Centre for
Data Analytics

Insight

Time Series Classification by Sequence Learning in All-Subsequence Space

Thach Le Nguyen, Severin Gsponer, Georgiana Ifrim

Insight Centre for Data Analytics, Dublin, Ireland

April 21, 2017

Outline

- Background
- Our Approach
- Evaluation
- Conclusion
- Future Work

Background

Time Series Classification

Temporal Data (aka Time Series)

- An ordered collection of numeric measurements gathered at equal time intervals.
- A numeric vector $V =< v_1, v_2, \ldots, v_L >$.

Time Series Classification

Temporal Data (aka Time Series)

- An ordered collection of numeric measurements gathered at equal time intervals.
- A numeric vector $V = < v_1, v_2, \ldots, v_L >$.

Example

Northern California Earthquake Data Center (the earthquake classification problem)

0	-0.26927	-0.26927	-0.26927	-0.26927	-0.26927
1	-0.46887	2.748	1.6263	-0.46887	-0.46887
0	2.2429	-0.39296	-0.39296	-0.39296	-0.39296
0	-0.45836	2.4229	-0.45836	2.5162	1.9876
0	-0.58609	-0.58609	-0.58609	-0.58609	-0.58609
0	1.8657	-0.44769	-0.44769	-0.44769	1.7914
0	1.3541	1.9638	-0.53962	-0.53962	-0.53962

Time Series Classification

An ideal time series classifier:

Time Series Classification

An ideal time series classifier:

- Accurate.

Time Series Classification

An ideal time series classifier:

- Accurate.
- Efficient.

Time Series Classification

An ideal time series classifier:

- Accurate.
- Efficient.
- Interpretable.

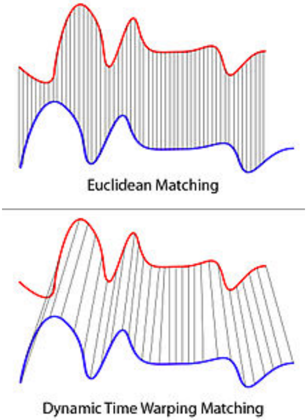
Time Series Classification: State-of-the-art

Shape-based classifiers (numeric data)

- 1NN - Euclidean
- 1NN - Dynamic Time Warping (DTW)

Structure-based classifiers (discretised data)

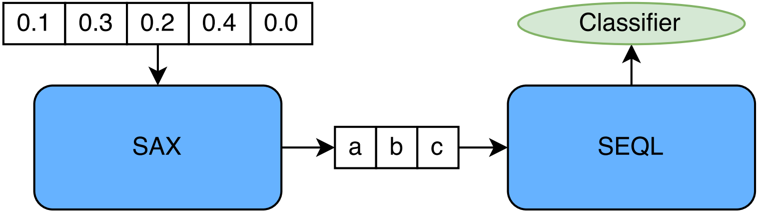
- SAX - VSM (time domain)[Senin and Malinchik, 2013]
- BOSS VS (frequency domain)[Schäfer, 2015]



Source of picture: https://sflscientific.com/blog/2016/6/3/dynamic-time-warping-time-series-analysis-ii

Our Approach: Time series classification with Sequence Learner

Our approach



Symbolic Representation of Time Series

- X-axis slicing: Number of segments as a parameter (w).

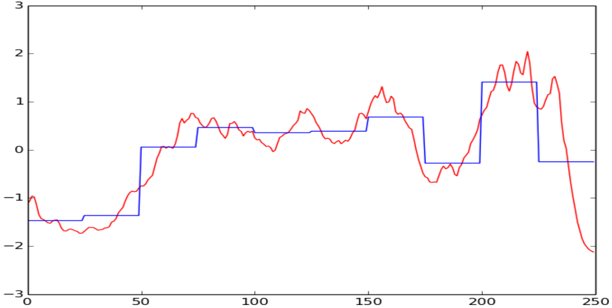


Figure: X-axis slicing: Original time series versus Piecewise Aggregate Approximation (PAA) segments.

Symbolic Representation of Time Series

- Y-axis slicing: Size of the alphabet as a parameter (a).

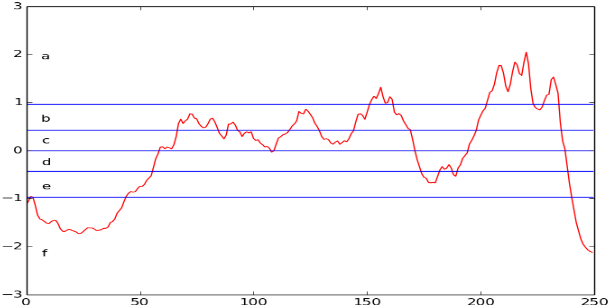


Figure: Y-axis slicing: Equal probability slicing with alphabet size $a = 6$.

Symbolic Representation of Time Series

- Both X and Y-axis slicing: Symbolic Aggregate Approximation (SAX) [Lin et al., 2003].

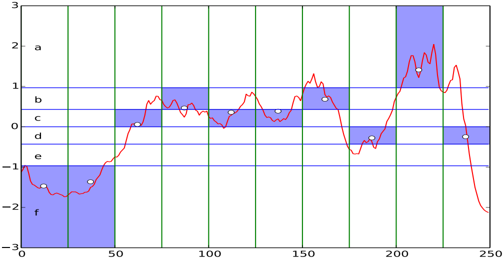
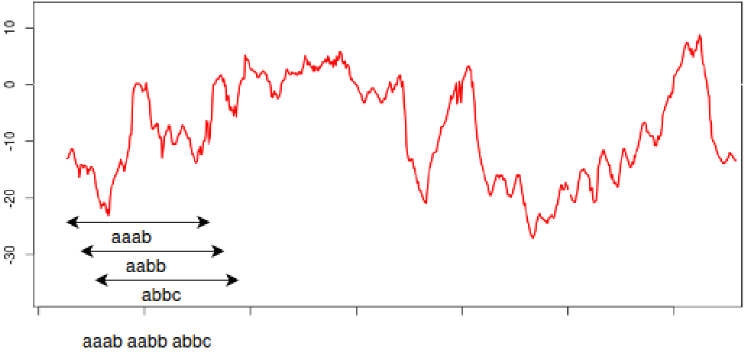


Figure: SAX transformation to symbolic sequence **ffcbccbdad**. Parameters $w = 10, a = 6$.

Symbolic Representation of Time Series

With a sliding window

Length of the window as a parameter (l). Result is a sequence of SAX words.



Symbolic Representation of Time Series

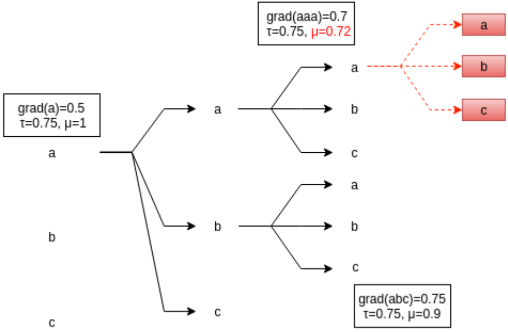
Potential issues

- Patterns can be shifted on either X or Y-axis.
- *False dismissal* [Rakthanmanon and Keogh, 2013]: two nearly equal numeric values discretized into 2 different symbols.
- Expensive training of the parameters (window length l, word length w and alphabet size a).

Time Series Classification with Sequence Learner

Sequence Learner [Ifrim and Wiuf, 2011]

A linear classifier on sequence data: SEQL learns discriminative subsequences from training data by exploiting the all-subsequence space using a coordinate gradient descent approach.



Why SEQL?

- Accuracy: Similar to state-of-the-art sequence classifiers (e.g., non-linear string-kernel-svm).
- Scalability: Sub-linear in number of features and linear in number of examples. Optimization with greedy gradient descent and branch-and-bound.
- Interpretability: List of human readable features ranked by optimized weights.

Time Series Classification with Sequence Learner

Adapted Sequence Learner

The original SEQL disregards the distance between two symbols thus it is vulnerable to the *false dismissal* problem. We introduce new techniques to address this problem:

(1) VSEQL (Variable SEQL)

- Process SAX words at character-level.
- Subsequences-of-subsequences learning. E.g., **abc**d **abc**c

Time Series Classification with Sequence Learner

Adapted Sequence Learner

The original SEQL disregards the distance between two symbols thus it is vulnerable to the *false dismissal* problem. We introduce new techniques to address this problem:

(1) VSEQL (Variable SEQL)
(2) FSEQL (Fuzzy SEQL)

- Match similar subsequences. E.g., **abcd abcc** abca
- Based on a distance function between subsequences.
  $D(s_1, s_2) = \sum_{i=1}^{L} d(s_{1i}, s_{2i})$ where
  $d(s_{1i}, s_{2i}) = |index(s_{1i}) - index(s_{2i})|$

Evaluation

Experiments Set-up

- Benchmark Data: UCR Archive.
- Test System: Linux PC with Intel Core i7-4790 Processor, 16GB 1600 MHz Memory and 256GB SSD storage.
- Code written in C++.
- Fixed parameters (if applicable): $l = 0.2 * L$, $w = 16$, $a = 4$

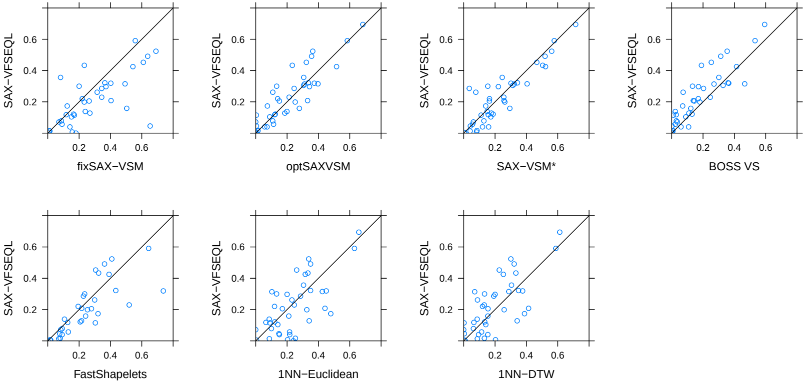
Classification Error Rates with different Symbolic Representation (without sliding window)

Dataset	Raw	DiscX(w)	DiscY(a)	SAXchars(w,a)
Coffee	0.4286	0.4643	**0.1071**	0.2143
Earthquakes	0.2019	**0.1801**	0.2453	0.2638
ECG200	0.63	0.64	**0.23**	0.25
ECGFiveDays	0.5006	0.5029	**0.3136**	0.338
FordA	0.4788	0.4874	**0.2285**	0.3602
FordB	0.4981	0.4884	**0.2162**	0.4686
Gun_Point	0.4667	0.4933	**0.06**	0.1333
Lighting2	0.4754	0.541	**0.3279**	0.3443
SonyAIBO	0.4309	0.4293	**0.3694**	0.3794
wafer	0.6723	0.5357	**0.016**	0.0386
yoga	0.4837	0.3123	**0.299**	0.3293

Classification Error Rates with different Symbolic Representation (with sliding window)

Dataset	SAX-SEQL	SAX-VSEQL	SAX-VFSEQL
Coffee	0.0714	**0.0**	0.0714
Earthquakes	0.2081	0.2112	**0.177**
ECG200	0.25	0.26	**0.22**
ECGFiveDays	0.0743	0.0244	**0.0081**
FordA	0.4685	0.1486	**0.1277**
FordB	0.5069	0.2175	**0.2046**
Gun_Point	0.04	**0.0133**	**0.0133**
Lighting2	0.377	0.2951	**0.2131**
SonyAIBO	**0.2363**	0.3694	0.2912
wafer	0.0081	0.0071	**0.0039**
yoga	**0.1977**	0.2313	0.206

Versus state-of-the-art



Efficiency: No need of tuning SAX parameters

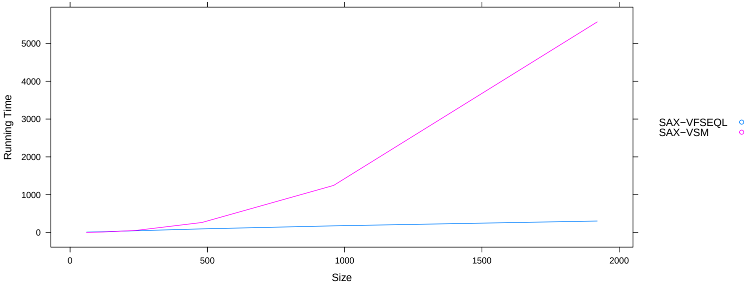
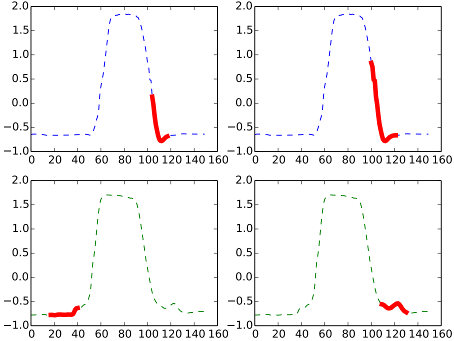


Figure: SAX-VFSEQL and SAX-VSM running time on synthetic dataset

Interpretability



- Gun (buttom panel) has bumps in the beginning and end of movement (lifting and putting Gun back in holster)
- Point (top panel) has a dip in the end of movement (hand inertia)

Conclusion & Future Work

Conclusion

- A new structure-based time series classification methods built on SAX transformation.
- Improve efficiency by reducing the need of tuning SAX parameters.
- Interpretable, yet accurate and efficient.

Future work

Multidimensional Time Series

- Training one SAX-SEQL classifier for each sensor. Output is a set of N models (lists of features)
- Using the features directly for prediction (sum all weights for each group, sum across all groups)

Thank you !

- Acknowledgment: This work was funded by Science Foundation Ireland (SFI) under grant number 12/RC/2289.
- Code and data can be found at https://github.com/lnthach/SAX-SEQL
- Please email thach.lenguyen@insight-centre.org if you have further questions.

References I

Ifrim, G. and Wiuf, C. (2011).
Bounded coordinate-descent for biological sequence classification in high dimensional predictor space.
In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 708–716, New York, NY, USA. ACM.

Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003).
A symbolic representation of time series, with implications for streaming algorithms.
In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, DMKD '03, pages 2–11, New York, NY, USA. ACM.

Rakthanmanon, T. and Keogh, E. (2013).
Fast shapelets: A scalable algorithm for discovering time series shapelets.
In *Proceedings of the thirteenth SIAM conference on data mining (SDM)*, pages 668–676. SIAM.

Schäfer, P. (2015).
The boss is concerned with time series classification in the presence of noise.
Data Mining and Knowledge Discovery, 29(6):1505–1530.

References II

Senin, P. and Malinchik, S. (2013).
Sax-vsm: Interpretable time series classification using sax and vector space model.
In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 1175–1180.