CASCC: A Fast Algorithm for Time Series Classification

Damian R. Eads[1,2], James Theiler[1], David Helmbold[2], Ethan L. Miller[2]

[1] Intelligent Searching of Images and Signals (ISIS) Team

Space and Remote Sensing Sciences Group, ISR-2, LANL

[2] Department of Computer Science, UCSC

CASCC is a new algorithm for classifying time series. It is highly competitive in terms of speed and accuracy compared to many other algorithms. It is inspired by another leading algorithm DTW-1NN however does not suffer the same computational limitations when applying the model to new time series.

MOTIVATION

The continual, real-time collection of time series signals makes manual analysis impractical demanding faster and more accurate analysis algorithms. We are developing algorithms for rapid clustering and classification of time series data sets.

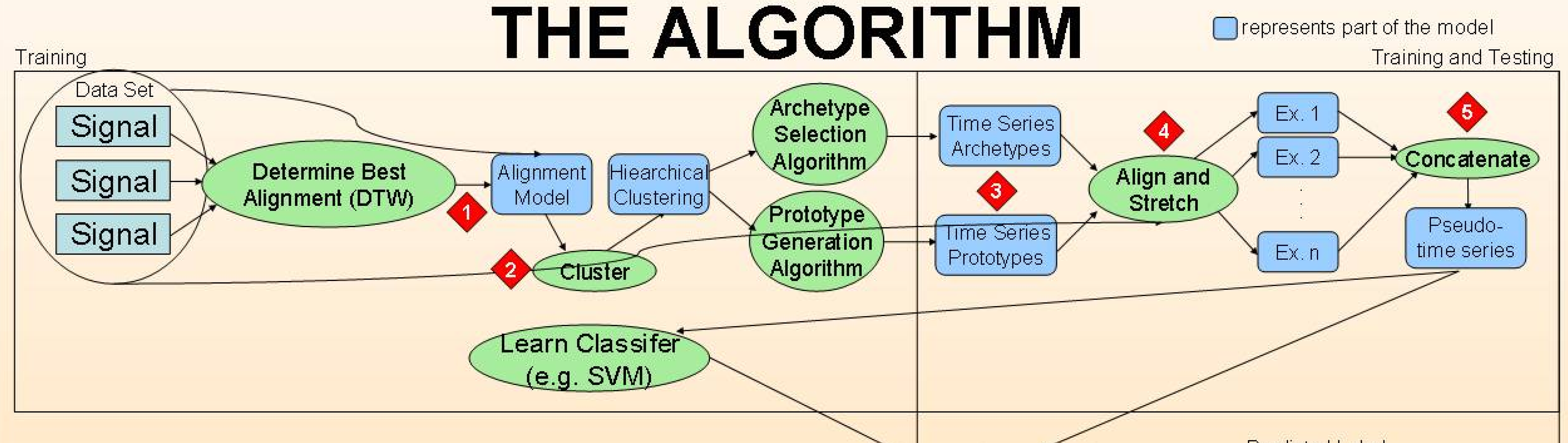
EXAMPLE TASKS

- Characterizing lightning strikes in signals acquired from the FORTE satellite.
- Identifying behavior of sea animals using data acquired from their bodies.
- Recognizing spoken words from audio signals.
- Identifying anomalous events in astronomical data.

BASIC IDEA

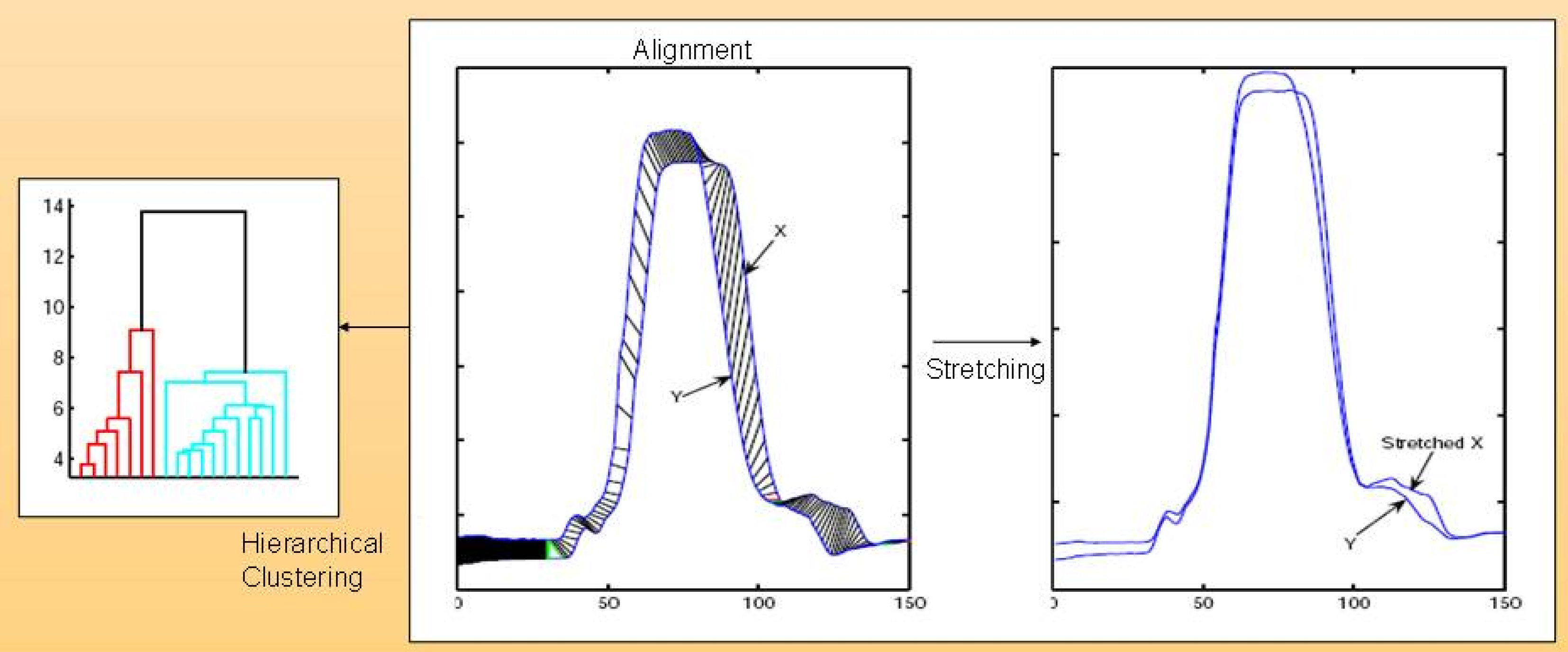
1. We generate prototypes for each class or large cluster so that each prototype resembles other time series in that class or cluster but not time series outside.
2. If we align a time series outside a prototype's class, there will be poor alignment and stretching. Otherwise the shape-based features will align very well. Furthermore, the features will be registered so we can forget about the pain of dealing with offset shift transformations.
3. Many conventional classifiers can handle properly registered features. The goodness of alignment+stretching can be used as discriminating features for the conventional classifier!
4. Big win: model application is independent of the number of training points, unlike the algorithm upon which this approach is based DTW-1NN.

THE ALGORITHM



1. Align shaped-based features of training time series using the Dynamic Time Warping (DTW) algorithm. Then, compute distances between each time series to produce a distance matrix used for clustering.
2. Use these computed distances to cluster time series (usually hierarchically).
3. Generate a set of prototype time series that roughly characterize specific clusters (or classes). Also, select training time series that roughly characterize major clusters. We have developed several algorithms for doing this.
4. Align and stretch each training time series against the prototype or archetype. The approach used here is chosen through cross validation.
5. Concatenate each of these results into a pseudo-time series for (6) input into a conventional classifier, such as an SVM.

Let's illustrate some of this!



EXPERIMENTS

The approach was tested against several other algorithms: baseline classifiers on raw time series, DTW-1NN, AdaBoost using Decision Stumps on DTW distances, and Continuous Hidden Markov Models (using Left-to-Right (LR), strict LR, and arbitrary transition matrices). We performed 10-fold cross validation 10 times, using the same folds for each of the methods tried. The average cross-validation error and standard deviation are reported. The best method for each experiment is shown in **bold**. The total computation for all the experiments is roughly 100 days on a 2.8 Ghz processor. Several prototype and archetype algorithms (RODS, CLTR, MADS, KC) and conventional classifiers (Linear SVM, SVM with RBF kernel) were tried. All parameters were adjusted through cross-validation.

		Data Sets				
	Method	FORTE-2	FORTE-7	GUN	SYN	LEAF
Raw	LSVM	29.20 ± 1.40	38.38 ± 3.10	6.06 ± 0.28	4.85 ± 0.21	49.60 ± 0.70
	RSVM	27.36 ± 1.67	**35.12 ± 1.27**	**2.00 ± 0.53**	**1.61 ± 0.18**	**28.59 ± 0.84**
	1NN	**25.62 ± 1.35**	35.66 ± 0.81	5.39 ± 0.21	7.89 ± 0.19	33.54 ± 0.42
DTW	1NN-DIST	**11.39 ± 1.15**	**22.38 ± 1.74**	**1.39 ± 0.46**	**0.39 ± 0.16**	**8.54 ± 0.33**
	ADA.M1 DSTUMP	20.20 ± 2.62	N/A	4.67 ± 0.53	N/A	N/A
CASCC	RODS-LSVM	25.44 ± 2.02	27.43 ± 2.28	3.61 ± 0.77	1.48 ± 0.31	22.12 ± 0.59
	RODS-RSVM	25.62 ± 3.37	26.73 ± 0.98	2.33 ± 0.53	0.87 ± 0.44	19.46 ± 0.44
	RODS-1NN	**21.95 ± 1.04**	30.69 ± 1.96	4.78 ± 0.25	2.04 ± 0.54	22.94 ± 1.51
	CLTR-LSVM	28.01 ± 5.35	24.01 ± 1.68	5.00 ± 0.78	1.80 ± 0.31	22.57 ± 0.72
	CLTR-RSVM	25.07 ± 1.56	23.31 ± 1.55	1.78 ± 0.48	0.80 ± 0.15	19.63 ± 0.57
	CLTR-1NN	25.71 ± 1.06	33.72 ± 2.00	4.72 ± 0.82	1.74 ± 0.18	24.60 ± 0.54
	MADS-LSVM	24.52 ± 1.95	22.92 ± 1.39	4.78 ± 0.42	1.11 ± 0.36	21.63 ± 0.78
	MADS-RSVM	23.42 ± 1.70	**22.92 ± 0.72**	**1.44 ± 0.37**	0.81 ± 0.09	16.01 ± 0.25
	MADS-1NN	25.07 ± 0.39	28.05 ± 1.21	3.28 ± 1.29	1.96 ± 0.38	23.56 ± 0.57
	KC-LSVM	25.71 ± 0.72	26.81 ± 1.61	4.89 ± 0.74	1.35 ± 0.20	21.69 ± 1.42
	KC-RSVM	23.88 ± 2.08	26.26 ± 1.19	1.78 ± 0.75	**0.72 ± 0.26**	**15.92 ± 0.97**
	KC-1NN	25.25 ± 2.31	32.40 ± 3.11	4.39 ± 0.21	2.19 ± 0.77	24.32 ± 1.79
CHMM	CHMM-LR	31.77 ± 1.91	51.44 ± 3.38	23.78 ± 3.22	1.83 ± 0.56	29.86 ± 1.09
	CHMM-SLR	30.76 ± 2.26	Unavailable †	25.39 ± 3.27	2.89 ± 0.86	25.79 ± 1.48
	CHMM-ARBITRARY	**30.58 ± 2.93**	**44.68 ± 3.08**	**19.67 ± 1.67**	**1.69 ± 0.61**	**23.95 ± 1.49**

"Unavailable": experiment could not finish due to a system crash.

CONCLUSIONS

- Better accuracy and training time than Hidden Markov Models and ADABoost+DStump on the data sets tried.
- Accuracy is competitive with the leading DTW+1NN algorithm. Performance is significantly better than baseline.
- Model application is much faster than DTW+1NN.

FUTURE WORK

- New prototype generation algorithms are being tested.
- Work is underway to extend the algorithm to handle anomaly detection, highly irregularly-sampled time series, hierarchical classification, and a large number of classes.

REFERENCES

- Eamonn Keogh. The UCR time series data mining archive, 2006.
- Chotirat (Ann) Ratanamahatana and Eamonn J. Keogh. Making time-series classification more accurate using learned constraints. In SDM, 2004.
- Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. pages 267–296, 1990.
- Kevin Murphy. Hidden Markov Model Toolbox, 2003.
- Damian Eads, et al. Grammar-guided feature extraction for time series classification. Technical Report LA-UR-05-4487
- Steve Michael. KDTree Range Search Toolbox, 2005.
- Junshui Ma, Yi Zhao, and Stanley Ahalt. OSU Support Vector Machine Toolbox, 2003.
- C. Myers, L. Rabiner, and A. Rosenberg. Performance tradeoffs in for isolated word recognition. IEEE TASSP. 28(6):623–635, 1980.