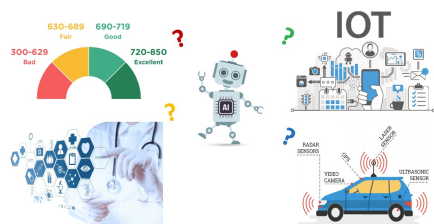


Designing Shapelets for Interpretable Data-Agnostic Classification

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.

BODY PARTS OF THE TIGER



BODY PARTS OF THE LION



Shapelets

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

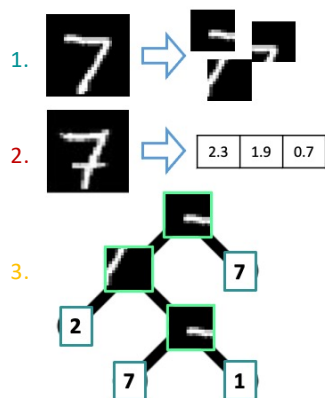
We extend the notion of shapelets to

- 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

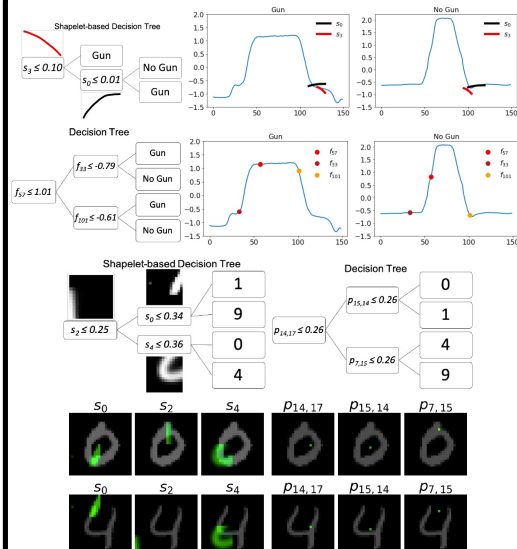


Table 2: Running time (in seconds) for shapelet discovery.

dataset	DASH	GDBo	BFA	RND
arrowhead	5.271*	8.121	25.835	1.435
ecp200	1.907	6.032	23.319	3.739*
electric	255.735*	85.784	> 1h	> 1h
gunpoint	7.177*	5.543*	200.825	4.683*
italypower	0.301	5.035	3.630	2.511*
phalanges	11.923	23.513*	3184.247	278.733

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

dataset	DASH	GDBo	BFA	RND
ecp200	0.044 ± 0.025	0.038 ± 0.029	0 ± 0	0.285 ± 0.233
gunpoint	0.006 ± 0.010	0.255 ± 0.260	0 ± 0	0.114 ± 0.175
italypower	0.033 ± 0.113	0.886 ± 0.002	0 ± 0	0.249 ± 0.273
phalanges	0.023 ± 0.055	0.183 ± 0.182	0 ± 0	0.101 ± 0.128

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

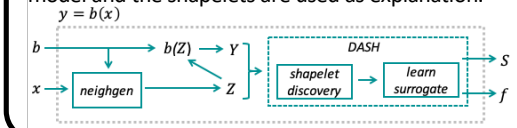
dataset	DT				INN				LR						
	DASH	GDBo	BFA	RND	DASH	GDBo	BFA	RND	DASH	GDBo	BFA	RND			
arrowhead	583*	474	623	554	585	446*	423	497	400	800	2600*	634	589	497	542
ecp200	800	640	740	780	804	700*	560	720	720	880	770	640	800*	800	720
electric	540*	548	588	484	444*	479	-	-	-	549	484*	513	1109	474	239
gunpoint	807*	820	800	780	815	867	513	793*	533	913	853	773	847*	747	566
italypower	803*	641	812	769	951	813	662	550	800*	955	791	561	764	766*	741
phalanges	693*	640	705	683	693	394*	519	387	614	761	615*	613	638	600	547

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

dataset	classes	DT				INN				LR				Time
		DASH	RND	NOSSH	NOSSH	DASH	RND	NOSSH	NOSSH	DASH	RND	NOSSH	DASH	
cifar10	all	180*	171*	218	218	153*	169	303	232	260*	131	154.083	2493.121	
	0,1,2,3	445	441*	425	425	395	296	301*	405*	435	401	93.327	217.718	
fashion	all	511	580	645	645	368	333	850	741	732	461	62.688	164.277	
	0,1,2,3	271*	271*	271*	271*	271	411	1140*	447*	591	92*	21.204	278.394	
mnist	all	431	378	596	596	236	240*	969	788	799	425	176.084	156.023	
	0,1,2,3	710*	703	911	911	623*	514	743	706*	761	978	27.483	311.008	

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.