Exploring Classification Score Profiles for Change Point Detection

Study Project Exposé

Submitter:	Arik Ermshaus
Submission Date:	Tuesday 21st January, 2020
Supervisor:	Dr. rer. nat. Patrick Schäfer
Supervisor:	Prof. Dr. rer. nat. Ulf Leser
Institution:	Department of Computer Science, Humboldt-Universität zu Berlin

Exploring Classification Score Profiles for Change Point Detection

Abstract. A time series is a temporally ordered sequence of measurements that could for example capture a physical process. In recent years, the amount of unlabelled sensor data has grown significantly through the increase in computational power and omnipresence of sensors such as in smart devices. The literature, however, contains a great selection of time series classification algorithms that require labelled datasets as a ground truth for training. We can adapt the applied principals to assist solving unsupervised time series problems like motif discovery, signal segmentation or change point detection. In this study project, we propose to study a self-supervised methodology that attaches hypothetical labels to consecutive time series segments and uses a time series classifier to discriminate self-similar regions. We divide a time series repeatedly at increasing offsets (split points) into two segments and attach the hypothetical label (-1) to the left segment and (+1) to the right one. We classify each split and create a classification score profile with the corresponding accuracies. The profile displays how self-similar the left and right region is for every split point. We want to explore classification score profiles for segmenting time series into self-similar regions where the highest accuracy could indicate abrupt shifts like change points. Future directions of this work include signal segmentation as well as anomaly detection.

1 Introduction

A time series is an ordered sequence of real-valued measurements that could for example capture a physical process. Time series analysis comprises a variety of problems that have become increasingly important in practical applications like medicine, finance or meteorology [26, 18, 12]. Typical tasks include signal segmentation as well as classification [11, 2]. The literature contains a broad catalogue of time series classification methods that include distance as well as feature-based techniques [13, 23]. For the current state of the art time series classification algorithms, see [2] for reference. These algorithms learn internal representations from labelled time series in order to classify new ones. Therefore, in their original formulation they are not applicable to solve unsupervised time series problems like motif or discord discovery, signal segmentation or change point detection. [4, 16, 27]. These tasks however, become more relevant as the amount of unlabelled sensor data grows. We intend to tackle unsupervised time series problems by adapting the applied principals used for time series classification.

We propose to study a novel self-supervised learning approach inspired by preceding work on change analysis from Hido, S. et al. [10]. The main idea is to split a time series into two consecutive segments at increasing offsets (split points) and attach the hypothetical label (-1) to the first segment and (+1) to the second segment for each split. Every segment is further divided into a set of labelled windows. This creates a labelled dataset with which a time series classification algorithm is trained and evaluated. By repeating this process with different splits, we create a classification score profile with the corresponding classification accuracies for each split point. The classifier should detect pure self-similar segments due to their related statistical properties and classify them with significantly higher classification accuracy compared to more diverse segments. Therefore, the profile should display the self-similarity for all splits and its minima and maxima may be useful for solving the aforementioned unsupervised learning tasks.

In change point detection, a time series captures a physical process that conceptually contains a number of discrete states that are separated by transitions. Discrete states are coherent regions in time series with recognizable statistical properties or temporal patterns. Transitions or change points are abrupt shifts in physical processes due to external events. The change point detection problem is further to locate the k change points in a time series that segment it into m coherent regions. Change point detection can be applied in a number of important use cases. In medical condition monitoring for instance, a patient's physiological variables like heart rate or electroencephalogram (EEG) are continuously monitored and analysed in order to detect trends and anomalies. Change point detection can assist the analysis to identify epilepsy or sleep problems [20, 25]. Another use case is human activity analysis where time series sensor data from accelerometers is used to capture human behaviour like "stay", "walk" or "jog". Change point detection can be used to identify the activities and gather insight on health status or provide activity-aware services [6].

In this study project, we investigate the change point detection problem with classification score profiles and different time series classifiers for one change point ($k = 1$) in order to implement the basic research needed to solve more complex unsupervised time series tasks. We expect the global maxima in a classification score profile to capture one change point as the split point with the highest classification accuracy divides a time series into the two most self-similar segments.

2 Background

In this section we formally define physical processes, time series and the change point detection problem.

1 A physical process $P = (S, C)$ has two or more discrete states $s_1, ..., s_m \in S$ that are pairwise separated by a change point $(s_i, s_k) \in C \subseteq S \times S$.

2 A time series $T = t_1, ..., t_n$ is a temporally ordered sequence of real-valued measurements from a sensor that measures an observable output of a process P. A window $T_{i,k} = t_i, ..., t_k$ is a subsequence from a time series T with $1 \leq i \leq k \leq n$. A segmentation of a time series T is an ordered sequence of consecutive windows $T_{1,i}, ..., T_{k,n}$ with splits at increasing offsets.

3 Given a time series T that captures a physical process $P = (S, C)$, the change point detection problem is to segment T into an ordered sequence of segments $T_{1,i}, ..., T_{k,n}$ so that the transition between two consecutive segments represents a change point in C.

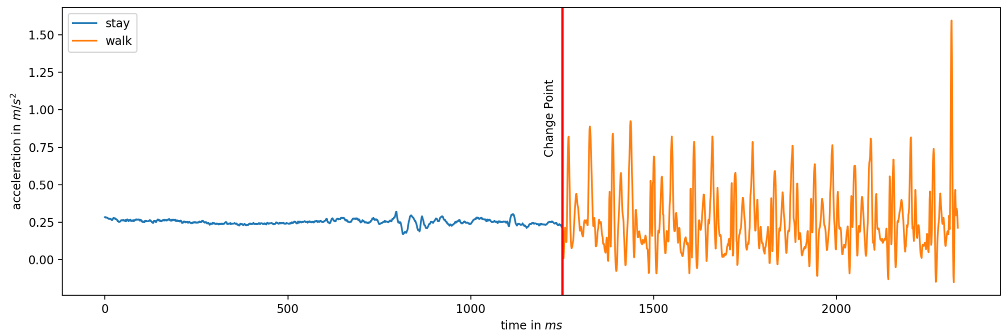

Figure 1: Accelerometer x-axis sensor data from the HASC Challenge 2011[14]

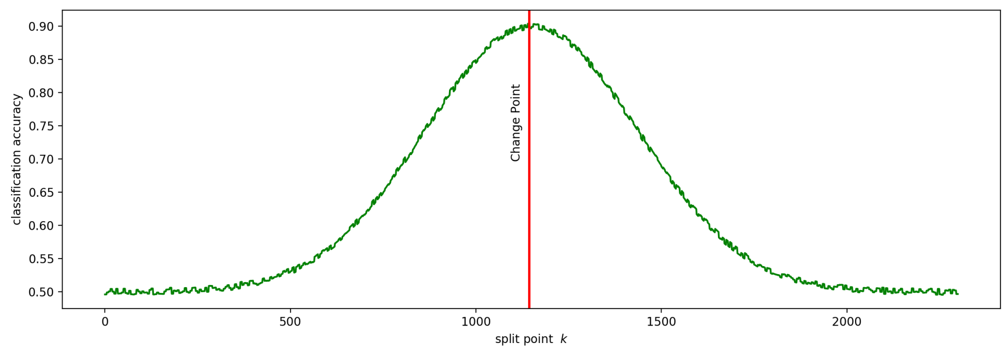

Figure 2: An exemplary classification score profile for Figure 1

As an example for Definitions 1-3 we consider the process of a human being who first "stays" and then "walks". The process $P_a = (S_a, C_a)$ contains the states $S_a = \{stay, walk\}$ and the change point $C_a = \{(stay, walk)\}$. Figure 1 shows a time series T_a from an accelerometer that captures P_a. The change point detection problem in this case would be to find the change point at 1250 ms that divides T_a into the ordered sequence of segments $T_{a1,1250}$ and $T_{a1251,2300}$. Figure 2 shows a synthetic classification score profile for Figure 1. The x-axis constitutes the increasing offsets k that split T_a into two consecutive segments and the y-axis illustrates the corresponding classification accuracies for each split point. The global maxima in Figure 2 approximately captures the change point in T_a because it divides T_a into the two pure self-similar segments "stay" and "walk". Split points left to the change point divide T_a into the first segment that captures "stay" and the second segment that captures "stay" and "walk". Therefore, a time series classification algorithm confuses windows from the first and second segment which leads to lower classification accuracy. Analogously, split points right to the change point lead to a decrease in accuracy as well.

3 Related Work

Unsupervised time series problems like motif or discord discovery as well as signal segmentation and change point detection have been mostly researched using unsupervised methods. For instance, a matrix profile creates an ordered set of all subsequences of length k for a given time series. For every subsequence, it finds the nearest neighbour subsequence (closest match) using euclidean distance and stores the results in a vector called matrix profile [27]. A number of algorithms have been proposed to efficiently calculate matrix profiles and to solve unsupervised learning tasks based on matrix profiles [28]. The nearest neighbour subsequences as well as the corresponding distances in matrix profiles can be leveraged to find time series motifs, discords or change points. Gharghabi, S. et al. for instance, use the proximity of nearest neighbour subsequence matches to create an arc curve from which self-similar regions and change points can be determined [8].

A variety of other change point detection methodologies have been studied before. A large fraction of the research is focused on specialised approaches in fields like motion capture or electrical power demand or domain-agnostic unsupervised methods that are commonly used to discover patterns in time series based on statistical features [19, 22]. Likelihood ratio methods for instance, split time series into consecutive segments and compare the corresponding probability distributions [15]. If these are significantly different, the segments are coherent regions separated by change points. Kernel based methods further split time series into sliding windows and use a kernel-based test statistic to determine the homogeneity between windows in order to detect change points [9]. Graph based methods infer graphs from time series by using the observations as nodes and distances between observations as edges [3]. Common graph structures include minimum spanning trees or nearest neighbour graphs [7]. These methods then split the graphs repeatedly at different observations into sub graphs and compare how mutually connected they are in order to assess if a split is a change point.

Supervised change point detection algorithms have not been studied as much. Such methods can use statistical or temporal features extracted from time series windows in order to train a classifier to recognize coherent regions separated by change points. For instance, binary classification methods train on features from pre-labelled time series segments that capture a state or a transition to detect transitions in unseen segments [5]. Multi-label classification methods further use the same idea but classify the exact state or transition between two states. Self-supervised change point detection methods have not been studied before to the best of our knowledge. Hido, S. et al. classify different datasets from the same feature space with hypothetical labels in order to detect concept drifts and interpret potential changes [10]. We propose to extend this methodology by attaching hypothetical labels to windows from consecutive time series segments and using a time series classifier to discriminate self-similar regions separated by change points.

A multitude of time series classification algorithms have been proposed that include distance as well as feature-based methods [2]. For instance, the one nearest neighbour algorithm with euclidean distance is a very simple distance-based whole series classification algorithm. Given a labelled training time series dataset, it calculates the euclidean distance for a time series query in combination with every training time series and returns the label from the closest match. Optimizations to this methodology use the dynamic time warping distance instead of euclidean distance in order to consider time lag [17]. A more sophisticated feature-based time series classification algorithm is WEASEL. It divides time series into subsequences for which it calculates the Fourier coefficients and selects a discriminative subset that is discretized with a fixed alphabet. A time series is thereby transformed into a set of noise-invariant words for which a $\tilde{\chi}^2$-filtered histogram is constructed and classified with a logistic regression [24].

4 Classification Score Profiles

A time series classification algorithm conceptually extracts statistical properties or temporal patterns as features from labelled training time series in order to discriminate different classes. If the extracted features for each class are statistically distinguishable, the classifier's predictive power and hence performance is high. We can use this principal in order to discover self-similar regions in unlabelled time series in the following way.

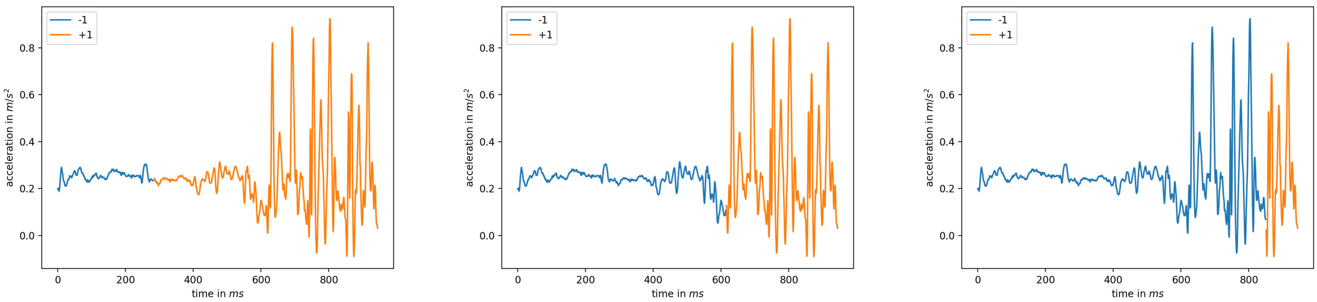

Figure 3: Exemplary time series segments $T_{1,k}$ and $T_{k+1,n}$ with increasing k, annotated with hypothetical labels -1/+1; accelerometer x-axis sensor data from the HASC Challenge 2011[14]

We split a time series T, $|T| = n$ into two consecutive segments $T_{1,k}$ and $T_{k+1,n}$ and attach the hypothetical label (-1) to the first segment and (+1) to the second segment. Going on, we divide each segment into a set of labelled windows with a fixed length and thereby create a two class dataset. We can further split the segments into a training and test subset and fit a time series classification algorithm with the training windows and assess its accuracy with the test windows. If the segments $T_{1,k}$ and $T_{k+1,n}$ are self-similar and the window length and stride is sufficient to capture the similarities, the time series classification accuracy for the test set should be significantly higher compared to more diverse segments. Thus, the classification accuracy should correlate with the homogeneity of the two segments. See Figure 3 (centre) as an example for self-similar segments $T_{1,k}$ and $T_{k+1,n}$.

We can repeat this process for all segments $T_{1,k}$ and $T_{k+1,n}$ with $1 < k < n$ and create a classification score profile $CSP = a_2, ..., a_{n-1}$ with the corresponding accuracies a_k for every split k. See Figure 3 again for an exemplary illustration. CSP is a time series itself and conceptually illustrates to which degree T can be split into two self-similar regions at position k. We can potentially use the minima and maxima in classification score profiles as a mean to detect anomalies and regularities in time series which can assist in solving the aforementioned unsupervised time series problems.

5 Change Point Detection with Classification Score Profiles

In this study project, we want to explore classification score profiles for change point detection. Going on, we specify which problem we want to research in detail and how we want to apply and evaluate classification score profiles as a solution.

5.1 Problem Definition and Solution

We consider a time series T, $|T| = n$ as an ordered sequence of real-valued measurements that captures a physical process $P = (S, C)$, $|C| = 1$ with two discrete states separated by one change point. In this instance, the change point detection problem is to locate the change point position k that segments T into the ordered sequence of segments $T_{1,k}$, $T_{k+1,n}$. We propose to solve this problem in the following way. We split T into all segments $T_{1,k}$ and $T_{k+1,n}$ with $1 < k < n$. We further divide each segment into a set of windows and annotate them with labels (-1) for the first segment and (+1) for the second segment. We classify the windows with a time series classification algorithm and determine the classification performance with a cross validation ranked with classification scores such as accuracy or the F1-score. We repeat this process for every k and create a classification score profile CSP with the corresponding classification performances. Lastly, we report the position of the global maxima in CSP as the change point. See Figure 2 again for a synthetic example.

5.2 Open Problems

The aforementioned solution contains a number of open problems that we must investigate to use classification score profiles to detect one change point in time series. The following list comprises an overview about our main research questions:

1 In order to divide time series segments into windows, we need to determine the **window length and stride** that is sufficient to capture the similarities in the segments. A typical solution to this problem is hyper parameter selection with cross validation. The cross validation must consider all classification scores in a profile to choose the optimal window length and stride.

2 The **time series classification algorithm** determines how self-similar two segments are. We have to choose a suitable classifier that is able to detect similar temporal patterns. Candidate algorithms include nearest neighbour classification with euclidean or dynamic time warping distance as well as BOSS and WEASEL [17, 23, 24].

3 The time series **classifier parameters** vary based on the selected classification algorithm and impact the runtime and classification performance. Hyper parameter choice is non-trivial for domain-agnostic time series applications and is typically solved with cross validation as well. We must consider the entire classification score profile in the cross validation in order to choose optimal classifier parameters.

4 The **classification score** is reported in the classification score profile as the degree of self-similarity for each split. Therefore, we must select a score that satisfies this property and considers class-imbalance as the segments have changing size which leads to an imbalanced amount of windows per segment. Commonly used classification scores include accuracy or the F1-score.

5 The **scalability and robustness** of unsupervised time series algorithms determine their applicability for real-world applications. Our design choices must be scalable and the classification performance should be invariant to time series superposed with white noise. For instance, the one nearest neighbour algorithm can use a precomputed matrix profile in order to predict hypothetical labels by re-labelling the matrix profile for each split point.

Aside from these open problems, there is also a multitude of optimizations we will look into like time series preprocessing, potential segmentation space pruning and parallelization. If the presented methodology solves the problem of identifying one change point in time series sufficiently, we believe that the generalization of detecting k change points and further finding change points with an unknown k in time series is possible.

5.3 Evaluation

In order to assess the performance of our solution, we implement it in Python and evaluate it with the change point detection dataset and the metric published by Gharghabi et al. [8].

Time Series Name	Length	Change Point
Cane	5340	2345
DutchFactory	8761	2184
EEGRat	2000	1000
EEGRat2	2000	1000
GrandMalSeizures	18432	8200
GrandMalSeizures2	10433	4550
InsectEPG1	17001	3802
InsectEPG2	10001	1800
InsectEPG3	7001	1710
InsectEPG4	19001	3160
NogunGun	7383	3000
PigInternalBleeding-DatasetAirwayPressure	14973	7501
PigInternalBleeding-DatasetArtPressureFluidFilled	14973	7501
PigInternalBleedingDatasetCVP	14973	7501
Powerdemand	7682	4500
PulsusParadoxusECG1	17521	10000
PulsusParadoxusECG2	17521	10000
PulsusParadoxusSP02	17521	10000
RoboticDogActivityX	12699	8699
RoboticDogActivityY	14699	10699
RoboticDogActivityZ	11000	4000
SuddenCardiacDeath2	12001	3250
SuddenCardiacDeath3	12001	3250
TiltABP	40000	25000
TiltECG	40000	25000

Table 1: Specification of the 25 time series in the change point detection dataset [8]

Table 1 specifies the 25 time series in the dataset. Every time series captures a biological, mechanical or synthetic process and has one change point in the start, centre or end region. We create a classification score profile CSP for each time series $T, |T| = n$ and calculate the normalized distance $d_{found,true}$ between the found and true change point position cp_{found} or rather cp_{true} in the following way:

$$d_{found,true} = \frac{|cp_{found} - cp_{true}|}{n}$$

We compare the performance of classification score profiles with FLOSS and Autoplait and rank the results based on the normalized distance $d_{found,true}$ [8, 21]. In addition, we evaluate the influence of the window length, the time series classification algorithm as well as the classification score. Furthermore, we present colour pump failure from the durst Gamma XD printer series as a predictive maintenance use case and discuss possible extensions to our solution we plan to investigate in future work.

References

[1] Aminikhanghahi, S., Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.

[2] Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Mining and Knowledge Discovery, 31(3), 606-660.

[3] Chen, H., Zhang, N. (2015). Graph-based change-point detection. The Annals of Statistics, 43(1), 139-176.

[4] Chiu, B., Keogh, E., Lonardi, S. (2003, August). Probabilistic discovery of time series motifs. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 493-498). ACM.

[5] Cook, D. J., Krishnan, N. C. (2015). Activity learning: discovering, recognizing, and predicting human behavior from sensor data. John Wiley & Sons.

[6] Feuz, K. D., Cook, D. J., Rosasco, C., Robertson, K., Schmitter-Edgecombe, M. (2014). Automated detection of activity transitions for prompting. IEEE transactions on human-machine systems, 45(5), 575-585.

[7] Friedman, J. H., Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. The Annals of Statistics, 697-717.

[8] Gharghabi, S., Yeh, C. C. M., Ding, Y., Ding, W., Hibbing, P., LaMunion, S., ... Keogh, E. (2019). Domain agnostic online semantic segmentation for multi-dimensional time series. Data mining and knowledge discovery, 33(1), 96-130.

[9] Harchaoui, Z., Vallet, F., Lung-Yut-Fong, A., Cappé, O. (2009, April). A regularized kernel-based approach to unsupervised audio segmentation. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1665-1668). IEEE.

[10] Hido, S., Idé, T., Kashima, H., Kubo, H., Matsuzawa, H. (2008, May). Unsupervised change analysis using supervised learning. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 148-159). Springer, Berlin, Heidelberg.

[11] Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H. T. (2001, November). Time series segmentation for context recognition in mobile devices. In Proceedings 2001 IEEE International Conference on Data Mining (pp. 203-210). IEEE.

[12] Itoh, N., Kurths, J. (2010). Change-point detection of climate time series by nonparametric method. In Proceedings of the world congress on engineering and computer science (Vol. 1, pp. 20-23).

[13] Jeong, Y. S., Jeong, M. K., Omitaomu, O. A. (2011). Weighted dynamic time warping for time series classification. Pattern Recognition, 44(9), 2231-2240.

[14] Kawaguchi, N., Ogawa, N., Iwasaki, Y., Kaji, K., Terada, T., Murao, K., Nishio, N., ... (2011, March). HASC Challenge: gathering large scale human activity corpus for the real-world activity understandings. In Proceedings of the 2nd augmented human international conference (p. 27). ACM.

[15] Kawahara, Y., Sugiyama, M. (2012). Sequential change-point detection based on direct density-ratio estimation. Statistical Analysis and Data Mining: The ASA Data Science Journal, 5(2), 114-127.

[16] Keogh, E., Lin, J., Fu, A. (2005, November). Hot sax: Efficiently finding the most unusual time series subsequence. In Fifth IEEE International Conference on Data Mining (ICDM'05) (pp. 8-pp). IEEE.

[17] Keogh, E., Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. Knowledge and information systems, 7(3), 358-386.

[18] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307-319.

[19] Kozey-Keadle, S., Libertine, A., Lyden, K., Staudenmayer, J., Freedson, P. S. (2011). Validation of wearable monitors for assessing sedentary behavior. Medicine & Science in Sports & Exercise, 43(8), 1561-1567.

[20] Malladi, R., Kalamangalam, G. P., Aazhang, B. (2013, November). Online Bayesian change point detection algorithms for segmentation of epileptic activity. In 2013 Asilomar Conference on Signals, Systems and Computers (pp. 1833-1837). IEEE.

[21] Matsubara, Y., Sakurai, Y., Faloutsos, C. (2014, June). Autoplait: Automatic mining of co-evolving time sequences. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data (pp. 193-204). ACM.

[22] Reinhardt, A., Christin, D., Kanhere, S. S. (2013, November). Predicting the power consumption of electric appliances through time series pattern matching. In Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (pp. 1-2). ACM.

[23] 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.

[24] Schäfer, P., Leser, U. (2017, November). Fast and accurate time series classification with weasel. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 637-646). ACM.

[25] Staudacher, M., Telser, S., Amann, A., Hinterhuber, H., Ritsch-Marte, M. (2005). A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. Physica A: Statistical Mechanics and its Applications, 349(3-4), 582-596.

[26] Yang, P., Dumont, G., Ansermino, J. M. (2006). Adaptive change detection in heart rate trend monitoring in anesthetized children. IEEE transactions on biomedical engineering, 53(11), 2211-2219.

[27] Yeh, C. C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... Keogh, E. (2016, December). Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In 2016 IEEE 16th international conference on data mining (ICDM) (pp. 1317-1322). IEEE.

[28] Zhu, Y., Zimmerman, Z., Senobari, N. S., Yeh, C. C. M., Funning, G., Mueen, A., ... Keogh, E. (2016, December). Matrix profile ii: Exploiting a novel algorithm and gpus to break the one hundred million barrier for time series motifs and joins. In 2016 IEEE 16th international conference on data mining (ICDM) (pp. 739-748). IEEE.