A shapelet transform for multivariate time series classification

Aaron Bostrom and Anthony Bagnall

University of East Anglia, Norwich, UK, NR47TJ.

Abstract. Shapelets are phase independent subsequences designed for time series classification. We propose three adaptations to the Shapelet Transform (ST) to capture multivariate features in multivariate time series classification. We create a unified set of data to benchmark our work on, and compare three multivariate ST variants with multivariate dynamic time warping variants. We demonstrate that multivariate shapelets are not significantly worse than other state-of-the-art algorithms.

1 Introduction

Multivariate time series classification (MTSC) has gained traction in recent years although the majority of work in time series classification (TSC) has focused on the univariate case. Where a single signal for univariate TSC is assigned a class label, multiple signals are recorded for one class with MTSC. MTSC has many practical applications. These can range from medical problems, such as electroencephalogram, finance, multimedia, human activity recognition and gesture recognition.

Recently a large experimental analysis of the state of univariate time series classification was conducted [1]. One of the most successful algorithms within that study was the Shapelet Transform (ST) [2,3]. Shapelets are discriminative phase-independent subsequences. The Shapelet Transform finds a set of good discriminatory shapelets then transforms the data into attributes representing minimal distance to shapelets. A heterogeneous ensemble is then employed to build a classifier. On univariate problems, ST is significantly better than the original Shapelet Tree algorithm [4] where the shapelets were used to form the rules within a decision tree. We are interested in testing out whether the high level of accuracy translates to MTSC problems.

We propose three simple multivariate ST approaches and compare performance to dynamic time warping variants on 22 MTSC data sets. We present a constrained version of ST which can be used on very large problems where enumeration is infeasible. We make the data, results and code publicly available [1].

[1] http://research.cmp.uea.ac.uk/multivariate_shapelets/
https://bitbucket.org/TonyBagnall/time-series-classification

2 Background

we define a MTSC dataset as a set of n time series, **MT**$= \{MT_1, MT_2, ..., MT_n\}$, where a single time series $MT = \{\{T_1, T_2, ..., T_d\}, c\}$ is a set of d time series with a single shared class label. Each series in a multivariate instance is described as $T_i = \{t_{i,1}, t_{i,2}, ..., t_{i,m}\}$ where we assume all series are equal length m and all real valued.

A large volume of research has been conducted on analyzing multivariate time series. The breadth of this work spans from gesture recognition to mining of historical documents and handwriting. Gesture and human activity recognition is one of the most popular areas of research with in this field [5]. Gesture recognition has also been extended to particular activities such as playing musical instruments. Multivariate research in the health domain has focused on health records, EEG and MEG classification, or balance and mobility sensor data for patients with Parkinsons disease.

Most of these research domains have focused on using dynamic time warping with a nearest neighbour classifier, mainly because until very recently it was considered the state of the art solution to time series classification [6]. There are two simple ways of adapting dynamic time warping for the multivariate case. Independent dynamic time warping ($\mathrm{DTW_I}$) finds the DTW distance of each dimension independently. So, given, two multivariate time series, Q and C with d dimensions, the independent DTW distance is

$$DTW_I(Q, C) = \sum_{i=1}^{d} DTW(Q_i, C_i).$$

Dependent dynamic time warping (DTW_D), performs a single warping, but uses all dimensions in specific distance calculation. If the distance matrix used by standard DTW is denoted M, then DTW_D calculates M as

$$M(Q, C)_{i,j} = \sum_{k=1}^{d} (q_{k,i} - c_{k,j})^2.$$

The choice between $\mathrm{DTW_I}$ and $\mathrm{DTW_D}$ is problem dependent. An adaptive form of DTW, $\mathrm{DTW_A}$, that chooses between $\mathrm{DTW_I}$ and $\mathrm{DTW_D}$ for each case based on train set evaluation was described in [7]. We use the three classifiers $\mathrm{DTW_I}$, $\mathrm{DTW_D}$ and $\mathrm{DTW_A}$ as benchmarks to compare our proposed shapelet algorithms.

2.1 Shapelets

There are numerous approaches to using shapelets to classify. All shapelet finding algorithms require the measuring of the distance between a candidate and a time series. This is done by sliding the candidate along the series and calculating the Euclidean distance at each position (after normalisation) to find the minimum.

The distance between a shapelet and a series is then given by Equation 1, where W is the set of all subsequences which are the same length as S in T, and *dist* is the Euclidean distance between two equal length series.

$$sDist(S,T) = min_{w \in W}(dist(S,w)). \tag{1}$$

The Shapelet Forest algorithm[8] can be adapted for MTSC as can the learned shapelet approach[9]. Ultimately we will compare with these algorithms. However, we do not yet have stable multivariate implementations and we believe following the standard TSC approach of benchmarking against DTW is a sensible first set of experiments.

2.2 Shapelet Transform

The original shapelet transform enumerated all possible shapelets. However, we have found that enumeration is very rarely required and sampling a tiny proportion of the shapelet space does not lead to a significant decrease on accuracy [10]. To summarise, for the datasets that cannot be fully enumerated in a pre-specified time, we randomly sample shapelets from the whole space until the algorithm runs out of time. We define this approach as a contract classifier, and the subsequent definitions will define a contracted Shapelet Transform.

Algorithm 1 describes the contract ST algorithm. The first stage (line 1) is to estimate how many candidate shapelets can be evaluated in the contracted time, defined by the user. This involves estimating the time taken for a fundamental operation on the hardware in question, then calculating the expected number of fundamental operations an average shapelet will take. The task then is to find k shapelets from r randomly sampled candidates. This is done by a round robin method that involves finding a shapelet for a different class at each iteration (line 4 and 5). The round robin method involves randomly choosing a case of the given class cv, then randomly selecting a starting position and shapelet length within the selected series. We have explored a range of sampling and heuristic search techniques for finding shapelets such as simulated annealing and tabu search. However, none as yet have proved significantly better than simply randomly sampling the shapelet space.

The distance between the candidate and all train series is found using Equation 1 and stored in the array D_S (line 6). The quality of the shapelet is then assessed from D_S (line 7). The standard quality metric is information gain, which is found by sorting the set of distances D_S then finding the best split of the data into class cv or not cv. This binary evaluation has been shown to be significantly better for problems with many class values [3]. The candidates are all stored (line 8) until we have evaluated our alloted r shapelets. Note that not all candidates are actually stored in the evaluation, but the algorithm is clearer if we assume they are. The selection of final shapelets is performed (line 10) so that the number of shapelets selected reflects the class frequency. This helps avoid the problem over shapelets from one class overwhelming other classes.

The k best shapelets are used to create a data transformation. By using the distance (as defined by Equation 1) between the shapelet and each series, we form an n by k feature matrix. One of the main advantages of using transformed data is that we can use any classifier, as opposed to using only a decision tree as per the original definition of the algorithm. The classifier typically used in conjunction with ST is the heterogeneous ensemble of simple classification algorithms (HESCA) [11,12].

The algorithm is simplified in comparison to the implementation for ease of explanation. This description omits the details of class balancing, binary shapelet methods, and the numerous refinements to speed up the process such as early abandon and reordering.

Algorithm 1 ShapeletTransform($\mathbf{T}$, min, max, k, t)

Input: A list of time series $\mathbf{T}$, min and max length shapelet, the desired number of shapelets and maximum run time t. c is the number of class values
Output: A list of k Shapelets

$r \leftarrow$contractNumCandidates(t)
candidates $\leftarrow \emptyset$
for $s \leftarrow 1$ to r **do**
    $cv \leftarrow s\%c$
    $S \leftarrow$ roundRobinSampleShapelet($\mathbf{T}$,min, max, cv)
    $D_S \leftarrow$findDistances(S,$\mathbf{T}$)
    $quality \leftarrow$assessBinaryCandidate($\mathbf{T}, S, D_S, cv$)
    candidates.add($S, quality, cv$)
$< k_1, \ldots, k_c >\leftarrow$ shapeletsPerClass($\mathbf{T}$,k)
kShapelets $\leftarrow$extractKBest(candidates,$< k_1, \ldots, k_c >$)
return kShapelets

3 Datasets

In Table 1 we present the list and the properties of the multivariate datasets we have collated from the literature. The datasets have a range of different sizes, number of instances, length of the series, the number of series and finally number of classes. To simplify and reduce the need for extensive dataset knowledge we have reduced some problems into sub problems. This is most notable with the AALTD problems. These were originally from a challenge dataset produced for the Second ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data (AALTD). The aim was classify six different gestures using eight spatial sensors placed on a person, resulting in 3 dimensional movement information for each sensor. We split the dataset into a separate classification problem for each sensor.

The final dataset is MVMotion, an example of Human Activity Recognition (HAR). There are three variants: MVMotionA; MVMotionG; and MVMotionAG. The MVMotion datasets are collected from a 3D accelerometer and a 3D gyroscope on a mobile device during a particular set of activities. All MVMotion datasets consist of four classes: walking; resting; running and badminton. Participants were required to record motion a total of five times, and the data is sampled once every tenth of a second, for a ten second period.

datasets	n	d	m	c	datasets	n	d	m	c
AALTD_0	90	3	52	6	CricketLeft	84	3	1198	12
AALTD_1	90	3	52	6	CricketRight	84	3	1198	12
AALTD_2	90	3	52	6	HandwritingA	150	3	153	26
AALTD_3	90	3	52	6	HandwritingG	500	3	153	26
AALTD_4	90	3	52	6	JapaneseVowels	270	12	30	9
AALTD_5	90	3	52	6	MVMotionA	40	3	101	4
AALTD_6	90	3	52	6	MVMotionAG	40	6	101	4
AALTD_7	90	3	52	6	MVMotionG	40	3	101	4
ArabicDigit	6599	13	94	10	PenDigits	7494	2	9	10
AWordLL	275	3	145	25	UWaveGesture	120	3	316	8
AWordT1	275	3	145	25	Epilepsy	137	3	207	4
AWordUL	275	3	145	25					

Table 1: A list of the datasets in the multivariate time series archive. Number of instances is denoted by n, number of dimensions is d, length of series is m, and number of classes is denoted by c.

4 Multivariate Shapelet Transform

In this section we describe the three shapelet methods developed for the Multivariate Shapelet Transform.

4.1 Multivariate Shapelet Transform (MST)

The first multivariate shapelet method is the simplest generalisation of the univariate approach, which we denote MST. This algorithm finds single dimension shapelets then assesses the shapelets quality against the other series via sliding the shapelet along the same dimension in the multivariate series. Once the k best shapelets have been found, they are used to transform the original dataset as with the univariate ST. A shapelet then is a subseries of a single dimension and is only ever compared to its dimension of origin.

4.2 Multivariate Dependent Shapelet Transform ($\mathbf{MST}_D$)

The second multivariate shapelet method is called MST_D. A shapelet now spans all dimensions, rather than a single dimension, and the distance calculation uses this distance over all dimensions as described in Equation 2. There are fewer candidate shapelets for MST_D, but the enforced alignment means it is susceptible to being deceived by slight phase shift across dimensions.

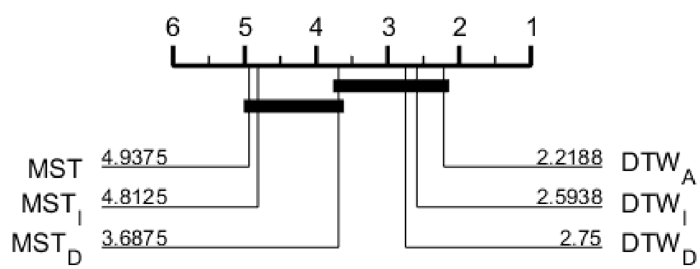
4.3 Multivariate Independent Shapelet Transform (MST_I)

The third multivariate shapelet method is called MST_I. This method also extracts multivariate shapelets. The difference comes in the distance calculation, where an approach similar to DTW_I is adopted (Equation 2). Rather than fix each dimension when sliding across a series to find the minimum distance, MST_I finds the closest match to each dimension independently, then sums the closes matches. This means that, for example, the closest match for dimension 1 need not be at the same position in time as the closest match to dimension 2. This allows for independent phase shift between dimensions.

5 Results

The experimental setup follows the same approach outlined in [1]. We perform 100 fold resampling on each data set, then average accuracy over test sets. For each algorithm presented we have performed 2,400 experiments. We compare multiple classifiers on multiple data sets with critical difference diagrams, which display the average ranks of the classifiers over all problems and group classifiers into cliques, within which there is no significant difference. Full results are omitted due to space, but can be downloaded from [2].

Fig. 1: Multivariate shapelet and dynamic time warping classifiers on 16 problems.



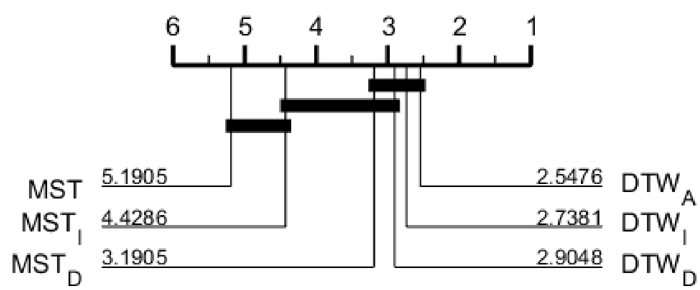
We present results for the three MTSC shapelet algorithms for full enumeration and for time constrained versions using random sampling. We were only able to complete full enumeration on 16 of the 21 datasets within a one week runtime limit on our HPC. Figure 1 shows the critical difference diagram for the three shapelet and three DTW classifiers. We observe that DTW_A has the highest average rank, but it is not significantly better than DTW_I, DTW_D or MST_D when compared with a pairwise Wilcoxon signed rank test. It is significantly

[2] http://research.cmp.uea.ac.uk/multivariate_shapelets/

better than the other two shapelet approaches. We conclude that the dependent approach is the most promising multivariate shapelet algorithm.

To demonstrate that shapelet finding is not necessarily computationally intensive, we repeated the experiments but constrain the shapelet search to approximately one hour of run time. For the large problems, this means we are evaluating only a tiny fraction of the entire shapelet space. Figure 2 shows the critical difference diagram for the constrained search. We note the identical pattern. MST_D is the best multivariate shapelet approach, and is not significantly worse than the benchmark.

Fig. 2: Multivariate ST constrained to one hour and dynamic time warping classifiers on 21 problems.



6 Conclusion

We have presented three new shapelet algorithms for MTSC and compared them to DTW based benchmarks. We found that finding shapelets from a single dimension (MST) was the worst method. We conclude that the dependent approach, MST_D, is the most promising, given that it is not significantly worse than DTW_A even when the search space is constrained to one hour of sampling. This is a starting point. The analysis would be more conclusive if we had more datasets to use in the evaluation, and a comparison to alternative multivariate shapelet approaches[8,9] is the next step.

Acknowledgement. This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/M015807/1]. The experiments were carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia and using a Titan X Pascal donated by the NVIDIA Corporation.

References

1. A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 31, no. 3, pp. 606–660, 2017.
2. J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining and Knowledge Discovery*, vol. 28, no. 4, pp. 851–881, 2014.
3. A. Bostrom and A. Bagnall, "Binary shapelet transform for multiclass time series classification," *Proc. 17th International Conference on Big Data Analytics and Knowledge Discovery (DAWAK)*, 2015.
4. L. Ye and E. Keogh, "Time series shapelets: a novel technique that allows accurate, interpretable and fast classification," *Data Mining and Knowledge Discovery*, vol. 22, no. 1-2, pp. 149–182, 2011.
5. O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors.," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
6. C. A. Ratanamahatana and E. Keogh, "Three myths about dynamic time warping data mining," in *Proceedings of the 2005 SIAM International Conference on Data Mining*, pp. 506–510, SIAM, 2005.
7. M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing dtw to the multi-dimensional case requires an adaptive approach," *Data Mining and Knowledge Discovery*, vol. 31, no. 1, pp. 1–31, 2017.
8. O. P. Patri, R. Kannan, A. V. Panangadan, and V. K. Prasanna, "Multivariate time series classification using inter-leaved shapelets," in *NIPS 2015 Time Series Workshop*, 2015.
9. J. Grabocka, M. Wistuba, and L. Schmidt-Thieme, "Fast classification of univariate and multivariate time series through shapelet discovery," *Fast classification of univariate and multivariate time series through shapelet discovery*, vol. 49, pp. 429–454, 2016.
10. A. Bostrom, A. Bagnall, and J. Lines, "Evaluating improvements to the shapelet transform," *Knowledge Discovery and Data Mining, in Workshop on Mining and Learning from Time Series*, 2016.
11. J. Lines, S. Taylor, and A. Bagnall, "Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles," *ACM Transactions on Knowledge Discovery from Data*, 2018.
12. J. Large, J. Lines, and A. Bagnall, "The Heterogeneous Ensembles of Standard Classification Algorithms (HESCA): the Whole is Greater than the Sum of its Parts," 2017.