Motif Alignment for Time Series Data Augmentation

Chapter · August 2023

DOI: 10.1007/978-3-031-39831-5_4

4 authors:

Omar Bahri
Utah State University
13 PUBLICATIONS **33** CITATIONS
SEE PROFILE

Peiyu Li
Utah State University
12 PUBLICATIONS **25** CITATIONS
SEE PROFILE

Soukaina Filali Boubrahimi
Utah State University
60 PUBLICATIONS **381** CITATIONS
SEE PROFILE

Shah Muhammad Hamdi
Utah State University
43 PUBLICATIONS **241** CITATIONS
SEE PROFILE

Motif Alignment for Time Series Data Augmentation

Omar Bahri, Peiyu Li, Soukaina Filali Boubrahimi, and Shah Muhammad Hamdi

Computer Science Department, Utah State University, Logan, UT, USA
`{omar.bahri,peiyu.li,soukaina.boubrahimi,s.hamdi}@usu.edu`

Abstract. In this paper, we propose MotifAug, a parameter-free, pattern mixing-based time series data augmentation method that improves previous approaches in the literature. MotifAug leverages the warping path constructed by MotifDTW, a novel alignment method that uses the Matrix Profile (MP) motif discovery mechanism and Dynamic Time Warping (DTW) to align two time series data instances.

Keywords: Data Augmentation · Time Series Data · Time Series Motifs

1 Introduction

In contrast to other domains such as image recognition, data augmentation techniques for time series datasets are still in their early stages. Iwana et al. [4] categorize the current efforts into four distinct groups: random transformations, generative models, decomposition methods, and pattern mixing. In this paper, we focus on the latter. Pattern mixing time series augmentation methods are based on the idea that by combining two data samples belonging to the same dataset, and particularly to the same dataset class, it is possible to generate new, realistic data samples. The main advantages of these methods is that they neither make assumptions about the validity of specific transformations in certain data domains, nor do they require extensive training to learn models that mimic the dataset distribution. Thus, they are domain agnostic and can be used on-the-fly. Notable examples include the Synthetic Minority Over-sampling Technique (SMOTE) [1], Weighted-Dynamic Barycenter Averaging (wDBA) [2], SuboPtimAl Warped time series geNEratoR (SPAWNER) [5], and guided warping [3].

The Matrix Profile (MP) time series data structure has been introduced recently by Yeh et al. [8] and was met with huge success in the time series mining community. By considerably reducing the time complexity required for the all-pairs similarity search for time series subsequences, the MP has been quickly leveraged to solve a plethora of problems and improve algorithms. However, to the extent of our knowledge, it has not been used yet in the context of data augmentation. In this paper, we introduce (1) MotifDTW, a novel algorithm that leverages the MP to match motifs between two time series instances and constructs a warping path by aligning them using DTW, and (2) MotifAug, a time series data augmentation approach that uses MotifDTW at its core to generate meaningful intra-class data samples from existing datasets.

2 Preliminaries

2.1 Matrix Profile

Let T be a time series vector of length l: $T = (t_1, t_2, ..., t_l)$ and T' a time series vector of length l': $T' = (t'_1, t'_2, ..., t'_{l'})$. If T is the query time series and T' is the reference time series, the aim of the MP is to provide fast access to the nearest neighbor of each subsequence of T of length m defined as $sub_{i,m} = (t_i, t_{i+1}, ..., t_{i+m})$ to every other subsequence $sub'_{j,m}$ of T' of the same length m. This is achieved by computing a distance profile vector $D_i = (d_{i,0}, d_{i,1}, ..., d_{i,l-m+1})$ for each query subsequence $sub_{i,m}$ s.t. $d_{i,j} = dist(z(sub_{i,m}), z(sub'_{j,m}))$, where z is the z-normalization function and $dist$ is the Euclidean distance. Then, the matrix profile vector $MP(T, T') = (mp_1, mp_2, ..., mp_{l-m+1})$, which provides direct access to the distance separating each $sub_{i,m}$ from its nearest neighbor in T', is created such that $mp_i = min(D_i)$. Similarly, the matrix profile index vector $MPI(T, T') = (mpi_1, mpi_2, ..., mpi_{l-m+1})$ with $mpi_i = argmin(D_i)$ indicates the location of each nearest neighbor subsequence.

2.2 Pan-Matrix Profile

Determining the window size parameter m for the MP is a challenging task that usually requires prior knowledge from experts. Therefore, Madrid et al. [6] introduced the Pan Matrix Profile (PMP), a new data structure that combines information from MPs with a range r of window sizes lengths, and proposed the Scalable KInetoscopic Matrix Profile (SKIMP) algorithm to compute it with a time complexity and space complexity of $O(l^2r)$ and $O(lr)$ respectively. Given a range of window lengths $m = [m_{min}, m_{max}]$ such that $m_{min} > 0$ and $m_{max} < l$, a query time series T, and a reference time series T', $PMP(T, T')$ is simply a matrix where each row consists in $MP(T, T')$ with window length m. Similarly, $PMPI(T, T')$ is the matrix where each row consists in $MPI(T, T')$.

2.3 DTW Alignment for Time Series Data Augmentation

Besides its use as a distance measure, DTW [7] can be adopted to align two time series instances. In the context of time series data augmentation, several methods have taken advantage of this feature to generate new data samples. wDBA [2] uses DBA, which aligns two or more time series samples before computing their medoid, to create new data samples. SPAWNER [5] generates new time series by selecting two intra-class instances, splitting them into two segments each, aligning the first two segments and the last two segments separately using DTW, averaging the result, and adding some noise to ensure that the instance generated is different than the original ones. As to Guided Warping [3], it generates time series instances by selecting two intra-class instances from the original data set and aligning one to the other using either DTW or a modified version of it described below.

One of the main drawbacks of aligning time series instances using DTW for time series data augmentation is that the result might contain sudden, non-smooth changes that disrupt the continuity of the time series. This is most relevant when the warping path maps two adjacent time steps to two distant points in the reference time series. Therefore, Iwana and Uchida [3] introduced shapeDTW, a modified version of DTW that takes into consideration high-level shape descriptors to optimize the warping path. To this end, shapeDTW proceeds similarly to DTW but considers the immediate neighborhood of each time series data point (time steps within $m/2$ on each side) in its distance calculation.

3 Proposed Method

3.1 Motif Mapping

By introducing data points' neighborhoods in distance calculations, shapeDTW aims to avoid abrupt changes in the warping path and preserve high-level shape descriptors. However, shapeDTW suffers from two main problems: **(1) It often fails to avoid abrupt changes.** Consider the case where the reference time series T' contains repeating motifs, i.e. quasi-identical shape descriptors, occurring at different time steps. There is no guarantee under the shapeDTW constraints that two adjacent data points t_i and $t_{i\pm1}$ in the query time series, whether or not they belong to the same shape descriptor, will not be mapped to two distant data points t'_j and t'_k belonging to two distant occurrences of the same motif in the target time series. See section 4.2 for visual examples. **(2) It is dependent on the window length parameter.** Determining the window length m that defines shape descriptors requires prior expert knowledge of the domain. Moreover, it might be necessary to use several lengths. Therefore, the fact that shapeDTW is highly dependent on m makes it vulnerable to issues resulting from the use of a wrong value. For example, using too large of a value will result in missing important shape descriptors, while setting it too small might emphasize meaningless intervals.

Thus, we propose MotifDTW, a novel time series alignment method that relies on motifs to construct a meaningful warping path. MotifDTW solves the issues described above by (1) being parameter-free and (2) mapping all the time steps of a motif in the query time series to the time steps of one single matching motif in the reference time series instance.

Algorithm 1 describes MotifDTW. Given query time series $T = (t_1, t_2, \ldots, t_l)$ and reference time series $T = (t'_1, t'_2, \ldots, t'_l)$, MotifDTW starts by initializing a motif alignment path to a vector $MAP(T, T') = (map_1, map_2, ..., map_l)$ of length l. Then, it computes $PMP(T, T')$ matrix. Since each row of $PMP(T, T')$ represents $MP(T, T')$ with a different window length m, the higher the length value m the higher the corresponding row values will be. Thus, MotifDTW divides the values in each row by their corresponding m. This allows ranking motifs of different lengths by according equal importance to longer ones. Then, starting with the motif $sub_{i,m}$ with the lowest value in $PMP(T, T')$, the time steps $(map_i, map_{i+1}, ..., map_{i+m})$ in $MAP(T, T')$ are set to the corresponding time steps of its nearest neighbor in T'. This is repeated for each motif, in descending order of importance until all the time steps are mapped. During this process, overlapping motif indices are resolved as follows. If the current overlapping motif extends a previously mapped one (with a shorter length), the non-mapped indices are introduced in $MAP(T,T')$. Otherwise, the overlapping motif is ignored. Then, the final warping path $WP(T,T') = (wp_1, wp_2, ..., wp_l)$ is constructed by aligning each mapped motif $(t_i, t_{i+1}, ..., t_{i+m})$ to its nearest neighbor subsequence $(T'_{map_i}, T'_{map_{i+1}}, ..., T'_{map_{i+m}})$ using DTW.

Algorithm 1 MotifDTW

```
Inputs: Time series samples $T$ and $T'$.
Output: Warping Path $WP(T,T')$.
Initialize the motif alignment path and the warping path:
$MAP(T,T') \leftarrow (map_1, map_2, ..., map_l)$ s.t. $\forall map_i = -1$
$WP(T,T') \leftarrow (wp_1, wp_2, ..., wp_l)$
Compute $PMP$ and $PMPI$ matrices:
$PMP(T,T'), PMPI(T,T') \leftarrow SKIMP(T,T')$        ▷ $SKIMP()$ from [6]
Divide each row ($MP$) in $PMP$ by its window length $m$:
for $MP_i(T,T')$ in $PMP(T,T')$
    $MP_i(T,T') \leftarrow MP_i(T,T')/m$
Map the motifs with the highest importance in $T$ (smallest value in $PMP(T,T')$)
to its nearest neighbor in $T'$ until all $T$ time steps are mapped:
while $\exists map \in MAP(T,T')$ s.t. $map = -1$
    $m, i \leftarrow argmin(PMP(T,T'))$
    $nn\_i \leftarrow PMPI(T,T')_{m,i}$
    if $\forall map \in (map_i, map_{i+1}, ..., map_{i+m})$, $map = -1$
     or $\forall map_{i+j} \in (map_i, map_{i+1}, ..., map_{i+m})$ s.t. $map_{i+j} > 0$, $map_{i+j} = nn_i + j$
    then
        $(map_i, map_{i+1}, ..., map_{i+m}) \leftarrow (nn\_i, nn\_i + 1, ..., nn\_i + m)$
    $PMP(T,T')_{m,i} \leftarrow \infty$
Record the DTW path between each mapped motif $sub_{i,m}$ in $MAP(T,T')$ and its
nearest neighbor
for $sub_{i,m}$ in $MAP(T,T')$
    $nn \leftarrow (T'_{map_i}, T'_{map_{i+1}}, ..., T'_{map_{i+m}})$
    $(wp_i, wp_{i+1}, ..., wp_{i+m}) \leftarrow DTW(sub_{i,m}, nn)$
return $WP(T,T')$
```

3.2 Time Series Augmentation

We introduce MotifAug, a novel parameter-free time series augmentation method that leverages this mechanism for time series data augmentation. Given a time series dataset containing N time series instances $\mathcal{D} = \{T_1, T_2, ..., T_N\}$ with each T_i belonging to a class $C_m \in \{C_1, C_2, ..., C_M\}$, MotifAug generates a new instance T_{aug} with class label C_m by selecting two intra-class instances from D, a query and a reference $T, T' \in C_m$. Then, it uses MotifDTW to construct the warping path $WP(T,T')$ and aligns T to T'. The result is T_{aug}, a new time series that has the time motifs of the query T aligned to the time steps of T'. Since $WP(T,T') \neq WP(T',T)$, for each dataset class C_m containing N_m data samples, MotifAug can generate a total of $N_m \times (N_m - 1)$ new instances.

4 Experimental Evaluation

Due to space restrictions, we encourage the reader to visit our project website[1] and repository[2] for additional details and to access the source code.

4.1 Setup

We compare the performance of MotifAug on the 110 fixed-length datasets from the UCR archive that contain at least two intra-class instances per class to the

---

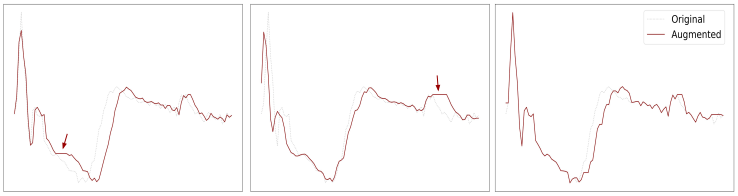
[1] `https://sites.google.com/view/MotifAug/home`

[2] `https://github.com/omarbahri/MotifAug`

following baseline methods (with default parameters): **SMOTE** [1], **wDBA** [2] with the ASD weighting scheme since it had the best results in the original paper, **SPAWNER** [5], Discriminative GW using DTW alignment (**DGW-D**), Random GW using DTW alignment (**RGW-D**), DGW using shapeDTW alignment (**DGW-sD**), and RGW using shapeDTW alignment (**RGW-sD**) [4].

4.2 Aligning time series using MotifDTW

The warping path constructed by MotifDTW between two time series instances allows their alignment in a realistic way, avoiding any abrupt changes by ensuring that all time steps contained in a single motif are aligned to adjacent time steps belonging to a single subsequence. In Fig. 1, we visualize the time series instances generated from two samples from ECG200, a dataset of of healthy and unhealthy electrocardiograms, using RGW-sD, DGW-sD, and MotifAug. The gray lines represent the original query time series and the red lines represents the generated instances. In Fig. 1.a and 1.b, the red arrows point to short horizontal subsequences in the data generated by RGW-sD and DGW-sD that are unusual in the electrocardiograms present in the dataset. The segments are a direct result of the abrupt changes in shapeDTW's warping path described in section 3.1. This problem is avoided by MotifAug as shown in Fig. 1.c.



(a) RGW-sD (b) DGW-sD c) MotifAug

Fig. 1: Visually inspecting data generated from the ECG200 dataset. The subsequences pointed to by the arrows represent unusual heartbeat segments

4.3 Performance Gain

To evaluate the performance of MotifAug in comparison to the time series data augmentation benchmarks, we adopt the classification performance gained from extending the training set with the augmented data samples to double the original size as the main criteria. We adopt two main state-of-the-art time series classification models: RandOm Convolutional KErnel Transform (ROCKET) and Residual Network (ResNet). For both classifiers, we used the default training and testing splits provided in the UCR archive and combine the original training data with the augmented data samples to form a larger training set. Then, we train the classification model and test its performance on the original test set. Since some datasets in the UCR archive suffer from data imbalance, our classification measure of choice is the f1-score. For ROCKET, we repeat this procedure ten times with different initialization seeds and consider the mean f1-scores of the ten trials. Since training ResNet is considerably more time-consuming, the mean f1-scores are taken from only 3 runs. In addition, we narrow down the number of datasets to 50 randomly selected ones.

In Table 1, we display the average ranks of the mean f1-scores of all methods including the classification results using the original datasets (without any augmentation). The detailed results can be found in our project website. An interesting result we noted using ROCKET is that only half of the augmentation methods —including MotifAug — have led to an overall increase in performance. MotifAug had the first average rank using both classifiers, followed by SMOTE with ROCKET and DGW-sD with ResNet.

Table 1: Classification comparison: average rank of f1-scores

Average Rank	**No Aug**	**SMOTE**	**wDBA**	**SPAWNER**	**RGW-D**	**DGW-D**	**RGW-sD**	**DGW-sD**	**MotifAug**
ROCKET	4.64	4.13	5.14	6.45	5.21	5.65	4.45	4.89	**4.00**
ResNet	5.96	4.36	4.44	5.94	4.9	4.1	3.56	3.88	**3.28**

5 Conclusion

We introduced MotifDTW, an alignment method for time series data that leverages the fast MP motif discovery process. MotifDTW allows constructing a meaningful warping path by ensuring that all adjacent motif time steps in the query data sample are mapped to adjacent time steps in the reference. We also introduced MotifAug, a time series augmentation algorithm that uses MotifDTW as its core. To our knowledge, this is the first effort to leverage the MP motif discovery for time series data augmentation.

Acknowledgments This project has been supported in part by funding from GEO Directorate under NSF awards #2204363, #2240022, and #2301397 and the CISE Directorate under NSF award #2305781.

References

1. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research **16**, 321–357 (jun 2002)
2. Forestier, G., Petitjean, F., Dau, H.A., Webb, G.I., Keogh, E.: Generating synthetic time series to augment sparse datasets. In: Proceedings - IEEE International Conference on Data Mining, ICDM. vol. 2017-Novem, pp. 865–870 (dec 2017)
3. Iwana, B.K., Uchida, S.: Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher (apr 2020)
4. Iwana, B.K., Uchida, S.: An empirical survey of data augmentation for time series classification with neural networks. PLOS ONE **16**(7), e0254841 (jul 2021)
5. Kamycki, K., Kapuscinski, T., Oszust, M.: Data Augmentation with Suboptimal Warping for Time-Series Classification. Sensors **20**(1), 98 (dec 2019)
6. Madrid, F., Imani, S., Mercer, R., Zimmerman, Z., Shakibay, N., Keogh, E.: Matrix profile XX: Finding and visualizing time series motifs of all lengths using the matrix profile. In: International Conference on Big Knowledge, ICBK 2019. pp. 175–182. Institute of Electrical and Electronics Engineers Inc. (nov 2019)
7. Sakoe, H., Chiba, S.: Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing **26**(1), 43–49 (1978)
8. Yeh, C.C.M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H.A., Silva, D.F., Mueen, A., Keogh, E.: Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets pp. 1317–1322 (feb 2017)