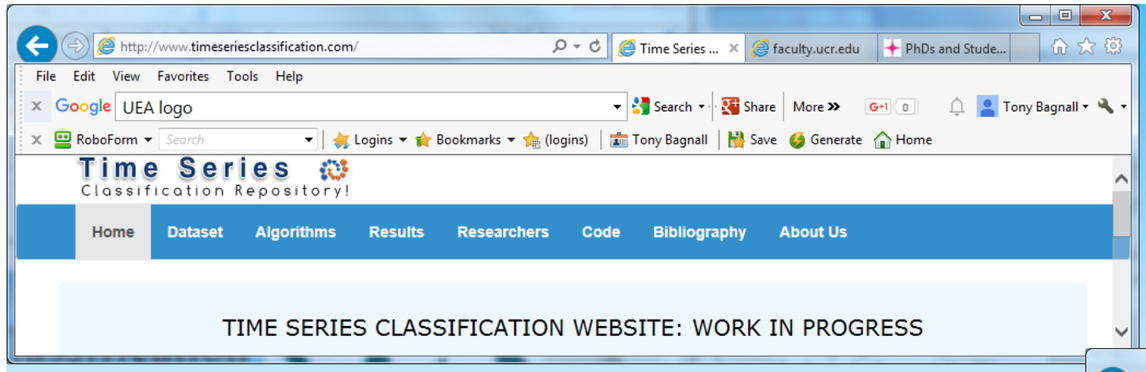
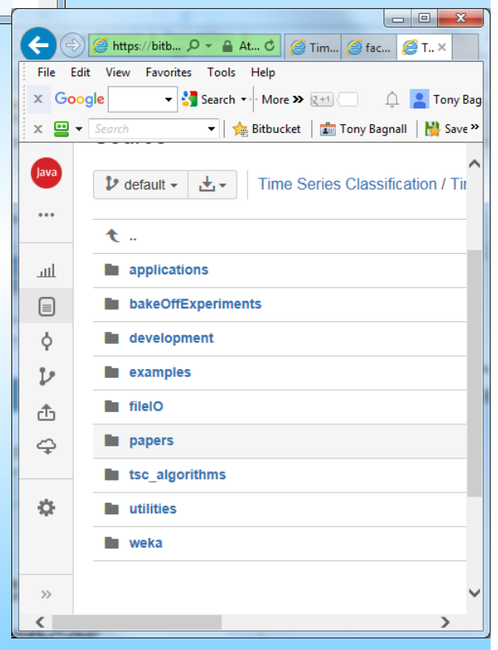
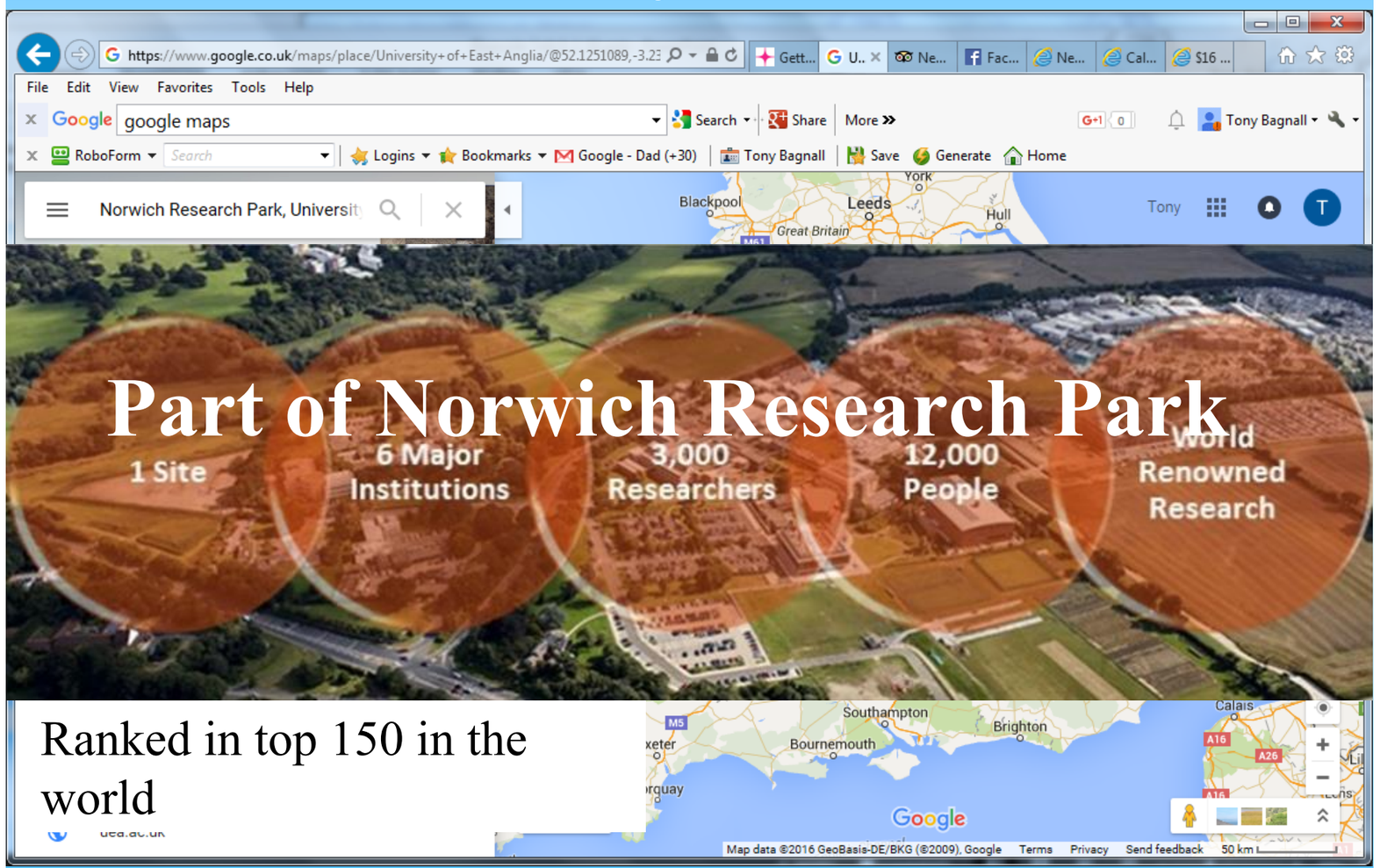




Tony Bagnall
School of Computing Sciences
University of East Anglia
Norwich
England

Some slides are recycled from talks given by Eamonn Keogh

University of where?



UEA Research in Cmp Sci

The school is organised into three laboratories:
- Computational Biology
- Graphics, Vision and Speech
- Machine Learning and Statistics

Within these labs research is carried out on computational biology, speech processing, computer graphics, computer vision, colour processing, data mining, optimisation, machine learning and statistics.

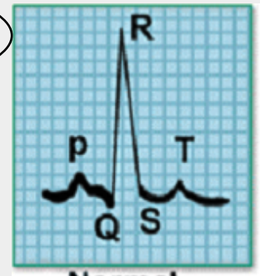
I am in the machine learning and statistics group. My main area of research interest is time series data mining, with particular focus on time series classification

Time Series Data Mining

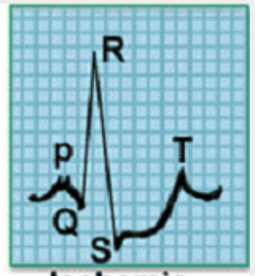
Tony's main interest

Clustering

Classification

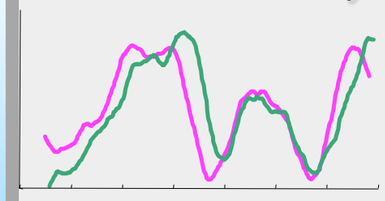


Normal



Ischemia

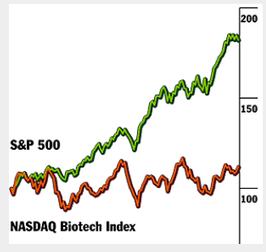
Motif Discovery



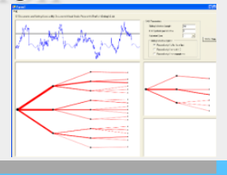
We generally leave time series forecasting to the statisticians

$c = 0.3$

ery by ntent



Visualization



Novelty Detection

Tony's Research Goals

1. Find the best algorithms for classifying time series without the application of domain knowledge
2. Develop tools so we can reproduce results and evaluate new algorithms on test problems
3. Make these tools accessible so we can implement solutions for new problems very quickly
3. Develop algorithms that can mitigate the need for domain knowledge
4. Find as many new applications in as wide a range of problems as possible

Time Series Classification

- Time series classification is the problem of building a classifier from a collection of labelled training time series, where each series has m ordered, real-valued observations, and a class label c_i

$$T_i = <t_{i1}, t_{i2}, \ldots, t_{im}, c_i>$$

- The problem is to find a function from the space of possible time series to the space of possible class labels.
  - Informally: given T_j, can we predict c_j?

TSC Variants

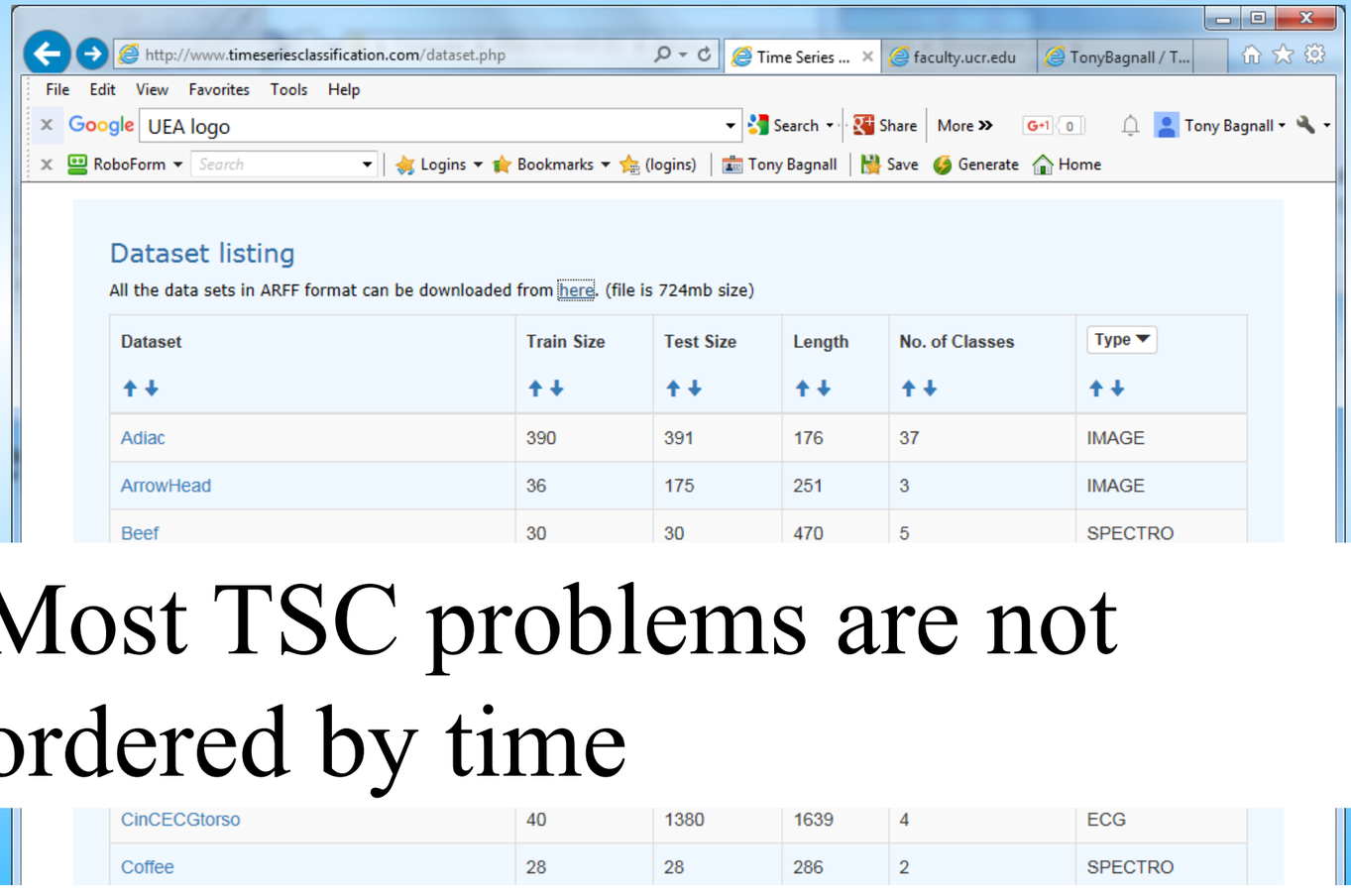
- Very long series
- Streaming data
- Different length series
- Multivariate series
- Ordinal series
- Semi-supervised learning etc

I don't do any of that yet

Assume independent series the same length. Univariate, real valued and tractable.

UCR/UEA TSC Repository

Currently contains 85 datasets from a wide range of problems



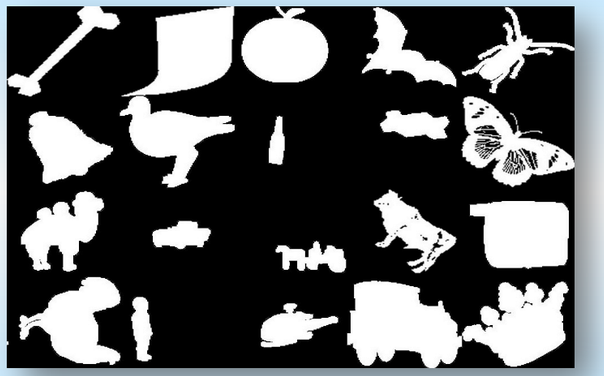
Most TSC problems are not ordered by time

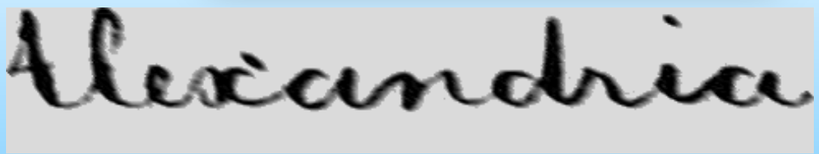
TSC UCR/UEA Repository

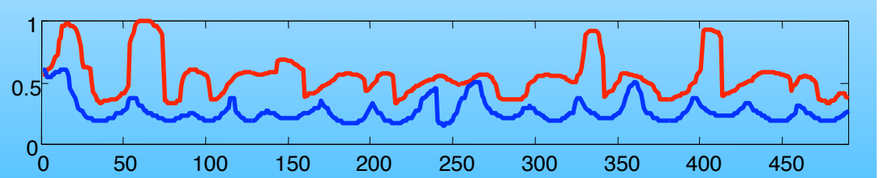
		Train Size		Length		Nos Classes
	Count	Min	Max	Min	Max	Max
IMAGE	29	16	1800	80	2709	60
SENSOR	18	20	3636	24	1639	39
MOTION	14	36	896	150	1882	12
SPECTRA	7	28	613	234	570	5
DEVICE	6	250	8926	96	720	7
ECG	6	23	1800	82	750	42
SIMU	5	20	1000	60	1024	8

Image Outlines (29)

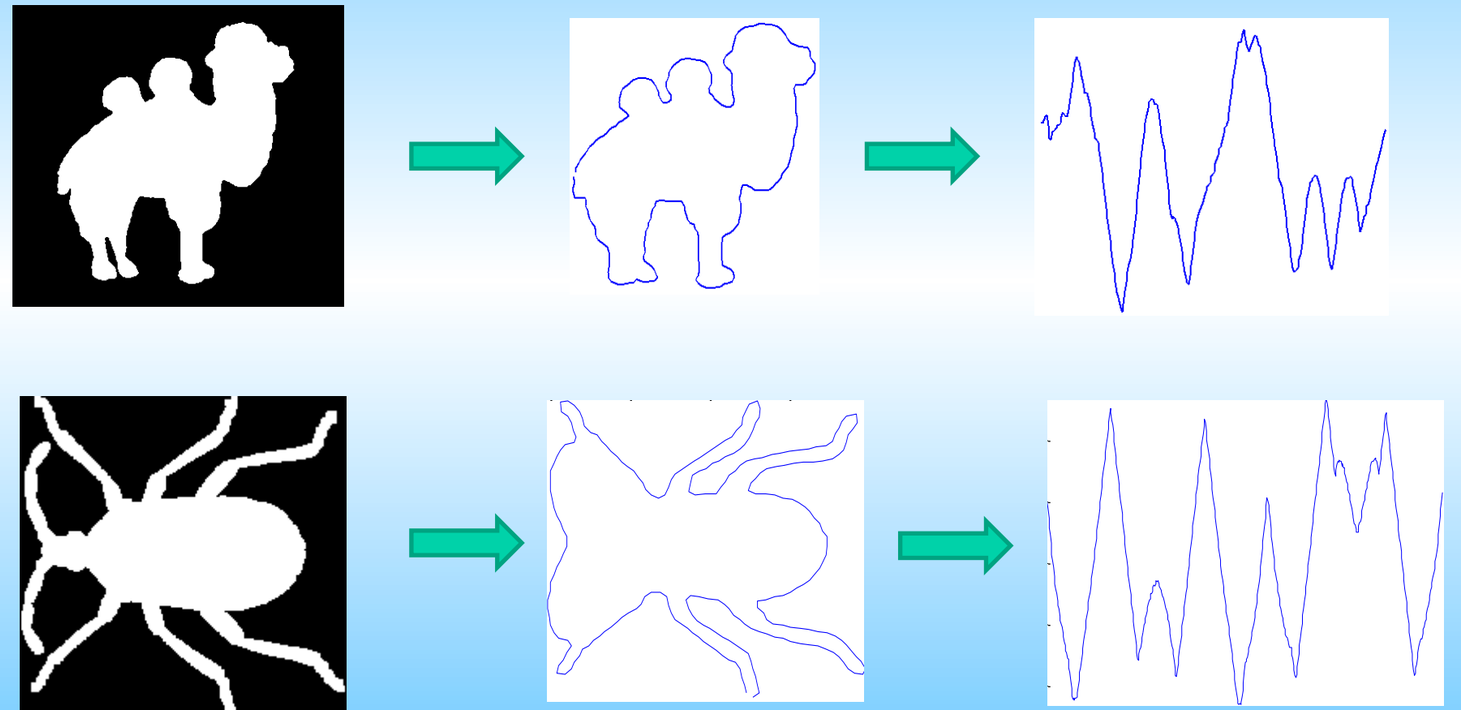
- Hand and bone outlines, Herring Otoliths, Faces, Leaves, Arrow Heads, Yoga, Words/letter, Shapes (MPEG7)



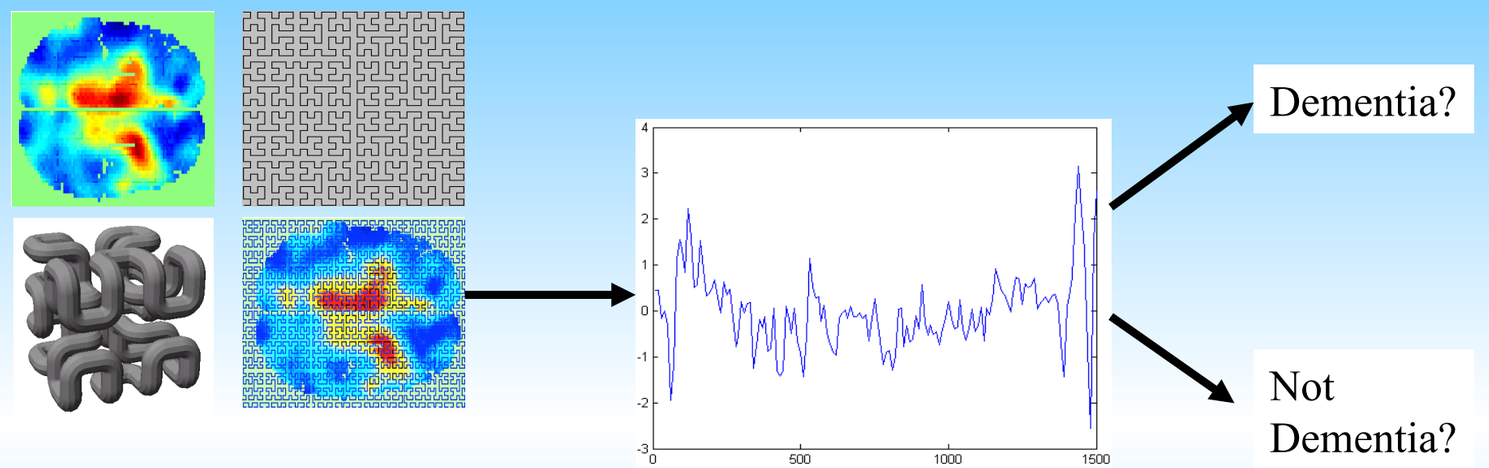




Creating an image outline



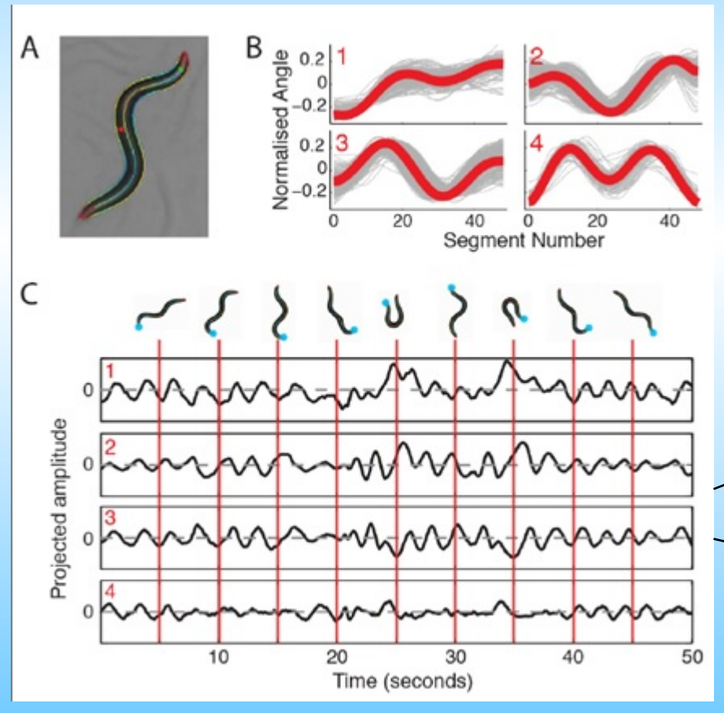
Brain scans (3D voxels), may best be thought of as time series…



Wang, Kontos, Li and Megalooikonomou ICASSP 2004

Sensor (18)

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



Worms and Worms Two Class

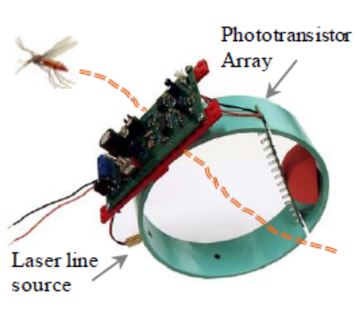
Classifying *C. elegans* mutants through motion traces with Andre Brown at Imperial

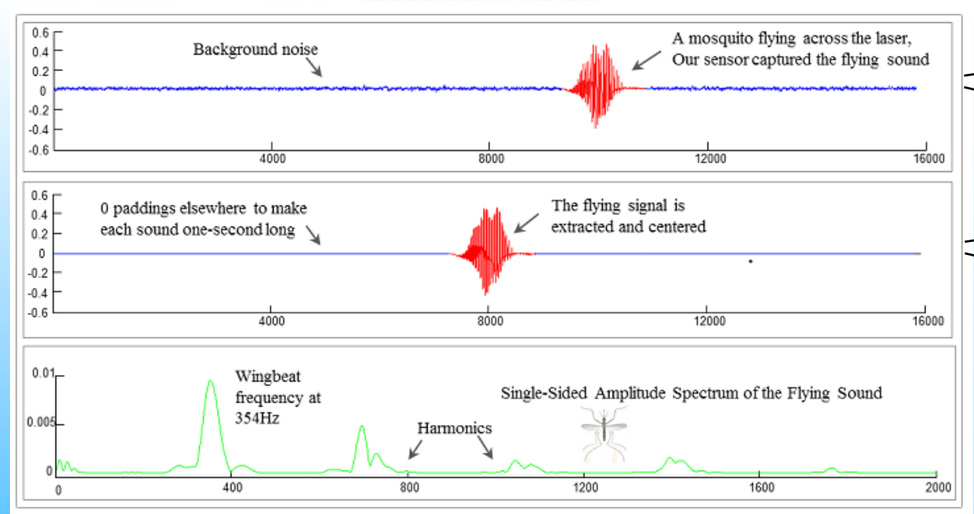
Mutant?

Wild Type?

Insect Classification

icipe
INTERNATIONAL CENTRE OF INSECT PHYSIOLOGY AND ECOLOGY
African Insect Science for Food and Health





Mosquito?

Not mosquito?

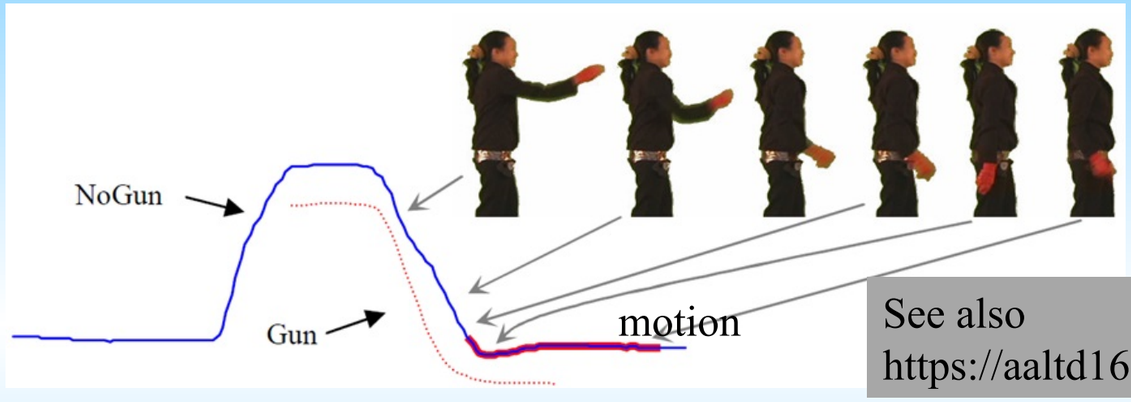
Male?

Female?

Species?

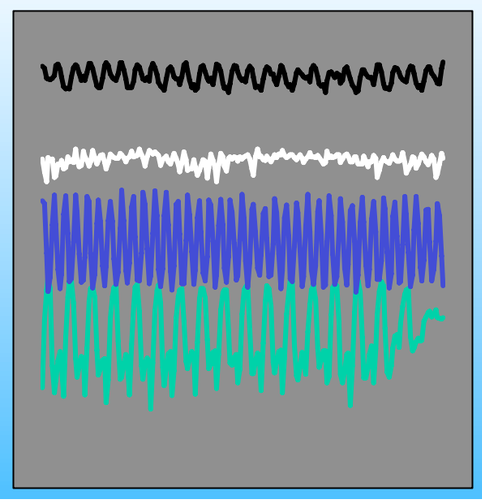
Motion (14)

- Gestures (Uwave), Cricket hand signals, Gun Point,



See also
https://aaltd16.irisa.fr/challenge/

Epilepsy detection with motion sensors from **"Generalized Models for the classification of abnormal movements in daily life and its applicability to epilepsy convulsions recognition"** by J. Villar, *et al*
University of Oviedo



Walking

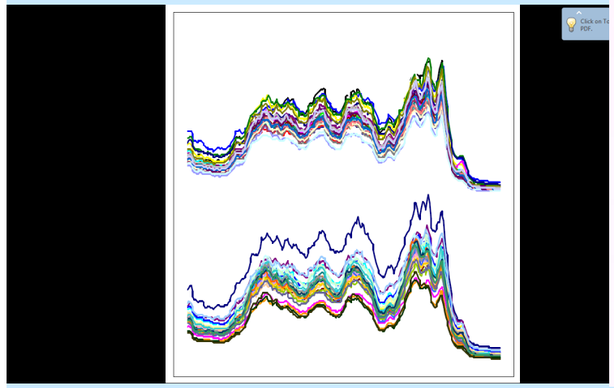
Seizure

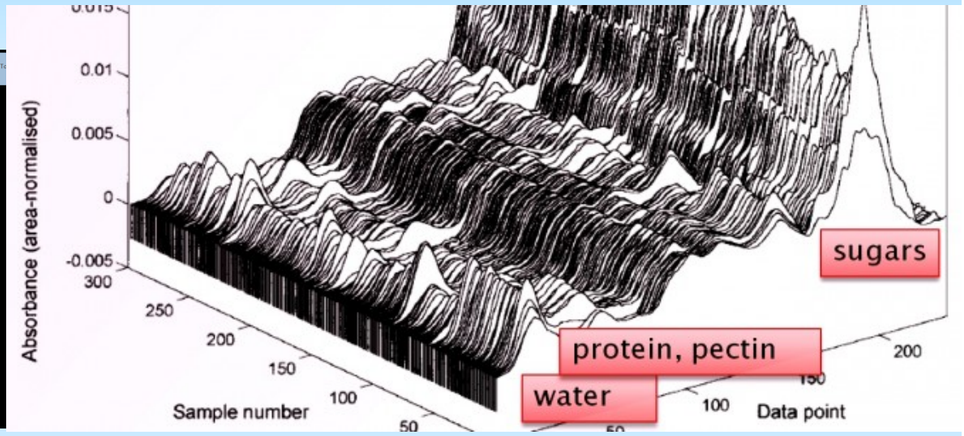
Sawing

Running

Spectra (7)

- Beef, Coffee, Ham, Meat, Olive Oil, Strawberry, Wine. All from BBSRC Institute of Food Research, Norwich Research Park

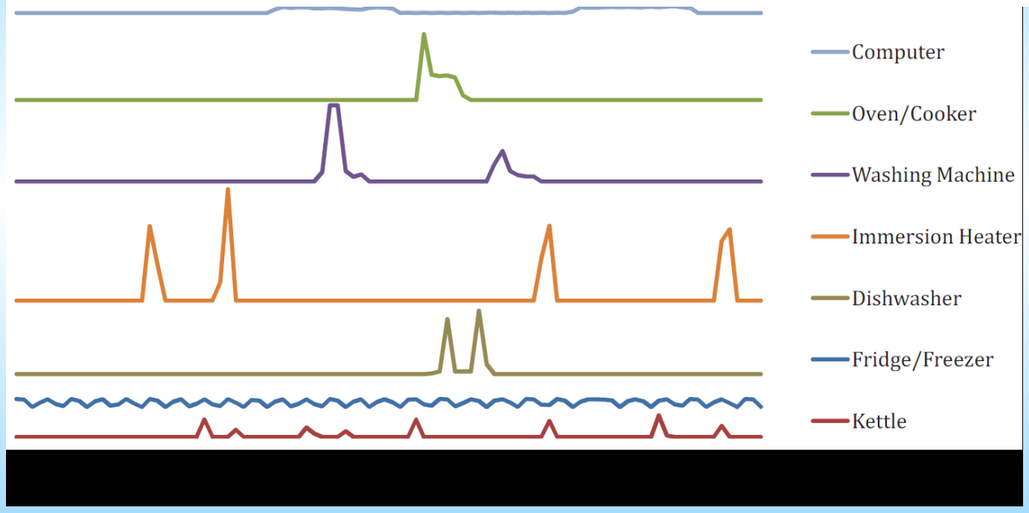




New project with the Scotch Whisky Research Institute on classifying whisky as forged or genuine

Electric Devices (6)

- Electric device measurements (from a study by the UK Energy Savings Trust "Powering the Nation") http://www.energysavingtrust.org.uk/resources/our-research-and-reports

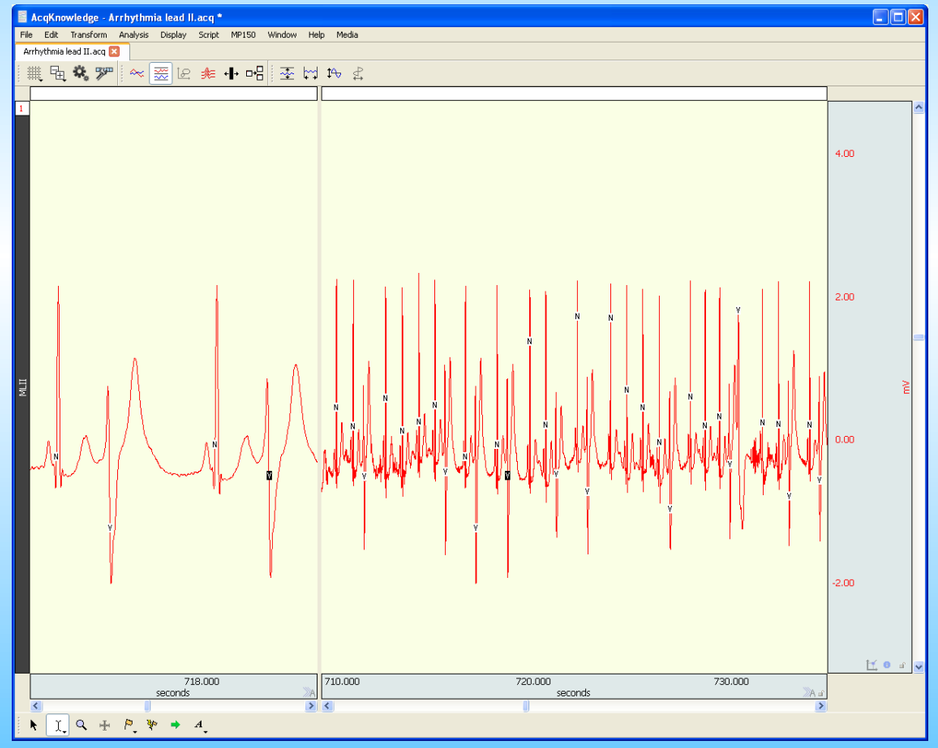


Vermont Energy Investment Corporation

New project: what level of granularity
of data do you need to detect faulty equipment?

ECG (6)

- ECG abnormality detection, computers in cardiography challenges, physionet data sets



So Why is TSC Hard?

- Large attribute space
- Correlated features
- Redundant features

Problems observable in all classification problems

- Autocorrelation structure
- Imbedded models, phase independent feature
- Misallignment
- Very diverse set of problems

Problems specific to ordered series

COTE the Collective of Transform Ensembles

3Year EPSRC responsive mode grant started May 2015

AIMS: Develop and assimilate new algorithms for time series classification

OBJECTIVE 1: Establish what is the best current technique for TSC

See "**The Great Time Series Classification Bake Off: An Experimental Evaluation of Recently Proposed Algorithms. Extended Version"** https://arxiv.org/abs/1602.01711 (under review)

Time Series Classification with COTE: The Collective of Transformation Based Ensembles IEEE Trans KDE http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7069254&tag=1

Approaches to TSC

Full series methods

Vector based, Elastic distance measures First order difference distance measures

Phase dependent subseries methods (intervals)

Find intervals of the series with discriminatory features.

Phase independent subseries methods (shapelets)

Find subsequences that can occur anywhere and define class membership

Subseries distribution methods (dictionary)

Count the occurrence of similar subsequences, then classify based on these counts

Model based approaches

Fit model (e.g. spectral, autoregressive, HMM), measure similarity between series as similarity between models

Ensemble techniques

– All of the above

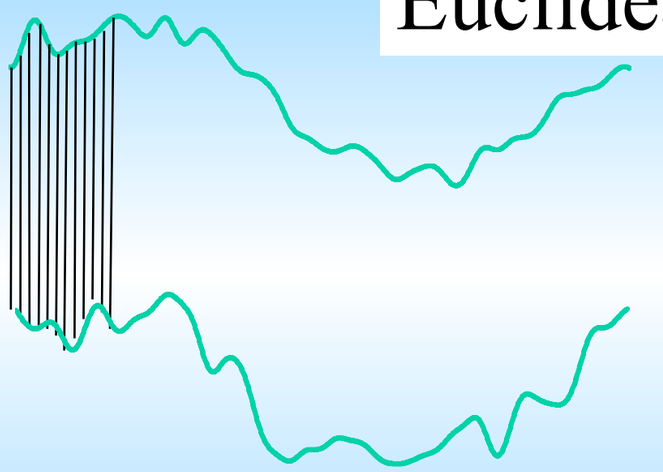
1. Full series methods

Match all of one series to another with a similarity measure

Whole series similarity measure used to classify

Euclidean distance

$$d(A,B) \equiv \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

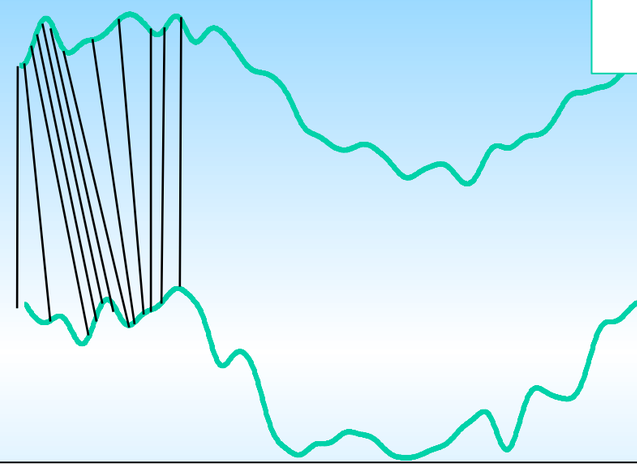


Often used as a benchmark, but really not very good.
Outperformed by standard classifiers

Used with a distance based classifier (e.g. nearest neighbour, support vector machine).

Two examples of starlight curves described in **Finding anomalous periodic time series**, Rebbapragada et al, Machine Learning, 2009

DTW Based



Dynamic Time Warping (DTW)

- Path found with dynamic programming.
- Used by most people as the standard benchmark
- Sensitive to warping window

"*1NN-DTW is very hard to beat*"

Xi et al. "Fast time series classification using numerosity reduction." *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006.

"*Dynamic Time Warping (DTW) is remarkably hard to beat as a time series distance measure*"

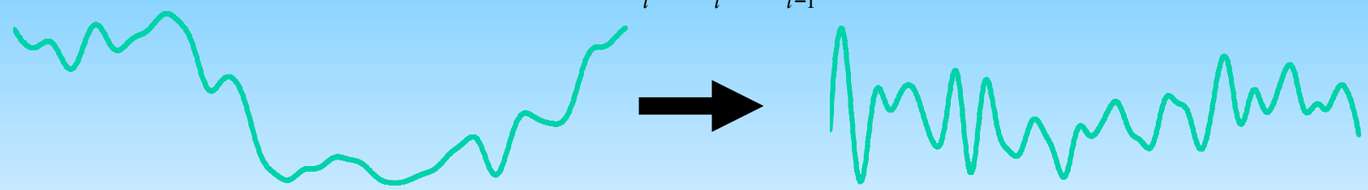
Shokoohi-Yekta et al. "On the Non-Trivial Generalization of Dynamic Time Warping to the Multi-Dimensional Case." *Proc. SIAM Int. Conf. on Data Mining, Vancouver, British Columbia, Canada, April*. 2015.

"*There is increasing evidence that the classic Dynamic Time Warping (DTW) measure is the best measure in most domains. It is difficult to overstate the ubiquity of DTW*"

Rakthanmanon et al. "Data Mining a Trillion Time Series Subsequences Under Dynamic Time Warping." *IJCAI*. 2013.

First Order Differences

$$d_i' = d_i - d_{i-1}$$



And then continue as before with some distance function

Variants:

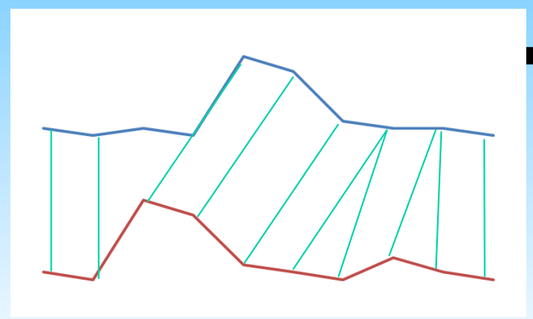
Derivative DTW (DD_DTW) Gorecki *et al*. Data Mining and Know. Disc., 2013

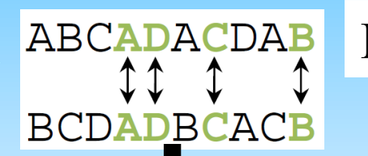
Weighted derivative DTW (WDDTW) Y. Jeong *et al*. Pattern Recognition, 2011

Transform differences (DTD_C). Gorecki *et al.* Data Mining and Know. Disc., 2014

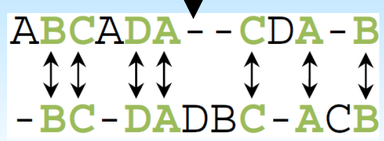
Complexity Invariant Distance(CID) Batista *et al.* Data Mining and Know. Disc., 2014

Edit Distance/Hybrids





Discretise



Transform one on to the other through edit operations

Classify based on the number of operations required

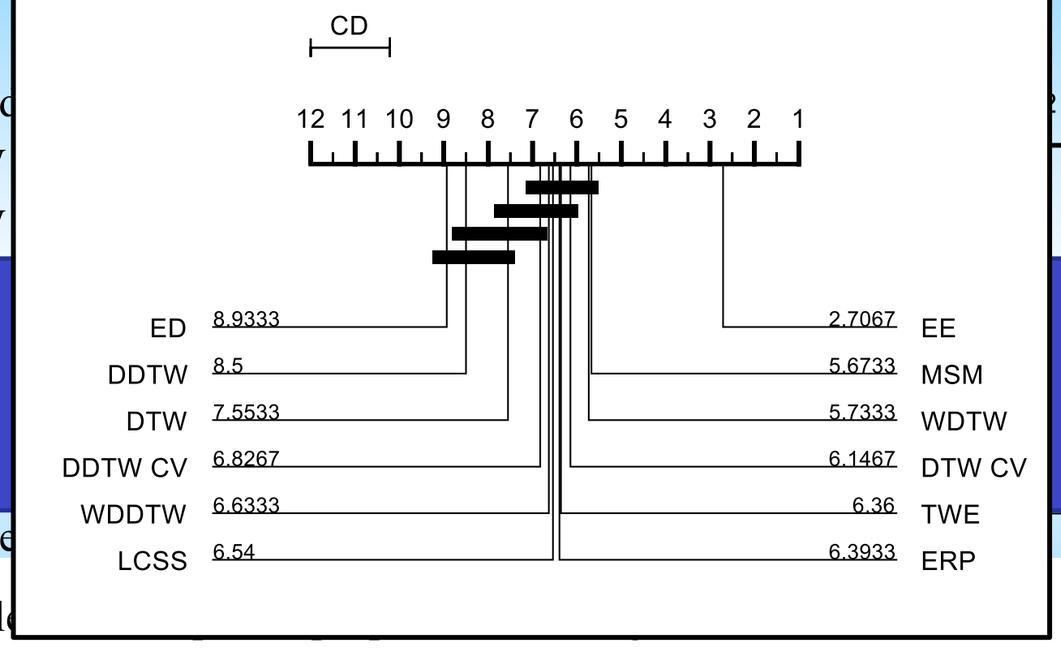
Time Warp Edit Distance Marteau, IEEE Trans. PAMI 2014

Move-Split-Merge Stefan *et al*. IEEE Trans. KDE 2013

Edit Distance with Real Penalty(ERP)

Longest Common Subsequence (LCSS)

Elastic Distance Measures



Euclid

DTW

DTW

Longe

LCSS

- Simple

- Constituents weighted by training cross validation accuracy

J. Lines and A. Bagnall. *"Ensembles of elastic distance measures for time series classification."* *SIAM International Conference on Data Mining,*524-532 (2014).

J. Lines and A. Bagnall. *"Time series classification with ensembles of elastic distance measures."* *Data Mining and Knowledge Discovery* 29.3 (2015): 565-592.

Approaches to TSC

Full series methods

Vector based, Elastic distance measures First order difference distance measures

Phase dependent subseries methods (intervals)

Find intervals of the series with discriminatory features.

Do we need the whole series?

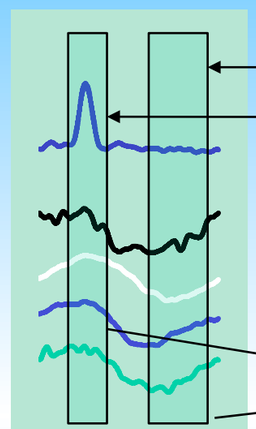
Take intervals

Calculate features on intervals

Construct classifiers

2. Phase dependent subseries methods

Find intervals of the series with discriminatory features.



Select features to construct classifiers

Your favourite classifier

Sample intervals multiple times and construct and ensemble of classifiers

Time Series Forest (TSF) Deng *et al.* Information Science, 2013

Time Series Bag of Features (TSBF) Baydogan *et al*. IEEE Trans. PAMI 2013

Learned Pattern Similarity (LPS) Baydogan *et al*. Data Mining and Know. Disc., 2015

Approaches to TSC

Full series methods

Vector based, Elastic distance measures First order difference distance measures

Phase dependent subseries methods (intervals)

Find intervals of the series with discriminatory features.

Phase independent subseries methods (shapelets)

Find subsequences that can occur anywhere and define class membership

Work derived from

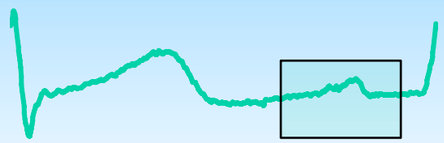
L. Ye and E. Keogh. Time series shapelets: A new primitive for data mining. In Proc. 15th ACM SIGKDD, 2009

A. Mueen, E. Keogh & N. Young, N. *Logical-shapelets: an expressive primitive for time series classification. In Proc. 17th ACM SIGKDD, 2011*

3. Phase independent subseries methods

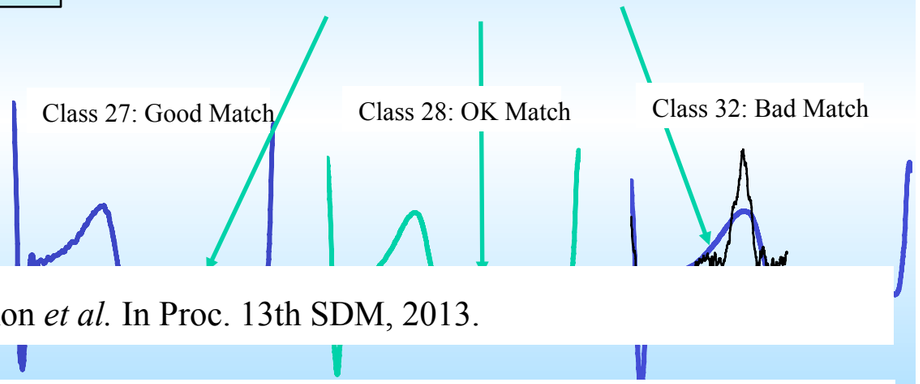
Find subsequences that can occur anywhere and define class membership (shapelets)

Non invasive fetal heartbeat data



Subseries extracted from series

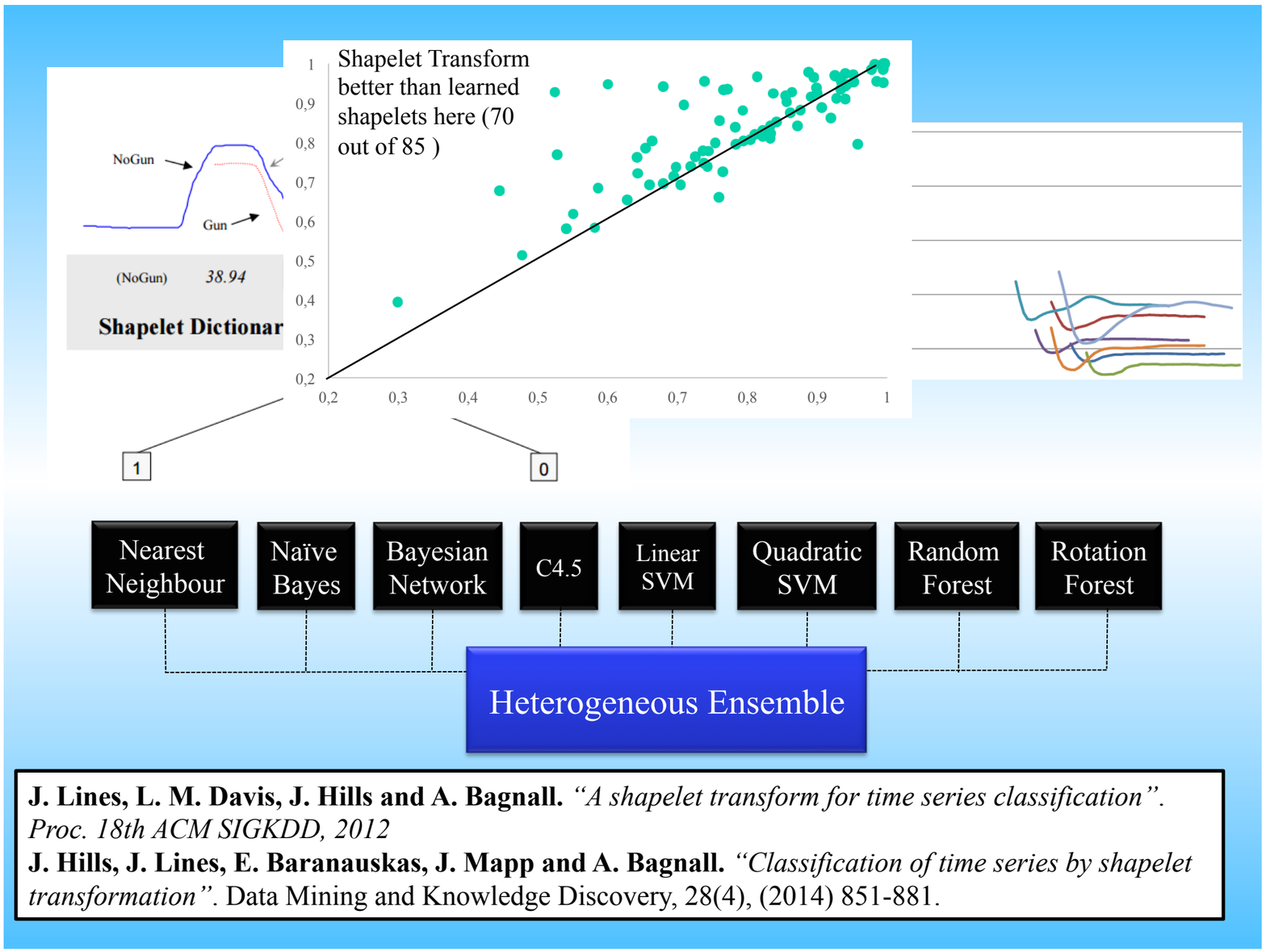
Classify based on distance between shapelets and series



Fast Shapelets (FS) Rakthanmanon *et al.* In Proc. 13th SDM, 2013.

Shapelet Transform (ST) Hills *et al.* Data Mining and Know. Disc., 2015

Learned Shapelets (LS) Grabocka *et al. Proc. 20th ACM SIGKDD, 2014*



J. Lines, L. M. Davis, J. Hills and A. Bagnall. *"A shapelet transform for time series classification"*. *Proc. 18th ACM SIGKDD, 2012*

J. Hills, J. Lines, E. Baranauskas, J. Mapp and A. Bagnall. *"Classification of time series by shapelet transformation"*. Data Mining and Knowledge Discovery, 28(4), (2014) 851-881.

Approaches to TSC

Full series methods

Vector based, Elastic distance measures First order difference distance measures

Phase dependent subseries methods (intervals)

Find intervals of the series with discriminatory features.

Phase independent subseries methods (shapelets)

Find subsequences that can occur anywhere and define class membership

Subseries distribution methods (dictionary)

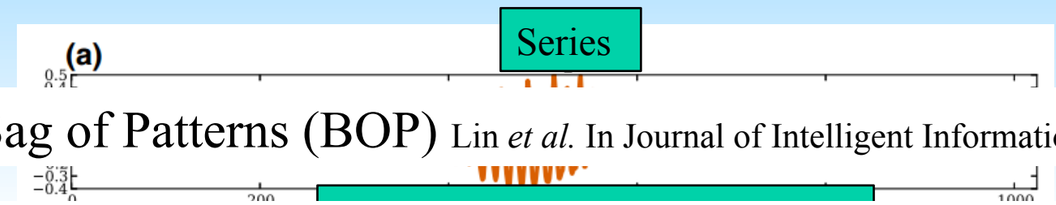
Count the occurrence of similar subsequences, then classify based on these counts

Techniques inspired by information retrieval

4. Subseries frequency methods (dictionary based)

Count the occurrence of similar subsequences, then classify based on these counts

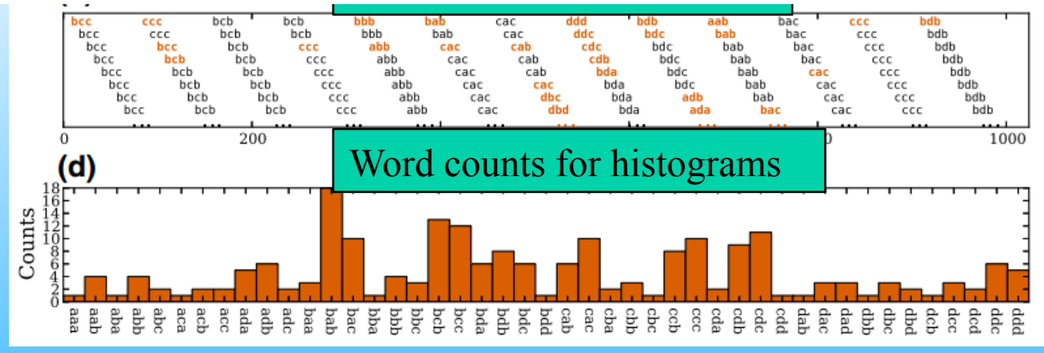
Series



Bag of Patterns (BOP) Lin *et al.* In Journal of Intelligent Information, 2012.

SAX-Vector Space Model (SAX-VSM) Senin *et al.* 13th ICDM, 2013

Bag-of-SFA-Symbols (BOSS) Schäfer, Data Mining and Knowledge Discovery, 2015



Word counts for histograms

Classification based on histogram similarity

Approaches to TSC

Full series methods

Vector based, Elastic distance measures First order difference distance measures

Phase dependent subseries methods (intervals)

Find intervals of the series with discriminatory features.

Phase independent subseries methods (shapelets)

Find subsequences that can occur anywhere and define class membership

Subseries distribution methods (dictionary)

Count the occurrence of similar subsequences, then classify based on these counts

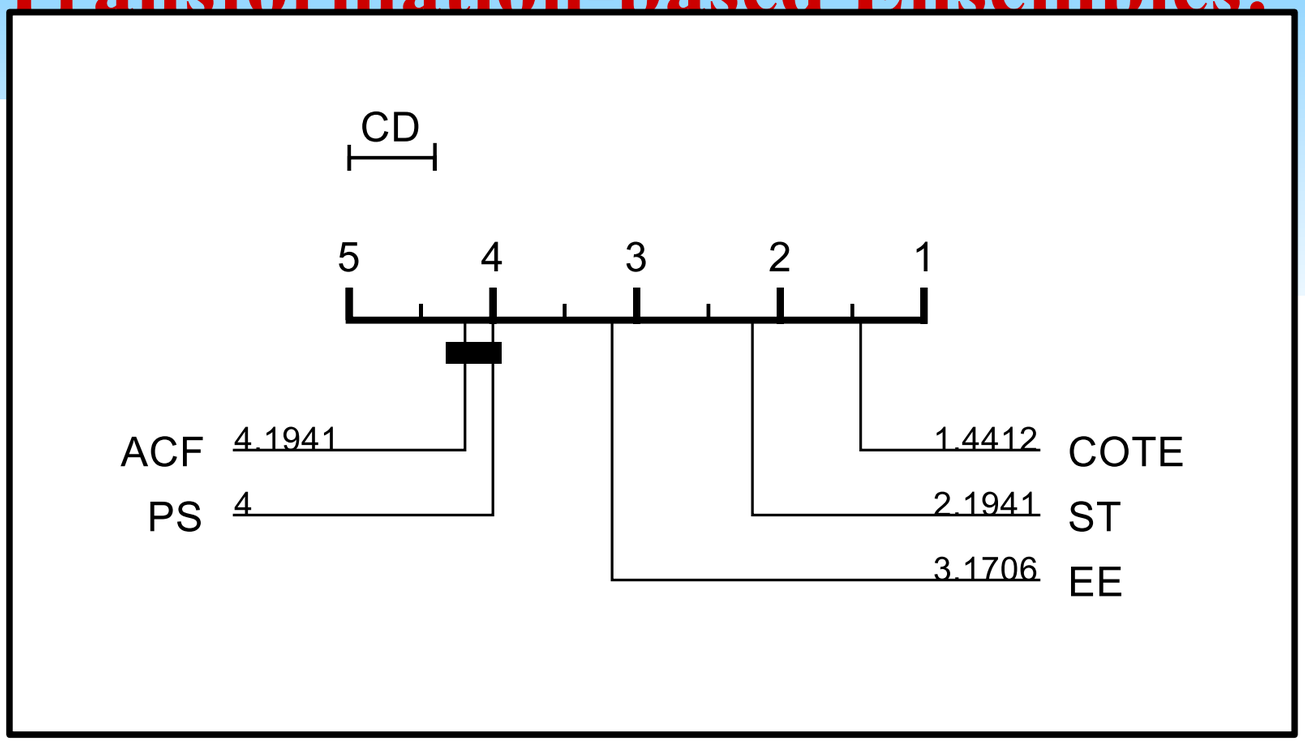
Model based approaches

Fit model (e.g. spectral, autoregressive, HMM), measure similarity between series as similarity between models

Ensemble techniques

– All of the above

The Collective of Transformation-based Ensembles:



A. Bagnall, J. Lines, J. Hills and A. Bostrom. "*Time-series classification with COTE: the collective of transformation-based ensembles*". *Knowledge and Data Engineering, IEEE Transactions on*, *27*(9), 2522-2535 (2015).

The Great Time Series Classification Bakeoff

- One of the **largest ever studies in** machine learning
- 22 competing TSC algorithms from top-tier journal and conference papers implemented in a common framework
- All run on 100 resamples of the 85 UCR datasets
- Over 30 million individual experiments
- Completely reproducible and transparent

All code and results publicly available: www.timeseriesclassification.com

Null hypothesis mean difference is zero

Proportion of problems better than DTW

Average difference in accuracy

Algo	P Value	Prop	Mean
COTE	0.00	96.47%	8.12%
EE	0.00	95.29%	3.51%
BOSS	0.00	82.35%	5.76%
ST	0.00	80.00%	6.13%
DTW_F	0.00	74.70%	2.65%
TSF	0.00	68.24%	1.91%
TSBF	0.00	65.88%	2.19%
MSM_1NN	0.00	62.35%	1.89%
LPS	0.01	61.18%	1.83%
WDTW_1NN	0.03	60.00%	0.20%
DTD_C	0.05	52.94%	0.79%
CID_DTW	0.03	50.59%	0.54%
DD_DTW	0.24	56.47%	0.42%
RotF	0.49	56.47%	-0.02%
TWE_1NN	0.83	49.41%	0.37%
LS	0.03	47.06%	-2.99%
SAXVSM	0.00	41.18%	-3.29%
BOP	0.00	37.65%	-3.05%
FS	0.00	30.59%	-7.40%

12 significantly better than DTW

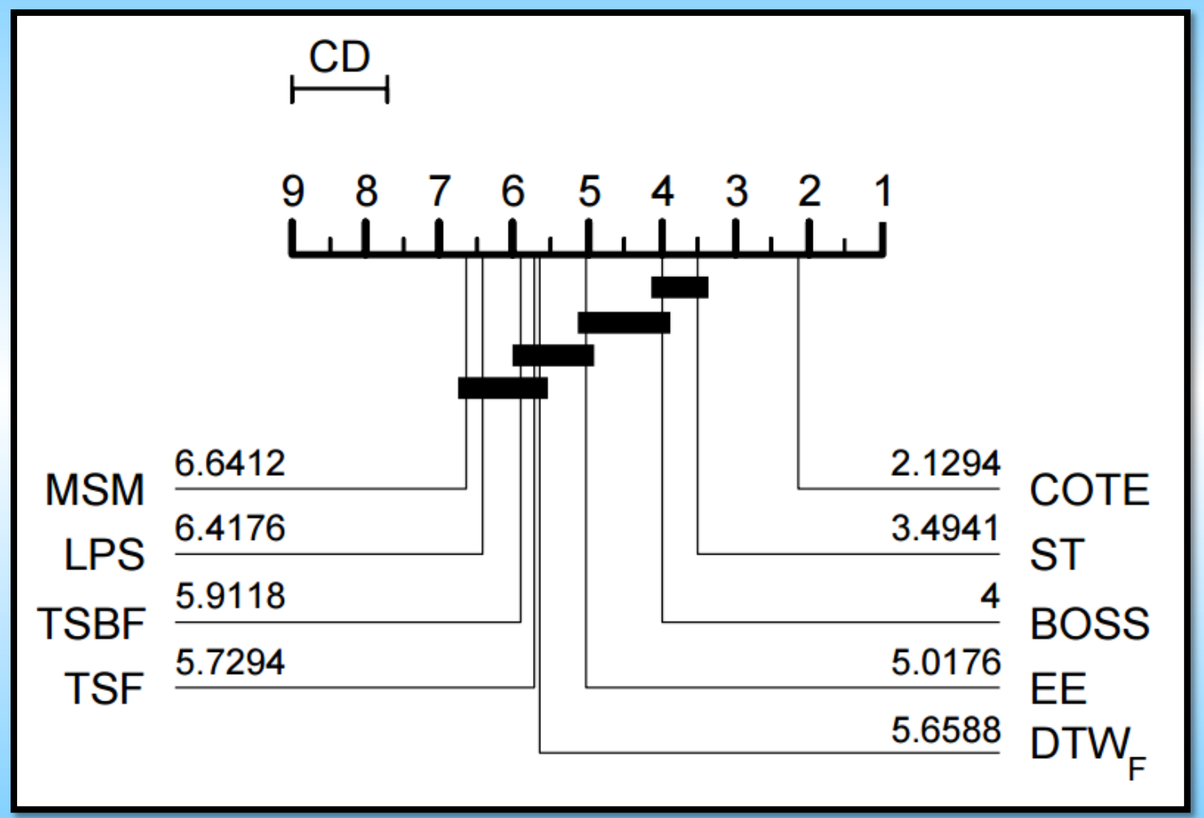
Large improvement
ST, BOSS, EE and COTE

Small improvement
LPS,MSM, TSF, TSBF and DTW_F

Very minor improvement
CID, DTD and WDTW

Significantly **worse** than DTW

The Great Time Series

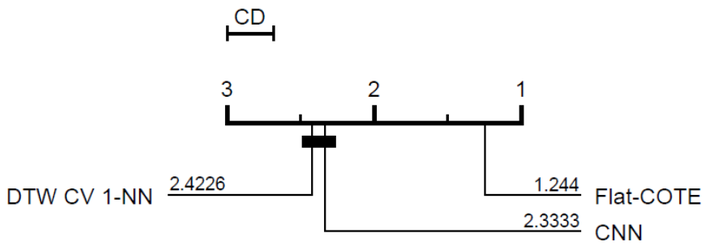


All code and results publicly available: www.timeseriesclassification.com

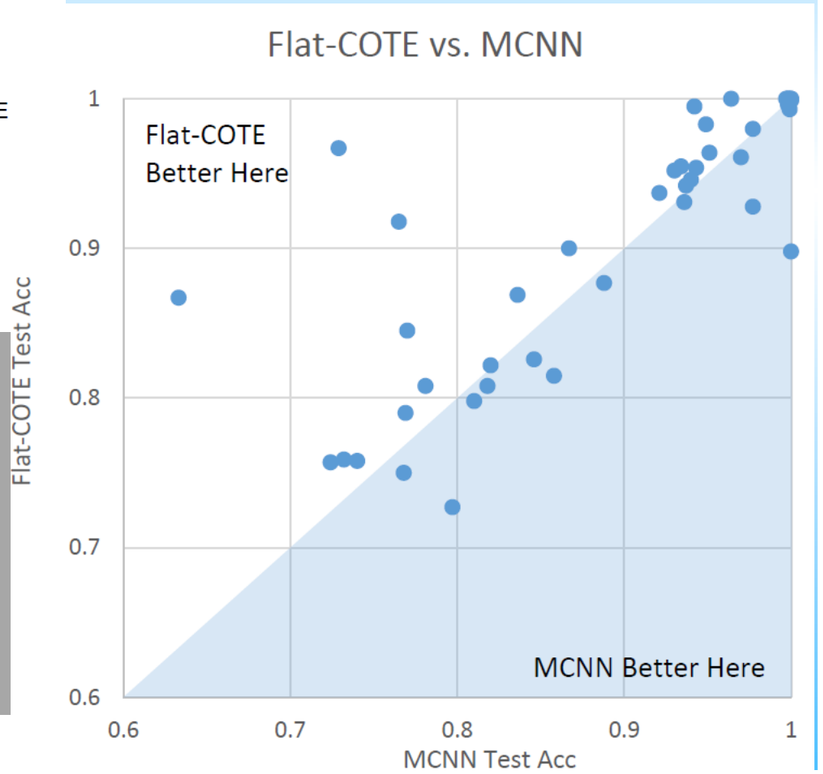
Unpublished results

Where Next for COTE?

Flat COTE vs Deep Learning



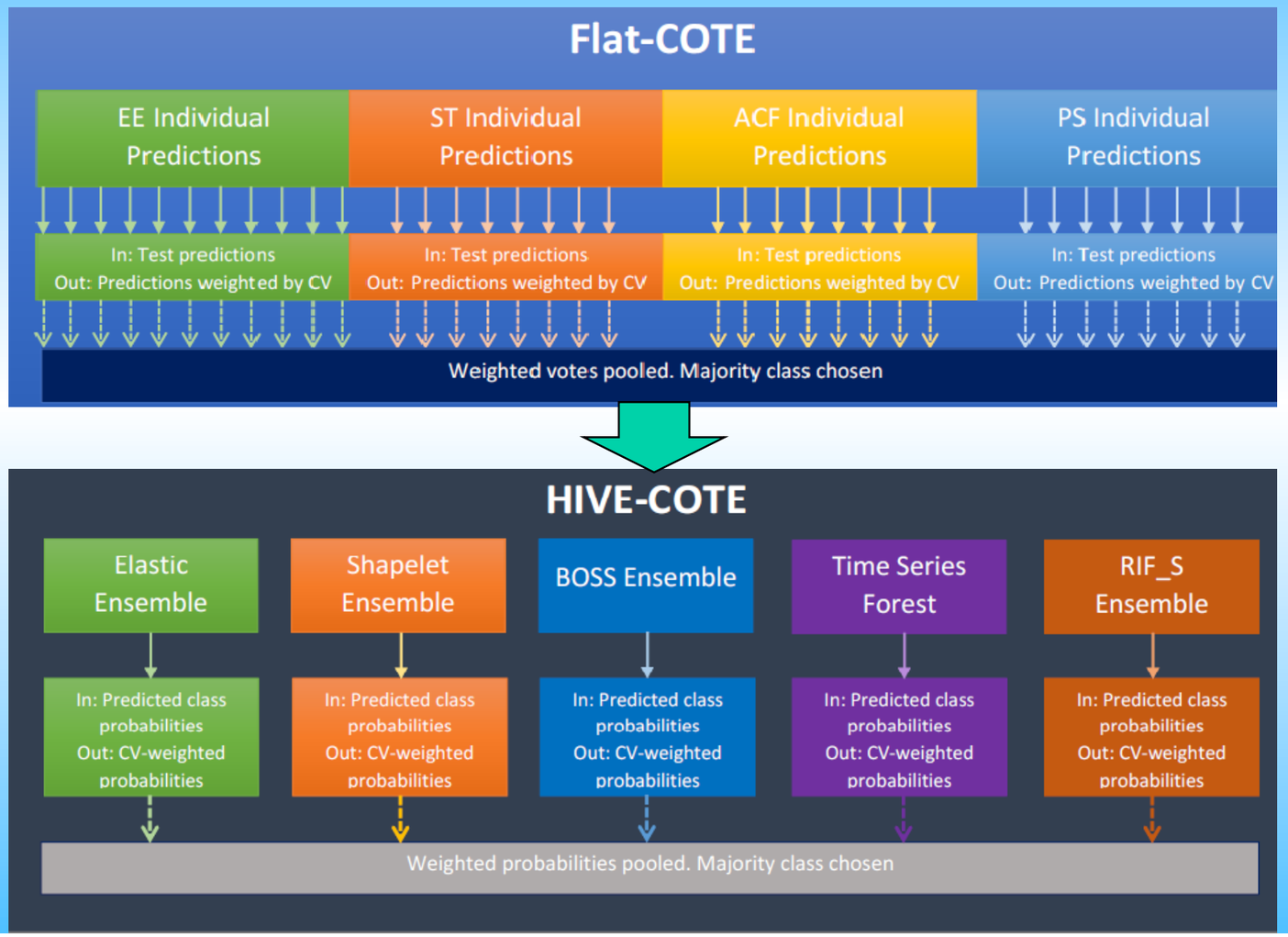
Multi-Scale Convolutional Neural Networks for Time Series Classification
Authors: Zhicheng Cui, Wenlin Chen, Yixin Chen
http://arxiv.org/abs/1603.06995



Where Next for COTE?

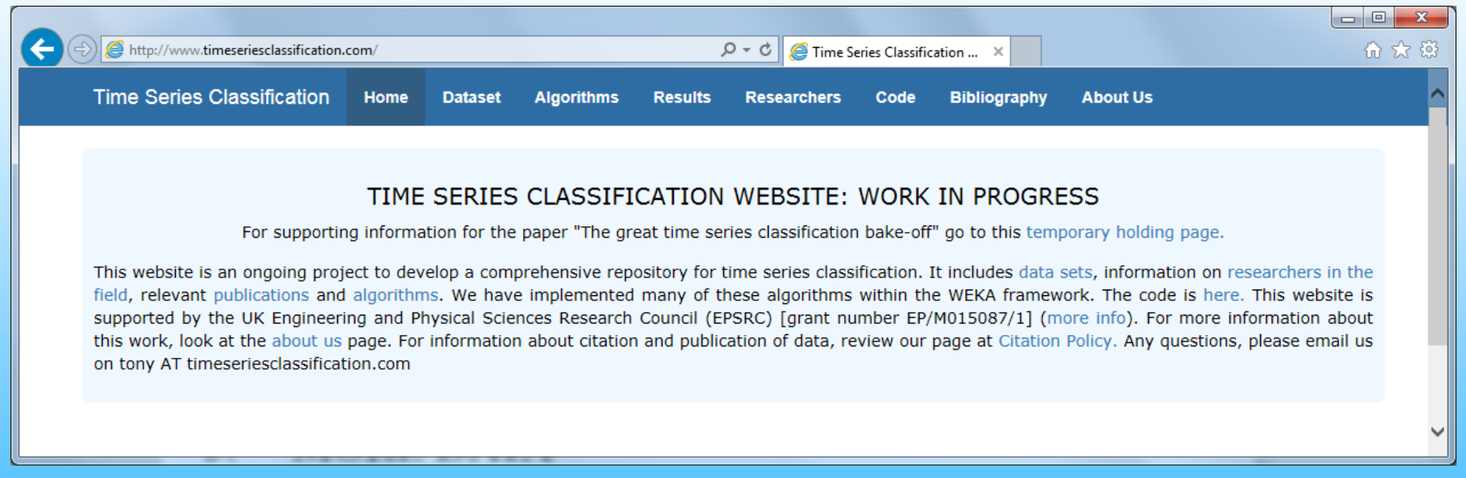
Unpublished results

Modular Structure and assimilate BOSS and TSF



Where Next for COTE?

1. MORE DATA
2. Better fusion strategy
3. Better use of autocorrelation function
4. Better weighted ensemble
5. Scalable
6. Multidimensional



Any Questions?