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

List of Tables

List of Abbreviations

DTW	Dynamic Time Warping
EE	Elastic Ensemble
ERP	Edit Distance with Real Penalty
eTSCA	Early Time Series Classification Algorithms
HM	Harmonic Mean
i.i.d	Independent and Identically Distributed
KNN	K-Nearest Neighbor Algorithm
KNNED	K-Nearest Neighbor using Euclidean Distance
LCSS	Longest Common Subsequence
MPL	Mimimum Prediction Length
MSM	Move-Split-Merge
PF	Proximity Forest
TSC	Time Series Classification
TSCA	Time Series Classification Algorithms
TWE	Time Warp Edit
WDTW	Weighted Dynamic Time Warping

1 Introduction

Time Series Classification is a field of machine learning that has grabbed the attention of many reasearches 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 [3], 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 occurence 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 Mimimum 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 [3], 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 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.

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, 3]. The existence of the time component makes time series an abundant form of data that covers various domains like; medicine, finance, engineering and biology [26]. A time series dataset is a collection of time series instances.

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

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* $T_i = [t_{i1}, t_{i2}, ...t_{iL}]$.

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* $T_i = [t_{i1}, t_{i2}, ...t_{iL}]$*. With* t_{ij} *as a real valued number.*

While the other form is when multiple measurments are captured by the observations. According to [27], 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* $T_i = [t_{i1}, t_{i2}, ...t_{iL}]$*. With* t_{ij} *having M dimensions, each is a univariate time series.*

While panel data is when the same kind of measurments 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) [27].

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 the unique property of dependency between adjacent features of instances [7]. The main goal of time series classification is to learn a function f, which given a training dataset T = $\{T_1, T_2, ..., T_n\}$ of time series instances along with their corresponding class labels Y=$\{y_1, y_2, ... y_n\}$ where $y_i \in \{1,2,...C\}$, can predict class labels for unseen instances [10].
Time series classification has been studied with different objectives, some papers focused on attaining the highest accuracy of classification as the main goal [22, 21, 6, 26, 38, 13], while other papers focused on attaining lower time complexity [35, 3, 41, 34, 37].
In this master thesis, we are more interested in assessing the results in terms of accuracy than time complexity. We define accuracy like [39]; as the percentage of correctly classified instances for a given dataset D, either being a training or testing dataset.

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

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[46]. 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 [32, 43, 47, 31]. 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 [30, 33, 24]. This is why eTSC has been applied in various domains like early medical diagnosis [17, 14], avoiding issues in network traffic flow [5], human activity recognition [48, 18] and early prediction of stock crisis [15].
We follow the definition of earliness mentioned by [39]; as the mean number of data points s after which a label is assigned.

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

As well as the objective measure, Harmonic mean (HM), mentioned by [14, 39], which includes both accuracy and earliness. For the problem we have, HM is a weighted average between accuracy and earliness.

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

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 -¿ time series data
         B. On multiple objects -¿ 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(M2) 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 instaces 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 prefered in cases of similar or higher accuracy due to it's 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 ensenble algorithms, but defeats them in time
It starts by using Proximity Forest, dictionary-based and interval-based algorithms to build an esnsemble 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 Time Series Classification Algorithms

This chapter will introduce different types of TSCA. There are multiple ways to divide TSCAs

3.1 Whole Time Series Algorithms

Whole time series similarity algorithms, also called distance-based algorithms, are methods that compare pairs of time series instances. An unlabeled time series instance is given the class label of the nearest instance in a training data set [22]. There are two main techniques for carrying out the comparison; either by comparing vector representations of the time series instances, or by combining a defined distance function with a classifier, KNN being the most common one [26]. Whole time series algorithms are best suited for problems where the unique features can exist anywhere along the whole time series[3].
One of the simplest forms of whole time series is 1-NN with Euclidean Distance [12], yet it can suprisingly attain high accuracy compared to other distance measures [45]. But Nearest Neighbor with Euclidean Distance (KNNED) is an easy to beat baseline, due to it's sensitivity for distortion and inability to handle time series of unequal lengths [45, 22, 26]. This lead many of the researchers to focus on defining more advanced and elastic distance measures that can compensate for misalignment between time series [1]. The standard and most common baseline classifier utilizing elastic distance measures is 1-NN with Dynamic Time Warping (DTW) [3]. In contrast to the idea that more powerful machine learning algorithms will be able to defeat the simple KNN and an elastic measure, DTW has proved to be a very tough opponent to other algorithms and other elastic distance measures as well [22, 25, 44]. But there were also other distance metrics that have been introduced in literature, these inlcude extensions of DTW on one hand like; Weighted Dynamic Time Warping (WDTW) which penalizes large warpings based on distance [21] and Derivative Dynamic Time Warping (DDTW) [23, 16] which uses derivatives of sequences as features rather than raw data to avoid singularities. On the other hand, Edit Distance with Real Penalty (ERP) [8], Time Warp Edit (TWE) [29], Longest Common Subsequence (LCSS) [9] and Move-Split-Merge (MSM) [40] are all alternatives for distance measures, yet multiple experiments have considered DTW to be relatively unbeatable [3, 1, 6]. To the extend of our knowledge, the most powerful whole time series classifiers are Elastic Ensemble (EE) [25] and Proximity Forest (PF) [28].

3.1.1 Nearest Neighbor with ED

The Euclidean distance is a remarkably simple technique to calculate the distance between time series instances. Given two instances $T_1 = [t_{11}, t_{12}, ... t_{1n}]$ and $T_2 = [t_{21}, t_{22}, ... t_{2n}]$, the euclidean distance between them can be determined as:

Definition 8 $ED(T_1,T_2)= \sqrt{\sum_{i=1}^{n}(t_{1i} - t_{2i})^2}$

Euclidean distnace has been preferred to other classifiers due to it's space and time efficiency, but it suffers from two main shortcomings [4, 21, 22]. The first one is that it cannot handle comparisons between time series of different lengths. While the second one is it's sensitivity to minor discrepancies between time series; it would calculate large distance values for small shiftings or misalignments. Although other metrics have been introduced to overcome the drawbacks of euclidean distance, experimental proof showed that this is only the case for small datasets, but for larger datasets the accuracy of other elastic measures converge with euclidean distance [19, 11, 2].

3.1.2 Nearest Neighbor with DTW

Dynamic Time Warping was a very strong baseline for time series classification for a long time [1, 3]. It was first introduced as a technique to recognize spoken words that can deal with misallignments between time series that Euclidean Distance couldn't handle [41].
To calculate the distance between two time series instances $T_1 = [t1_1, t1_2, ...t1_{\mathrm{m}}]$ and $T_2 = [t2_1, t2_2, ...t2_{\mathrm{m}}]$; a distance matrix $M(T_1, T_2)$, of size m × m, is calculated for T_1 and T_2. With $M_{\mathrm{i,j}}(t1_{\mathrm{i}}, t2_{\mathrm{j}})$ representing the distance between $t1_{\mathrm{i}} \in T_1$ and $t2_{\mathrm{j}} \in T_2$. The goal of DTW is to find an optimal path that minimizes the cumulative distance between points of T_1 and T_2.
A candidate path $P = [p_1, p_2, ...p_{\mathrm{p}}]$ is to be found by traversing M. For a path to be valid it must conform to some conditions:

- $p_1 = (t1_1, t2_1)$
- $p_{\mathrm{p}} = (t1_{\mathrm{m}}, t2_{\mathrm{m}})$
- for all i <m :
  - $0 \leq t1_{\mathrm{i}+1} - t1_{\mathrm{i}} \leq 1$
  - $0 \leq t2_{\mathrm{i}+1} - t2_{\mathrm{i}} \leq 1$

Finding an optimal path under DTW can be computationally expensive with complexity of $O(n^2)$ for a time series of length n [37, 34]. Consequently it is usual to use a contraint with the path; to prevent comparison of points outside a certain window [41]; like the famous Sakoe-Chiba Band [36], Itakura Parallelogram [20] and Ratanamahatana-Keogh Band [35]. Typically DTW can make use of it's warping window to handle distortion in time series, but still it is vulnerable to cases where the differene in length between instances length is larger than the warping window [42].

3.1.3 Neares Neighbor with MSM Distance

3.1.4 Elastic Ensemble

4 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. "Transformation based ensembles for time series classification". In: *Proceedings of the 2012 SIAM international conference on data mining*. SIAM. 2012, pp. 307–318.

[3] 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.

[4] Mustafa Gokce Baydogan, George Runger, and Eugene Tuv. "A bag-of-features framework to classify time series". In: *IEEE transactions on pattern analysis and machine intelligence* 35.11 (2013), pp. 2796–2802.

[5] Laurent Bernaille et al. "Traffic classification on the fly". In: *ACM SIGCOMM Computer Communication Review* 36.2 (2006), pp. 23–26.

[6] Aaron Bostrom and Anthony Bagnall. "A shapelet transform for multivariate time series classification". In: *arXiv preprint arXiv:1712.06428* (2017).

[7] 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`.

[8] Lei Chen and Raymond Ng. "On the marriage of lp-norms and edit distance". In: *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. 2004, pp. 792–803.

[9] Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. "Finding similar time series". In: *European Symposium on Principles of Data Mining and Knowledge Discovery*. Springer. 1997, pp. 88–100.

[10] Houtao Deng et al. "A time series forest for classification and feature extraction". In: *Information Sciences* 239 (2013), pp. 142–153.

[11] Hui Ding et al. "Querying and mining of time series data: experimental comparison of representations and distance measures". In: *Proceedings of the VLDB Endowment* 1.2 (2008), pp. 1542–1552.

[12] Christos Faloutsos, Mudumbai Ranganathan, and Yannis Manolopoulos. "Fast subsequence matching in time-series databases". In: *Acm Sigmod Record* 23.2 (1994), pp. 419–429.

[13] Hassan Ismail Fawaz et al. "Inceptiontime: Finding alexnet for time series classification". In: *Data Mining and Knowledge Discovery* 34.6 (2020), pp. 1936–1962.

[14] 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.

[15] 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.

[16] Tomasz Górecki and Maciej Łuczak. "Using derivatives in time series classification". In: *Data Mining and Knowledge Discovery* 26.2 (2013), pp. 310–331.

[17] 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.

[18] Ashish Gupta et al. "A Fault-Tolerant Early Classification Approach for Human Activities using Multivariate Time Series". In: *IEEE Transactions on Mobile Computing* (2020).

[19] Jon Hills et al. "Classification of time series by shapelet transformation". In: *Data Mining and Knowledge Discovery* 28.4 (2014), pp. 851–881.

[20] Fumitada Itakura. "Minimum prediction residual principle applied to speech recognition". In: *IEEE Transactions on acoustics, speech, and signal processing* 23.1 (1975), pp. 67–72.

[21] 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.

[22] 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.

[23] Eamonn J Keogh and Michael J Pazzani. "Derivative dynamic time warping". In: *Proceedings of the 2001 SIAM international conference on data mining*. SIAM. 2001, pp. 1–11.

[24] 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.

[25] Jason Lines and Anthony Bagnall. "Time series classification with ensembles of elastic distance measures". In: *Data Mining and Knowledge Discovery* 29.3 (2015), pp. 565–592.

[26] 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).

[27] Markus Löning et al. "sktime: A unified interface for machine learning with time series". In: *arXiv preprint arXiv:1909.07872* (2019).

[28] Benjamin Lucas et al. "Proximity forest: an effective and scalable distance-based classifier for time series". In: *Data Mining and Knowledge Discovery* 33.3 (2019), pp. 607–635.

[29] Pierre-François Marteau. "Time warp edit distance with stiffness adjustment for time series matching". In: *IEEE transactions on pattern analysis and machine intelligence* 31.2 (2008), pp. 306–318.

[30] 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.

[31] 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.

[32] Usue Mori et al. "Early classification of time series using multi-objective optimization techniques". In: *Information Sciences* 492 (2019), pp. 204–218.

[33] Nathan Parrish et al. "Classifying with confidence from incomplete information". In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 3561–3589.

[34] 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.

[35] 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.

[36] Hiroaki Sakoe and Seibi Chiba. "Dynamic programming algorithm optimization for spoken word recognition". In: *IEEE transactions on acoustics, speech, and signal processing* 26.1 (1978), pp. 43–49.

[37] 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.

[38] Patrick Schäfer and Ulf Leser. "Multivariate time series classification with WEASEL+ MUSE". In: *arXiv preprint arXiv:1711.11343* (2017).

[39] Patrick Schäfer and Ulf Leser. "TEASER: early and accurate time series classification". In: *Data Mining and Knowledge Discovery* 34.5 (2020), pp. 1336–1362.

[40] Alexandra Stefan, Vassilis Athitsos, and Gautam Das. "The move-split-merge metric for time series". In: *IEEE transactions on Knowledge and Data Engineering* 25.6 (2012), pp. 1425–1438.

[41] 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.

[42] Chang Wei Tan et al. "Time series classification for varying length series". In: *arXiv preprint arXiv:1910.04341* (2019).

[43] 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.

[44] Xiaoyue Wang et al. "Experimental comparison of representation methods and distance measures for time series data". In: *Data Mining and Knowledge Discovery* 26.2 (2013), pp. 275–309.

[45] Zhengzheng Xing, Jian Pei, and Eamonn Keogh. "A brief survey on sequence classification". In: *ACM Sigkdd Explorations Newsletter* 12.1 (2010), pp. 40–48.

[46] 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.

[47] Zhengzheng Xing, Jian Pei, and S Yu Philip. "Early classification on time series". In: *Knowledge and information systems* 31.1 (2012), pp. 105–127.

[48] Omolbanin Yazdanbakhsh and Scott Dick. "Multivariate Time Series Classification using Dilated Convolutional Neural Network". In: *arXiv preprint arXiv:1905.01697* (2019).

5 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