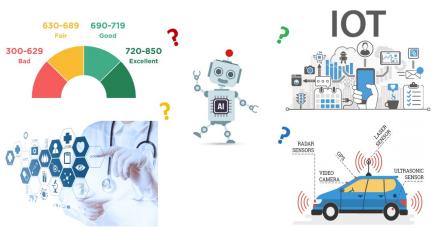


Designing Shapelets for Interpretable Data-Agnostic Classification

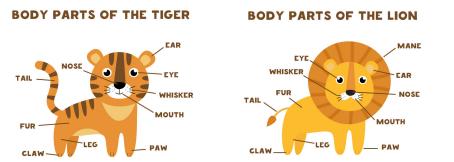
Riccardo Guidotti, Anna Monreale - University of Pisa, Italy

Introduction & Motivations

Explainable Artificial Intelligence (XAI) and Interpretable Machine Learning are widely debated nowadays. This is due to the diffusion of AI systems in many applications for which both the predictive accuracy and the comprehensibility of the system reasoning are important.



If the AI is interpretable as human we trust more a decision process using a logic similar to that one of a human being, rather than a reasoning that we can understand but that is outside the human way of thinking. For instance, we recognize a cat in an image by the presence of a tail, pointed ears, and mustaches, not from pixels having certain values.



Shapelets

Shapelets are time series subsequences particularly suitable for separating instances of different classes.

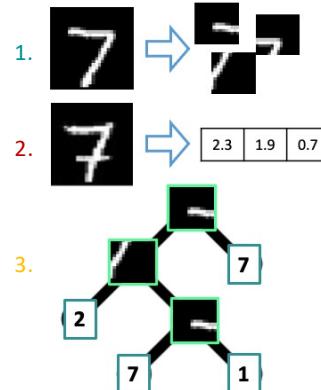
We extend the notion of shapelets to

- Images
- Text
- Tabular data

DASH

1. Extracts shapelets from data prototypes obtained with a clustering process
2. Represent data as distances from shapelets
3. Trains an interpretable classifier on the distances from shapelets

The classification model returned by DASH can be used as an interpretable AI system.



Results

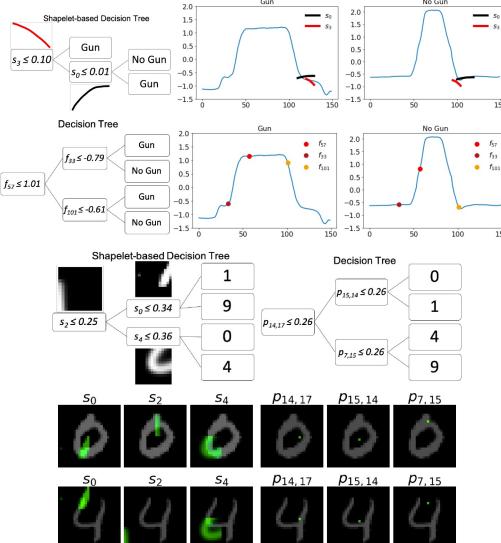


Table 2: Running time (in seconds) for shapelet discovery.

dataset	DASH	GDBD	BPA	RND
arrowhead	5.271*	4.01	6.223	5.54
ecg200	8.00	6.032	22.319	1.435
electric	25.95*	8.535	8.534	3.329*
gunpoint	7.17*	8.543	29.825	4.683
italpower	0.991	5.038	3.630	2.511*
phalanges	11.923	23.513	3184.247	278.733

Table 3: Shapelet instability as mean and standard deviation of distances across ten shapelet discovery runs.

dataset	DASH	GDBD	DT	NN	LR	RND	NOSH
arrowhead	387*	410	623	554	385*	409	380
ecg200	800	740	780	204	200*	520	720
electric	540*	548	484	444*	479	765	549
gunpoint	820	808	808	808	808*	770	640
italpower	112	768	851	662	750	800*	747
phalanges	603*	640	705	683	935	701	561

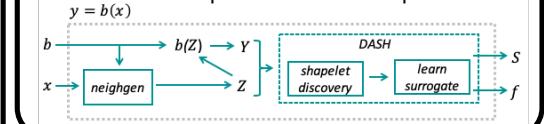
Table 4: Accuracy for different interpretable classifiers and shapelets-based methods on images.

dataset	classes	DASH	RND	NOSH	DASH	RND	NOSH	DT	NN	LR	RND	NOSH	Time
cifar10	0, 1, 2, 3	189*	171	218	149	303	212	266*	137	154.482	240.917	131	1.17
cifar10	0, 1	762	733	745*	695	557	607*	742	735*	708	1.169	47.412	-
fashion	0, 1, 2, 3	721	725	873	771	611	711*	744	732	741	62.668	168.777	27.77
mnist	0, 1	991	910	996	965	828	978	990	910	999	18.486	68.276	10.558

Table 5: Accuracy and running time for different interpretable classifiers and shapelets-based methods on images.

Shapelet-based Explanations

A local shapelet-based explainer takes as input the black box b and the instance to explain x . First, it generates the neighborhood Z with $neighgen$. Then, it labels the synthetic instances with the black box $Y = b(Z)$. Finally, a DASH classifier is used as surrogate model and the shapelets are used as explanation.



Summary & Future Works

DASH is an interpretable data-agnostic classification approach based on shapelets defined on time series, images, texts and tabular data that exploits prototypes with a clustering process to speed-up and stabilize the shapelet computation. The decision process based on human-understandable parts can be easily accepted from humans.

Several research directions can be mentioned as future research directions. First, extension for the classification method of alternative types of data like item sequences, mobility trajectories, genomics sequences, etc. Second, further speed up.

Acknowledgements

This work has been partially supported by the European Community H2020 programme under the funding schemes: G.A. 871042 SoBigData++, G.A. 952026 HumanE AI Net, G.A. 825619 AI4EU, and G.A. 952215 TAILOR.