WEASEL 2.0 - A Random Dilated Dictionary Transform for Fast, Accurate and Memory Constrained Time Series Classification

Patrick Schäfer
Humboldt-Universität zu Berlin
Berlin, Germany
patrick.schaefer@hu-berlin.de

Ulf Leser
Humboldt-Universität zu Berlin
Berlin, Germany
leser@informatik.hu-berlin.de

Source Codes

TSC

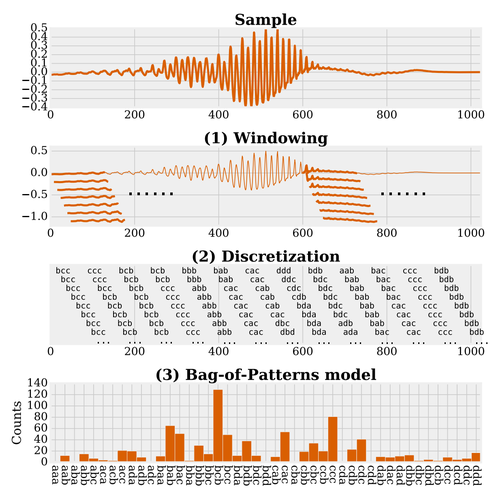
Time Series Classification (TSC)

- Assign label to a time series (TS) based on model trained from labelled samples
- State of the art TSC can be divided into:
  - Dictionary (D): BOSS, cBOSS, WEASEL, TDE;
  - Hybrid (H): HiveCote 2.0, HiveCote 1.0, TS CHIEF,
  - Deep-Learning (DL): InceptionTime;
  - Shapelets (S): R-DST, MrSQM SFA k5;
  - Kernel (K): Arsenal, MiniRocket, Multi- Rocket, Rocket, Hydra

Dictionary Models

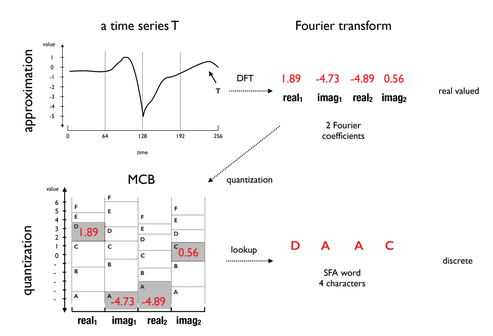
Dictionary Models

- Aim: Generate vector of word counts as features
- Rationale: Distinguishes TS by the **frequency of occurrence of substructures**
- Words are generated using **discretization** from substructures
- Are part of all SotA Hybrid-methods



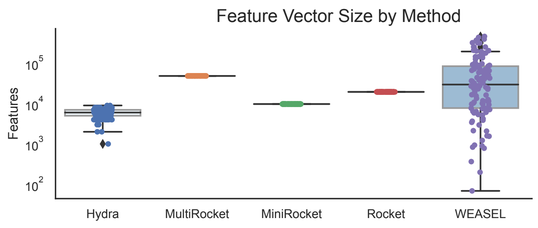
Word Generation

- Aim: Transform real-valued window into sequence of symbols
- Building blocks:
  - **Approximation** using Fourier Transform (for noise reduction)
  - **Binning** (equal-depth, equal-width) for obtaining symbols



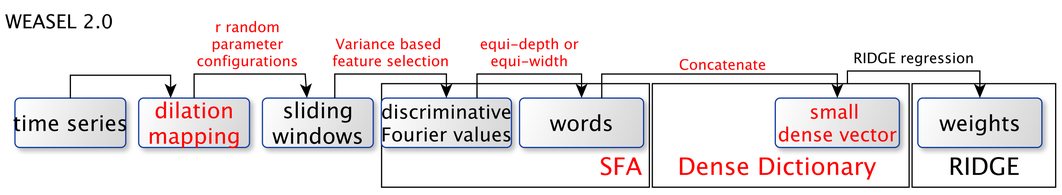
Limitations of Dictionary Models

1. Typically have **huge memory-footprint** (size of feature vector), due to a large variance in generated words
2. Lower accuracy than SotA



WEASEL 2.0 - Building Blocks

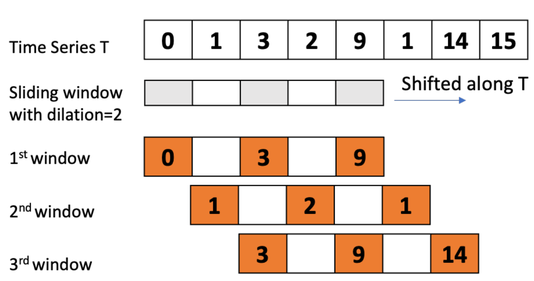
- Apply novel **dilation mapping** prior to downstream processing
- **Constrained memory footprint**
  - Fixed-Size Dictionary: only 128 to 256 distinct words for each dictionary
    - Word length $w \in \{7,8\}$ and alphabet size=2
  - Ensemble over random hyper-parameters to increase variance: ~70k features total
    - window length, dilation factor, word length, binning, first order differences
- Ridge Regression classifier



Dilation

Dilation

- Dilation **increases receptive field** by inserting gaps
  - Similar to **down-sampling**
- But: keep the **total number of values** in each window **constant**



- First used in kernel-based approaches (convolution, Rocket et al.)

WEASEL 2.0

Dilation Mapping

Dilation Mapping

- Novel **dilation mapping** as transformation
- Applied as pre-processing step by three lines of python code

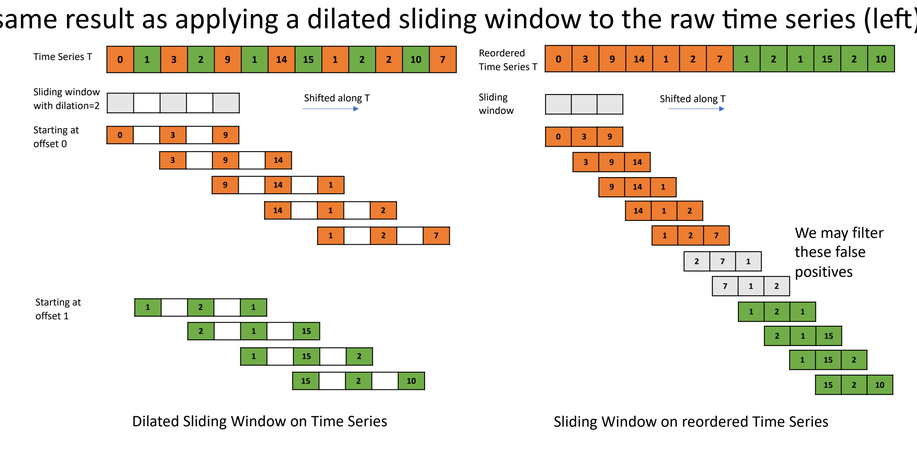
```
def dilation(X, d):
(1)   X^d = X[:, 0::d]
(2)   for i in range(1, d):
(3)       X^d = np.concatenate((X^d, X[:, i::d]), axis=1)
```

- It results in a dilation operation applied by down-stream classification model

Example for d=2

Dilation Mapping – How it Works

- Sliding a sliding window along the transformed time series (right) yields the same result as applying a dilated sliding window to the raw time series (left).

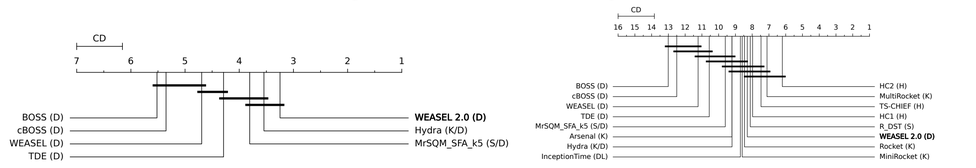


Experimental Evaluation

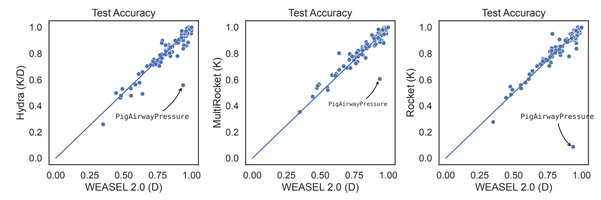
Experimental Evaluation

- Compare WEASEL 2.0 to 15 SotA TSC methods on 114 UCR datasets
  - Used implementations available in aeon/sktime, or published by the authors
- All reported numbers are **test accuracy** on **test split**

Rank on Test Accuracy on 114 UCR problems

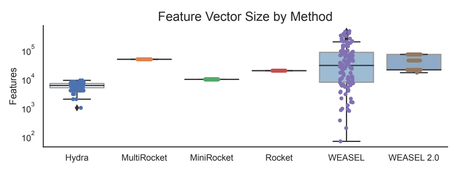


- WEASEL 2.0 is the **most accurate dictionary-based TSC**
- It is **not significantly worse than SotA** in test accuracy
- Most better methods are hybrids (ensembles)
- Best on datasets, for which frequency of patterns is important for classification

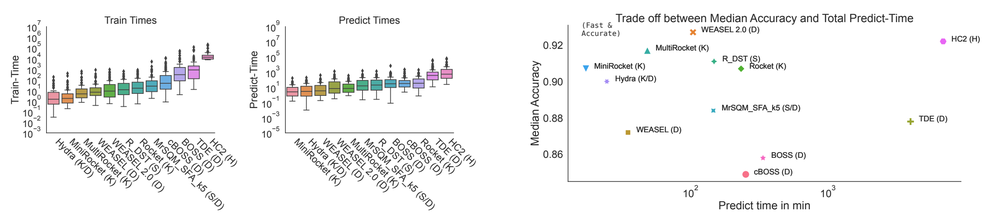


Constrained Memory Footprint (Vector Size)

- WEASEL 2.0 has up to 70k features - comparable to Rocket et al.



Runtime



- WEASEL 2.0 is fast in train and test times, yet, accurate

Conclusion

- Dictionary methods are part of all hybrid state of the art classifiers
- WEASEL 2.0 is a novel TSC method following the dictionary approach
  - Combines randomization over hyper-parameters with dilation
  - Proposes a novel dilation mapping, applicable as pre-processing step
  - It is the most accurate dictionary method
  - Excellent initial choice when datasets contain repetitive, phase-invariant, and noisy patterns