Time Series Shapelets

Lexiang Ye and Eamonn Keogh

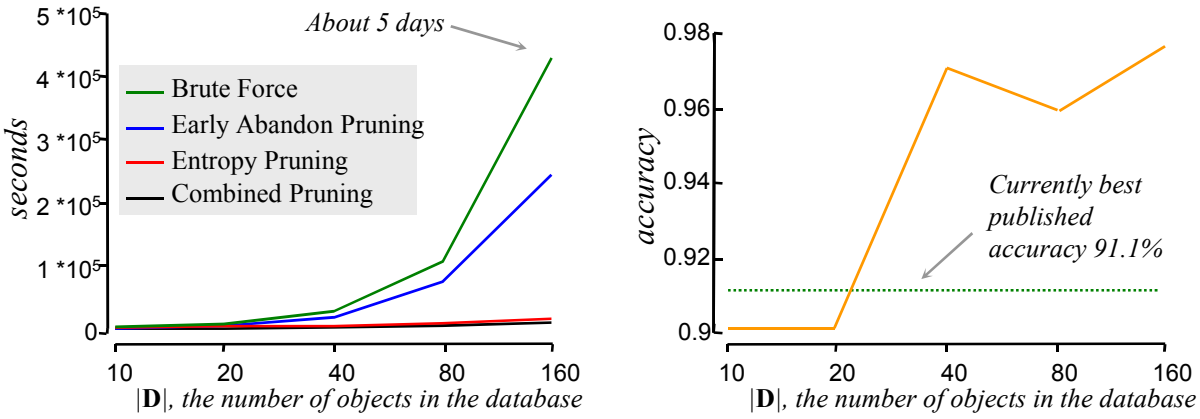
This file contains augmented versions of the figures in our paper, plus additional experiments and details that were omitted due to space limitations

Please recall that you can visit our webpage for even more information, including *all* code and *all* datasets.

www.cs.ucr.edu/~lexiangy/shapelet.html Until the paper is accepted, the webpage is password-protected with the username/password: *kdd* / *riverside*

Recall that there are many tools to allow anonymous surfing, we suggest you Google “anonymous surfing”

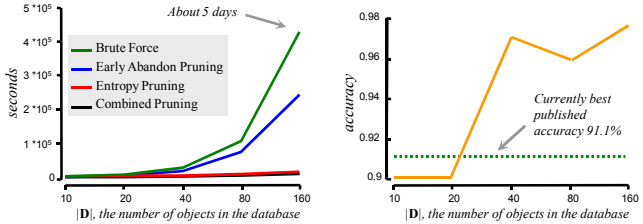
Scalability Test Experiment: Part 1



The dataset came from Jeffery, C. (2005). Synthetic Lightning EMP Data. http://public.lanl.gov/eads/datasets/emp/index.html Los Alamos National Laboratory.

This dataset is also mentioned in http://www.cs.ucr.edu/~eamonn/iSAX/iSAX_weighted.ppt

Figure in Paper

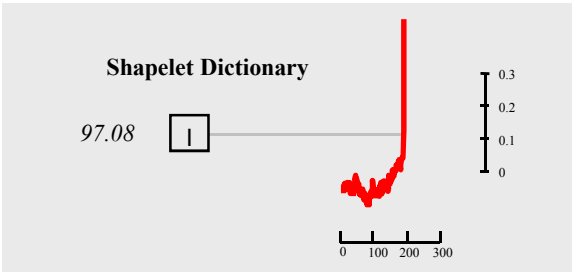


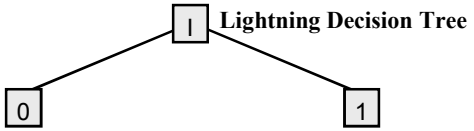
Raw Numbers

Size	Full Speedup	Distance Early Abandon	Entropy Prune	Brute Force	Accuracy
10	499	1108	739	1664	0.9036
20	1062	4524	1571	6662	0.9036
40	2236	18362	3241	26736	0.9688
80	4356	74360	6324	107000	0.958
160	8868	244900	12932	429000	0.9718

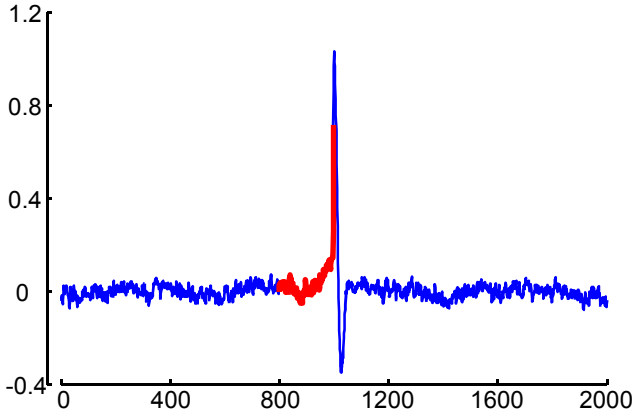
Scalability Test Experiment: Part 2

Here is the decision tree learned on this problem

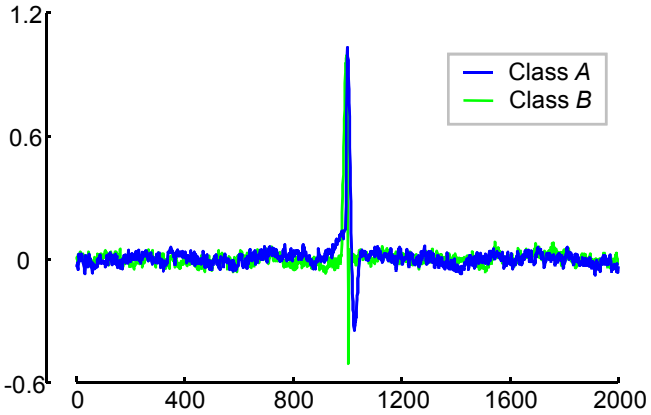




From the shapelet, we can see the most distinguishing feature for time series objects from two classes is: the lightning in class *A* has a shallow slope followed by a steep one; the lightning part in class B has a directly steep slope.

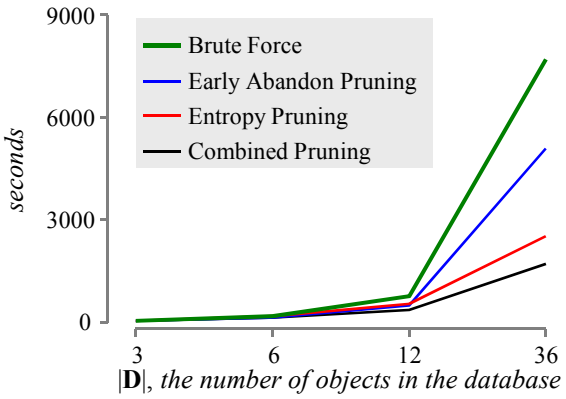


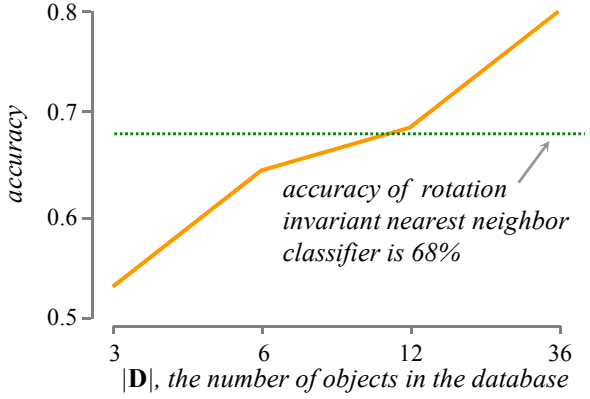
Shapelet in the time series object



Comparison of time series objects from two classes

Projectile Points Experiment: Part 1





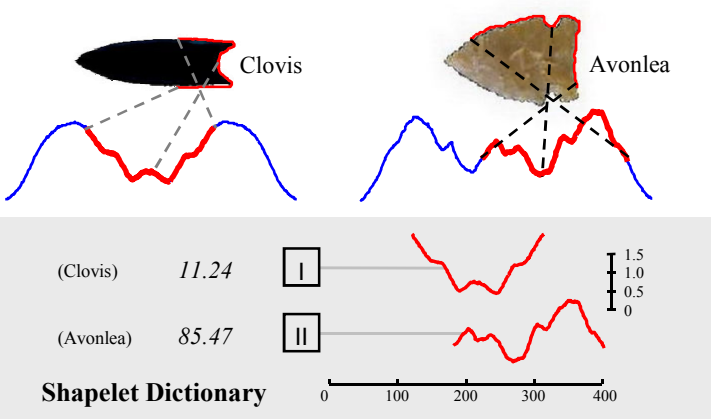
The dataset can be downloaded from the website: http://www.cs.ucr.edu/~lexiangy/shapelet.html

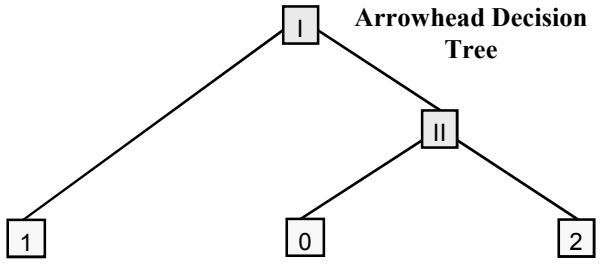
Raw Numbers

Size	Combined Pruning	Distance Early Abandon	Entropy Prune	Brute Force	Accuracy
3	29.81	29.95	49.28	49.51	0.5257
6	114.78	118.85	184.86	193.75	0.64
12	352	486.78	541.37	779.54	0.6828
36	1722	5100.94	2509.74	7708.42	0.8

Projectile Points Experiment: Part 2

Here is the decision tree learned on this problem





the ***Clovis*** projectile points can be distinguished from the others by an un-notched hafting area near the bottom connected by a deep concave bottom end.
After distinguishing the Clovis projectile points, the ***Avonlea*** points are differentiated from the mixed class by a small notched hafting area connected by a shallow concave bottom end.

trainingc		36	
test		175	
length		≈ 450*	
shapelet	**length**	**Index**	**start pos**
I	194	35	123
II	221	25	184

*: the length of the arrowheads varies, since it is not normalization by lengths.

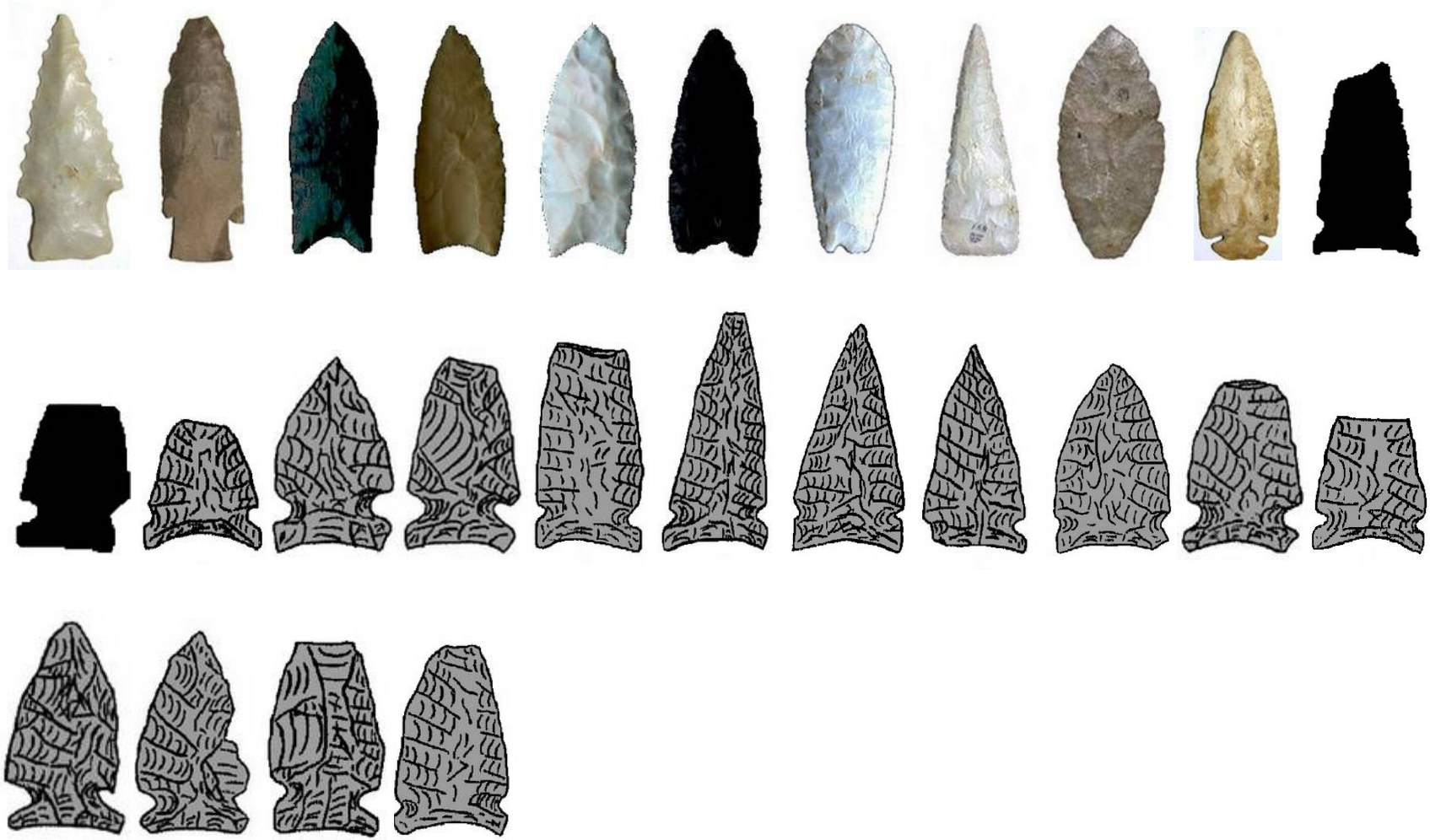
**: the corresponding shapelet can be gotten using the index and start pos in the training dataset. Index starts from 1.

Projectile Points Experiment: Part 3

Method	Break Ties Specify Dist. Measure	Accuracy	Classification time
Shapelet	Longest	0.749	
	Shortest	0.469	
	Max Separation	0.721	
	Max Mean Separation	0.800	0.332 sec.
Nearest Neighbor Rotation invariant	Euclidean Dist	0. 680	1013 sec.
	DTW		

Projectile Points Experiment: Part 4

Here are the projectile points that the ***shapelet decision tree classifier*** classifies correctly and the ***rotation-invariant nearest neighbor classifier*** does not.



Projectile Points Experiment Image process

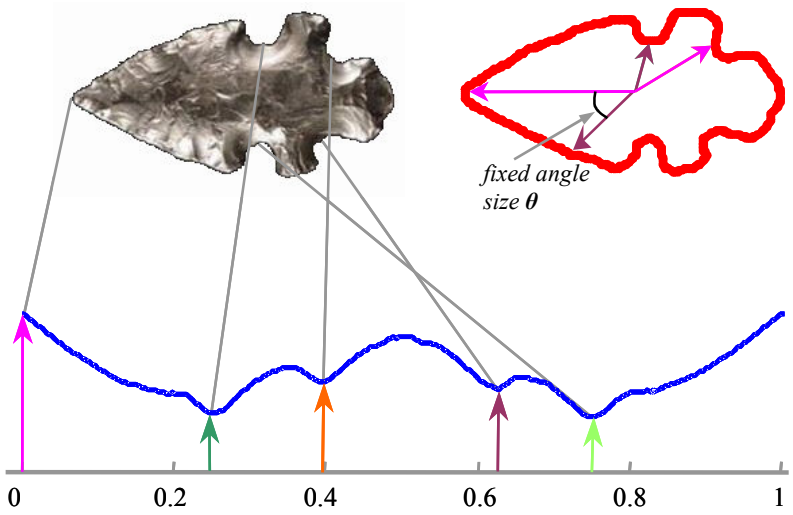
Adjust the images to approximately same size

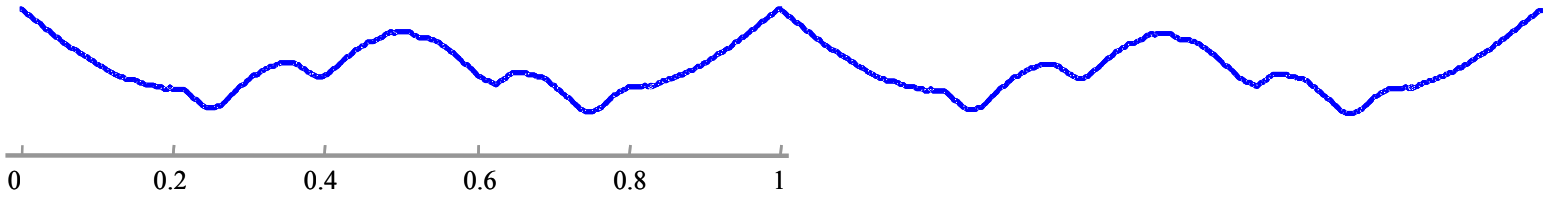
Convert to time series using the angle-based method

Concatenate a copy of the time series to itself. In this way, we make it ***rotation invariant***

Why not normalize the length?

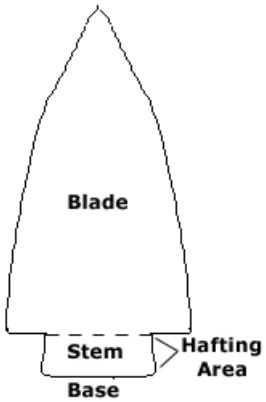
Normalized to the same length makes the similar part of outline of different length very different → unmatched



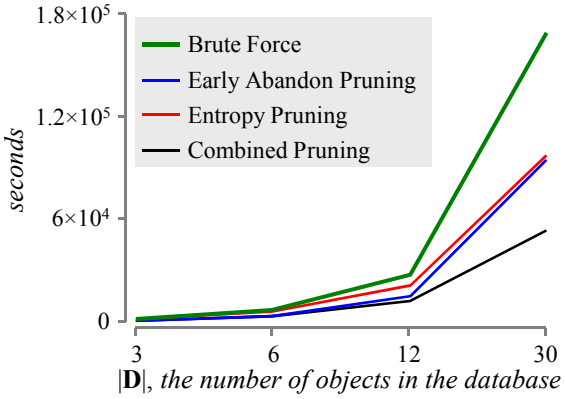


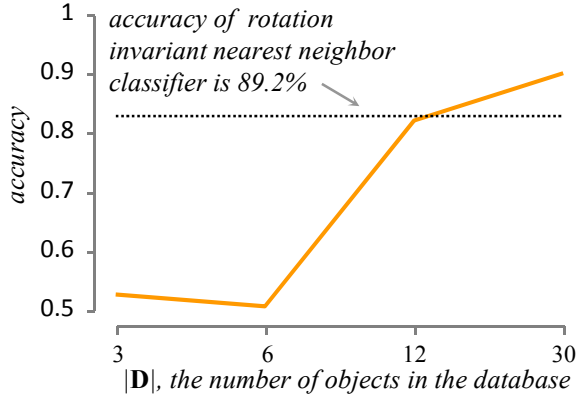
Projectile Points Reference

- In the paper we considered the problem of projectile point (arrowhead) classification. Space limitations prevented an extensive discussion of the topic, a detailed review of the literature on projectile point classification can be downloaded from a link on our webpage.
- More detailed information is from http://www.museum.state.il.us/ismdepts/anthro/proj_point/points_glossary.html



Mining Historical Documents Experiment: Part 1





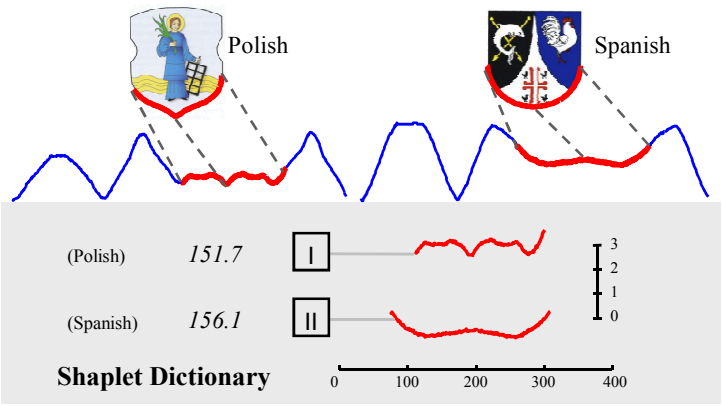
The dataset can be downloaded from the website: http://www.cs.ucr.edu/~lexiangy/shapelet.html

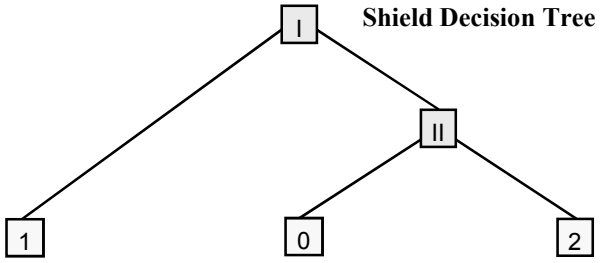
Raw Numbers

Size	Combined Pruning	Distance Early Abandon	Entropy Prune	Brute Force	Accuracy
3	856.84	857.66	1554.69	1551.42	0.5271
6	3653.53	3793.27	6580.92	6799.02	0.5039
12	12292.2	15697.6	21944.8	27737	0.8217
30	54359.6	95654.17	97823.7	170512.5	0.8992

Mining Historical Documents Experiment: Part 2

Here is the decision tree learned on this problem





training / testing: 30 / 129

From the shapelets, we can see the most distinguishing features for time series objects for three classes is:
The ***Polish*** examples are distinguished by a curved edge which has a sharp angle at the very bottom;
the unique semi-circular bottom of the ***Spanish*** crest is used in node II to discriminate it from the French examples.

trainingc		30	
test		129	
length		≈ 1100	
Shapelet**	**length**	**Index**	**start pos**
I	164	11	267
II	234	12	275

*: the length of the shields varies, since it is not normalization by lengths

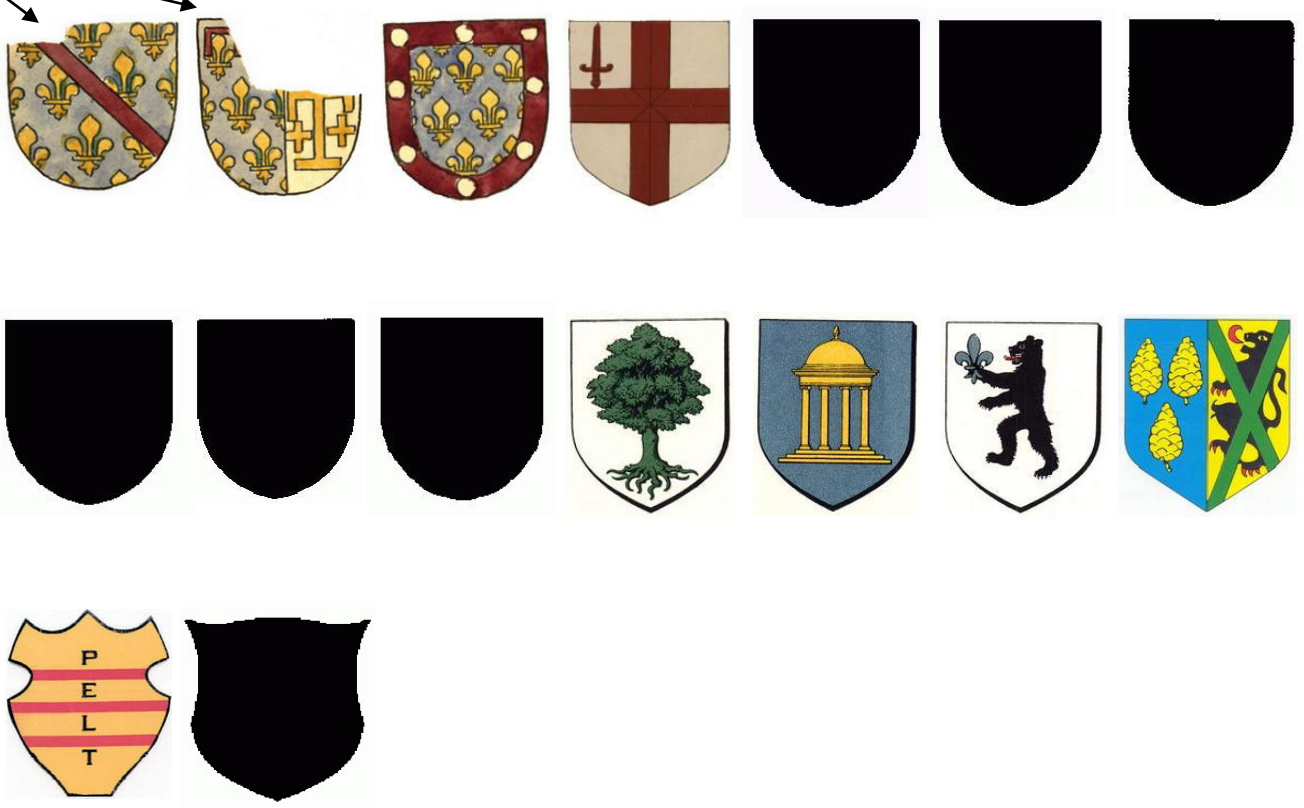
**: the corresponding shapelet can be gotten using the index and start pos in the training dataset. Index starts from 1.

Mining Historical Documents Experiment: Part 3

Method	Break Ties Specify Dist. Measure	Accuracy	Classification time
Shapelet	Longest	0.767	
	Shortest	0.698	
	Max Separation	0.860	
	Max Mean Separation	0.899	0.480 sec.
Nearest Neighbor Rotation invariant	Euclidean Dist	0.829	15600 sec.
	DTW		

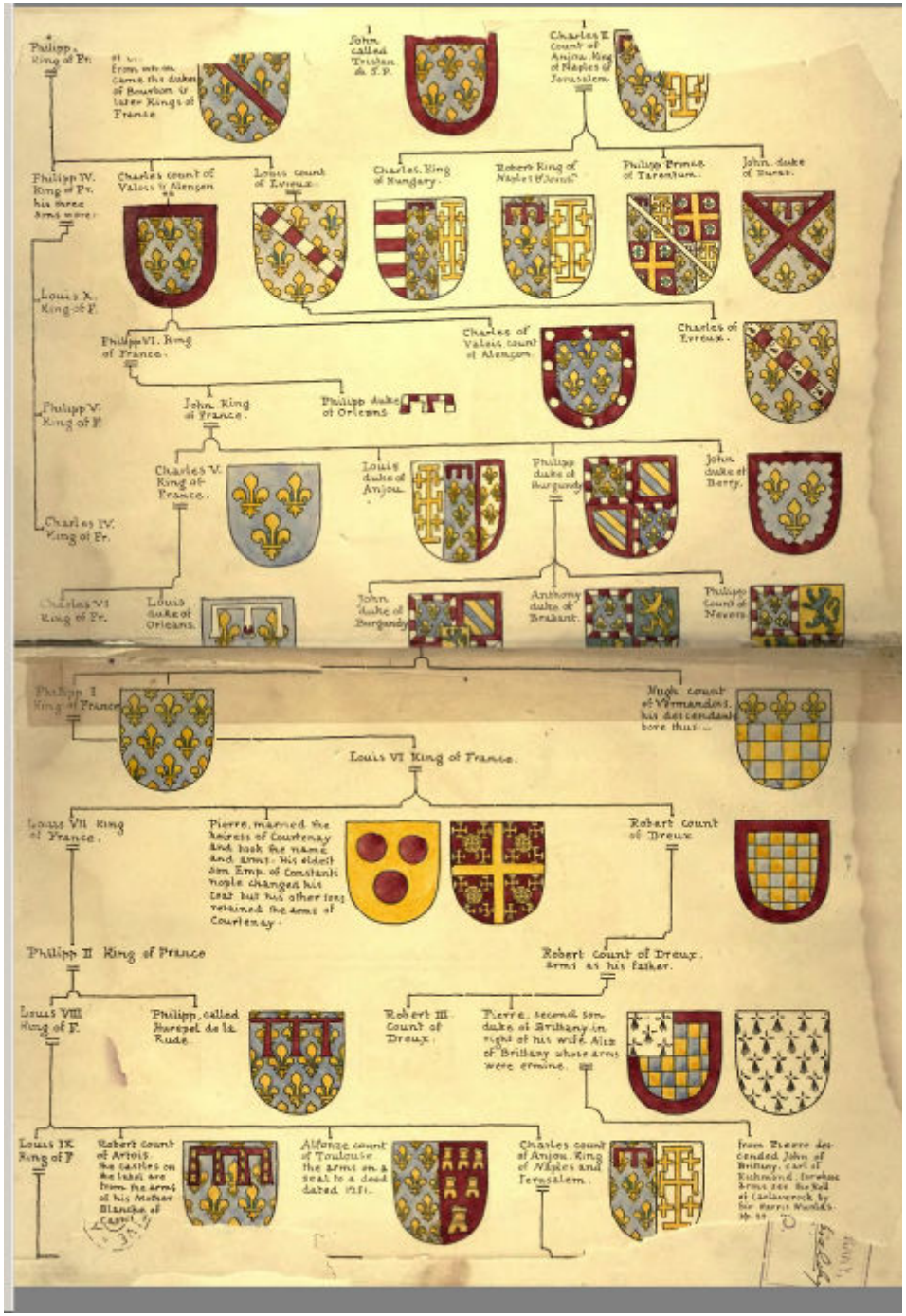
Mining Historical Documents Experiment: Part 4

Here is the shields that the ***shapelet decision tree classifier*** classifies correctly and the ***rotation-invariant nearest neighbor classifier*** does not. Note it can handle broken shapes.



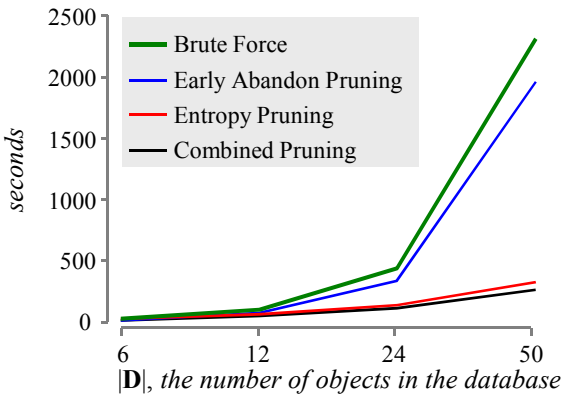
Historical Documents Reference

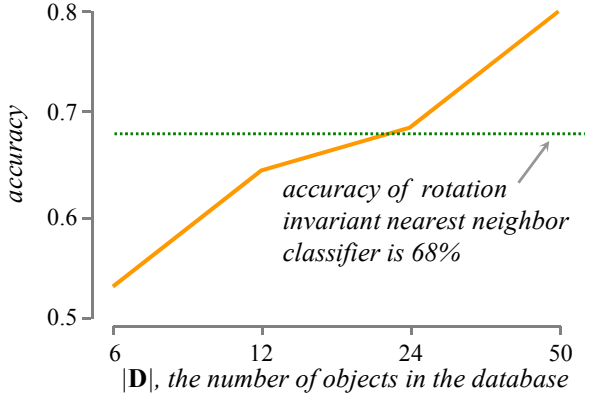
- Here is the full version of the cropped page from [a] which appears in the paper



- [a] Montagu, J.A. (1840). A guide to the study of heraldry. Publisher: London : W. Pickering. Online version www.archive.org/details/guidetostudyofhe00montuoft

Gun / NoGun Experiment: Part 1





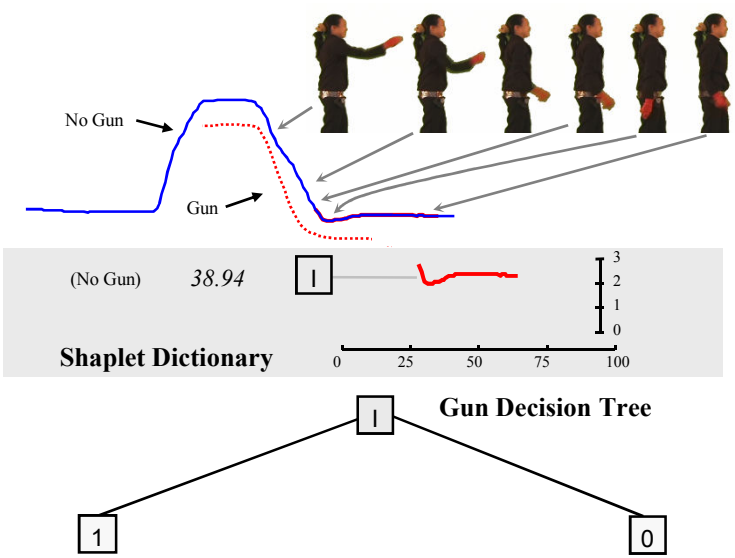
The dataset can be downloaded from the website: http://www.cs.ucr.edu/~lexiangy/shapelet.html

Raw Numbers

Size	Combined Pruning	Distance Early Abandon	Entropy Prune	Brute Force	Accuracy
3	29.81	29.95	49.28	49.51	0.5257
6	114.78	118.85	184.86	193.75	0.64
12	352	486.78	541.37	779.54	0.6828
36	1722	5100.94	2509.74	7708.42	0.8

Gun / NoGun Experiment: Part 2

Here is the decision tree learned on this problem



training / testing: 50 / 150

the *NoGun* class has a "dip" where the actor put her hand down by her side, and inertia carries her hand a little too far and she is forced to correct for it (a phenomenon known as "***overshoot***"). In contrast, when the actor *has the gun*, she returns her hand to her side more carefully, feeling for the gun holster, and no dip is seen.

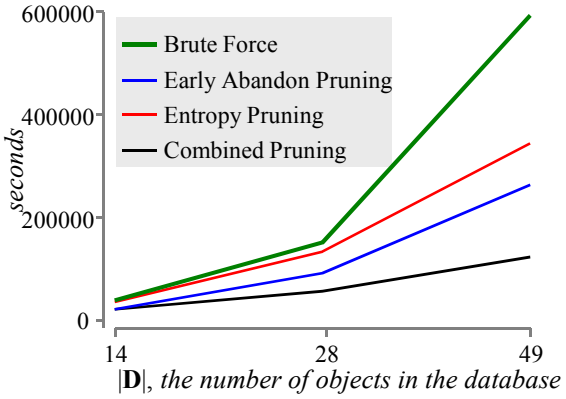
trainingc		50	
test		150	
length		150	
Shapelet*	**length**	**Index**	**start pos**
I	38	45	109

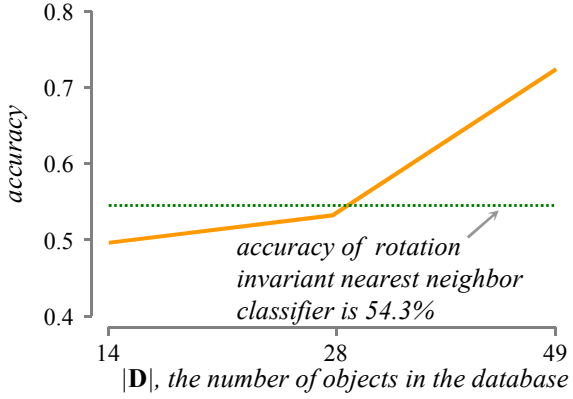
*: the corresponding shapelet can be gotten using the index and start pos in the training dataset. Index starts from 1.

Gun / NoGun Experiment: Part 3

Method	Break Ties Specify Dist. Measure	Accuracy	Classification time
Shapelet	Longest	0.933	0.016 sec.
	Shortest	0.609	
	Max Separation	0.933	
	Max Mean Separation	0.893	
Nearest Neighbor	Euclidean Dist	0.913	0.064 sec.
	DTW best warping window	0.913	
	DTW no warping window	0.907	

Wheat Spectrography Experiment: Part 1





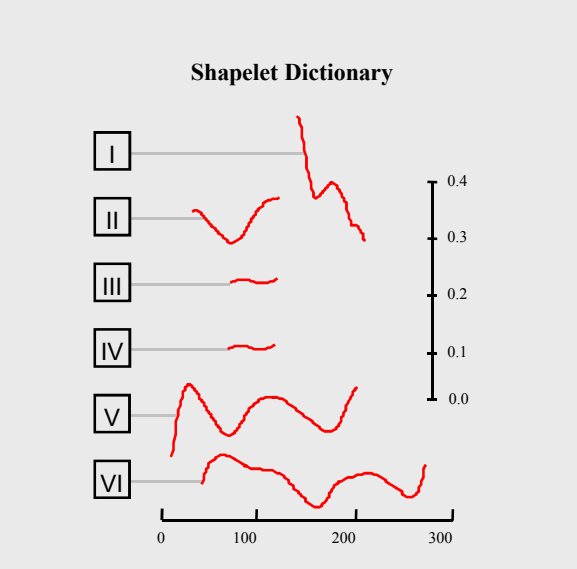
The dataset can be downloaded from the website: http://www.cs.ucr.edu/~lexiangy/shapelet.html

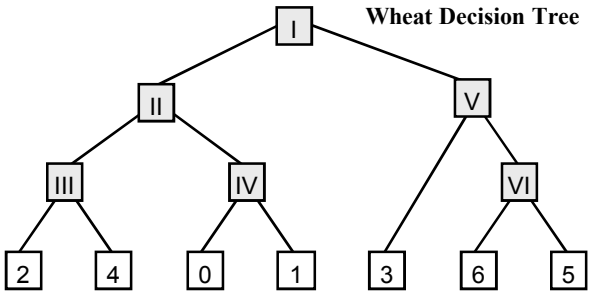
Raw Numbers

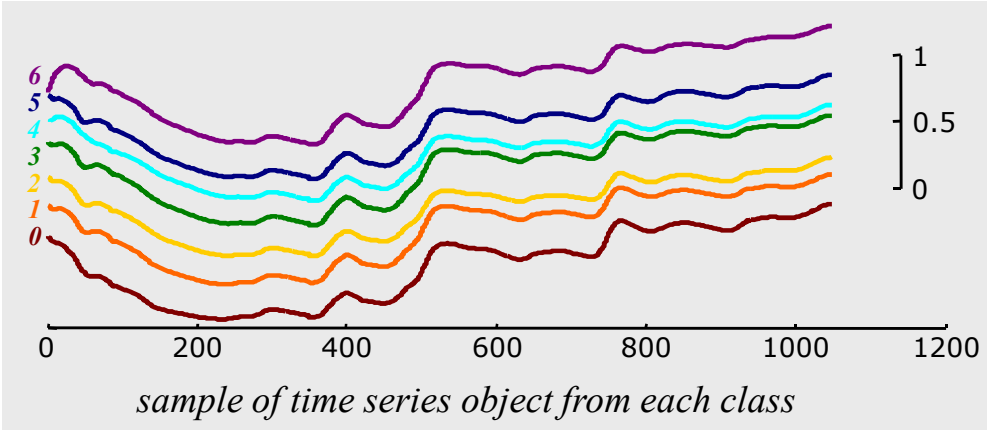
Size	Combined Pruning	Distance Early Abandon	Entropy Prune	Brute Force	Accuracy
14	20100.96	22326.3	35678.2	37345.1	0.497
28	55716.31	90200.7	132786.8	149992.1	0.533
49	122456	264211.5	343211.7	789533	0.720

Wheat Spectrography Experiment: Part 2

Here is the decision tree learned on this problem







sample of time series object from each class

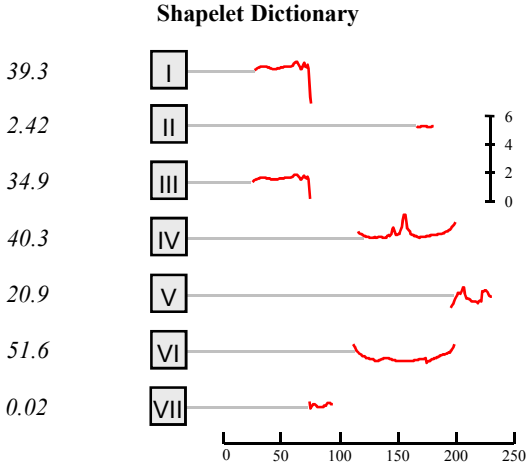
trainingc		49	
test		276	
length		1050	
Shapelet	**length**	**Index**	**start pos**
I	100	21	7
II	80	47	764
III	50	33	958
IV	50	8	958
V	20	7	59
VI	260	42	496

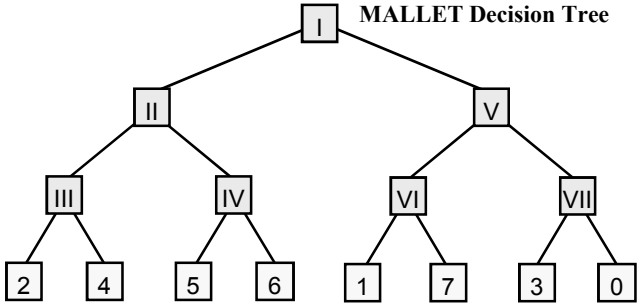
Wheat Spectrography Experiment: Part 3

Method	Break Ties Specify Dist. Measure	Accuracy	Classification time
Shapelet	Longest	0.530	
	Shortest	0.547	
	Max Separation	0.720	1.428 sec.
	Max Mean Separation	0.707	
Nearest Neighbor	Euclidean Dist	0.543	0.36 sec.

MALLET Experiment: Part 1

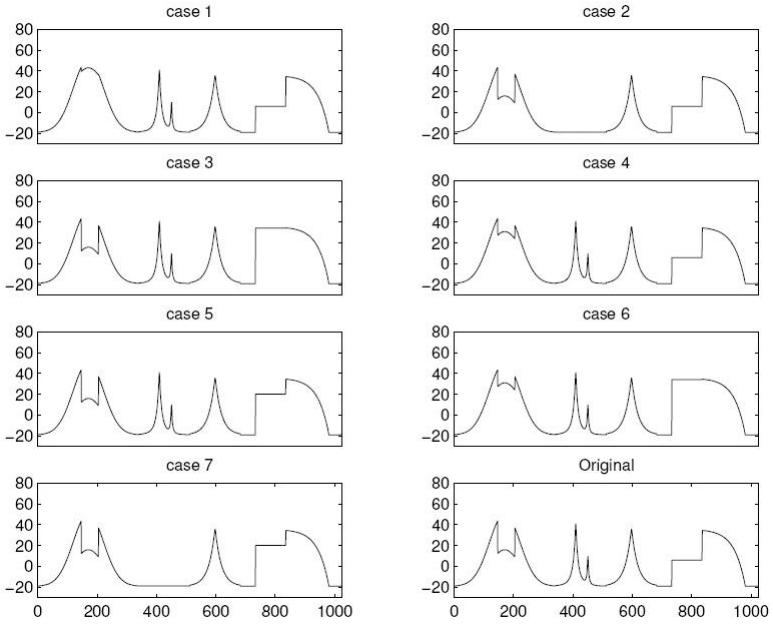
Here is the decision tree learned on this problem





The ***first*** shapelet covers the first rectangular dip. All the data objects on the left has either no or shallow dip, while objects on the right of the tree has deeper dip. Similar on the ***third*** shapelet The ***fourth*** shapelet shows that the most critical difference between the objects in class **5** and **6** is whether the second and the third peaks are present. Similar difference class **1** and **7** are interpreted by the ***sixth*** shapelet. The ***fifth*** shapelet presents the difference in the last peak. The objects on the left have a deep dip and those on right have no or shallow dip.

data sample from each class*



**** The figures are in the reverse order of time series data.***

trainingc		80	
test		2320	
length		256	
Shapelet	**length**	**Index**	**start pos**
I	49	75	27
II	15	13	165
III	50	27	25
IV	84	14	115
V	36	58	194
VI	87	18	111
VII	21	28	74

MALLET Experiment: Part 2

Method	Break Ties Specify Dist. Measure	Accuracy	Classification time
Shapelet	Longest	0.909	
	Shortest	0.740	
	Max Separation	0.945	
	Max Mean Separation	0.947	
Nearest Neighbor	Euclidean Dist	0.959	
	DTW best warping window	0.963	
	DTW no warping window	0.951	