Univariate Dictionary-based Classifiers

BoP, BOSS, WEASEL

Erik Gottschalk, Xing Wang

SE "Landnutzungsklassifikation – als Wettbewerb"
Humboldt-Universität zu Berlin

February 1, 2019

Structure

Structure

Introduction

- **Bag of Patterns**
- BOSS
- BOSS VS
- WEASEL
- ...

Introduction

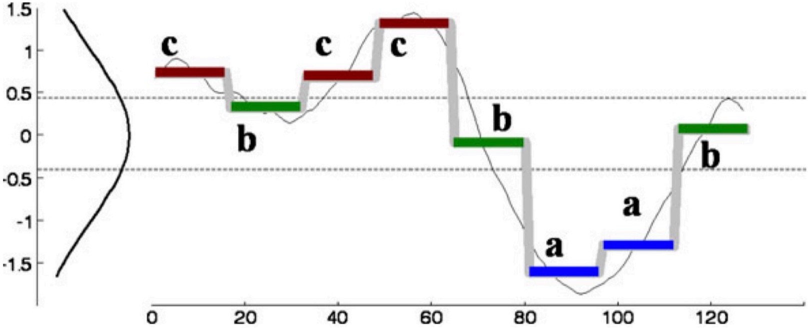


Figure: Bag of Patterns [1]

Introduction

- Bag of Patterns
- BOSS
- BOSS VS
- WEASEL
- . . .

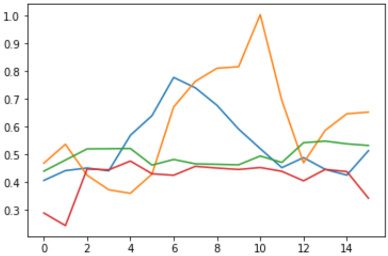
Structure

WEASEL

- Data Preprocessing (NDVI)
- **WEASEL Feature Selection/Transformation to Bag-of-Words**
- Logistic Regression/SVM... Classifier

- Fast Fourier Transformation (FFT)
- F-Test for discriminative Fourier coefficients
- Binning with Information Gain (Bag-of-Words)
- Chi-Square for Dimensionality Reduction

- Input: Preprocessed Time Series NDVI
- Output: Fourier Coefficients



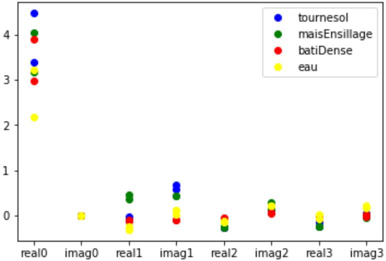


Figure: Example Time Series: 16 time steps

WEASEL

F-Test for discriminative Fourier coefficients

- Choose the f most discriminative Fourier coefficients
- f is the word length
- Example for $f = 1$ and one Fourier coefficient:

sample	class0	class1
0	0	2
1	3	5

- Choose the f most discriminative Fourier coefficients
- f is the word length
- Example for $f = 1$ and one Fourier coefficient:

sample	class0	class1
0	0	2
1	3	5
mean	1.5	3.5

- Total mean: 2.5

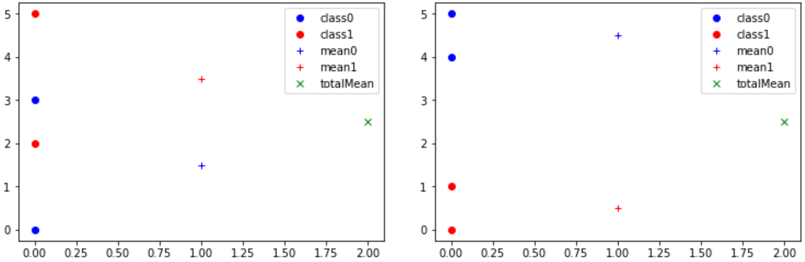


Figure: Plot of sample data: real0 (left), real1 (right)

WEASEL

F-Test for discriminative Fourier coefficients

$$F = \frac{\text{MSB}}{\text{MSW}}$$

$$\begin{aligned}
\text{MSB}_1 &= (1.5 - 2.5)^2 + (3.5 - 2.5)^2 = 1 \\
\text{MSW}_1 &= (0 - 1.5)^2 + (3 - 1.5)^2 + (2 - 3.5)^2 + (5 - 3.5)^2 = 4.5 \\
F_1 &= \frac{1}{4.5} = 0.\overline{2}
\end{aligned}$$

$$\begin{aligned}
\text{MSB}_2 &= (0.5 - 2.5)^2 + (4.5 - 2.5)^2 = 8 \\
\text{MSW}_2 &= (0 - 0.5)^2 + (1 - 0.5)^2 + (4 - 4.5)^2 + (5 - 4.5)^2 = 1 \\
F_2 &= \frac{8}{1} = 8
\end{aligned}$$

- Get the s most optimal split points for symbols
- s is the amount of the symbol alphabet
- Y_i : class labels

$$Ent(Y) = \sum_{(s_i, y_i) \in Y} -p_{y_i} \log_2 p_{y_i}$$

$$Ent(Y, x) = \frac{|Y_{Left}|}{|Y|} \cdot Ent(Y_{Left}) + \frac{|Y_{Right}|}{|Y|} \cdot Ent(Y_{Right})$$

$$InformationGain = Ent(Y) - Ent(Y, x)$$

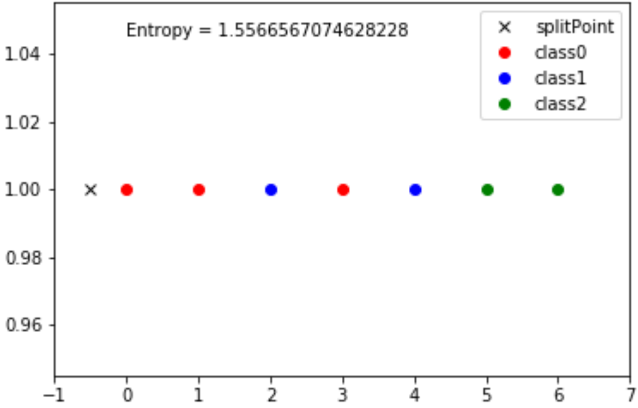


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

WEASEL

Binning with Information Gain

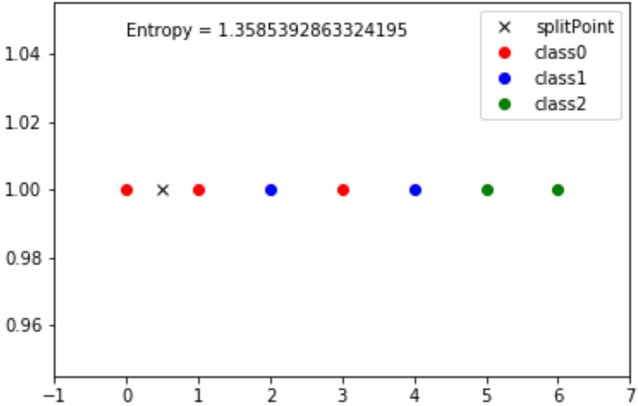


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

WEASEL

Binning with Information Gain

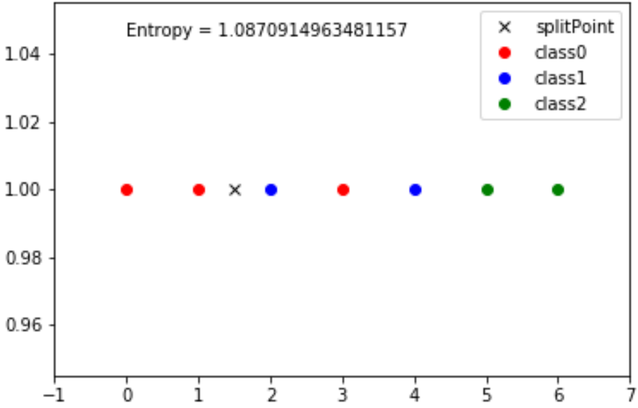


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

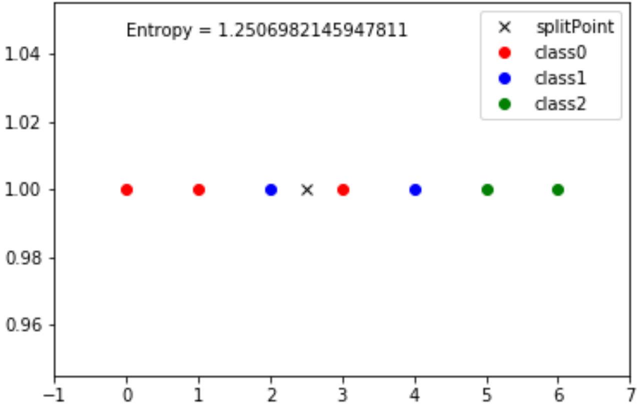


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

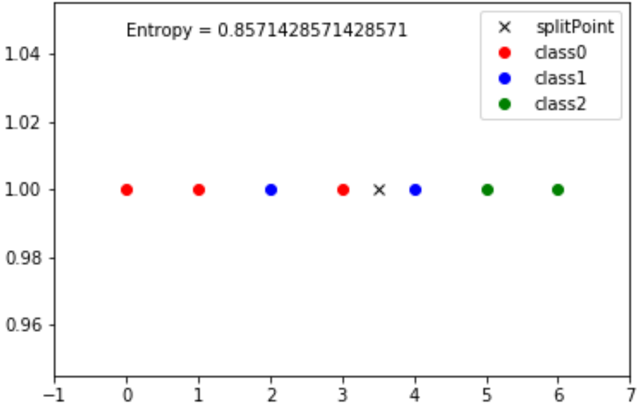


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

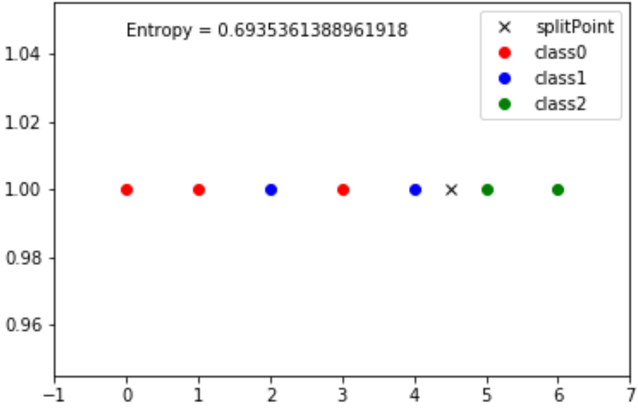


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

WEASEL

Binning with Information Gain

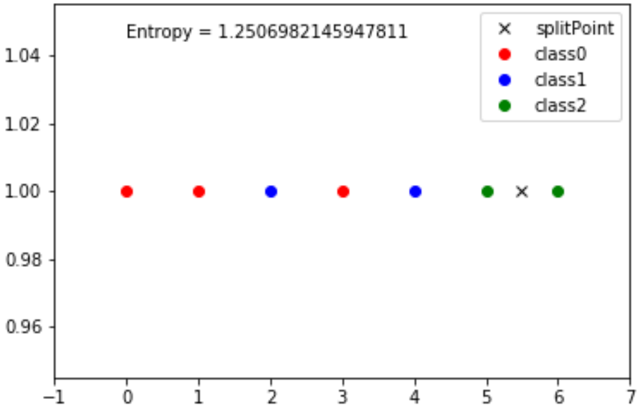


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

WEASEL

Binning with Information Gain

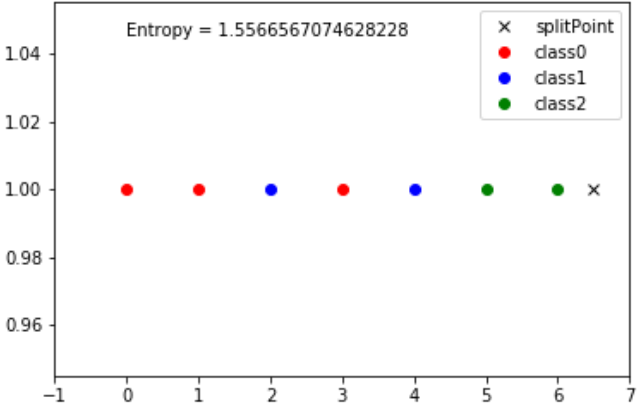


Figure: Fourier Coefficient real1: Find one split point, $s = 2$

WEASEL

Binning with Information Gain

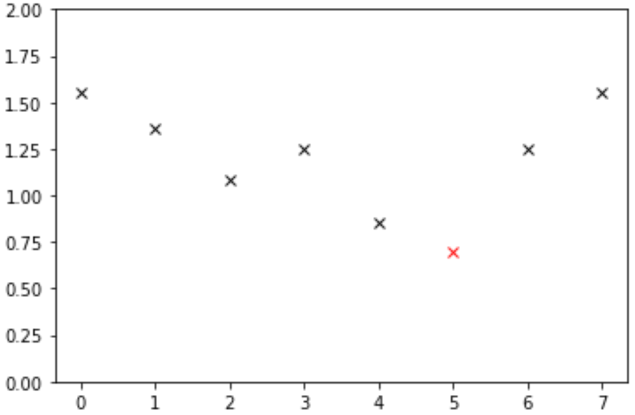


Figure: Fourier Coefficient real1: Max. Information Gain/Min. Entropy

WEASEL

Binning with Information Gain

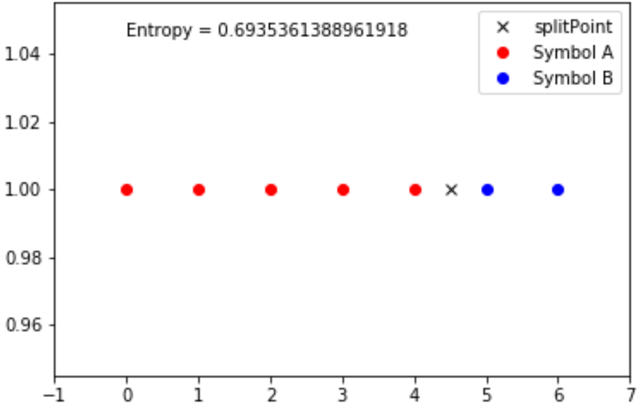


Figure: Fourier Coefficient real1: Splitting point into symbols 'A' & 'B'

- Dimensionality of Feature Space
- $\mathcal{O}(c^l)$ amount of words
- c is amount of symbols, l is word length
- using bigrams and $\mathcal{O}(n \cdot \log(n))$ non-overlapping windows yields $\rightarrow \mathcal{O}(c^{2l} \cdot n \cdot \log(n))$ features

	AA	AB	BA	BB	
1	350	50	50	0	450
2	150	50	50	50	300
3	100	0	0	100	200
4	0	0	0	50	50
	600	100	100	200	1000

Table: absolute frequency of words and classes

- degrees of freedom $= (4-1) \cdot (4-1) = 9$
- χ^2 values for:
  - AA:

$$(80/270)^2 + (30/180)^2 + (20/120)^2 + (30/30)^2 < 2$$

  - AB and BA:

$$(5/45)^2 + (20/30)^2 + (20/20)^2 + (5/5)^2 > 2$$

  - BB:

$$(90/90)^2 + (10/60)^2 + (60/40)^2 + (40/10)^2 > 19$$

Structure

Implementation

Preprocessing

- Univariate $\rightarrow$ NDVI

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$$

- NaNs $\rightarrow$ Interpolation: *Python Pandas*
  - Linear
  - Zero Filling
  - Quadratic Spline
  - Qubic Spline

Implementation

Preprocessing

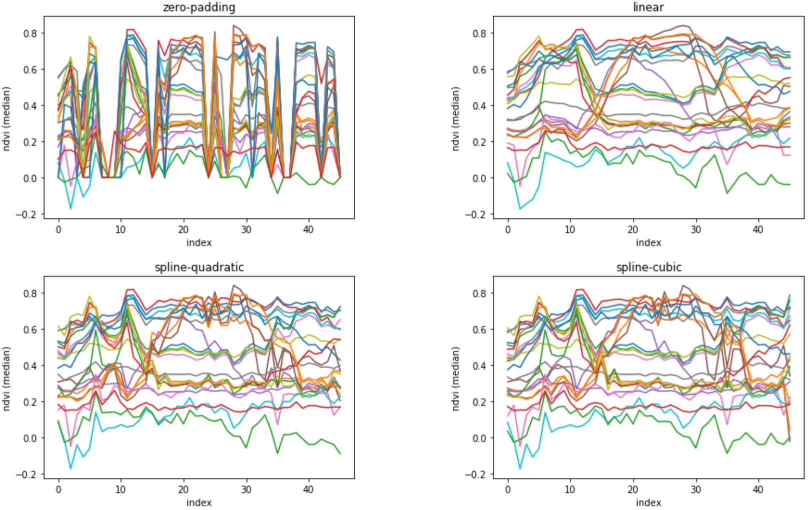


Figure: Imputation Methods; $N = 100000$

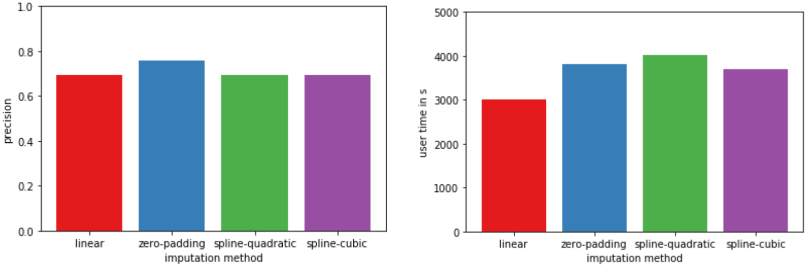


Figure: Influence of Imputation on Accuracy (left) and User Time (right), Train Set Size $N = 100000$

- Java Library for Time Series Data Analytics
  `https://github.com/patrickzib/SFA`
- Hyperparameters $\rightarrow$ Grid Search 90/10-Split
  - Min/Max Window Size
  - Max Alphabet Size

minF =maxF	maxS			
	2	4	8	16
2	0.583	0.579	0.618	0.632
4	0.598	0.643	0.643	0.635
6	0.624	0.654	0.645	0.640
8	0.616	0.667	0.645	0.608
16	0.641	0.636	0.624	0.060

	2	4	8	16
2	7071s	4769s	1247s	388s
4	2075s	395s	337s	251s
6	742s	282s	275s	296s
8	472s	282s	320s	323s
16	955s	691s	730s	130s

Table: Top: Accuracy, Bottom: User Time – Train Set Size: $N = 10000$

Implementation

Classification

minF =maxF	maxS			
	2	4	8	16
2				
4		0.662	0.675	0.668
6		0.684	0.687	0.683
8		0.689	0.694	0.680
16		0.686	0.685	0.166

	2	4	8	16
2				
4		10920s	4535s	4197s
6		5067s	3681s	3249s
8		3804s	3015s	3112s
16		3677s	3226s	846s

Table: Top: Accuracy, Bottom: User Time – Train Set Size: $N = 100000$

Structure

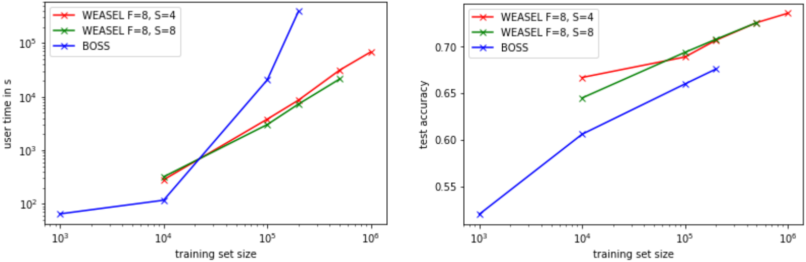


Figure: Influence of Train Set Size on User Time (left) and Accuracy (right)

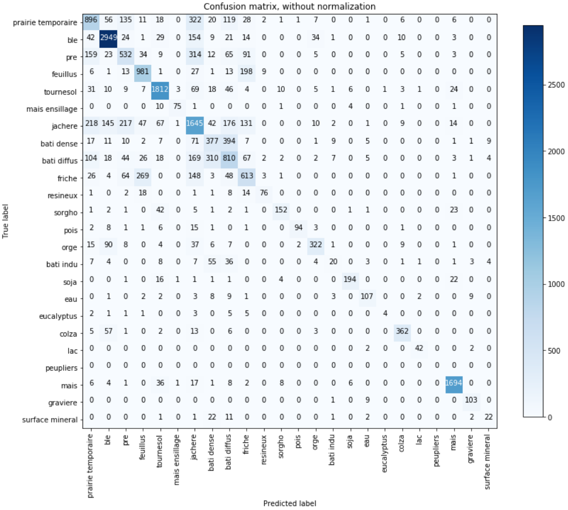
Evaluation

Per-Class Accuracy

	precision	recall	f1-score	support
prairie temporaire	0.58	0.55	0.57	1629
ble	0.87	0.90	0.88	3291
pre	0.50	0.42	0.46	1252
feuillus	0.70	0.78	0.74	1250
tournesol	0.87	0.88	0.87	2060
mais ensillage	0.93	0.81	0.86	93
jachere	0.54	0.60	0.57	2725
bati dense	0.42	0.41	**0.42**	922
bati diffus	0.45	0.51	0.48	1592
friche	0.52	0.52	0.52	1180
resineux	0.83	0.63	0.71	121
sorgho	0.85	0.66	0.74	232
pois	0.97	0.71	0.82	133
orge	0.81	0.64	0.72	502
bati indu	0.43	0.13	**0.20**	154
soja	0.92	0.80	0.86	241
eau	0.79	0.73	0.76	147
eucalyptus	0.80	0.18	**0.30**	22
colza	0.89	0.81	0.85	449
lac	0.91	0.91	0.91	46
peupliers	0.00	0.00	0.00	0
mais	0.94	0.95	0.95	1784
graviere	0.85	0.91	0.88	113
surface mineral	0.56	0.35	**0.44**	62
weighted avg	**0.70**	**0.69**	**0.69**	**20000**

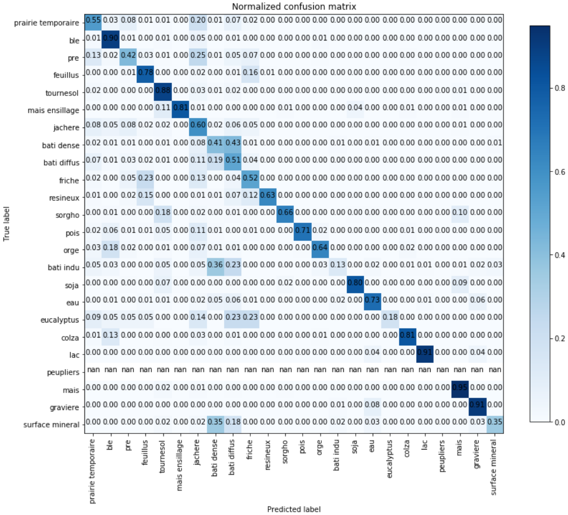
Evaluation

Confusion Matrix



Evaluation

Confusion Matrix



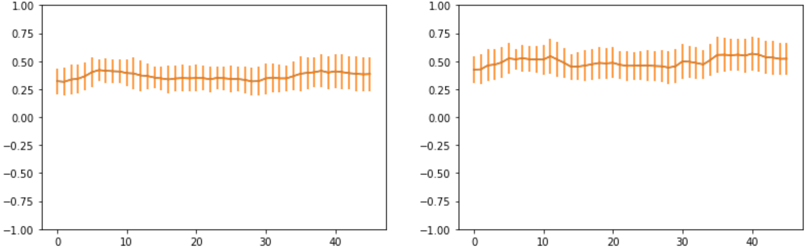


Figure: bati dense vs. bati diffuse

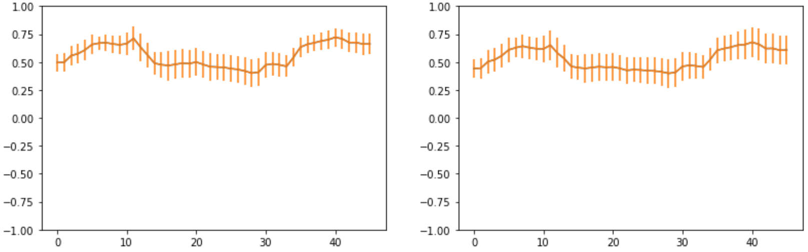


Figure: pre vs. jachere

Structure

Conclusion

- general method for time series classification
- few code adjustments
- lots of hyperparameters
- good scalability
- optimization: majority vote of multiple models (Ensemble)

References I

[1] Lin, J., Khade, R. & Li, Y.
Rotation-invariant similarity in time series using bag-of-patterns representation.
Journal of Intelligent Information Systems, 39(2):287–315, 2012.

[2] Schäfer, P.
The BOSS is concerned with time series classification in the presence of noise.
Data Mining and Knowledge Discovery, 29(6):1505–1530, 2015.

[3] Schäfer, P. & Leser, U.
Fast and accurate time series classification with WEASEL.
Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1505–1530, 2015.