Time Series Machine Learning
with the aeon toolkit

Tony Bagnall
A.J.Bagnall@soton.ac.uk

School of Electronics and Computer Science
University of Southampton

University of East Anglia

University of Southampton

Talk Structure

What is time series machine learning?

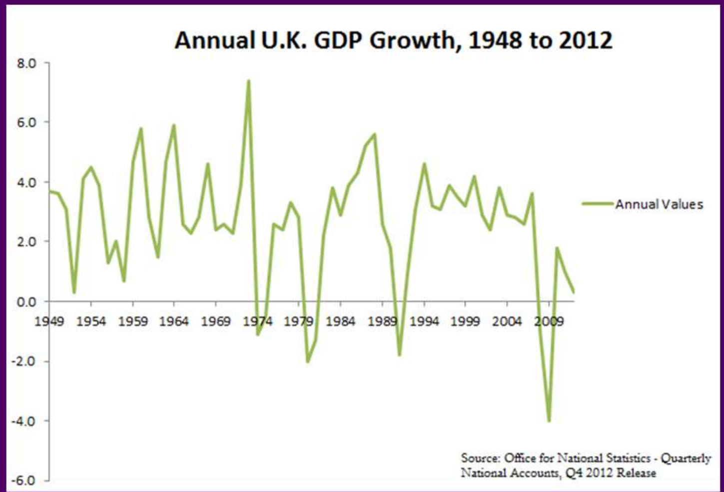
- Classification
- Clustering
- Regression

What is a time series?

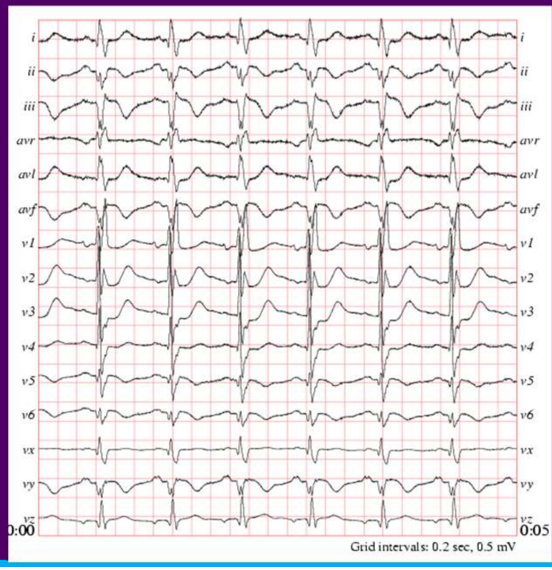
A time series is an ordered list of observations of real valued variable

If each observation is a scalar, we call it a univariate time series

If each observation is a vector of observations we call it a multivariate time series



The data does not need to be ordered in time (sometimes called a data series)



Time Series Machine Learning Repository
www.timeseriesclassification.com

Introduced in 2002 and expanded several times since, the archive datasets have been used in thousands of papers

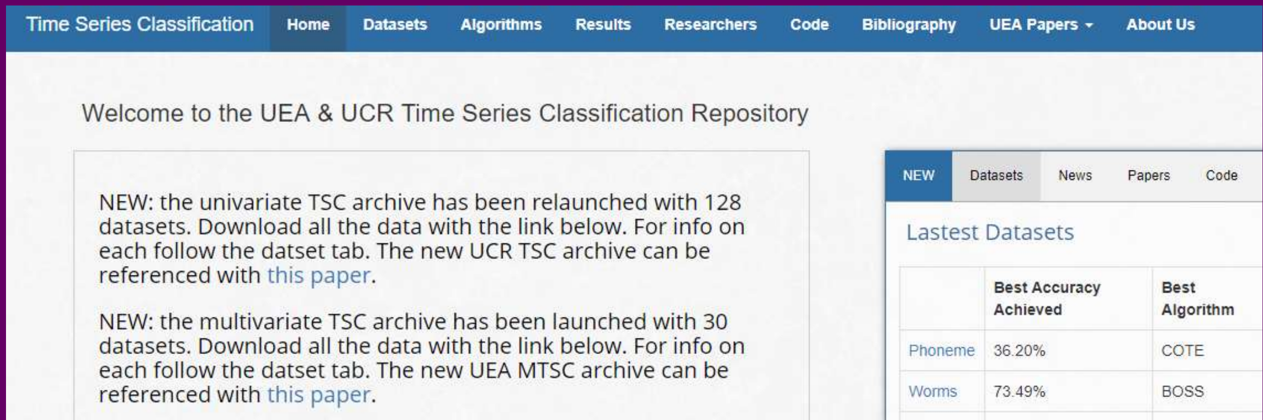
About half donated by the TSML group at UEA/Southampton

Expanded in 2018 to 128 datasets
Multivariate Archive introduced in 2018 with 30 datasets

University of East Anglia

University of Southampton

UNIVERSITY OF CALIFORNIA UCRIVERSIDE



We would like to thank everyone who donates and helps maintain these archives

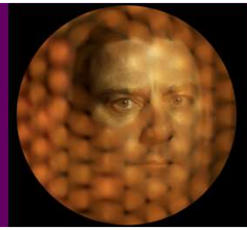
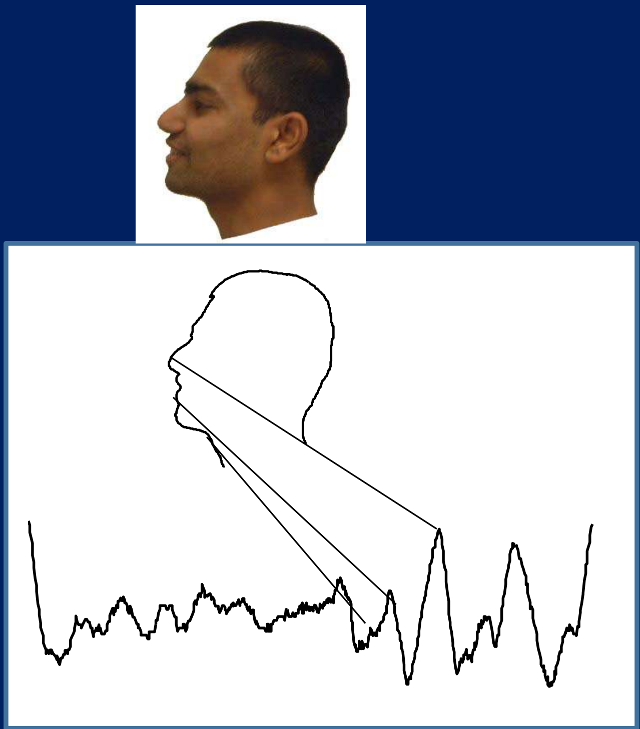
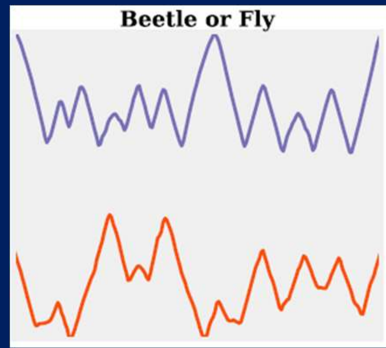
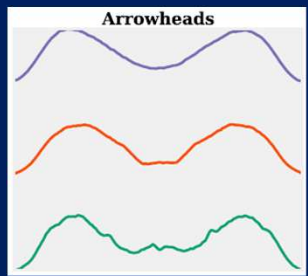


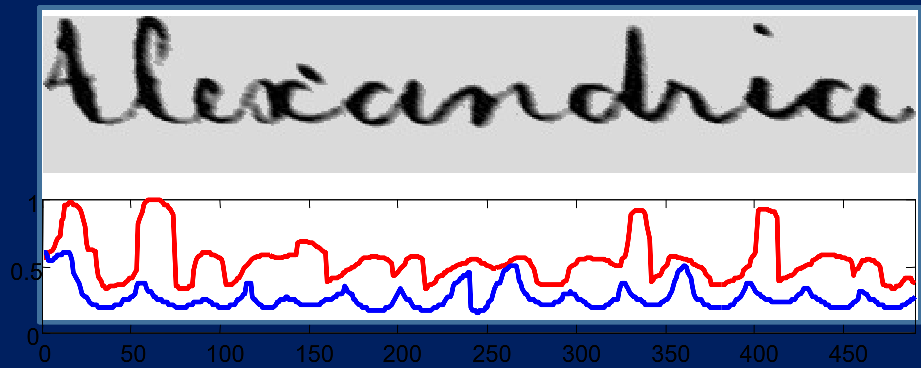
Image Outlines

Hand and bone outlines, Herring Otoliths, Faces, Leaves, Arrow Heads, Yoga, Words/letter, Shapes (MPEG7)





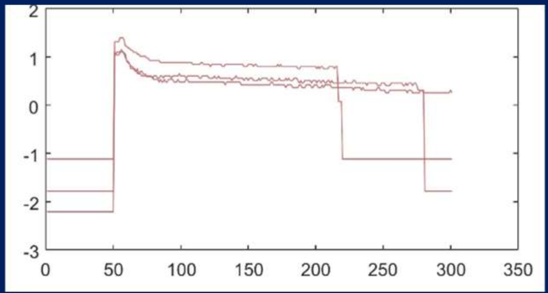


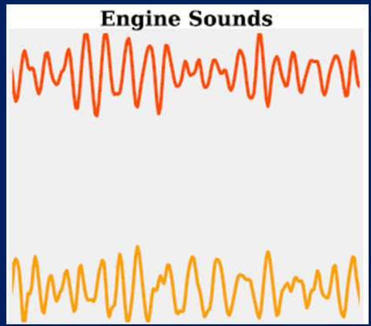


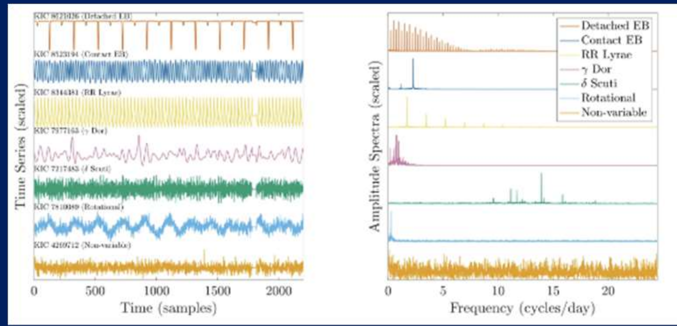
Sensor Data

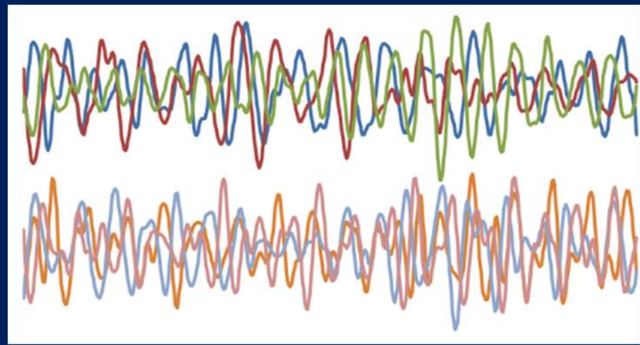
Car engines, light curves, lightning, electric devices

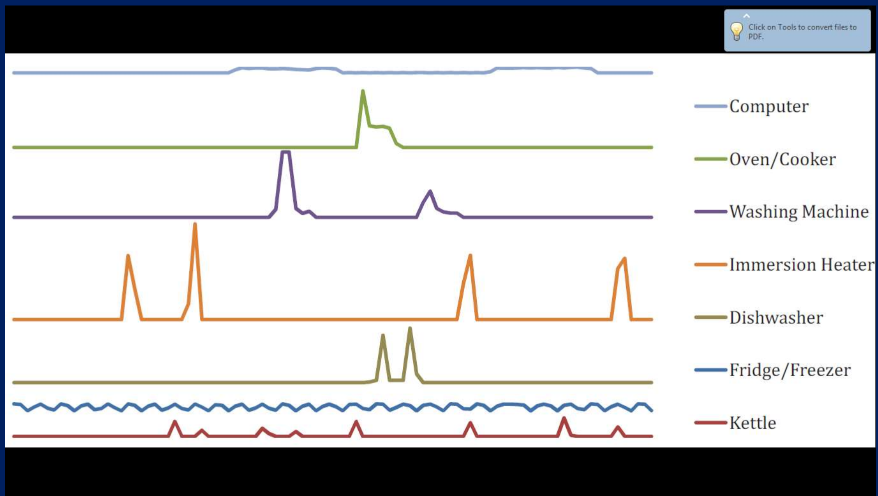
Insect wing beats, Car Engines, Phonemes (sound), Worm Motion

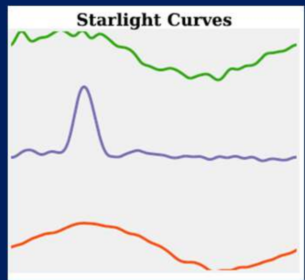


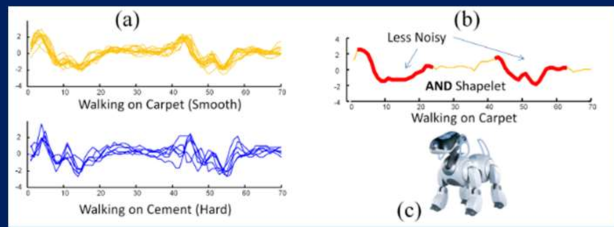






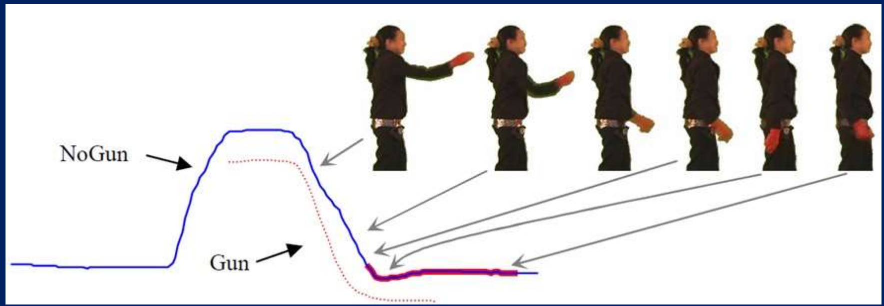


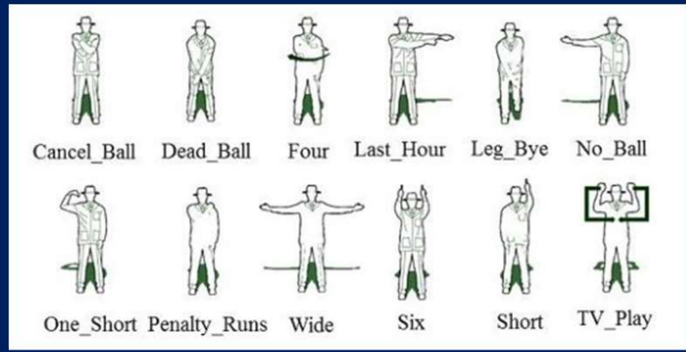


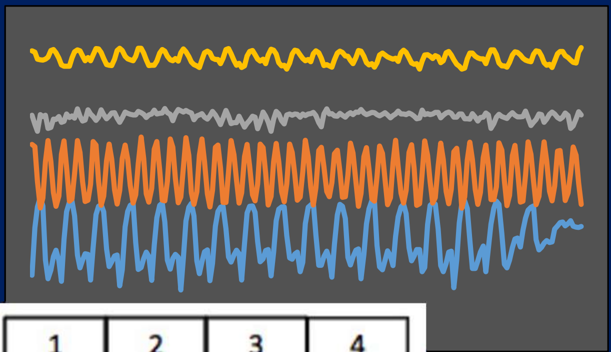


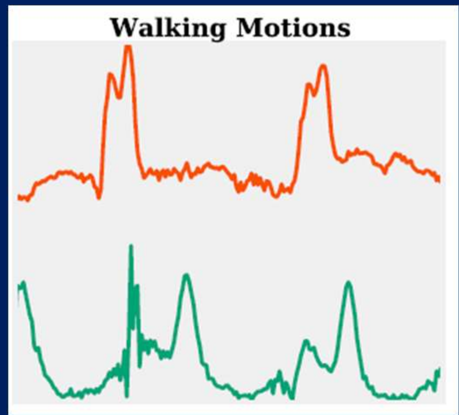
Human Activity Recognition

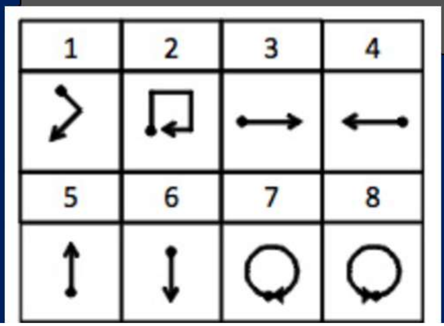
Gestures (Uwave), Cricket hand signals, Gun Point, Asphalt road condition, inline skating

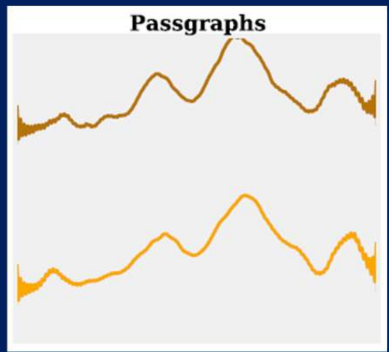


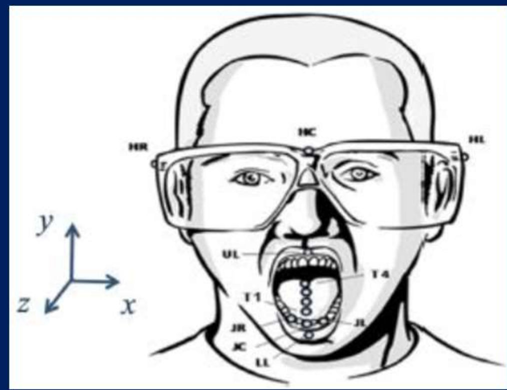






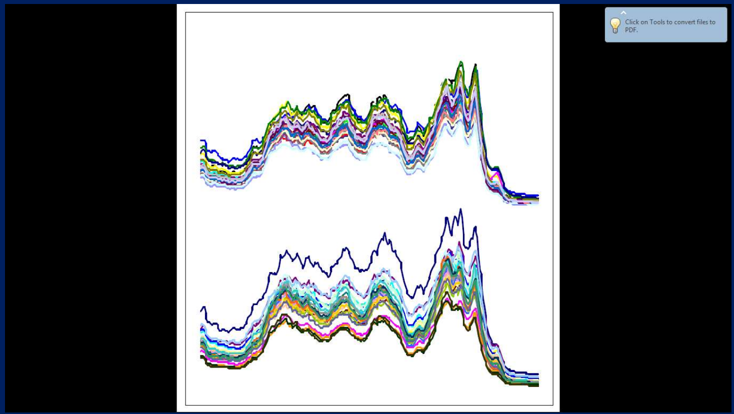


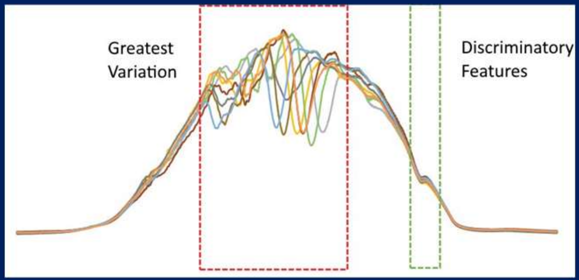


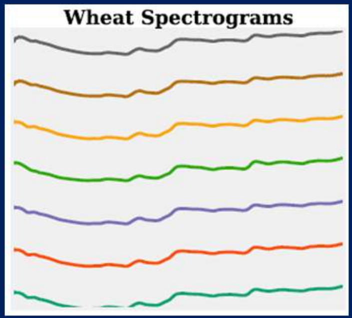


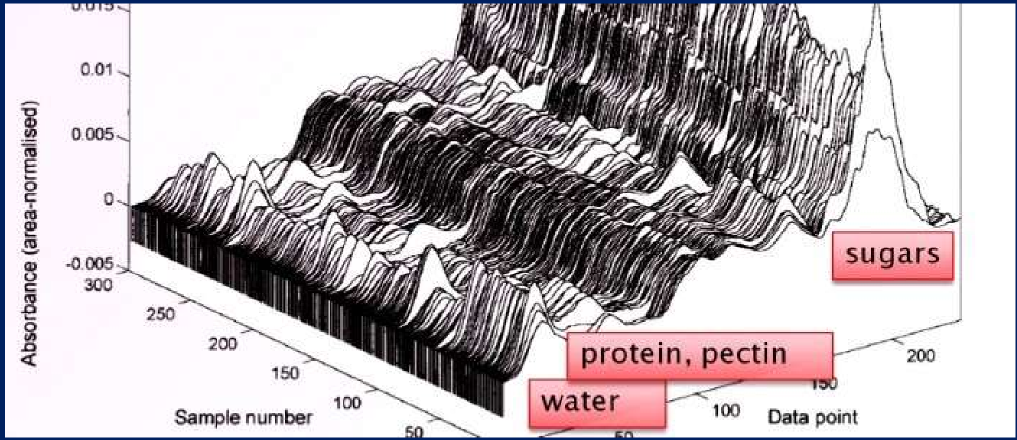
Food Spectrographs

Beef, Coffee, Ham, Meat, Olive Oil, Strawberry, Wine, Whisky



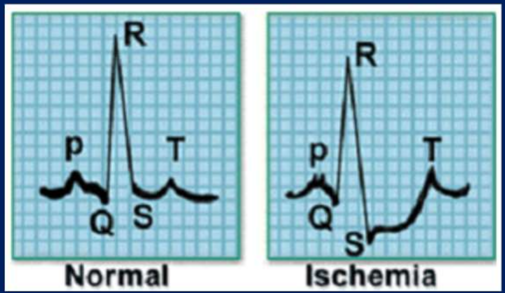


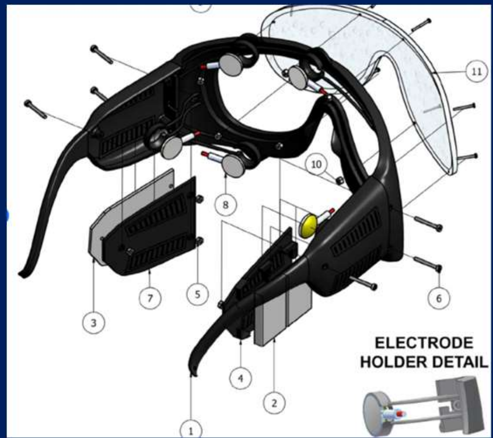


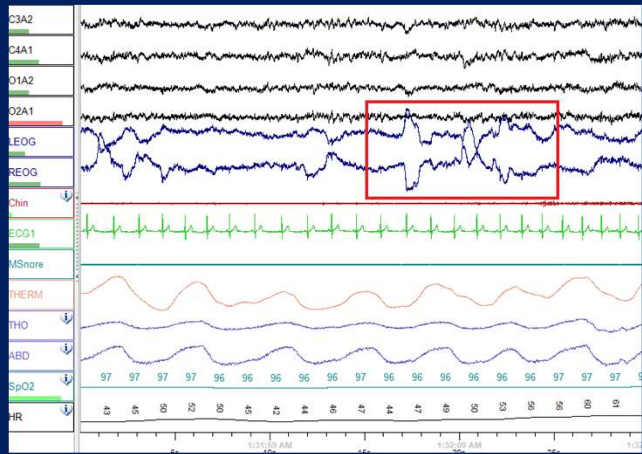


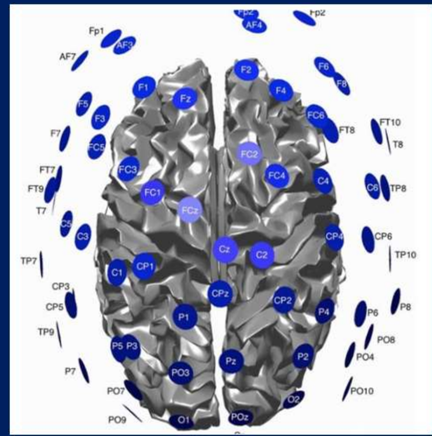
Biomedical signals

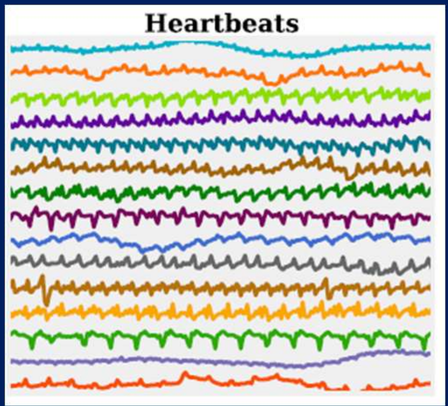
ECG, EEG, MEG, EOG

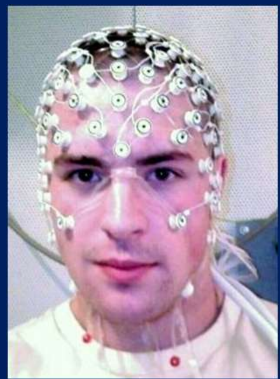


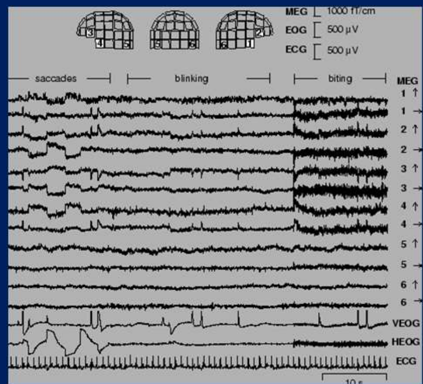


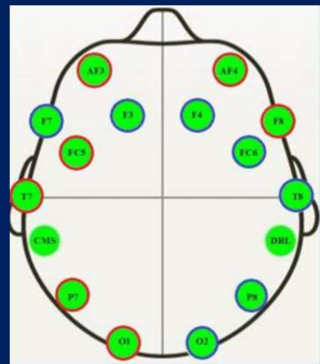




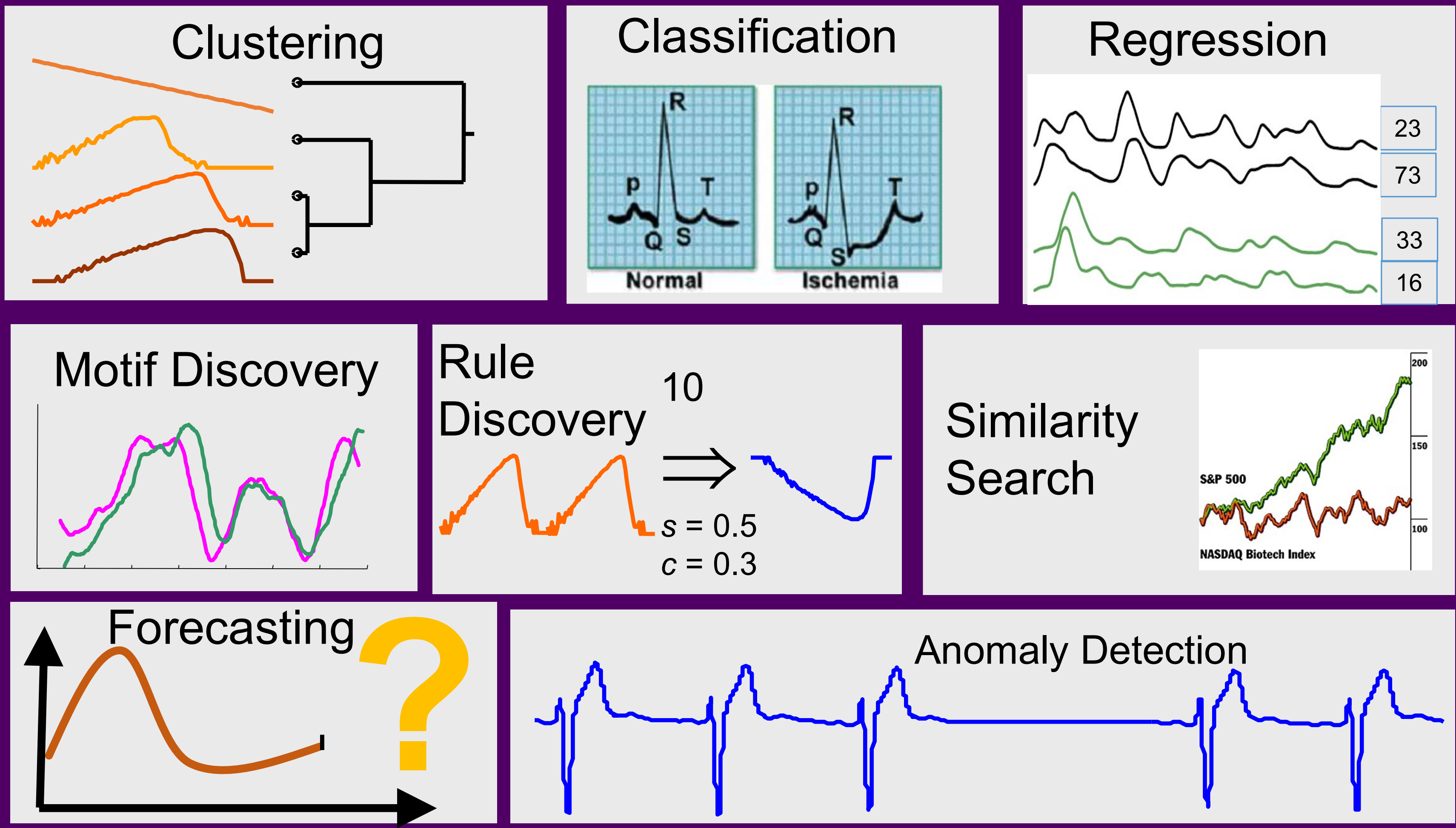




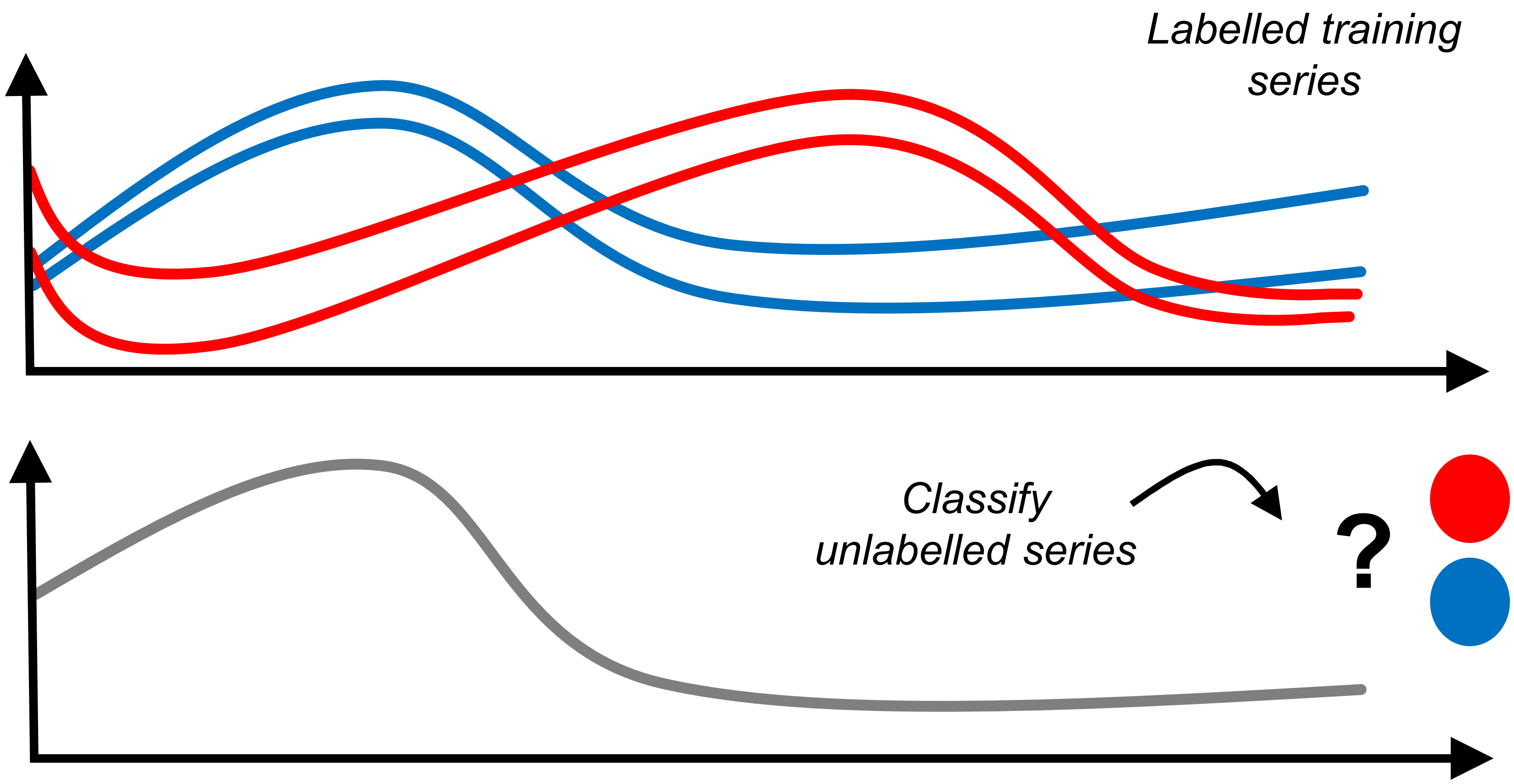




Time Series Machine Learning Tasks



Time Series Classification (TSC)



Home > Data Mining and Knowledge Discovery > Article

The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances

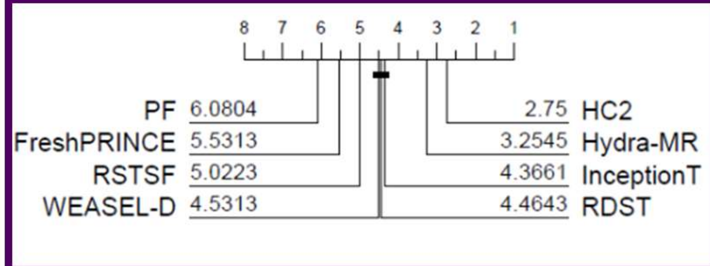
Open access | Published: 23 November 2016 | **31**, 606–660 (2017)

02-03-2019

Deep learning for time series classification: a review

Authors: Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, Pierre-Alain Muller

Published in: Data Mining and Knowledge Discovery | Issue 4/2019



arXiv > cs > arXiv:2304.13029

Search...

Help | A

Computer Science > Machine Learning

[Submitted on 25 Apr 2023]

Bake off redux: a review and experimental evaluation of recent time series classification algorithms

Matthew Middlehurst, Patrick Schäfer, Anthony Bagnall

aeon tsml toolkit

https://www.aeon-toolkit.org/

The Alan Turing Institute

aeon C:\Code\aeon
- .binder
- .github
- .pytest_cache
- aeon
  - annotation
  - anomaly_detection
  - base
  - benchmarking
  - classification
  - clustering
  - datasets
  - datatypes
  - distances
  - forecasting
  - local
  - networks
  - performance_metrics
  - pipeline
  - registry
  - regression
  - segmentation
  - similarity_search
  - tests
  - transformations

aeon-toolkit.org/en/latest/

Search

USING AEON

Installation

Getting Started

API Reference

Examples

DEVELOPING AEON

Get Involved

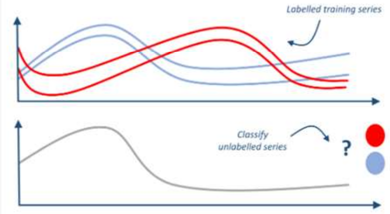
Development

Welcome to aeon

aeon is a scikit-learn compatible toolkit for time series tasks such as forecasting, classification and clustering.

- Provides a broad library of time series algorithms, including the latest advances.
- Efficient implementation of time series algorithms using numba.
- Interfaces with other time series packages to provide a single framework for algorithm comparison.

Get started with time series forecasting.



Get started with time series classification.

EPSRC
Pioneering research and skills

Engineering and Physical Sciences Research Council

GoW Search

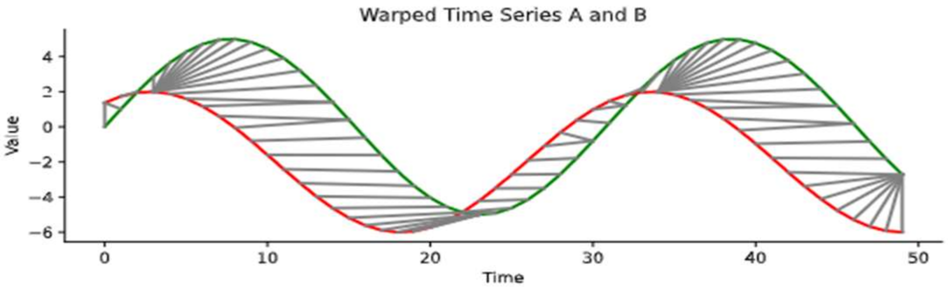
Home | GoW Home | Back | Research Areas | Topic | Sector | Scheme | Region | Theme | Organisation | Partners

Details of Grant

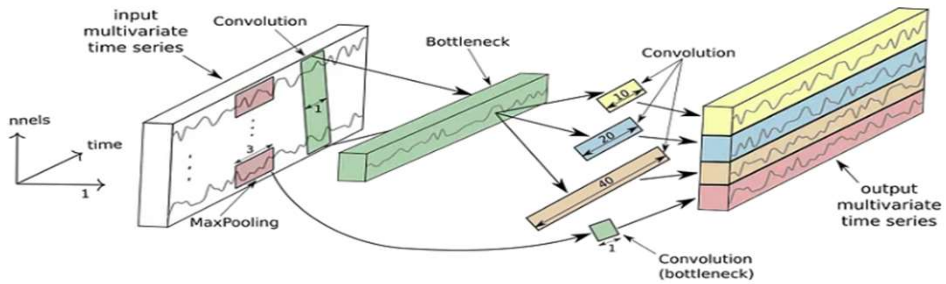
EPSRC Reference:	EP/W030756/2				
Title:	aeon: a toolkit for machine learning with time series				
Principal Investigator:	Bagnall, Professor A				
Other Investigators:	Sambrook, Dr TD	Sami AK, Dr S		Renoult, Dr L	
Researcher Co-Investigators:					
Project Partners:	GlaxoSmithKline plc (GSK) The Alan Turing Institute University of Cambridge	Mercedes-Benz AG UCL		Monash University University of California Riverside	
Department:	Electronics and Computer Science				
Organisation:	University of Southampton				
Scheme:	Standard Research				
Starts:	01 August 2023	Ends:	30 September 2025	Value (£):	403,617

Taxonomy of Time Series Classification Algorithms Part I

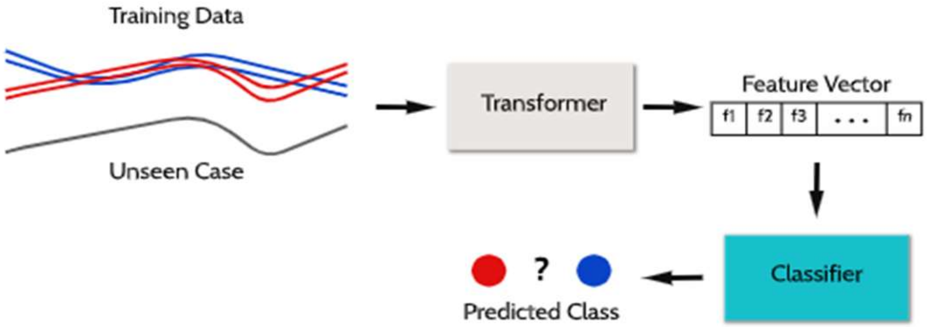
Distance based



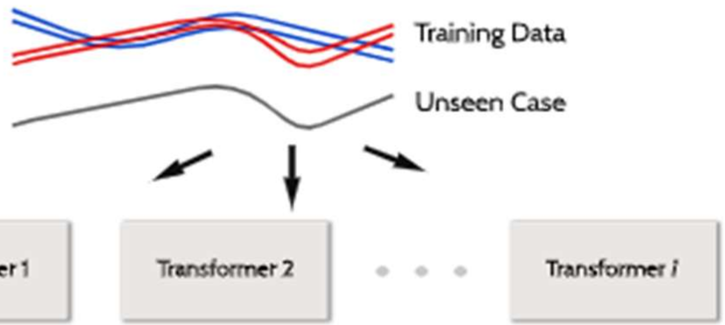
Deep learning



Feature based



Interval based

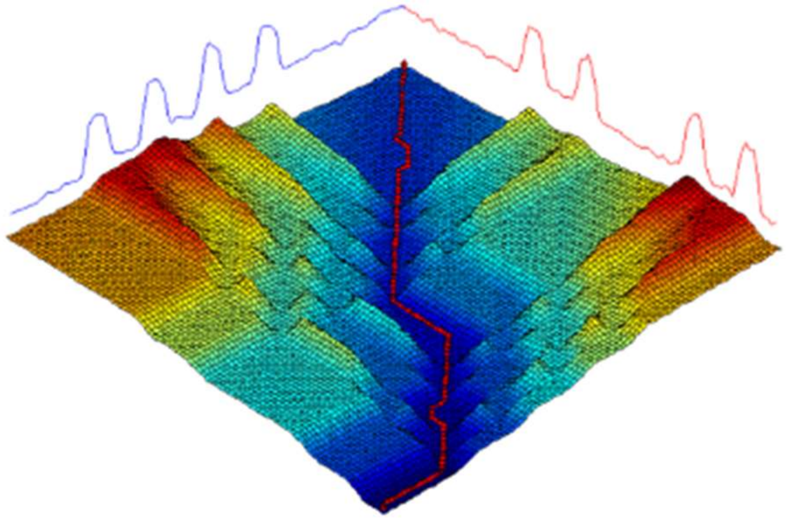


Distance Based Classifiers

Algorithm 1 DTW ($\mathbf{a}, \mathbf{b}$, (*both series of length* m), w (*window proportion, default value* $w \leftarrow 1$), M (*pointwise distance matrix*))

Let C be an $(m+1) \times (m+1)$ matrix initialised to zero, indexed from zero.
for $i \leftarrow 1$ to m **do**
  for $j \leftarrow 1$ to m **do**
    if $|i-j| < w \cdot m$ **then**
      $C_{i,j} \leftarrow M_{i,j} + \min(C_{i-1,j-1}, C_{i-1,j}, C_{i,j-1})$
return $C_{m,m}$

DTW has no explicit penalty for moving off the diagonal.

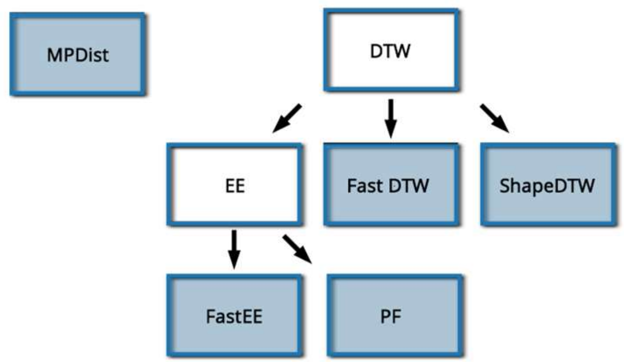


Algorithm 5 MSM($\mathbf{a}, \mathbf{b}$ (*both series of length* m), c (*minimum cost*), d, (*pointwise distance function*))

Let D be an $m \times m$ matrix initialised to zero.
$D_{1,1} = d(a_1, b_1)$
for $i \leftarrow 2$ to m **do**
  $D_{i,1} = D_{i-1,1} + C(a_i, a_{i-1}, b_1)$
for $i \leftarrow 2$ to m **do**
  $D_{1,i} = D_{1,i-1} + C(b_i, a_1, b+i-1)$
for $i \leftarrow 2$ to m **do**
  for $j \leftarrow 2$ to n **do**
    $match \leftarrow D_{i-1,j-1} + d(a_i, b_j)$
    $insert \leftarrow D_{i-1,j} + C(a_i, a_{i-1}, b_j)$
    $delete \leftarrow D_{i,j-1} + C(b_j, b_{j-1}, a_i)$
    $D_{i,j} \leftarrow \min(match, insert, delete)$
return $D_{m,m}$

$$C(x,y,z) = \begin{cases} c \textbf{ if } y \le x \le z \textbf{ or } y \ge x \ge z \\ c + min(|x-y|, |x-z|) \textbf{ otherwise}. \end{cases}$$

MSM uses a constant penalty if values are within a threshold, and a data dependent penalty otherwise



Distance Based	
DTW	Dynamic Time Warping [Ratanamahatana and Keogh 2005]
Fast DTW	Fast Dynamic Time Warping [Tan et al. 2018]
EE	Elastic Ensemble [Lines and Bagnall 2015]
FastEE	Fast Elastic Ensemble [Oastler and Lines 2019]
PF	Proximity Forest [Lucas et al. 2019]
Shape DTW	Shape based Dynamic Time Warping [Zhao and Itti 2018]
MPDist	Matrix Profile Distance [Gharghabi et al. 2020]

Elastic Ensemble (EE)

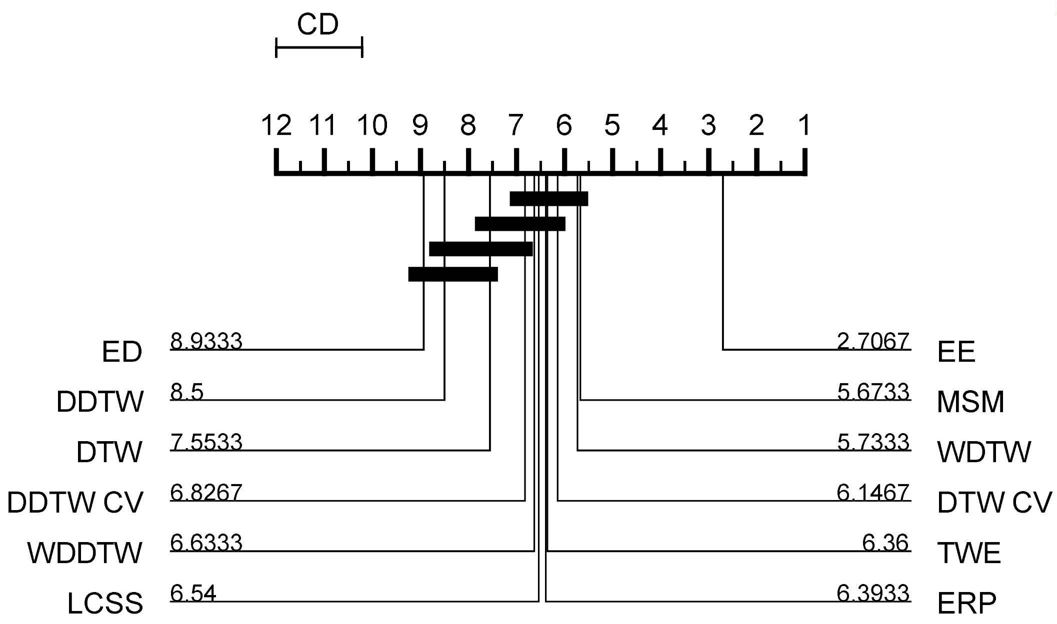
Published: 28 June 2014

Time series classification with ensembles of elastic distance measures

Jason Lines & Anthony Bagnall

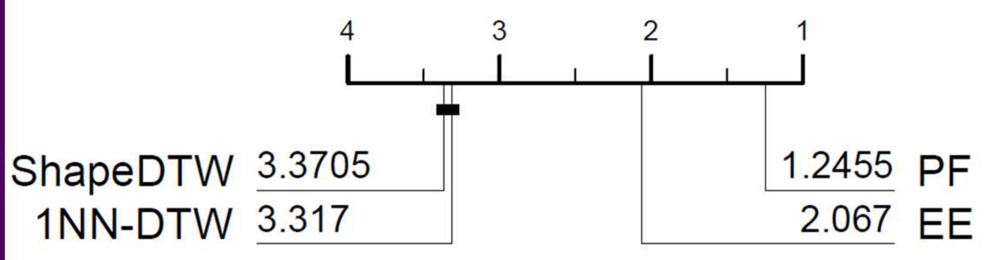
Data Mining and Knowledge Discovery **29**, 565–592 (2015) | Cite this article

Euclidean D
DTW (full)
DTW (cv wir
Derivati
Derivati
Weighte
Weighte
Longest
Edit Distanc
Move-Split-M
Time-warp Edit Distance



- Simple and transparent proportional voting scheme
- Constituents weighted by training cross-validation accuracy
- First reported TSC algorithm to **significantly outperform DTW** on the UCR datasets

Best in class: Proximity Forest (PF)



Published: 06 February 2019

Proximity Forest: an effective and scalable distance-based classifier for time series

Benjamin Lucas, Ahmed Shifaz, Charlotte Pelletier, Lachlan O'Neill, Nayyar Zaidi, Bart Goethals, François Petitjean & Geoffrey I. Webb

Data Mining and Knowledge Discovery **33**, 607–635 (2019) | Cite this article

MONASH University

BUT ….

Computer Science > Machine Learning

[Submitted on 12 Apr 2023 (v1), last revised 13 Apr 2023 (this version, v2)]

Proximity Forest 2.0: A new effective and scalable similarity-based classifier for time series

Matthieu Herrmann, Chang Wei Tan, Mahsa Salehi, Geoffrey I. Webb

We currently only have a Java implementation of PF, we would love to include it in aeon

Deep Learning Time Series Classifiers

Name	Year	Code
Disjoint-CNN	2021	y
Inception-FCN	2021	y
KDCTime	2022	n
Multi-Stage-Att	2020	n
CT_CAM	2020	n
CA-SFCN	2020	y
RTFN	2021	n
LAXCAT	2021	n
MACNN	2021	y
T2	2021	y
GTN	2021	y
TRANS	2021	n
FMLA	2022	n
AutoTransformer	2022	n
BENDER	2021	y
TST	2021	y
TARNET	2022	y

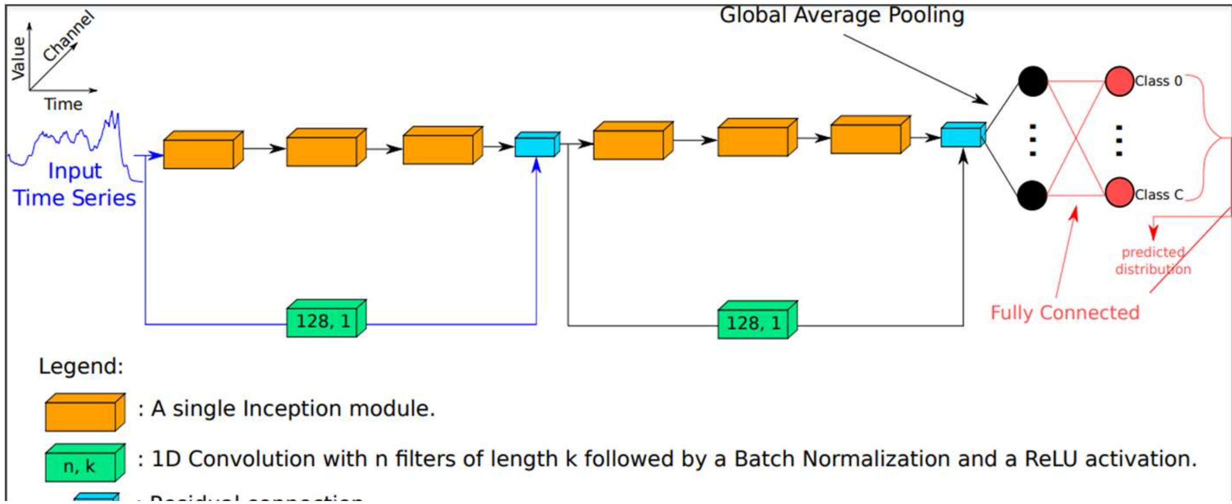
- There has been a huge international research effort to develop deep learners for TSC
- A recent survey references 246 papers, most of which have been published in the last three years.

[Submitted on 6 Feb 2023]

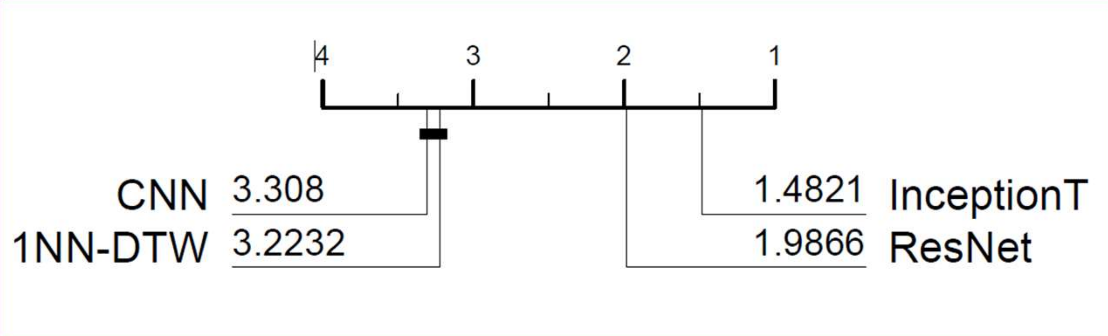
Deep Learning for Time Series Classification and Extrinsic Regression: A Current Survey

Navid Mohammadi Foumani, Lynn Miller, Chang Wei Tan, Geoffrey I. Webb, Germain Forestier, Mahsa Salehi

Generally poorly evaluated,not reproducible and self referential.



Best in class: InceptionTime



Published: 07 September 2020

InceptionTime: Finding AlexNet for time series classification

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller & François Petitjean

Data Mining and Knowledge Discovery **34**, 1936–1962 (2020) | Cite this article

InceptionTime is an ensemble of inception based deep learners

UNIVERSITÉ HAUTE-ALSACE

BUT ….

Deep Learning For Time Series Classification Using New Hand-Crafted Convolution Filters

Publisher: IEEE Cite This PDF

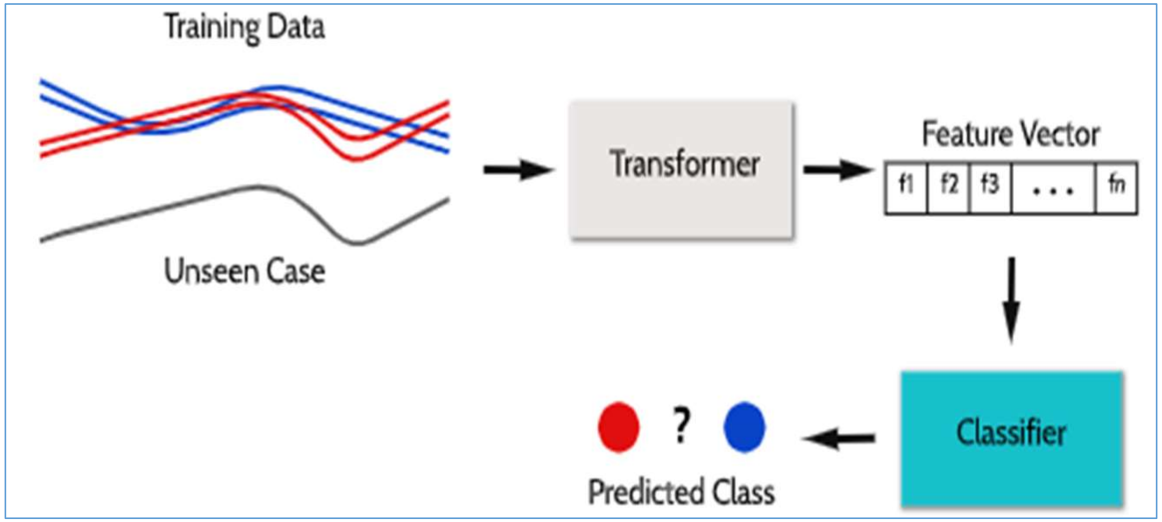
Published in Big Data 2022

Ali Ismail-Fawaz ; Maxime Devanne ; Jonathan Weber ; Germain Forestier **All Authors**

Feature Based Pipelines

The simplest transformation based approach is to create summary features then use a standard classifier

There are many toolkits that create summary features



tsfresh

MPF

MatrixProfile

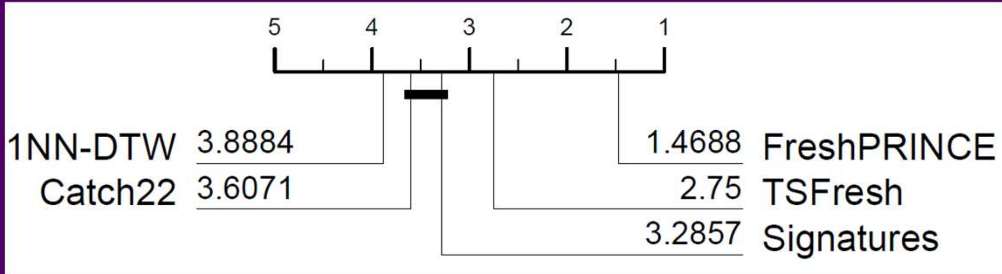
hctsa

THE UNIVERSITY OF SYDNEY

catch22

Explanatory Transformations	
hctsa	Highly Comparative Time-series Analysis [Fulcher and Jones 2017]
catch22	Canonical Time Series Characteristics [Lubba et al. 2019]
tsfresh	TS Feature Extraction based on Scalable Hypothesis Tests [Christ et al. 2018]
FreshPRINCE	Fresh Pipeline with Rotation Forest Classifier [Middlehurst and Bagnall 2022]
Signatures	Generalised Rough Path Signatures [Morrill et al. 2020]
MP	Matrix Profile Transform [Yeh et al. 2018]

Best in class: FreshPRINCE



The FreshPRINCE: A Simple Transformation Based Pipeline Time Series Classifier

Authors: Matthew Middlehurst, Anthony Bagnall Authors Info & Claims

Pattern Recognition and Artificial Intelligence: Third International Conference, ICPRAI 2022, Paris, France, June 1–3, 2022, Proceedings, Part II • Jun 2022 • Pages 150–161 • https://doi.org/10.1007/978-3-031-09282-4_13

The FreshPRINCE is a pipeline classifier combining TSFresh and the RotationForest classifier (FreshPRINCE).

BUT ….

orange™

arXiv > cs > arXiv:2308.01071

Computer Science > Machine Learning

[Submitted on 2 Aug 2023]

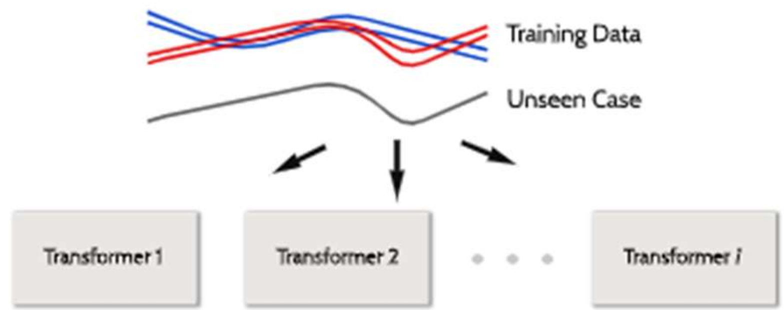
Automatic Feature Engineering for Time Series Classification: Evaluation and Discussion

Aurélien Renault, Alexis Bondu, Vincent Lemaire, Dominique Gay

Published in IJCNN in 2023

Interval Based Ensembles

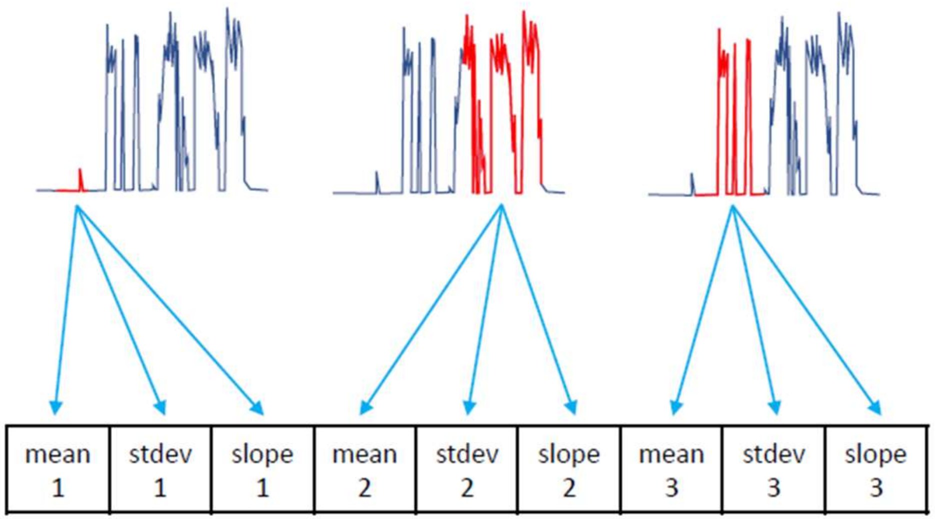
Ensembles of trees built on features extracted from randomised intervals

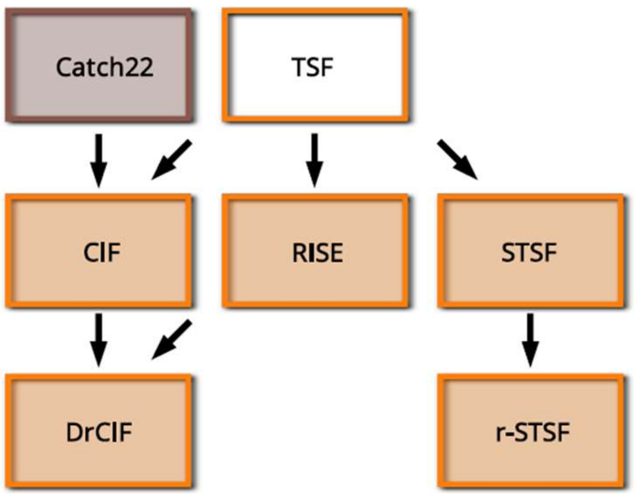


For each tree:

1. Randomly select intervals to apply to all series

2. Extract mean, standard deviation and slope from each interval



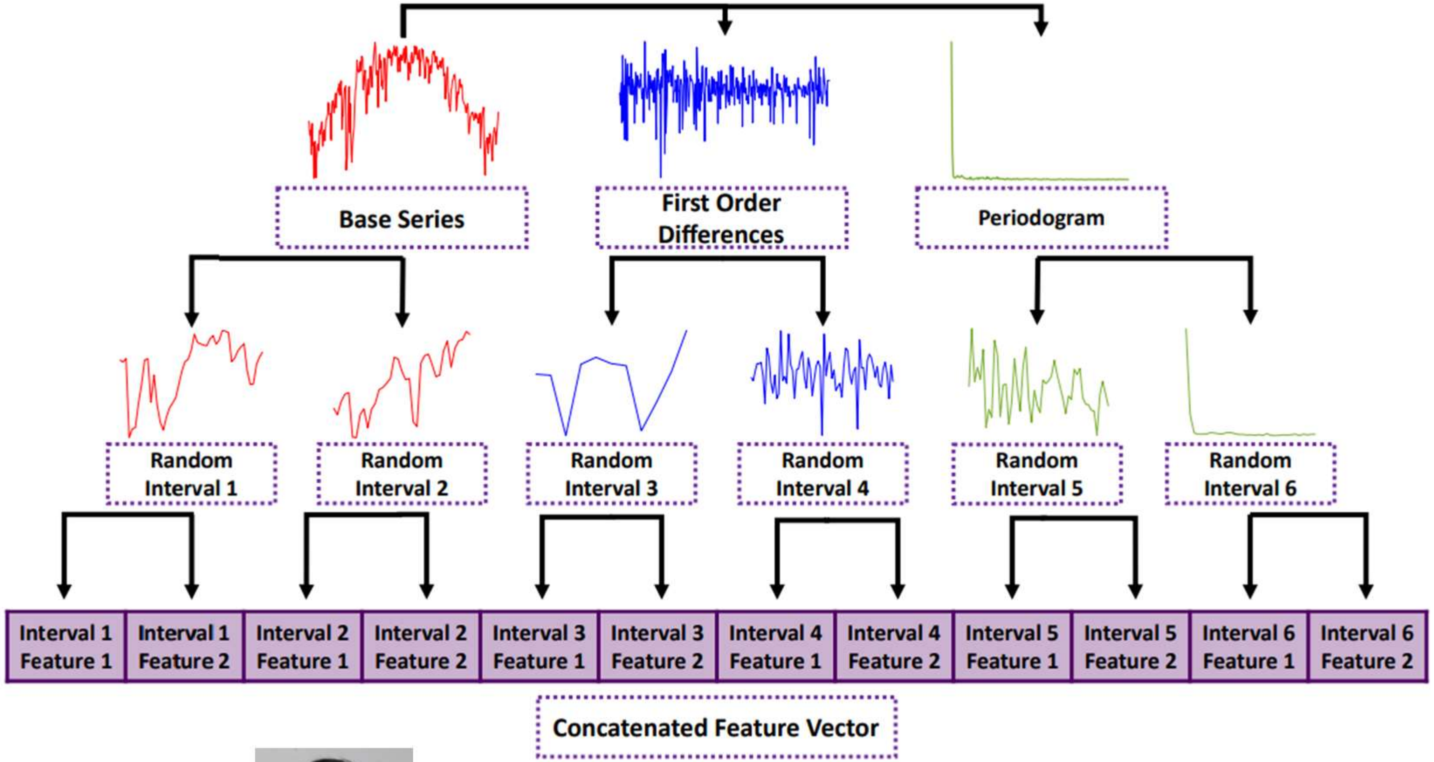


Time Series Forest (TSF) (2013)

Interval Based	
TSF	Time Series Forest [Deng et al. 2013]
RISE	Random Interval Spectral Ensemble [Flynn et al. 2019]
CIF	Canonical Interval Forest [Middlehurst et al. 2020a]
DrCIF	Diverse Representation Canonical Interval Forest [Middlehurst et al. 2021]
STSF	Supervised Time Series Forest [Cabello et al. 2020]
r-STSF	Randomised-Supervised Time Series Forest [Cabello et al. 2021]

Diverse Representation Canonical Interval Forest (DrCIF)

Diverse Representation: use raw data, the periodograms and first order differences
Canonical Interval Forest: derive random set of summary features (catch22) on each interval, concatenate into a new feature space for each tree



Conferences > 2020 IEEE International Confe...

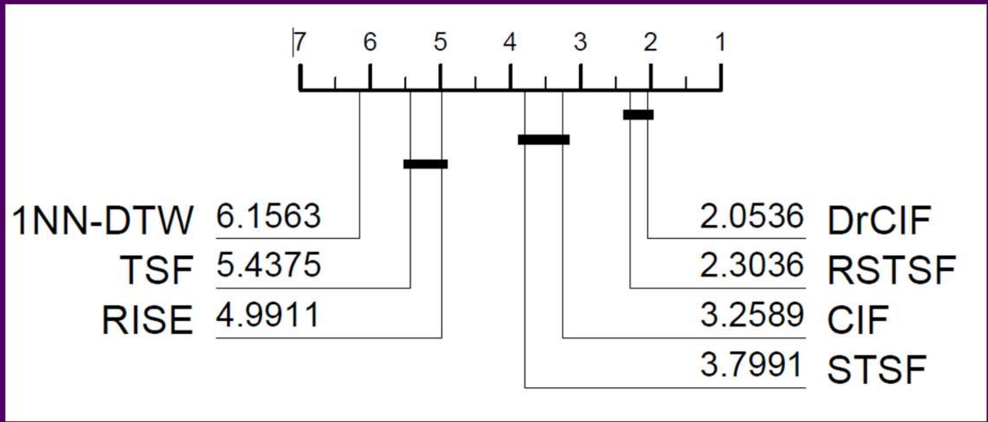
The Canonical Interval Forest (CIF) Classifier for Time Series Classification

Publisher: IEEE Cite This PDF

Matthew Middlehurst ; James Large ; Anthony Bagnall **All Authors**

University of East Anglia

Best in class: DrCIF/STSF

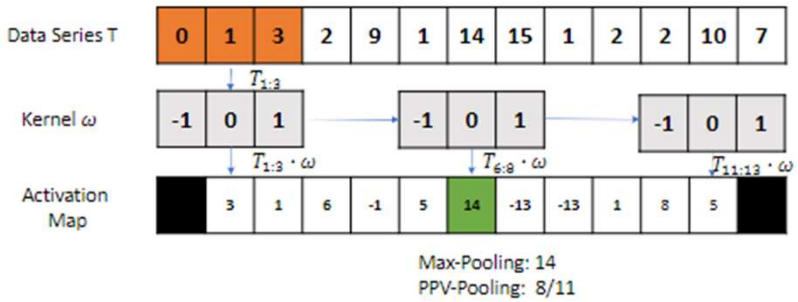


Randomised Supervised Time Series Forest (RSTF) is is an interval based tree ensemble that includes a supervised method for extracting intervals

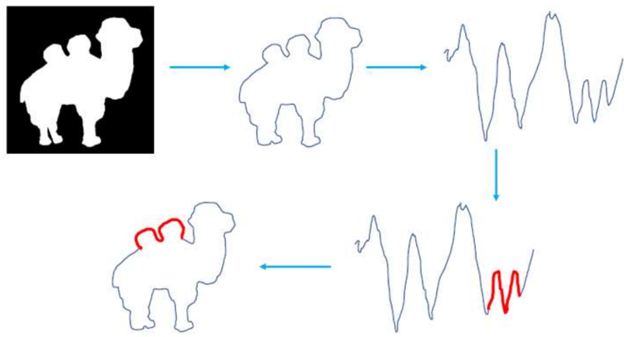
THE UNIVERSITY OF MELBOURNE

Taxonomy of Time Series Classification Algorithms Part II

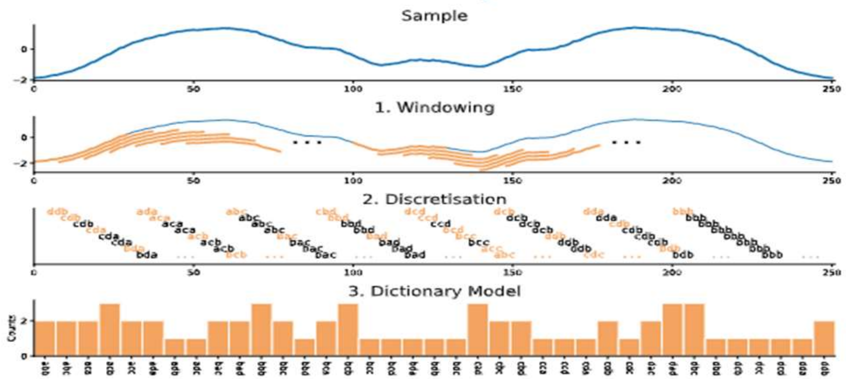
Convolution based



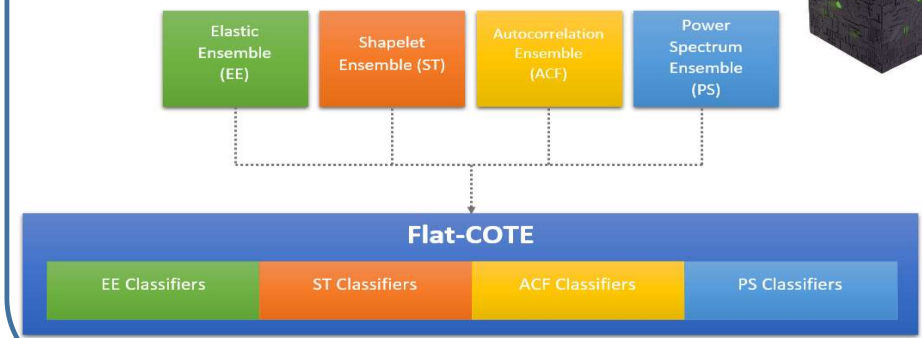
Shapelet based



Dictionary based



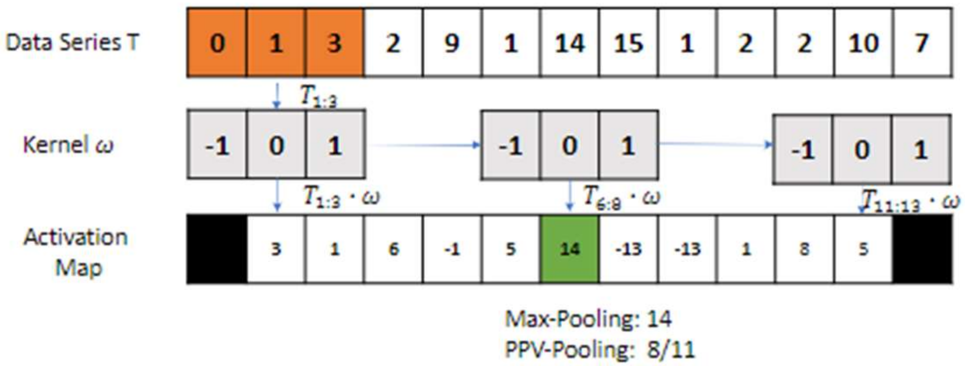
Hybrids



Convolution/kernel based:ROCKET

Geoff Webb's group in Monash proposed a simple approach to TSC that does surprisingly well

1. Create a large number of random convolutions
2. Create feature vectors by pooling operations
3. Fit a linear classifier



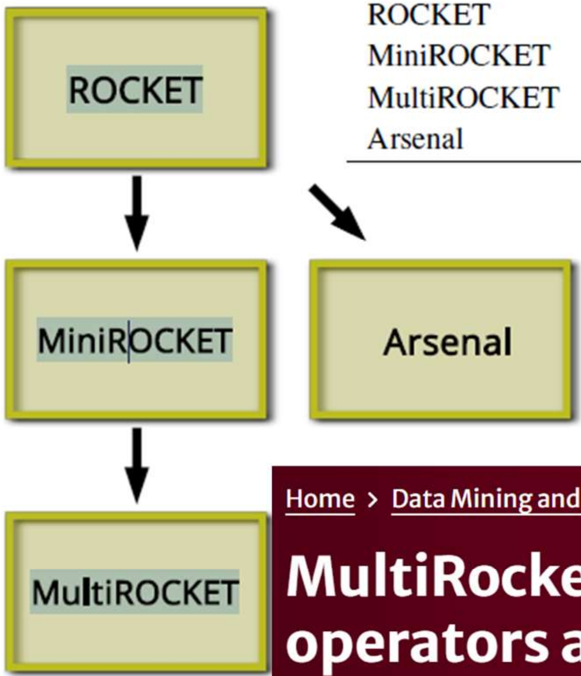
Published: 13 July 2020

ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels

Angus Dempster, François Petitjean & Geoffrey I. Webb

MONASH University

ROCKET Family: Convolution-Based



Kernel/Convolution Based

Hybrid STC	Hybrid Shapelet Transform Classifier [Guijo-Rubio et al. 2019]
ROCKET	Random Convolutional Kernel Transform [Dempster et al. 2020]
MiniROCKET	MINImally RandOm Convolutional KErnel Transform [Dempster et al. 2021]
MultiROCKET	MiniRocket with multiple pooling operators and transformations [Tan et al. 2022
Arsenal	Arsenal [Middlehurst et al. 2021]

Home > Data Mining and Knowledge Discovery > Article

MultiRocket: multiple pooling operators and transformations for fast and effective time series classification

Open access | Published: 29 June 2022 | **36**, 1623–1646 (2022)

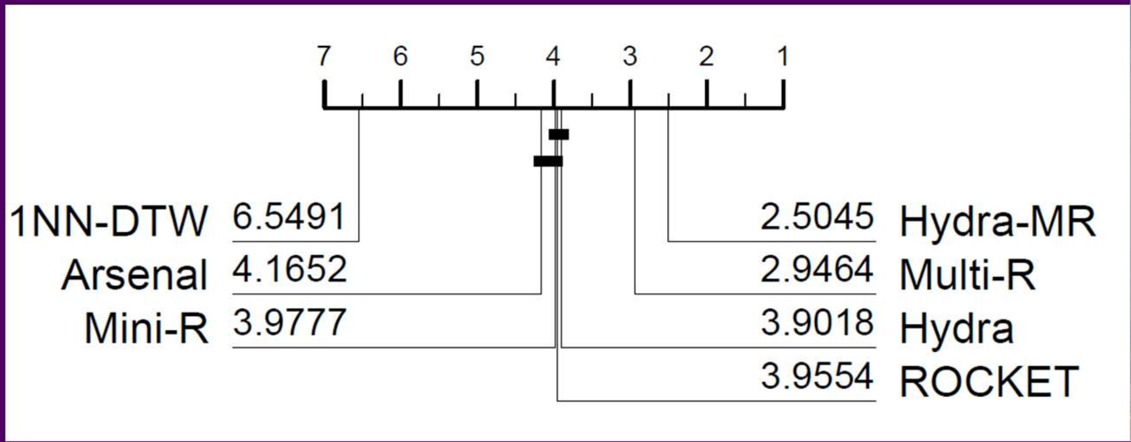
KDD > Proceedings > KDD '21 > MiniRocket: A Very Fast (Almost) Deterministic Transform

RESEARCH-ARTICLE

MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification

Authors: Angus Dempster, Daniel F. Schmidt, Geoffrey I. Webb Authors Inf

Best in class: Multi-Rocket-Hydra



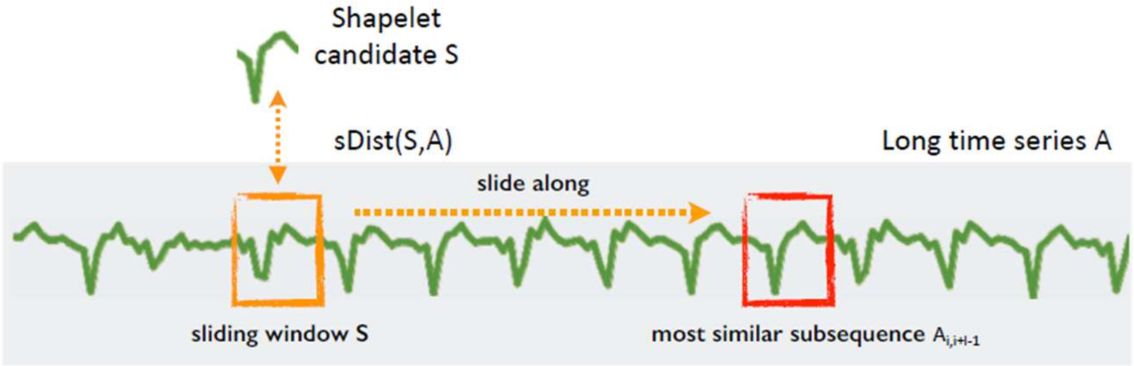
We got this one.

Home > Data Mining and Knowledge Discovery > Article

Hydra: competing convolutional kernels for fast and accurate time series classification

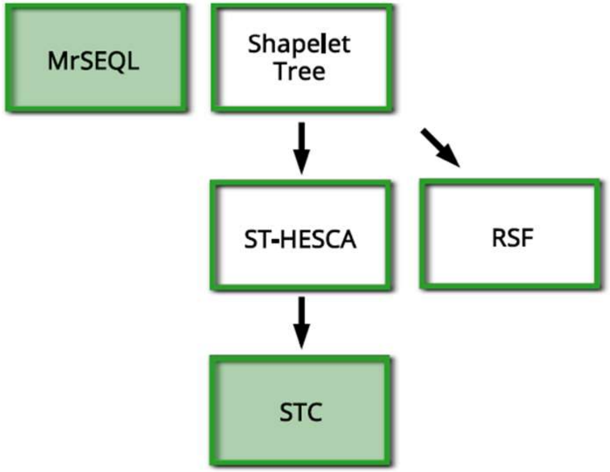
Open access | Published: 16 May 2023 | **37**, 1779–1805 (2023)

Shapelet Based Classifiers



Shapelets are discriminatory phase independent subseries taken from the train data

STC is a pipeline: select a set of shapelets, transform into distance to shapelet then build standard classifier



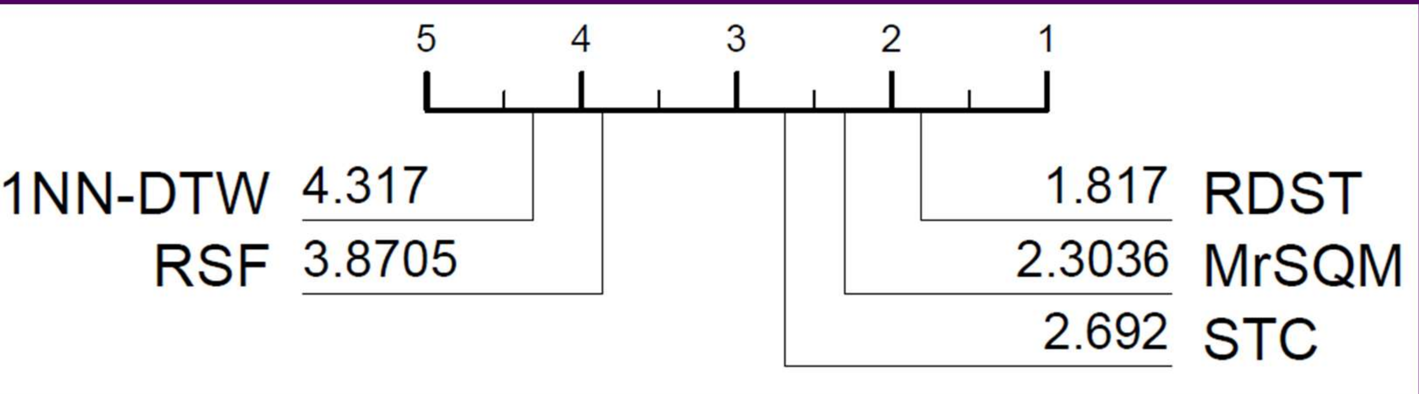
A shapelet transform for time series classification

Authors: Jason Lines, Luke M. Davis, Jon Hills, Anthony Bagnall Authors Info & Claims

KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining • August 2012 • Pages 289–297 • https://doi.org/10.1145/2339530.2339579

	Shapelet Based
Shapelet Tree	Shapelet Tree [Ye and Keogh 2011]
RSF	Generalised Random Shapelet Forest [Karlsson et al. 2016]
ST-HESCA	Shapelet Transform with HESCA Base Classifier [Bostrom and Bagnall 2017]
STC	Shapelet Transform Classifier [Bagnall et al. 2020]
MrSEQL	Multiple Representation Sequence Learner [Nguyen et al. 2017]

Best in class: RDST



Inspired by ROCKET, RDST employs dilation to improve performance.

Université d'ORLÉANS

Random Dilated Shapelet Transform: A New Approach for Time Series Shapelets

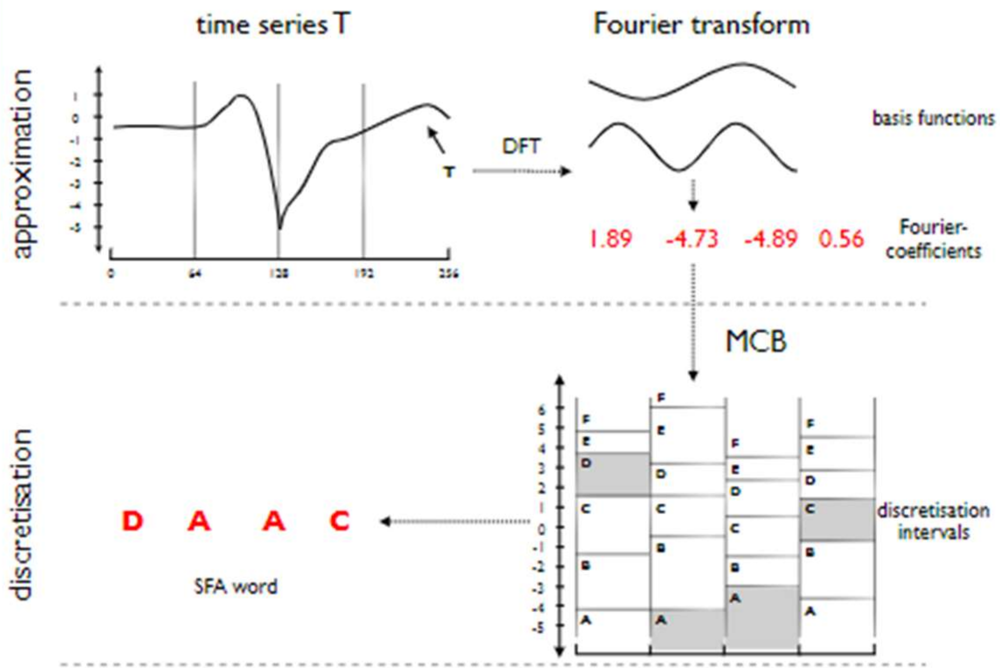
Authors: Antoine Guillaume, Christel Vrain, Wael Elloumi Authors Info & Claims

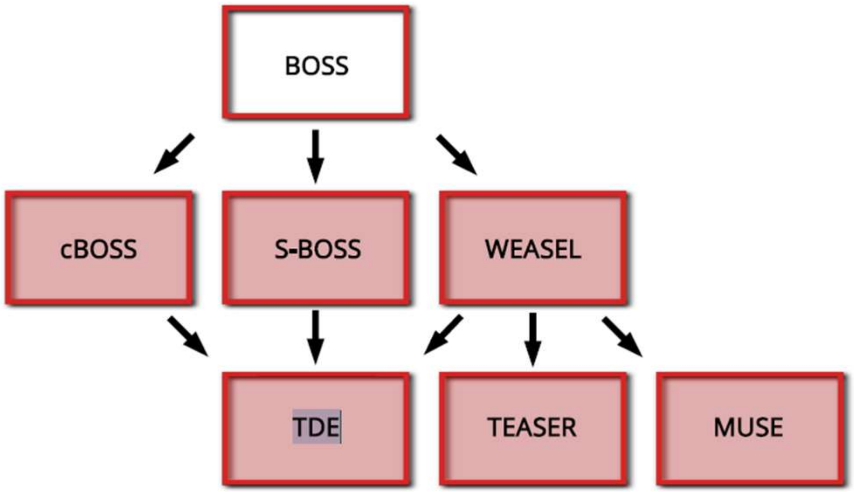
Pattern Recognition and Artificial Intelligence: Third International Conference, ICPRAI 2022, Paris, France, June 1–3, 2022, Proceedings, Part I • Jun 2022 • Pages 653–664 • https://doi.org/10.1007/978-3-031-09037-0_53

Published: 01 June 2022 Publication History

Dictionary Based Classifiers

Bag of words approaches create histograms of word counts.
The stages are
1) Windowing
2) Binning and
3) Discretisation





Classifiers built on histograms have historically been either nearest neighbour classifiers or linear classifiers

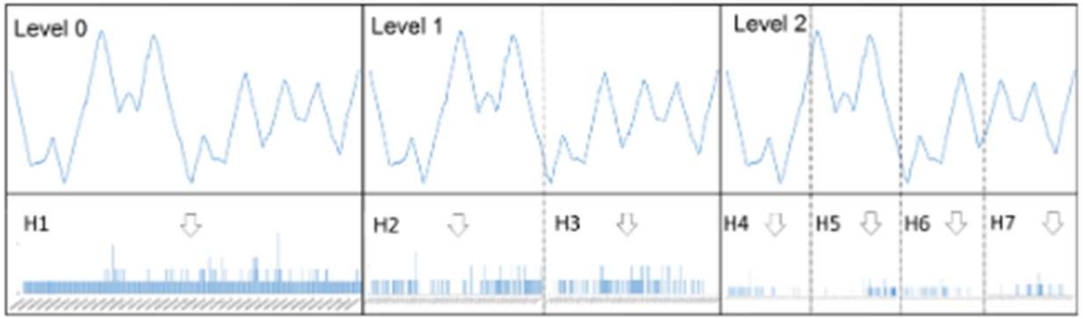
Dictionary Based	
BOSS	Bag of Symbolic Fourier Approximation Symbols [Schäfer 2015]
WEASEL	Word Extraction for Time Series Classification [Schäfer and Leser 2017a]
MUSE	Multivariate Symbolic Extension [Schäfer and Leser 2017b]
cBOSS	Contractable BOSS [Middlehurst et al. 2019]
S-BOSS	Spatial BOSS [Large et al. 2019a]
TDE	Temporal Dictionary Ensemble [Middlehurst et al. 2020b]
TEASER	Two-tier Early and Accurate Series Classifier [Schäfer and Leser 2020]

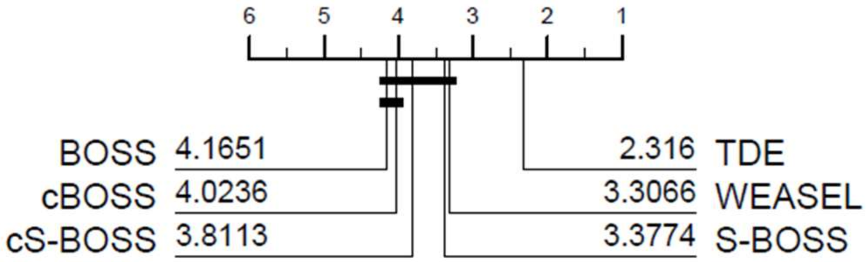
Temporal Dictionary Ensemble (TDE)

TDE is an ensemble on NN classifiers diversified by the bag of words parameters

It uses Spatial Pyramids to capture some location information

It employs an adaptive Gaussian process model to search the parameter space for ensemble members





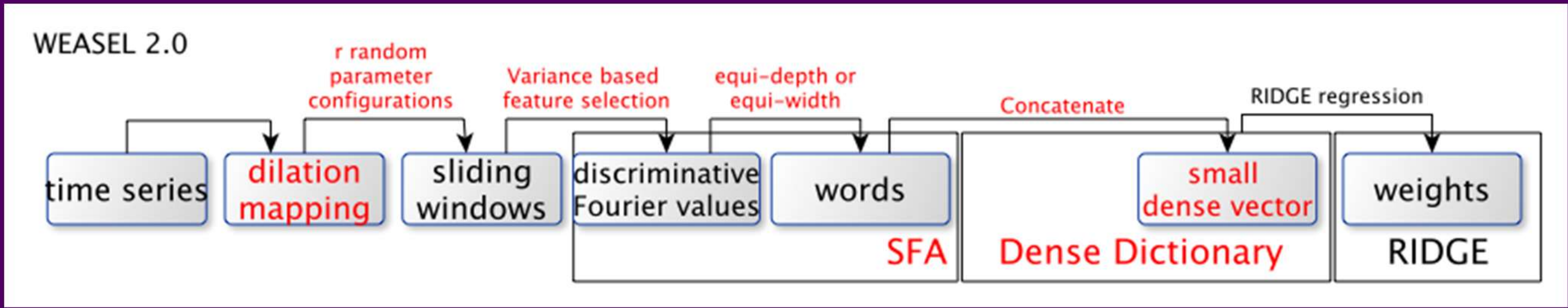
University of East Anglia

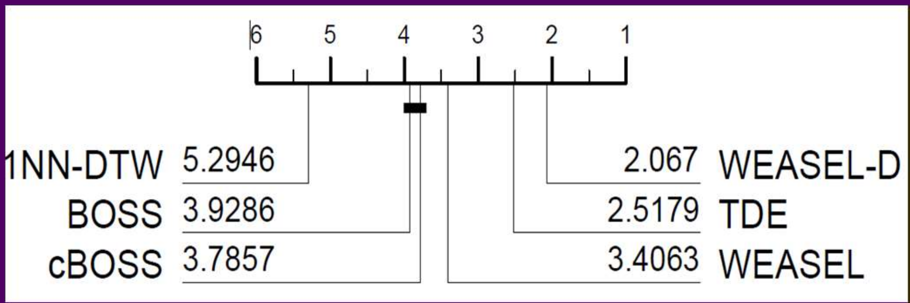
The Temporal Dictionary Ensemble (TDE) Classifier for Time Series Classification

Authors: Matthew Middlehurst, James Large, Gavin Cawley, Anthony Bagnall Authors Info & Claims

Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part I • Sep 2020 • Pages 660–676 • https://doi.org/10.1007/978-3-030-67658-2_38

Best in class: WEASEL 2





Home > Machine Learning > Article

WEASEL 2.0: a random dilated dictionary transform for fast, accurate and memory constrained time series classification

Open access Published: 19 September 2023 (2023)

Machine Learning

Dilated Sliding Window

Time Series T	0	1	3	2	9	1	14	15
Sliding window with dilation=2								
1st window	0		3		9			
2nd window		1		2		1		
3rd window			3		9		14	

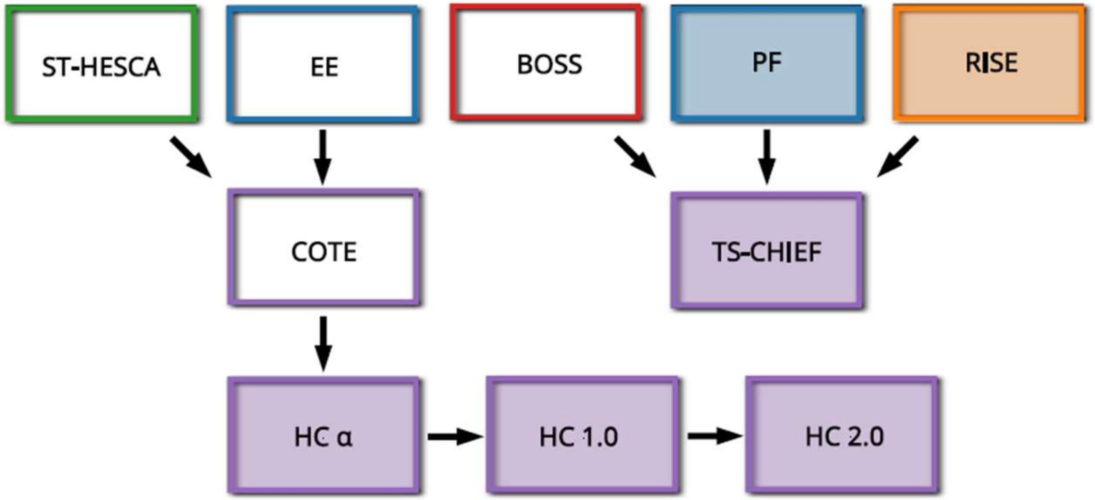
Shifted along T

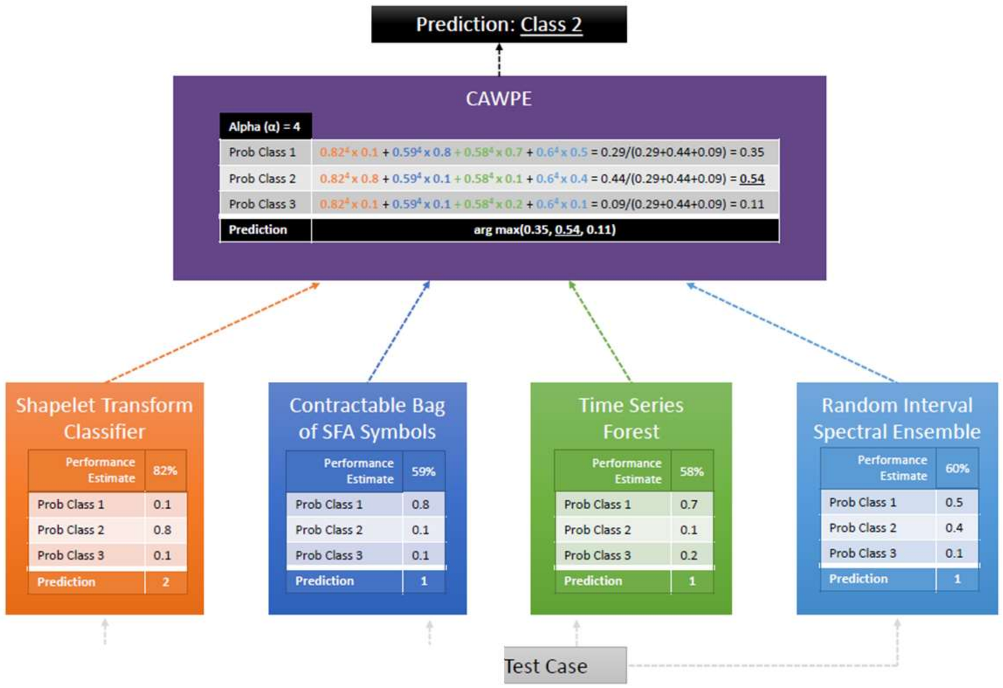
Example: d = 2 inserts gap of 1, down-samples by 2, doubles size of the receptive field

Improves:

- Rocket (convolution based)
- RDST (shapelet based)
- WEASEL (dictionary based)

Hybrids: Combine Approaches





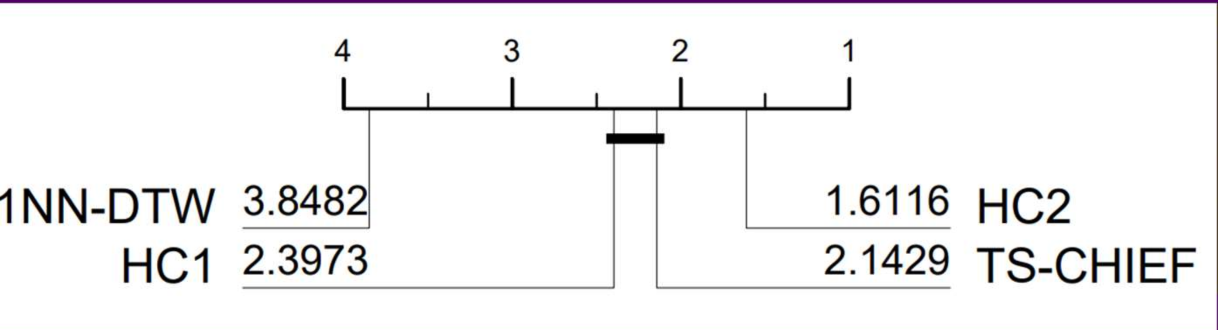
Journals & Magazines > IEEE Transactions on Knowledg... > Volume: 27 Issue: 9

Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles

Publisher: IEEE Cite This PDF

Anthony Bagnall ; Jason Lines ; Jon Hills ; Aaron Bostrom All Authors

Best in class: HC2



Hierarchical Vote Collective of Transformation based Ensembles

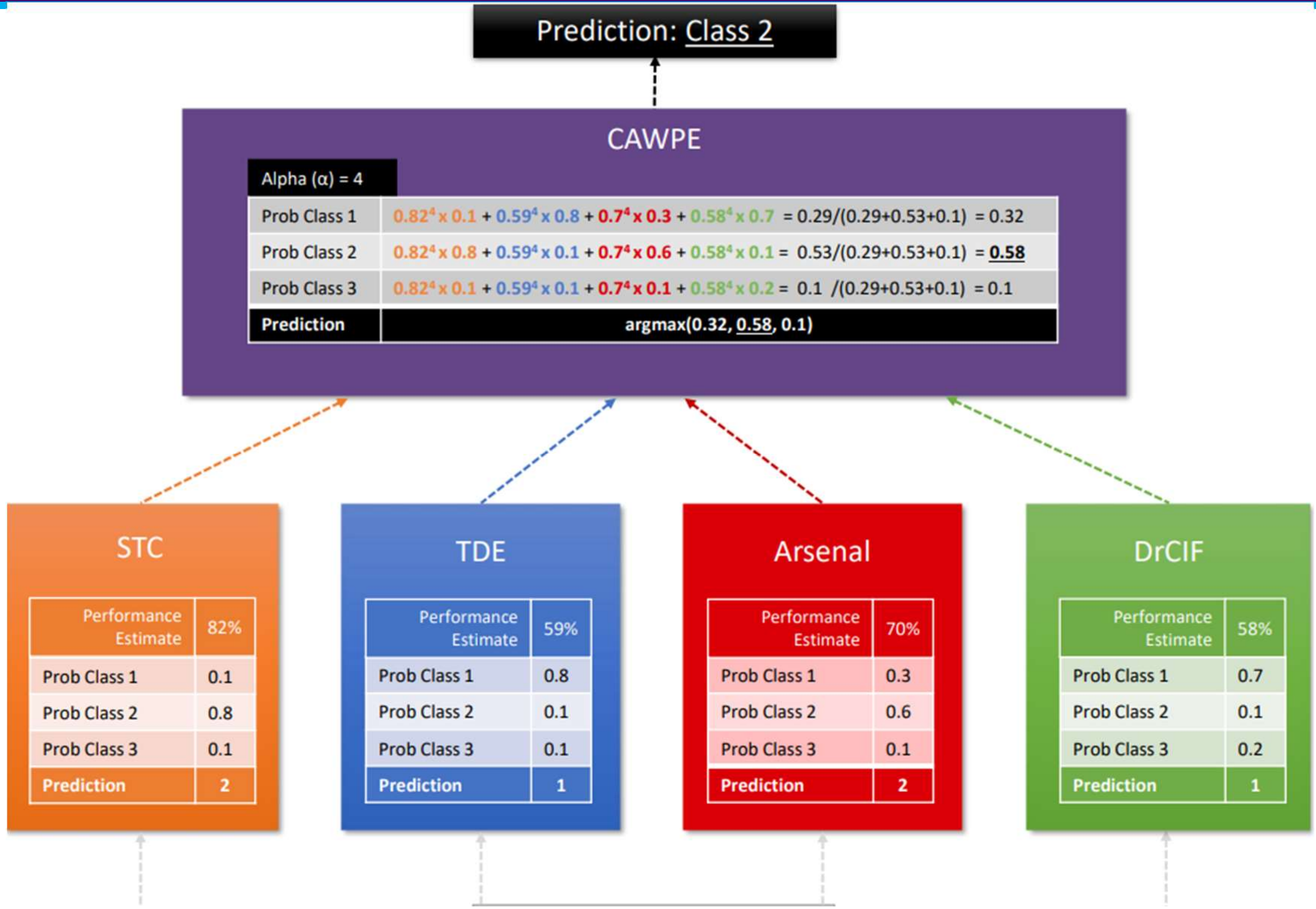
COTE - Flat COTE: ensemble of EE and STC

HIVE-COTE (alpha) - EE/STC/RISE/BOSS/TSF

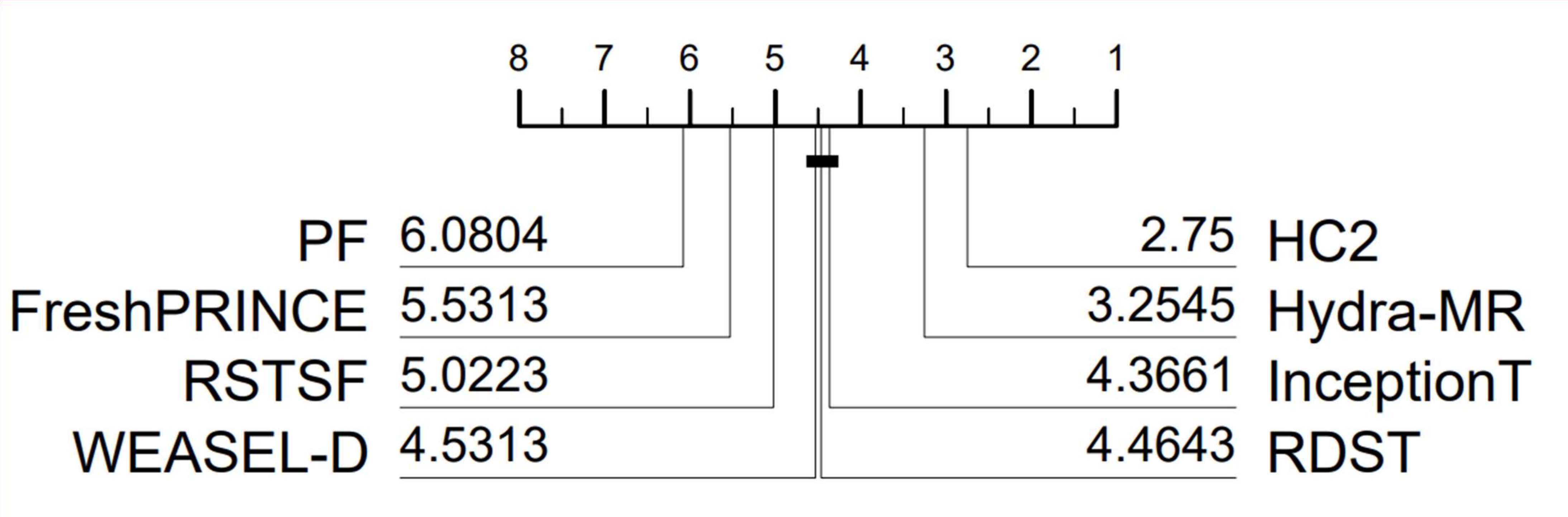
HIVE-COTE V1 - STC/RISE/CBOSS/TSF

HIVE-COTE V2 – STC/The Arsenal/TDE/DrCIF

Hierarchical Vote Collective of Transformation based Ensembles (HIVE-COTE V2)



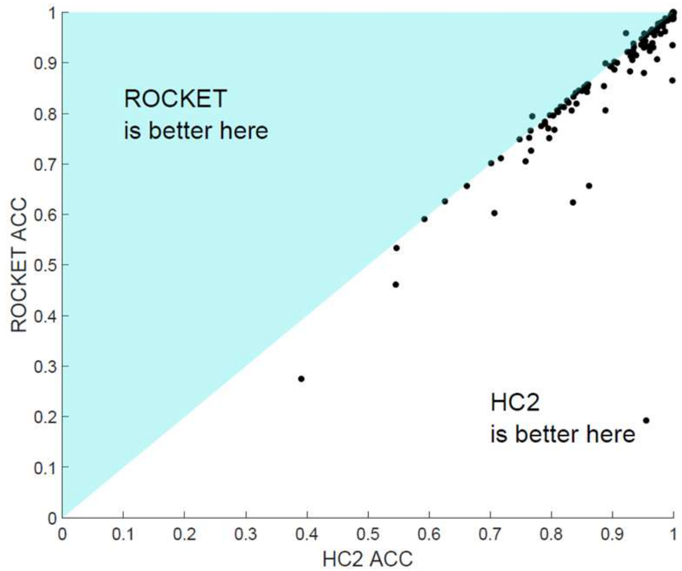
Bake off Redux: compare best in class



Is the Progress Real?

Performance on UCR datasets vs HC2

Algorithm	Mean Accuracy Diff	HC2 Wins	HC2 Loses
HC1	-1.06%	77	29
ROCKET	-2.49%	74	32
InceptionTime	-1.69%	82	25
TS-CHIEF	-1.36%	97	11

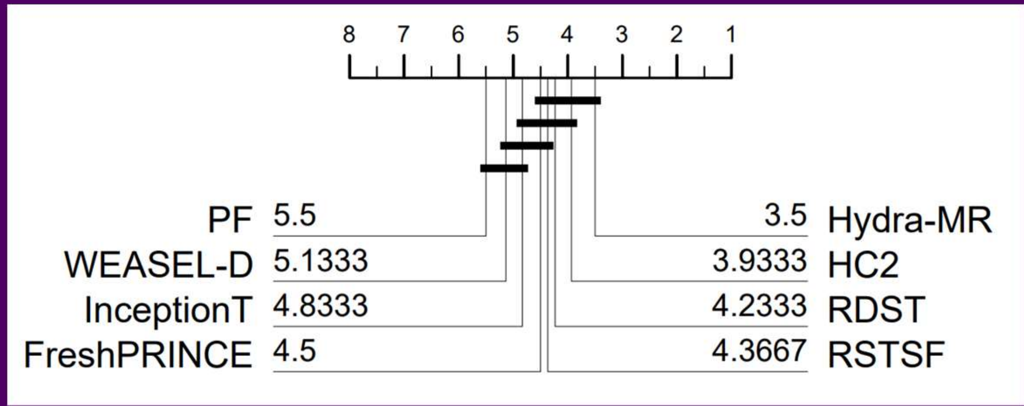


HC2 average accuracy is 89.11%

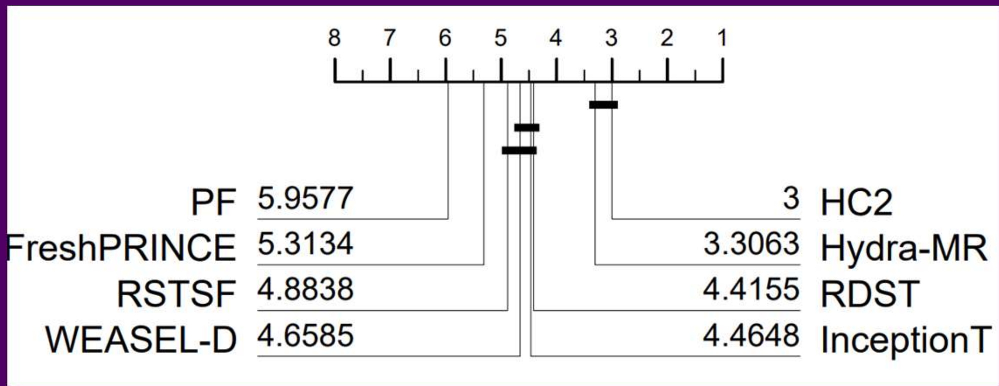
On average, over 112 UCR problems, HC2 is 12.37% more accurate than 1-NN DTW (wins on 107, ties on 2, loses on 3)

Are we all just overfitting the UCR archive?

30 new datasets



142 old+new datasets



Is the Progress Real? Case Study: Detecting Fraudulent Alcohol

Can we detect the methanol contents of spirits bottles non-invasively?

Algorithm	Accuracy and standard error (LOBO)
HC2	63.82% +/- 2.72%
ROCKET	52.72% +/- 2.69%
InceptionTime	44.93% +/- 2.67%
PLS	13.14% +/- 0.9%

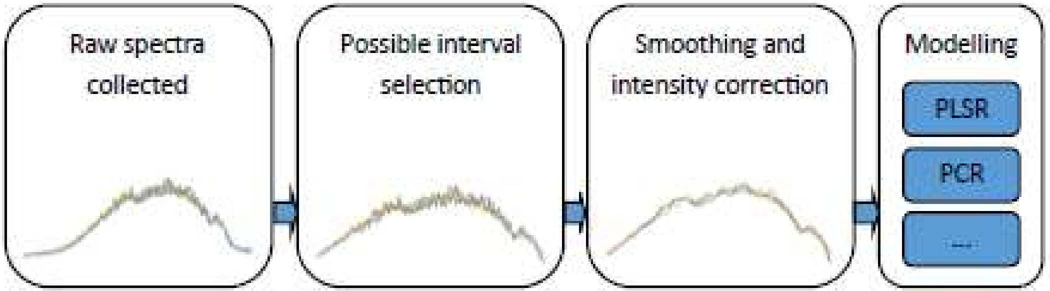
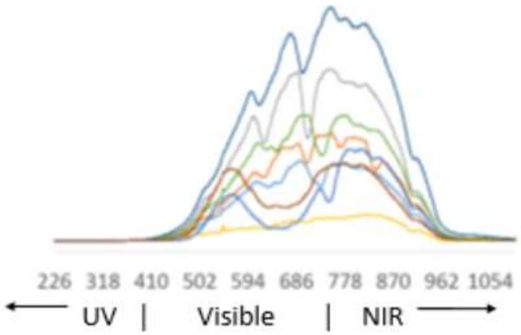


Fig. 2.3 An overview of a standard chemometric pipeline, applied to example spectra.



THE Scotch Whisky RESEARCH INSTITUTE

Time Series Clustering (TSCL)

Home > Knowledge and Information Systems > Article

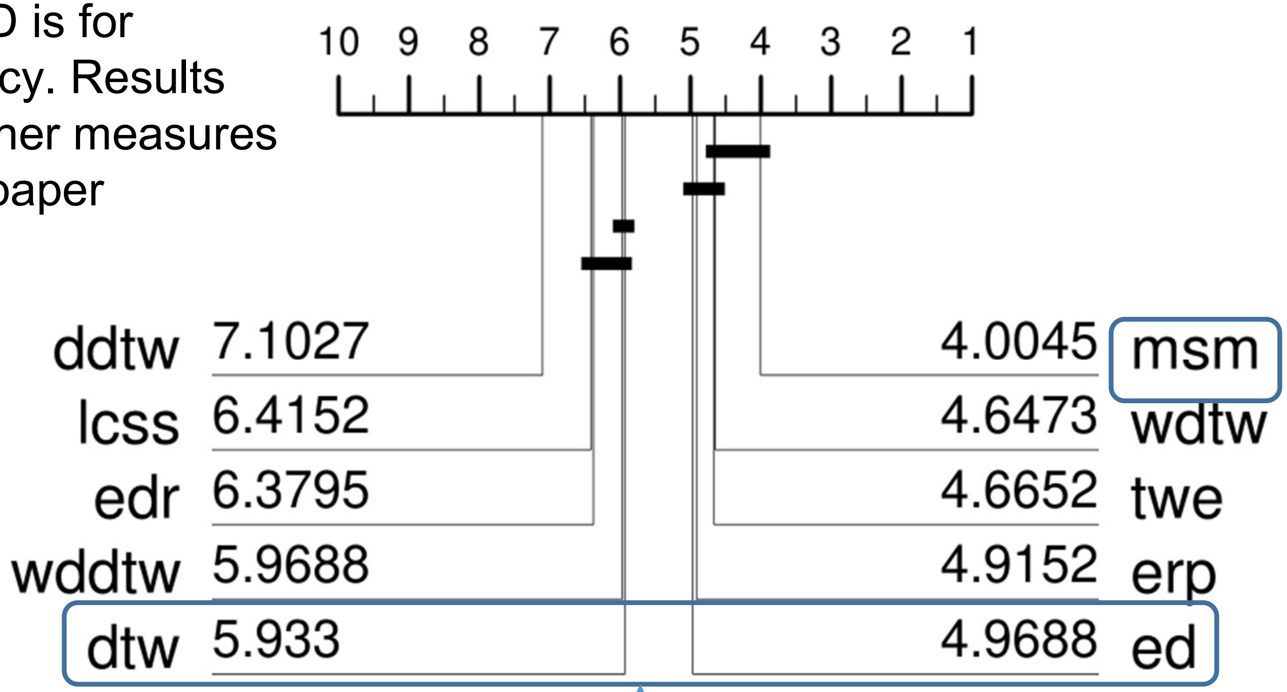
A review and evaluation of elastic distance functions for time series clustering

Review | Open access | Published: 07 September 2023 | (2023)

Compare the 11 elastic distance functions used in the EE classifier for means and median based clustering

Comparison of different distance functions for k-means TSCL

The CD is for accuracy. Results with other measures in the paper

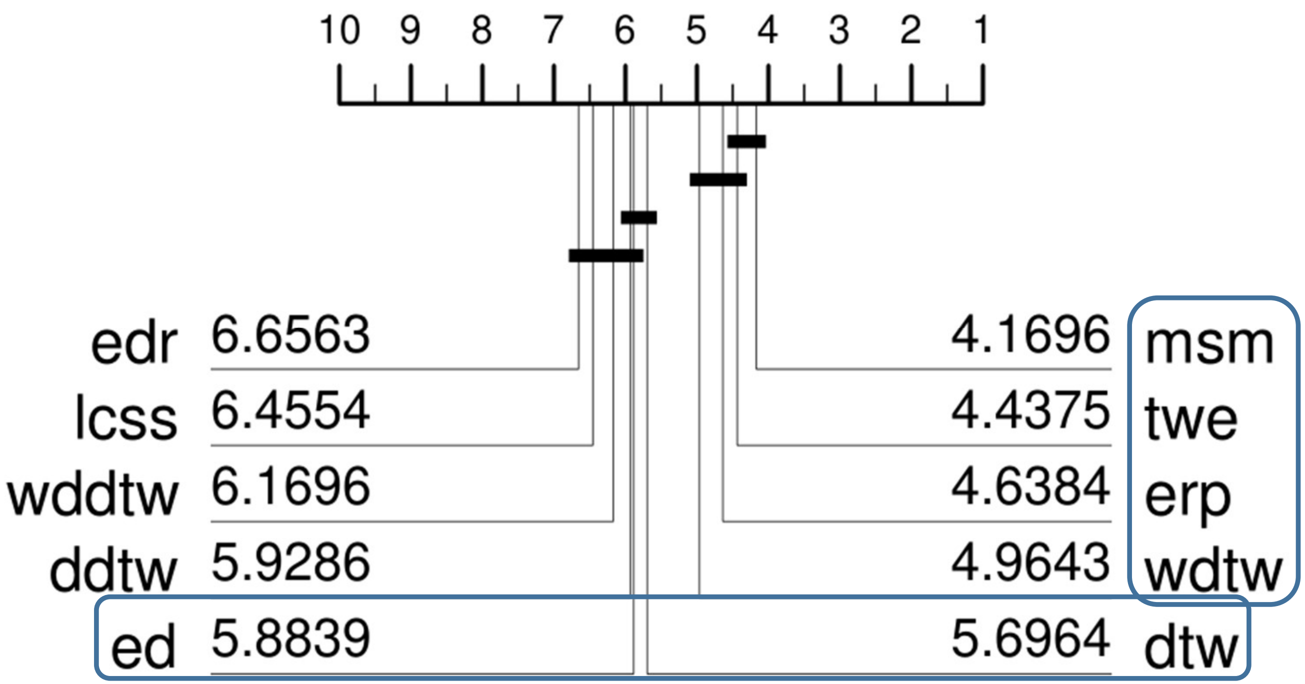


MSM only algorithm significantly better than Euclidean distance

DTW significantly worse than Euclidean distance!

Comparison of different distance functions K-medoids TSCL

K-means averages to find centroids. An alternative is to use medoids as centres (members of the cluster)



DTW not worse than Euclidean distance using medoids

Distances with explicit warp penalty all better than Euclidean distance

Barycentre Averaging (DBA) [4] for K-Means

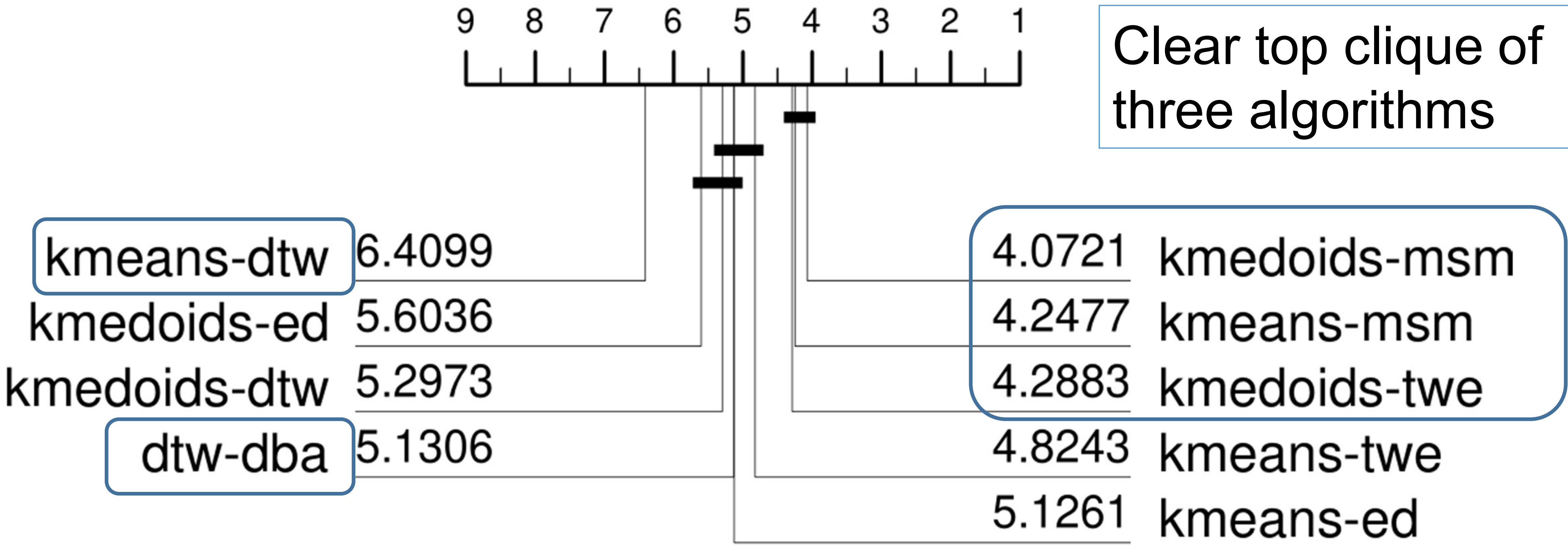
K-means averages to find centroids. An alternative is to warp cluster members onto each other.

Algorithm 7 DTW Barycentre Averaging($\mathbf{c}$, *the initial average sequence*, $\mathbf{X_p}$, *p time series to average.*

Let dtw_path be a function that returns the a list of tuples that contain the indexes of the warping path between two time series.
Let W be a list of empty lists, where W_i stores the values in $\mathbf{X_p}$ of points warped onto centre point c_i.
for $x \in \mathbf{X_p}$ **do**
    $P \leftarrow dtw\_path(\mathbf{x}, \mathbf{c})$
    for $(i, j) \in P$ **do**
        $W_i \leftarrow W_i U x_j$
for $i \leftarrow 1$ to m **do**
    $c_i \leftarrow mean(W_i)$
return c

Find centroids by averaging over realigned values

Comparison of DBA, k-means and k-medoids



Clear top clique of three algorithms

dba significantly improves DTW

Time Series Extrinsic Regression (TSER)

arXiv > cs > arXiv:2305.01429

Search...

Help | A

Computer Science > Machine Learning

[Submitted on 2 May 2023]

Unsupervised Feature Based Algorithms for Time Series Extrinsic Regression

David Guijo-Rubio, Matthew Middlehurst, Guilherme Arcencio, Diego Furtado Silva, Anthony Bagnall

UNIVERSIDAD Ð CÓRDOBA

University of Southampton

Universidade de São Paulo

Time Series Regression

- To most people, TSR means reducing forecasting to regression with a sliding window
- There is another sort of regression that aligns more with standard regression: use a time series to predict an external variable

Time series extrinsic regression

Predicting numeric values from time series data

Chang Wei Tan, Christoph Bergmeir, François Petitjean & Geoffrey I. Webb

Data Mining and Knowledge Discovery **35**, 1032–1060 (2021) | Cite this article

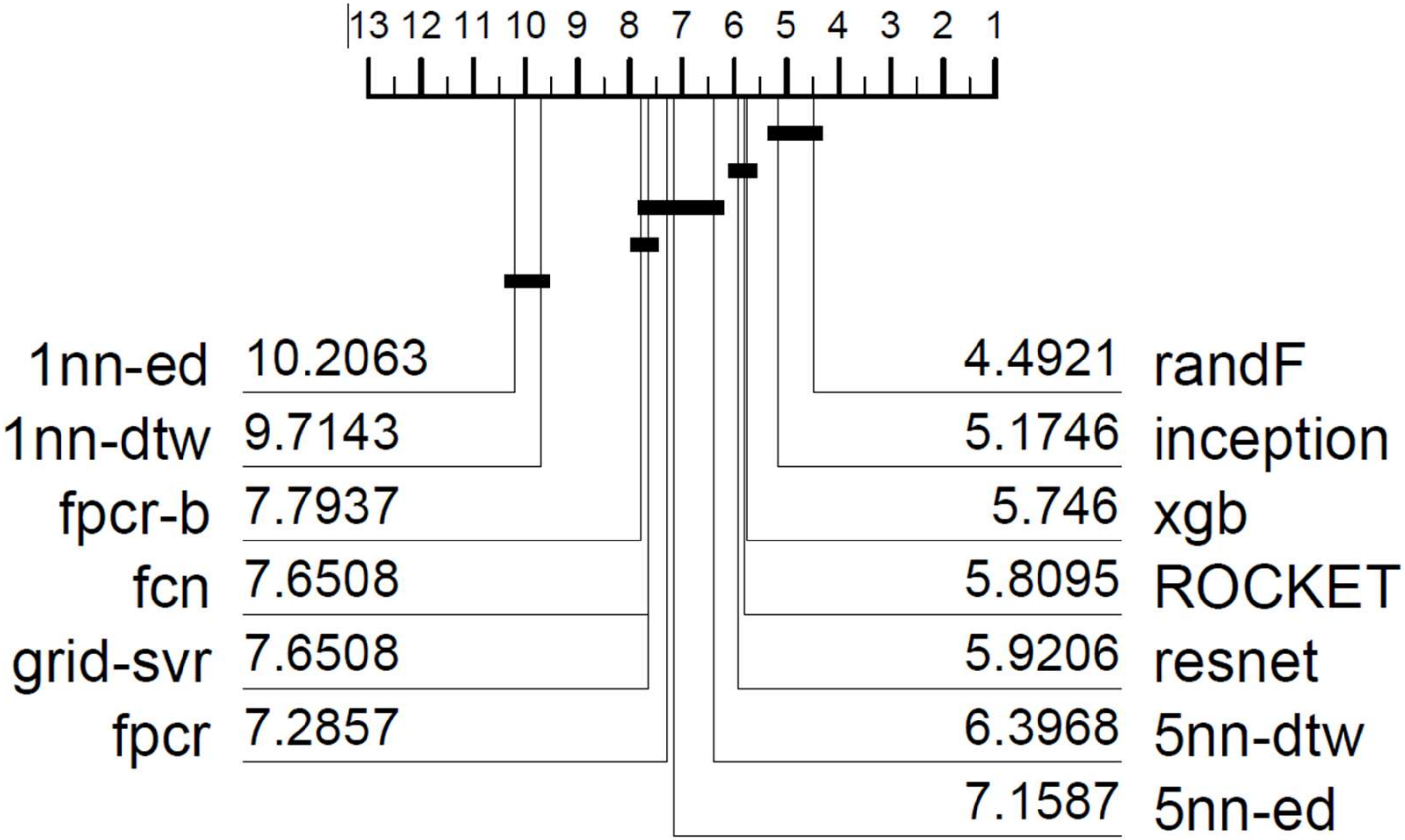
Time Series Regression Archive

We have expanded the archive from 19 datasets to 63

Name	Prediction problem (response variable)
Economic Analysis	
DailyOilGasPrices	Daily gas price with time series of oil prices.
Energy Monitoring	
Energy building predictors	Estimate the energy consumption of different sorts on buildings.
OccupancyDetectionLight	The average hourly occupancy of an office from sensor measurements [43].
SolarRadiationAndalusia	The average hourly solar radiation from atmospherical measurements.
TetuanEnergyConsumption	The daily average power consumption in three areas of Tetouan from atmospherical measurements [44].
WindTurbinePower	The daily power output of a wind turbine based on time series of torque measurements.
Environment monitoring	
AcousticContaminationMadrid	The 1st percentile of sound pressure levels from LAeq.
Africa Soil Chemistry	A set of 12 problems derived from the Africa Soil Information Service (AfSIS) Soil Chemistry.
BeijingIntAirportPM25Quality	The daily average of particulate matter in the Airport of Beijing from atmospherical data. [45].
DailyTemperatureLatitude	The latitude of a city based on the annual time series of daily temperature[10].
DhakaHourlyAirQuality	The Air Quality Index in Dhaka using localised particular matter time series[11].
MadridPM10Quality	The weekly average of particulate matter in the city of Madrid, Spain, from measurements of gases[12].
MetroInterstateTrafficVolume	The daily average traffic volume of a road in the USA from atmospherical variables[13].
ParkingBirmingham	The daily occupancy rate from the hourly total number of parked cars [46].
PrecipitationAndalusia	The yearly average of rainfall on Andalusia, Spain, from meterological measurements[14].
SierraNevadaMountainsSnow	The amount of snow based on temperature time series [47].
Equipment monitoring	
ElectricMotorTemperature	The temperature of an electric motor based on time series of torque readings[15].
LPGas/Methane MonitoringHomeActivity	The liquefied petroleum and methane concentration from gas sensors [48].
GasSensorArray Ethanol/Acetone	The concentrations of two analytes, acetone based on 16 metal-oxide sensors [49].
WaveTensionData	The tension of a string based on wave elevation time series[16].
Health Monitoring	
BarCrawl6min	The transdermal alcohol content by using an accelerometer [50].
Covid19Andalusia	The rate of deceased/contagions people from number of contagions in Andalusia, Spain [51].
VentilatorPressure	The pressure of the inspiratory solenoid valve from control input and output of the same valve[17].
Sentiment Analysis	
Crypto Sentiment	The sentiment of four cryptocurrencies based on the same days hourly price[18].
NaturalGasPriceSentiment	Sentiment scores about natural gas prices [52] based on the daily natural gas prices.

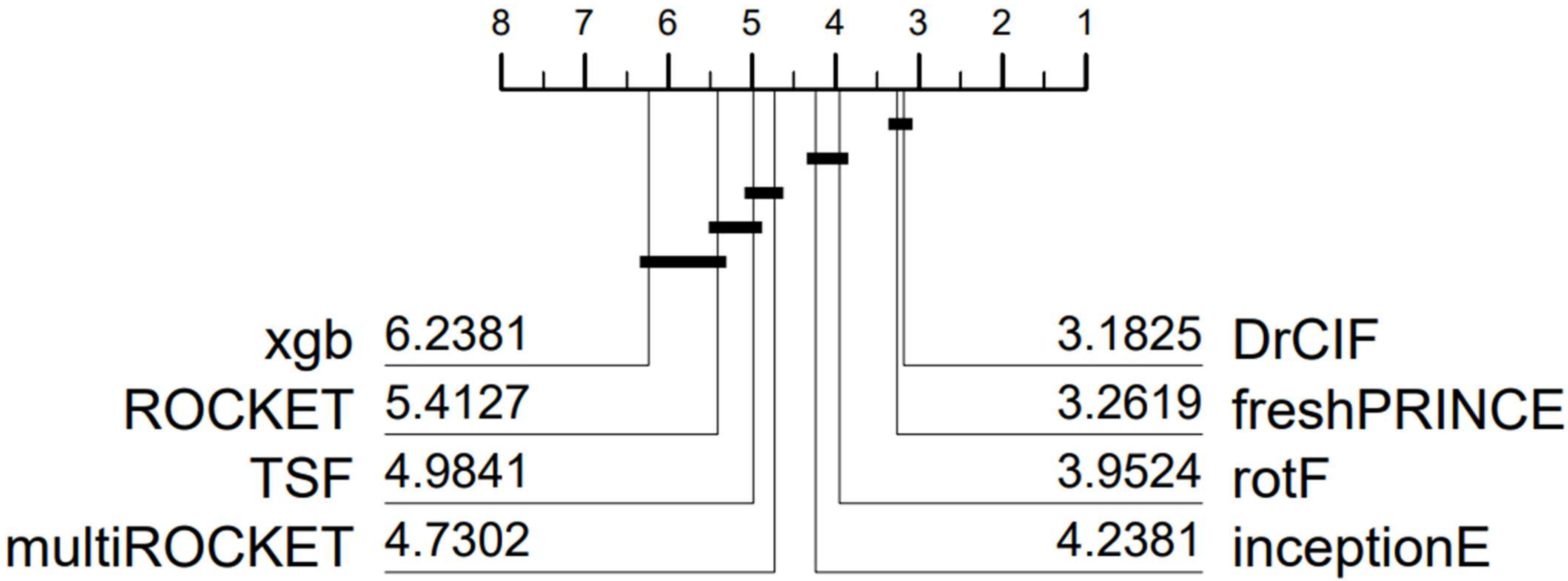
TSER Bakeoff

Current algorithms are not any better than standard regressors



New TSER Algorithms

Adapted DrCIF and FreshPRINCE are significantly better than all other approaches used



Open Source Software

https://twitter.com/aeon_toolkit https://github.com/aeon-toolkit/aeon

Scikit learn compatible machine learning toolkit with (nearly) all the functionality described in this talk available

Future Directions

Classification: scalability, explainability, more use cases (unequal length, streaming etc), more applications (EEG, vitals, industry)

Clustering: Bake off, cluster ensembles

Regression: HC2 for regression, forecasting regression

Similarity search, anomaly detection, segmentation …

Thank you for listening, any questions?

Evaluating a clustering

- Cluster quality is subjective and problem specific
- We use a range of measures to assess quality
- Accuracy
- Mutual Information
- Rand index
- … many others
- Some people evaluate on the train data, some on test data
- The results I present here are all accuracy on test data
- We have also reported results for other measures and on train data

Dynamic time warping (DTW/WDTW) distances

Algorithm 1 DTW $(\mathbf{a}, \mathbf{b},$ (*both series of length* m), w (*window proportion, default value* $w \leftarrow 1$), M (*pointwise distance matrix*))

Let C be an $(m+1) \times (m+1)$ matrix initialised to zero, indexed from zero.
for $i \leftarrow 1$ to m **do**
  for $j \leftarrow 1$ to m **do**
    if $|i - j| < w \cdot m$ **then**
      $C_{i,j} \leftarrow M_{i,j} + \min(C_{i-1,j-1}, C_{i-1,j}, C_{i,j-1})$
return $C_{m,m}$

DTW has no explicit penalty for moving off the diagonal. The warping window controls

Weighted DTW [1] (wdtw) has a weight penalty for moving off the diagonal

$$w(a) = \frac{w_{max}}{1 + e^{-g \cdot (a - m/2)}}$$

$$M^w_{i,j} = w(|i - j|) \cdot (a_i - b_j)^2$$

edit distance on real sequences (EDR) [2]

Algorithm 3 EDR ($\mathbf{a}, \mathbf{b}$, (*both series of length* m), ϵ (*equality threshold*)

Let E be an $(m+1) \times (m+1)$ matrix initialised to zero, indexed from zero.
for $i \leftarrow 1$ to m **do**
    for $j \leftarrow 1$ to m **do**
        if $i = 0 \vee j = 0$ **then**
            $E_{i,j} \leftarrow m$
        else
            if $|a_i - b_j| < \epsilon$ **then**
                $c \leftarrow 0$
            else
                $c \leftarrow 1$
            $match \leftarrow E_{i-1,j-1} + c$
            $insert \leftarrow E_{i-1,j} + 1$
            $delete \leftarrow E_{i,j-1} + 1$
            $E_{i,j} \leftarrow \min(match, insert, delete)$
return $E_{m,m}$

EDR is an adaptation of longest common subsequence that applies a constant penalty for mismatches

Edit distance with real penalty (ERP) [3]

Algorithm 4 ERP ($\mathbf{a}, \mathbf{b}$ (*both series of length m*), g, (*penalty value*), d, (*point-wise distance function*))

Let E be an $(m+1) \times (m+1)$ matrix initialised to zero, indexed from zero.
for $i \leftarrow 1$ to m **do**
    for $j \leftarrow 1$ to m **do**
        if i = 0 **then**
            $E_{i,j} \leftarrow \sum_{k=1}^{m} d(b_k, g)$
        else if j = 0 **then**
            $E_{i,j} \leftarrow \sum_{k=1}^{m} d(a_k, g)$
        else
            $match \leftarrow E_{i-1,j-1} + d(a_i, b_j)$
            $insert \leftarrow E_{i-1,j} + d(a_i, g)$
            $delete \leftarrow E_{i,j-1} + d(g, b_j)$
            $E_{i,j} \leftarrow \min(match, insert, delete)$
return $E_{m,m}$

ERP imposes a penalty for moving off diagonal (insert and delete) based on distance to a constant parameter g

Move split merge (MSM) [4]

Algorithm 5 MSM($\mathbf{a}, \mathbf{b}$ (*both series of length m*), c (*minimum cost*), d, (*point-wise distance function*))

Let D be an $m \times m$ matrix initialised to zero.
$D_{1,1} = d(a_1, b_1)$
for $i \leftarrow 2$ to m **do**
  $D_{i,1} = D_{i-1,1} + C(a_i, a_{i-1}, b_1)$
for $i \leftarrow 2$ to m **do**
  $D_{1,i} = D_{1,i-1} + C(b_i, a_1, b + i - 1)$
for $i \leftarrow 2$ to m **do**
  for $j \leftarrow 2$ to n **do**
    $match \leftarrow D_{i-1,j-1} + d(a_i, b_j)$
    $insert \leftarrow D_{i-1,j} + C(a_i, a_{i-1}, b_j)$
    $delete \leftarrow D_{i,j-1} + C(b_j, b_{j-1}, a_i)$
    $D_{i,j} \leftarrow \min(match, insert, delete)$
return $D_{m,m}$

$$C(x, y, z) = \begin{cases} c \textbf{ if } y \le x \le z \textbf{ or } y \ge x \ge z \\ c + min(|x - y|, |x - z|) \textbf{ otherwise.} \end{cases}$$

MSM uses a constant penalty if values are within a threshold, and a data dependent penalty otherwise

Time Warp Edit (TWE) [5]

Algorithm 6 TWE($\mathbf{a}, \mathbf{b}$ (*both series of length m*), λ (*edit cost*), ν (*warping penalty factor*), d, (*pointwise distance function*))

```
Let D be an m + 1 × n + 1 matrix initialised to 0
D_{0,0} = 0
for i ← 1 to m do
    D_{i,0} = ∞
    D_{0,i} = ∞
for i ← 1 to m do
    for j ← 1 to n do
        match = D(i − 1, j − 1) + d(a_i, b_j) + d(a_{i−1}, b_{j−1}) + 2ν(|i − j|)
        delete = D(i − 1, j) + d(a_i, a_{i−1}) + λ + ν
        insert = D(i, j − 1) + d(b_j, b_{j−1}) + λ + ν
        D(i, j) = min(match, insert, delete)
return D(m, n)
```

TWE uses a stiffness penalty (*nu*) for warping and an edit penalty for insert and delete (*lambda*)

Other distances in the comparison

- Euclidian distance (ED)
- Longest common subseq (LCSS)
- Derivate DTW (ddtw) [6]
- Derivate weighted DTW (dwdtw)

[1] Y. Jeong, M. Jeong, and O. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern Recognition*, 44:2231–2240, 2011.

[2] Lei Chen, M. Tamer Ozsu, and Vincent Oria. Robust and fast similarity search for moving object trajectories. In *proceedings of the ACM SIGMOD International Conference on Management of Data*, 2005.

[3] L. Chen and R. Ng. On the marriage of Lp-norms and edit distance. In *proceedings of the 30th International Conference on Very Large Data Bases*, 2004.

[4] A. Stefan, V. Athitsos, and G. Das. The Move-Split-Merge metric for time series. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1425–1438, 2013.

[5] P. Marteau. Time warp edit distance with stiffness adjustment for time series matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):306–318, 2009.

[6] E. Keogh and M. Pazzani. Derivative dynamic time warping. In *proceedings of the 1st SIAM International Conference on Data Mining*, 2001.