mentioned by [25]; as the mean number of data points s after which a label is assigned.

Definition 6 *Earliness* = $\frac{\sum_{T_i \in D} \frac{s}{len(T_i)}}{|D|}$

As well as the objective measure, Harmonic mean (HM), mentioned by [8, 25], which includes both accuracy and earliness. For the problem we have, HM is a weighted average between accuracy and earliness.

Definition 7 $F_\beta = (1 + \beta^2) \frac{\text{accuracy}(1 - \text{earliness})}{\beta^2(1 - \text{accuracy}) + \text{earliness}}$

The value of β can be used to give higher importance to one of the aspects over the other, but we use equal weights for both.

2.4 notes

General notes about time series classification problem:

1. Types of Data

(a) Static Data

Data which describe characteristics of the studied instances or properties that won't change with time. These could be the date of birth of a patient, or the species of an animal.

(b) Dynamic Data

i. Several variables from one or more objects observed in a series of time

A. On one object - i time series data

B. On multiple objects - i Panel data

ii. Data Balance

A. Balanced: Observation carried out thoroughly on all objects in a series of time

B. Unbalanced: When several variables of each object can't be full observed in the same timeframe

2. Types of Models

(a) Univariate Model

We Predict an object with some characteristics will be in which group based on a categorical variable

(b) Multivariate Model

Same as univariate but on multiple categorical variables simultaneously

3. Different Time Series Classification Techniques

(a) Similarity-based techniques

(b) Interval-based techniques

(c) Shapelet-based techniques

(d) Dictionary-based techniques

(e) Combination of transformations

1. A time series forest for classification and feature extraction:

Deng, Houtao, et al. "A time series forest for classification and feature extraction." Information Sciences 239 (2013): 142-153.

The paper introduces a new Tree ensemble classifier for time series data called the Time Series Forest (TSF).

It tries to overcome the shortcomings of time-series instance based classifiers; like 1-NN with Euclidean distance and Nearest Neighbor with Dynamic Time Warping (NNDTW) because they provide few insights on the temporal features which are important for distinguishing different time series classes.

It uses simple summary statistics features (mean, std, slope) but outperforms the others.

It introduces a new technique for choosing the best split called Entrance gain (Entropy & distance) which is better than and cheaper than previous techniques.

It has lower computational complexity of $O(M)$ instead of $O(M^2)$ from the previous methods.

Can be extended by using more complex features like wavelets.

It assumes that input time series are of the same length, so it can be extended by using techniques that align time series with different lengths like Dynamic Time Warping (DTW)

2. A Shapelet Transform for Time Series Classification

Lines, Jason, et al. "A shapelet transform for time series classification." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012.

The paper tries to improve using shapelets for time series classification

Shapelets are subsequences of a time series that are considered representatives

Shapelets are easily interpretable, compact and classify new instances fastly, allow for the detection of phase-independent shape-based similarity of subsequences.

It proposes a shapelet transformation. Which is separating the process of finding shapelets from the classification step (this is how it is done in the original technique).

This allows for using any classifier.

Shapelet transform, tries to reduce complexity of original algorithm by choosing top K candidate shapelets instead of keeping all of them.

Then the candidate shapelets are used to transform data instances into a number of features, that can be used with any classifier.

It also proposes a new shapelet evaluation method to use with multi class problems (Compare F-Statistic with Information gain)

Can be extended by doing clustering for the extracted shapelets and not using top K , because there were a lot of similar shapelets.

3. Early Prediction on Time Series: A Nearest Neighbor Approach

Xing, Zhengzheng, Jian Pei, and S. Yu Philip. "Early prediction on time series: a nearest neighbor approach." Twenty-First International Joint Conference on Artificial Intelligence. 2009.

The paper introduces a new concept called Minimum Prediction Length (MPL), which allows for Early Classification of Time Series (ECTS) using 1-Nearest Neighbor

ECTS should be able to make earlier predictions using shorter time series than normal 1-NN on Time series data using full-length time series. While retaining accuracy.

It compares to 1-NN with Euclidean distance as distance measure, because it has proved itself to be one of the best techniques in Time Series clustering.

It keeps using shorter subsequences as long as they give the same accuracy of the full time series.

Using 1NN and 1RNN (reverse nearest neighbor) they try to identify the most confident minimum prediction length (MPL) for early prediction

Can be extended for streaming data

4. Faster and More Accurate Classification of Time Series by Exploiting a Novel Dynamic Time Warping Averaging Algorithm

Petitjean, François, et al. "Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm." Knowledge and Information Systems 47.1 (2016): 1-26.

The paper tries to extend 1NN with Dynamic Time Warping (DTW)

It uses the Nearest Centroid Classifier (NCC), an algorithm that generalizes Nearest Neighbor by introducing representative prototype (center of mass) for each class. This allows for way cheaper and faster classification $O(1)$ instead of $O(N)$.

NCC in some scenarios offer higher accuracy than 1NN. So NCC is preferred in cases of similar or higher accuracy due to its less resource requirements

The problem with creating centroid using the traditional DTW is that the resulting average can be a value that is not even a representative of any existing instance

Instead the paper uses DTW Barycenter Averaging (DBA), one of the best averaging algorithm for time series. It defines an average sequence and iteratively refines it following an expectation maximization scheme

5. TS-CHIEF: a scalable and accurate forest algorithm for time series classification Shifaz, Ahmed, et al. "Ts-chief: A scalable and accurate forest algorithm for time series classification." *Data Mining and Knowledge Discovery* (2020): 1-34.

A very recent technique

The new state of the art in time series classification

An ensemble classifier that competes with the previous HIVE-COTE and FLAT-COTE ensemble algorithms, but defeats them in time

It starts by using Proximity Forest, dictionary-based and interval-based algorithms to build an ensemble of classification trees. The splits of these trees are a set of time series references, an object would go down the path of the most similar reference.

At each node candidate splits are created, then the best split is selected using weighted Gini index

For classification, After a time series instance is passed down to the leaf nodes of the trees. The final classification of instance is made using a majority vote of K trees.

TS-CHIEF has an overall almost linear complexity in respect to training size

Can be extended for multivariate time series data and variable length datasets

3 Appendix

References

- [1] Amaia Abanda, Usue Mori, and Jose A Lozano. “A review on distance based time series classification”. In: *Data Mining and Knowledge Discovery* 33.2 (2019), pp. 378–412.
- [2] Anthony Bagnall et al. “The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances”. In: *Data Mining and Knowledge Discovery* 31.3 (2017), pp. 606–660.
- [3] Laurent Bernaille et al. “Traffic classification on the fly”. In: *ACM SIGCOMM Computer Communication Review* 36.2 (2006), pp. 23–26.
- [4] Aaron Bostrom and Anthony Bagnall. “A shapelet transform for multivariate time series classification”. In: *arXiv preprint arXiv:1712.06428* (2017).
- [5] Aaron Bostrom and Anthony Bagnall. “Binary Shapelet Transform for Multiclass Time Series Classification”. In: *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXII: Special Issue on Big Data Analytics and Knowledge Discovery*. Ed. by Abdelkader Hameurlain et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 24–46. ISBN: 978-3-662-55608-5. DOI: 10.1007/978-3-662-55608-5_2. URL: https://doi.org/10.1007/978-3-662-55608-5_2.
- [6] Houtao Deng et al. “A time series forest for classification and feature extraction”. In: *Information Sciences* 239 (2013), pp. 142–153.
- [7] Hassan Ismail Fawaz et al. “Inceptiontime: Finding alexnet for time series classification”. In: *Data Mining and Knowledge Discovery* 34.6 (2020), pp. 1936–1962.
- [8] Mohamed F Ghalwash and Zoran Obradovic. “Early classification of multivariate temporal observations by extraction of interpretable shapelets”. In: *BMC bioinformatics* 13.1 (2012), p. 195.
- [9] Mohamed F Ghalwash, Vladan Radosavljevic, and Zoran Obradovic. “Utilizing temporal patterns for estimating uncertainty in interpretable early decision making”. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2014, pp. 402–411.
- [10] M Pamela Griffin and J Randall Moorman. “Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis”. In: *Pediatrics* 107.1 (2001), pp. 97–104.
- [11] Ashish Gupta et al. “A Fault-Tolerant Early Classification Approach for Human Activities using Multivariate Time Series”. In: *IEEE Transactions on Mobile Computing* (2020).
- [12] Young-Seon Jeong, Myong K Jeong, and Olufemi A Omitaomu. “Weighted dynamic time warping for time series classification”. In: *Pattern recognition* 44.9 (2011), pp. 2231–2240.

- [13] Rohit J Kate. “Using dynamic time warping distances as features for improved time series classification”. In: *Data Mining and Knowledge Discovery* 30.2 (2016), pp. 283–312.
- [14] Yu-Feng Lin et al. “Reliable early classification on multivariate time series with numerical and categorical attributes”. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2015, pp. 199–211.
- [15] Jason Lines, Sarah Taylor, and Anthony Bagnall. “Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles”. In: *ACM Transactions on Knowledge Discovery from Data* 12.5 (2018).
- [16] Markus Löning et al. “sktime: A unified interface for machine learning with time series”. In: *arXiv preprint arXiv:1909.07872* (2019).
- [17] Usue Mori et al. “Early classification of time series by simultaneously optimizing the accuracy and earliness”. In: *IEEE transactions on neural networks and learning systems* 29.10 (2017), pp. 4569–4578.
- [18] Usue Mori et al. “Early classification of time series using multi-objective optimization techniques”. In: *Information Sciences* 492 (2019), pp. 204–218.
- [19] Usue Mori et al. “Reliable early classification of time series based on discriminating the classes over time”. In: *Data mining and knowledge discovery* 31.1 (2017), pp. 233–263.
- [20] Nathan Parrish et al. “Classifying with confidence from incomplete information”. In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 3561–3589.
- [21] François Petitjean et al. “Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm”. In: *Knowledge and Information Systems* 47.1 (2016), pp. 1–26.
- [22] Chotirat Ann Ratanamahatana and Eamonn Keogh. “Making time-series classification more accurate using learned constraints”. In: *Proceedings of the 2004 SIAM international conference on data mining*. SIAM. 2004, pp. 11–22.
- [23] Patrick Schäfer and Ulf Leser. “Fast and accurate time series classification with weasel”. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017, pp. 637–646.
- [24] Patrick Schäfer and Ulf Leser. “Multivariate time series classification with WEASEL+ MUSE”. In: *arXiv preprint arXiv:1711.11343* (2017).
- [25] Patrick Schäfer and Ulf Leser. “TEASER: early and accurate time series classification”. In: *Data Mining and Knowledge Discovery* 34.5 (2020), pp. 1336–1362.
- [26] Chang Wei Tan, François Petitjean, and Geoffrey I Webb. “FastEE: Fast Ensembles of Elastic Distances for time series classification”. In: *Data Mining and Knowledge Discovery* 34.1 (2020), pp. 231–272.

- [27] Romain Tavenard and Simon Malinowski. “Cost-aware early classification of time series”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2016, pp. 632–647.
- [28] Zhengzheng Xing, Jian Pei, and S Yu Philip. “Early classification on time series”. In: *Knowledge and information systems* 31.1 (2012), pp. 105–127.
- [29] Zhengzheng Xing, Jian Pei, and S Yu Philip. “Early prediction on time series: a nearest neighbor approach”. In: *Twenty-First International Joint Conference on Artificial Intelligence*. Citeseer. 2009.
- [30] Omolbanin Yazdanbakhsh and Scott Dick. “Multivariate Time Series Classification using Dilated Convolutional Neural Network”. In: *arXiv preprint arXiv:1905.01697* (2019).

4 Declaration of Authorship

I hereby declare that I have written this thesis "TITLE TITLE TITLE" without any help from others and without the use of documents and aids other than those stated above. Furthermore, I have mentioned all used sources and have cited them correctly according to the citation rules. Moreover, I confirm that the paper at hand was not submitted in this or similar form at another examination office, nor has it been published before.

Magdeburg, DATE, SIGNATURE