Evaluating Explanation Methods for Multivariate Time Series Classification

Davide Italo Serramazza (davide.serramazza@ucdconnect.ie), Thu Trang Nguyen, Thach Le Nguyen, Georgiana Ifrim @ University College Dublin, Dublin, Ireland

Motivation

MTS classification has received a lot of attention, while explaining the classification has not.
We **compare saliency-based MTSC explanation methods** that provide a **2D heat map** highlighting the importance of each time step in each channel.
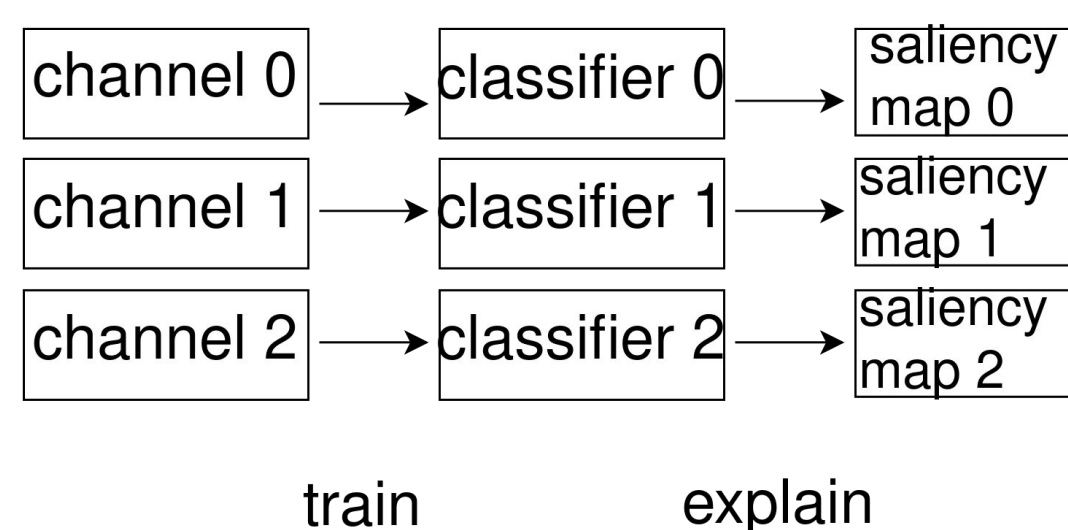
MTSC explanation methods state-of-the-art

- Few **bespoke methods** for explaining MTSC, mostly for deep learning models (dCAM)
- Some adaptations of univariate methods, e.g., SHAP to provide multivariate explanation (SHAP)
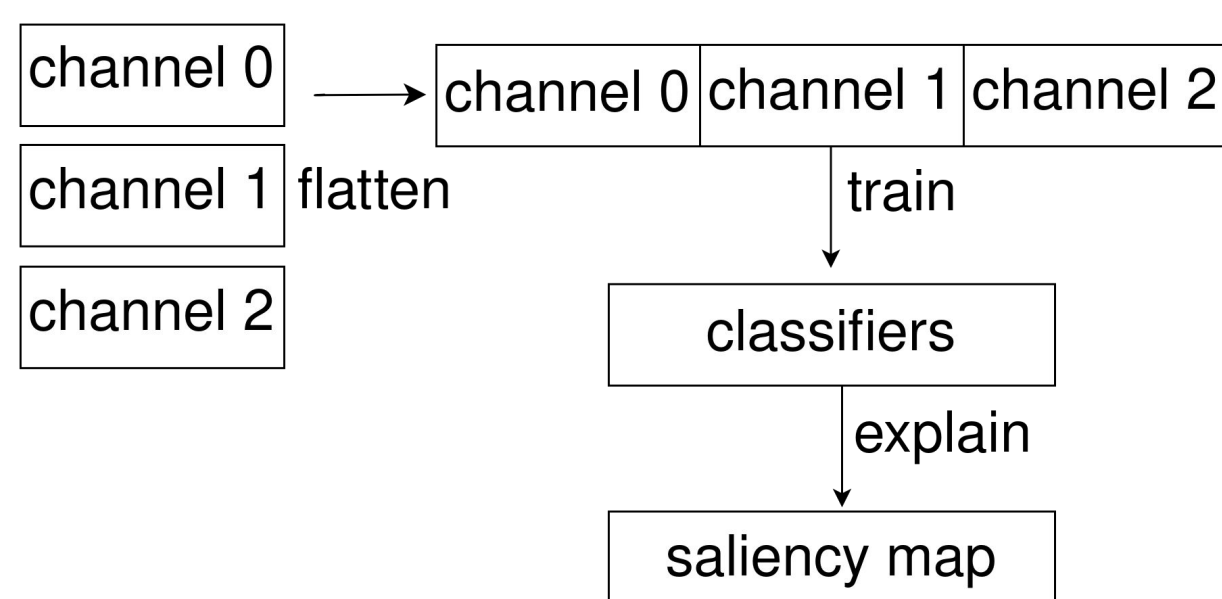
Methods

Classifier + explanation method:
- Rocket explained using SHAP (timeXplain)
  - Channel by channel

  - Concatenated channels

- dResNet explained using dCAM
- Ridge classifier and learned weights (baseline)
- Random explanation (sanity check)

Classifier Accuracy Results

Datasets:
- 3 synthetic (PseudoPeriodic, Gaussian, AutoRegressive); RidgeCV has best results
- 2 real-world (CMJ, MilitaryPress): ROCKET better than dResNET

Classifier/Dataset	PseudoPeriodic	Gaussian	AutoRegressive	CMJ	MilitaryPress
dResNet multivariate	1.0	0.83	0.82	0.82	0.79
dResNet concatenated	1.0	0.89	0.81	0.91	0.68
ROCKET multivariate	1.0	0.93	0.87	0.87	0.87
ROCKET concatenated	1.0	0.72	0.73	0.88	0.83
ROCKET ch-by-ch	0.99	0.72	0.95	0.85	0.65
RidgeCV	1.0	1.0	1.0	0.44	0.61

Explanation Results

For **synthetic datasets** we compare to the ground truth explanation
Findings:
- Ridge has the best explanation results and the best accuracy
- SHAP has better results than dCAM especially for Recall, but higher computational cost

For **real-world datasets (CMJ and MP, table below):**
- We use AMEE to rank and compare explanations (see paper)
- SHAP has better results than dCAM, but higher computational costs

Dataset	XAI method	Explanation Power	Rank	Explanation Time
CMJ-concat	SHAP concat	1.00	1	7.15h
	Random	0.99	2	0s
	dCAM	0.39	3	30s
	SHAP ch-by-ch	0.05	4	7.5h
	Ridge	0.0	5	0s
MP-concat	SHAP concat	1.00	1	24h
	dCAM	0.33	2	15m
	Random	0.07	3	0s
	Ridge	0.0	4	0s
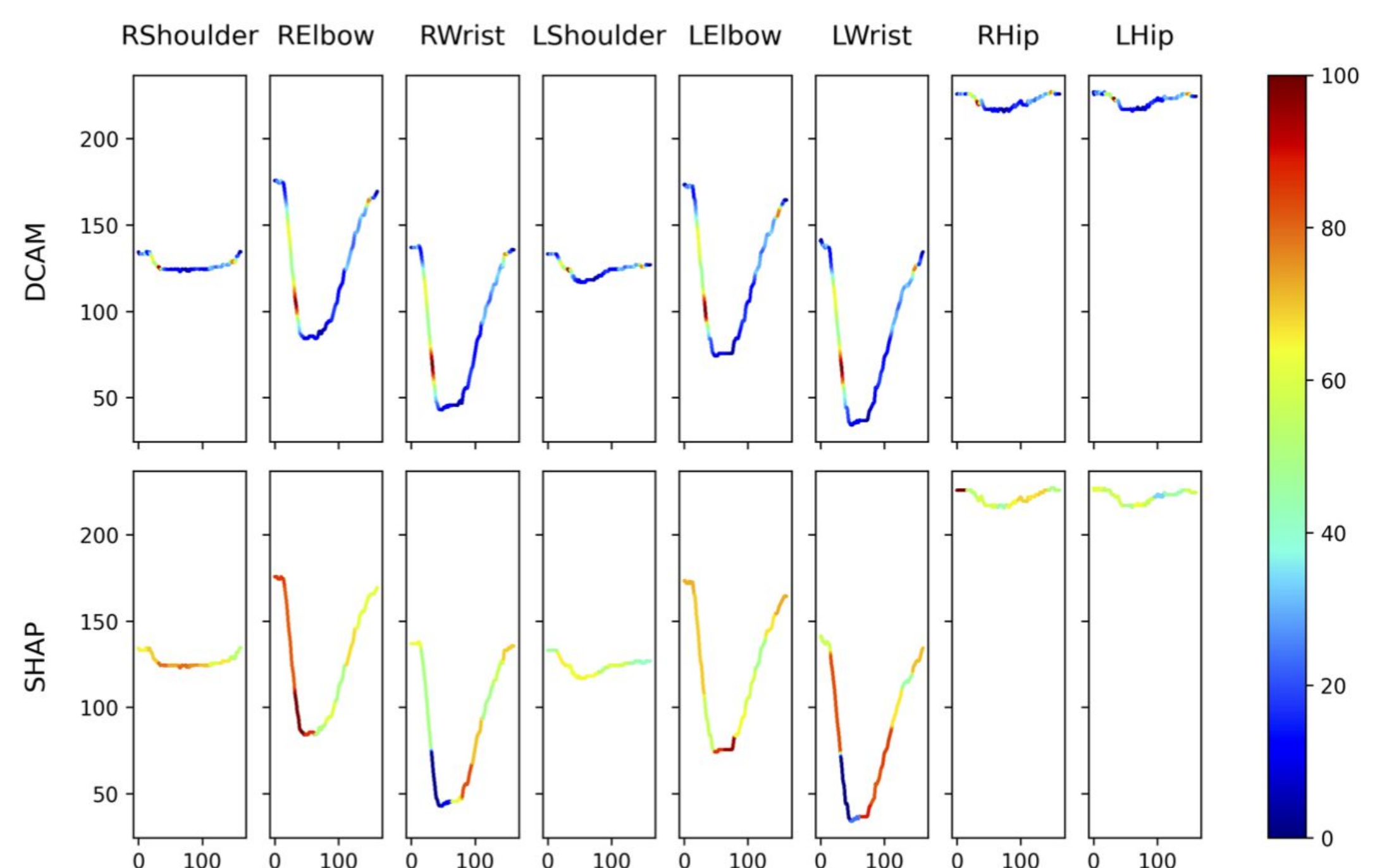
Conclusions and Future Work

SHAP provides better explanations at higher cost

Popular synthetic TSC benchmark not suitable for time series analysis

Classifier accuracy on Concatenated datasets is comparable to that on Multivariate datasets

We plan to study more datasets, explanation methods and evaluation methodologies

https://github.com/mlgig/Evaluating-Explanation-Methods-for-MTSC